# The detection of mycoviral sequences in grapevine using next-generation sequencing

by

Yolandi Espach

Thesis presented in partial fulfilment of the requirements for the degree
Master of Science in Genetics at Stellenbosch University

Supervisor: Prof. J.T. Burger
Co-supervisors: Dr. H.J. Maree and Dr. L. Mostert

Faculty of Science
Department of Genetics
March 2013

Declaration

By submitting this thesis electronically, I declare that the entirety of the work contained therein is my own, original work, that I am the sole author thereof (save to the extent explicitly otherwise stated), that reproduction and publication thereof by Stellenbosch University will not infringe any third party rights and that I have not previously in its entirety or in part submitted it for obtaining any qualification.

March 2013

Abstract

Metagenomic studies that make use of next-generation sequencing (NGS) generate large amounts of sequence data, representing the genomes of multiple organisms of which no prior knowledge is necessarily available. In this study, a metagenomic NGS approach was used to detect multiple novel mycoviral sequences in grapevine phloem tissue. Individual sequencing libraries of double-stranded RNA (dsRNA) from two grapevine leafroll diseased (GLD) and three shiraz diseased (SD) vines were sequenced using an Illumina HiScanSQ instrument. Over 3.2 million reads were generated from each of the samples and these reads were trimmed and filtered for quality before being *de novo* assembled into longer contigs. The assembled contigs were subjected to BLAST (Basic Local Alignment Search Tool) analyses against the NCBI (National Centre for Biotechnology Information) database and classified according to database sequences with which they had the highest identity. Twenty-six putative mycovirus species were identified, belonging to the families *Chrysoviridae*, *Endornaviridae*, *Narnaviridae*, *Partitiviridae* and *Totiviridae*. Two of the identified mycoviruses, namely grapevine-associated chrysovirus (GaCV) and grapevine-associated mycovirus 1 (GaMV-1) have previously been identified in grapevine while the rest appeared to be novel mycoviruses not present in the NCBI database. Primers were designed from the *de novo* assembled mycoviral sequences and used to screen the grapevine dsRNA used for sequencing as well as endophytic fungi isolated from the five sample vines. Only two mycoviruses, related to sclerotinia sclerotiorum partitivirus S and chalara elegans endornavirus 1 (CeEV-1), could be detected in grapevine dsRNA and in fungus isolates. In order to validate the presence of mycoviruses in grapevine phloem tissue, two additional sequencing runs, using an Illumina HiScanSQ and an Applied Biosystems (ABI) SOLiD 5500xl instrument respectively, were performed. These runs generated more and higher quality sequence data than the first sequencing run. Twenty-two of the putative mycoviral sequences initially detected were detected in the subsequent sequence datasets, as well as an additional 29 species not identified in the first HiScanSQ sequence datasets. The samples harboured diverse mycovirus populations, with as many as 19 putative species identified in a single vine. This indicates that the complete virome of diseased grapevines will include a high number of mycoviruses. Additionally, the complete genome of a novel endornavirus, for which we propose the name grapevine endophyte endornavirus (GEEV), was assembled from one of the second HiScanSQ sequence datasets. This is the first complete genome of a mycovirus detected in grapevine. Grapevine endophyte endornavirus has the highest sequence similarity to CeEV-1 and is the same virus that was previously detected in fungus isolates using the mycovirus primers. The virus was detected in two fungus isolates, namely *Stemphylium* sp. and *Aureobasidium pullulans,* which is of interest since mycoviruses are not known to be naturally associated with two distinctly different fungus genera. Mycoviral sequence data generated in this study can be used to further investigate the diversity and the effect of mycoviruses in grapevine.

Opsomming

Metagenomiese studies, wat gebruik maak van volgende-generasie volgordebepalingstegnologie, het die vermoë om die genetiese samestelling van veelvoudige onbekende organismes te bepaal deurdat dit groot hoeveelhede data genereer. Die bogenoemde tegniek was in hierdie studie aangewend om 'n aantal nuwe mikovirusse in die floëem weefsel van wingerd te identifiseer. Dubbelstring-RNS was gesuiwer vanuit twee druiwestokke met rolbladsiekte en drie met shiraz-siekte en 'n Illumina HiScanSQ instrument is gebruik om meer as 3.2 miljoen volgorde fragmente te genereer van elk van die monsters. Lae-kwaliteit volgordes was verwyder en die oorblywende kort volgorde fragmente was saamgestel om langer konstrukte te vorm wat met behulp van BLAST soektogte teen die NCBI databasis geïdentifiseer kon word. Ses-en-twintig mikovirus spesies, wat aan die families *Chrysoviridae*, *Endornaviridae*, *Narnaviridae*, *Partitiviridae* en *Totiviridae* behoort, was geïdentifiseer. Twee van die geïdentifiseerde mikovirusse, naamlik grapevine-associated chrysovirus (GaCV) en grapevine-associated mycovirus 1 (GaMV-1), was voorheen al in wingerd gekry terwyl die res nuwe mikovirusse is wat tans nie in die NCBI databasis voorkom nie. Inleiers was ontwerp vanaf die saamgestelde mikovirus basisvolgordes en gebruik om wingerd dubbelstring-RNS sowel as swamme wat vanuit die wingerd geïsoleer is te toets vir die teenwoordigheid van hierdie mikovirusse. Slegs twee mikovirusse, wat onderskeidelik verwant is aan sclerotinia sclerotiorum partitivirus S en chalara elegans endornavirus 1 (CeEV-1), kon deur middel van die inleiers in wingerd en swam isolate geïdentifiseer word. Twee addisionele volgordebepalingsreaksies, wat gebruik gemaak het van die Illumina HiScanSQ en ABI SOLiD 5500xl volgordebepalingsplatforms, was gebruik om die teenwoordigheid van mikovirusse in wingerd te bevestig. 'n Groter hoeveelheid volgorde fragmente was geprodusser wat ook van 'n hoër gehalte was as dié van die eerste volgordebepalingsreaksie. Twee-en-twintig mikovirus spesies kon weer geïdentifiseer word, sowel as 29 spesies wat nie in die eerste HiScanSQ basisvolgorde datastelle gevind was nie. Die wingerdstokke wat in hierdie studie ondersoek was, het 'n hoë diversiteit van mikovirusse bevat aangesien daar tot 19 mikovirus spesies in 'n enkele wingerdstok geïdentifiseer was. Dit is 'n aanduiding dat volledige virus profiele van siek wingerdstokke 'n aantal mikovirusse sal insluit. Die vollengte genoomvolgorde van 'n voorheen onbekende endornavirus was saamgestel vanuit een van die tweede HiScanSQ volgorde datastelle. Dit is die eerste mikovirus wat in wingerd gevind word waarvan die volledige genoomvolgorde bepaal is en ons stel die naam grapevine endophyte endornavirus (GEEV) voor vir hierdie virus. Grapevine endophyte endornavirus is die naaste verwant aan CeEV-1 en is dieselfde virus wat voorheen in wingerd dubbelstring-RNS en swam isolate gevind was deur middel van die mikovirus inleiers. Swam isolate waarin GEEV gevind is, was geïdentifiseer as *Stemphylium* sp. en *Aureobasidium pullulans*. Dit is van belang dat GEEV in twee swam isolate gevind is wat aan verskillende genusse behoort aangesien hierdie verskynsel nog nie voorheen in die natuur gevind is nie. Mikovirus nukleiensuurvolgordes wat in hierdie studie bepaal was kan

gebruik word in toekomstige studies om die verskeidenheid en impak van mikovirusse in wingerd verder te ondersoek.

Acknowledgements

I would like to extend my sincere gratitude towards the following people and institutions for their assistance and contributions during this study:

- Prof. Johan Burger, for guidance and support and for the opportunity to be part of his research group.
- Dr. Hans Maree, for intellectual input and guidance throughout this study and for allowing me to be part of his research.
- Dr. Lizel Mostert for guidance and assistance with the fungus work.
- The Department of Plant Pathology (Stellenbosch University) for use of their facilities during the fungalus work.
- Beatrix Coetzee and Rachelle Bester, for their assistance during the study.
- The owners of the various wine farms from which samples were collected for this study (ARC-Infruitec/Nietvoorbij, Plaisir de Merle and Fairview).
- Carel van Heerden, Dr. Ruhan Slabbert, Anelda van der Walt and the Central Analytical Facility of Stellenbosch University for the SOLiD 5500xl sequencing.
- Winetech for research funding.
- The NRF and Stellenbosch University for personal funding. The financial assistance of the National Research Foundation (NRF) towards this research is hereby acknowledged. Opinions expressed and conclusions arrived at, are those of the author and are not necessarily to be attributed to the NRF.
- SASEV and Whitehead Scientific, for travel funding to attend the 17th meeting of the ICVG in Davis, California, USA.
- Colleagues and friends in the Vitis laboratory, for the entertaining and supportive working environment that they created.
- My husband Christo, family and friends for endless love and support.

Table of contents

List of abbreviations

aa	Amino acid
ABI	Applied Biosystems
AMV	Avian myeloblastosis virus
ARC-BP	Agricultural Research Council Biotechnology Platform
BLAST	Basic Local Alignment Search Tool
BPEV	Bell pepper endornavirus
BSA	Bovine serum albumin
BWT	Burrows-Wheeler Transform
cDNA	Complimentary deoxyribonucleic acid
CeEV-1	Chalara elegans endornavirus 1
CTAB	Cetyltrimethylammonium bromide
CThTV	Curvularia thermal tolerance virus
cv	Cultivar
DAS-ELISA	Double antigen sandwich enzyme-linked immunosorbent assay
dNTP	Deoxynucleotide triphosphate
ddNTP	2',3'-dideoxynucleotide triphosphate
DNA	Deoxyribonucleic acid
DOI	Digital object identifier
dsDNA	Double-stranded deoxyribonucleic acid
dsRNA	Double-stranded ribonucleic acid
DTT	Dithiothreitol
ELISA	Enzyme-linked immunosorbent assay
EtBr	Ethidium bromide
EtOH	Ethanol
GaCV	Grapevine-associated chrysovirus

GaMV-1	Grapevine-associated mycovirus 1
GEEV	Grapevine endophyte endornavirus
GES	Glysine-NaOH/EDTA/sodium
GFLV	Grapevine fanleaf virus
GLD	Grapevine leafroll disease
GLRaV-3	Grapevine leafroll-associated virus 3
GLRaV-9	Grapevine leafroll-associated virus 9
GRSPaV	Grapevine rupestris stem pitting-associated virus
GRVFV	Grapevine rupestris vein-feathering virus
GSyV-1	Grapevine syrah virus 1
GVA	Grapevine virus A
GVE	Grapevine virus E
GYSVd	Grapevine yellow speckle viroid
HSVd	Hop stunt viroid
ICTV	International Committee for Taxonomy of Viruses
kbp	Kilobase pairs
kDa	Kilodalton
MAQ	Mapping and Assembly with Quality
mRNA	Messenger ribonucleic acid
NCBI	National Centre for Biotechnology Information
NGS	Next-generation sequencing
nt	Nucleotide
ORF	Open reading frame
PDA	Potato dextrose agar
PGM	Personal genome machine
PMWaV-1	Pineapple mealybug wilt-associated virus 1
PVP-40	Polyvinylpyrrolidone-40

RNA	Ribonucleic acid
Rpm	Revolutions per minute
rRNA	Ribosomal ribonucleic acid
SDS	Sodium dodecyl sulfate
STE	Sodium/Tris/EDTA
RdRp	RNA-dependent RNA polymerase
RT-PCR	Reverse transcription polymerase chain reaction
SD	Shiraz disease
SMRT	Single-molecule real-time
SOLiD	Sequencing by Oligo Ligation Detection
ssDNA	Single-stranded deoxyribonucleic acid
ssRNA	Single-stranded ribonucleic acid
TAE	Tris/Acetate/EDTA
UPGMA	Unweighted pair-group method with arithmetic mean
UTR	Untranslated region
v/v	Volume per volume
w/v	Weight per volume

List of figures

List of tables

Chapter 1: Introduction

1.1 General Introduction

Grapevine (*Vitis vinifera*) is a valuable fruit crop that is cultivated on six continents (Burger et al., 2009), with 7.6 million hectares under vines globally (www.oiv.int). South Africa is ranked as having the 12th largest area under vines (131 000 hectares) and as the 8th largest wine producing country, contributing 3.6% to the global wine production in 2011 (http://www.sawis.co.za/info/annualpublication.php). There are nine wine producing regions in South Africa, namely Stellenbosch, Paarl, Robertson, Malmesbury, Breedekloof, Olifants River, Worcester, Orange River and Little Karoo (named in descending order, in terms of total hectares under wine grape vines in 2011). Collectively, these areas produced 831 million litres of wine in 2011, of which 357 million litres was exported (http://www.sawis.co.za/info/annualpublication.php). The wine industry contributes significantly to the South African Gross Domestic Product, with R26.2 billion generated in 2008 (http://www.sawis.co.za/info/macro_study2009.php).

Grapevine-infecting agents like viruses, viroids, fungi, bacteria and phytoplasmas limit the growth and development of the wine industry as many of these agents affect the quality and yield of infected vines. The number of recorded grapevine infecting viruses has steadily increased over the last few years, and grapevine is now regarded as the woody crop that hosts the highest number of viruses (Martelli, 2012). In 2003, the International Council for the Study of Virus and Virus-like Diseases of the Grapevine (ICVG) recognised 55 grapevine infecting viruses (Martelli, 2003). This number grew to 58 in 2006 (Martelli and Boudon-Padieu, 2006), 60 in 2009 (Martelli, 2009) and 63 in 2012 (Martelli, 2012). This number does not include viruses that infect fungi, known as mycoviruses, which were recently identified in grapevine for the first time (Coetzee et al., 2010).

A number of the viruses that infect grapevine cause either negligible symptoms or no symptoms at all (Martelli, 2006). Others cause symptoms that are only visible during a specific time in the growing season, cause different symptoms depending on the variant of the virus present (Monette and James, 1990) or modify the symptoms caused by co-infecting viruses (Komar et al., 2007). This means that the symptoms displayed by a vine, or the lack thereof, is not an accurate indication of the virus(es) present in the vine.

Routinely-used virus detection techniques, like reverse transcription reverse transcription-polymerase chain reaction (RT-PCR) and enzyme-linked immunosorbent assay (ELISA), require prior knowledge of the virus to be detected and are thus inadequate for the detection of mycoviruses in grapevine, of which very little is known to date. A metagenomic approach, coupled with NGS, has the power to determine the complete virus community or virome of a vine, including mycoviruses. The unbiased nature and sensitivity of the approach ensures that known, unknown and low titre viruses are detected. Next-generation sequencing has been used successfully to

investigate grapevine infecting viruses and to identify novel viruses in grapevine, including mycoviruses (Alabi et al., 2012; Al Rwahnih et al., 2009; Al Rwahnih et al., 2011; Coetzee et al., 2010; Giampetruzzi et al., 2012; Pantaleo et al., 2010).

1.2 Aims and Objectives

The aim of this study was to determine the mycovirus complexity in individual diseased grapevines using NGS and to validate the presence of the identified mycoviruses in subsequent sequencing analyses. The following objectives were set out to achieve this aim:

- To identify diseased grapevine plants and extract high quality dsRNA.
- To prepare cDNA sequencing libraries and sequence the libraries using an Illumina HiScanSQ instrument.
- To perform *de novo* assemblies and identify mycoviral sequences present in assembled data.
- To screen the grapevine samples and endophytic fungi for the identified mycoviruses.
- To verify the presence of mycoviruses in the samples using additional sequencing runs.
- To further characterise novel mycoviruses that were identified.

1.3 Chapter Layout

This thesis is divided into five chapters: a general introduction, a literature overview, two research chapters and a general conclusion. Each chapter is introduced and referenced separately.

Chapter 1: Introduction
A general introduction to the study, including aims and objectives, an overview of the chapter layout and research outputs generated by the study.

Chapter 2: Literature review
An overview of literature relating to grapevine infecting agents, associated diseases and detection techniques is given with a focus on mycoviruses and next-generation sequencing.

Chapter 3: Determination of the mycovirus complexity in five grapevine samples.
The identification of mycoviral sequences in five next-generation sequencing datasets, generated from diseased grapevines, is described. The detection of the identified mycoviruses in grapevine and endophytic fungi using PCR-based techniques is also described.

Chapter 4: Confirmation of mycovirus complexity through next-generation sequencing, with a focus on a novel endornavirus.
The use of two additional next-generation sequencing runs to validate the mycovirus findings of Chapter 3 is described. The first complete genome assembly of a mycovirus identified in grapevine endophytes is also reported in this chapter.

Chapter 5: Conclusion

General concluding remarks and future prospects of the study.

1.4 Research Outputs

The following publication and conference contributions were produced during this study:

Publication

- **Espach, Y.**, Maree, H.J., Burger, J.T., 2012. Complete genome of a novel endornavirus assembled from next-generation sequence data. J. Virol. 86(23), 13142.
  This Genome Announcement publication forms part of Chapter 4.

International conference contributions

- Coetzee, B., Maree, H.J., **Nel[†], Y.**, Rees, D.J.G., <u>Burger, J.T</u>. Next Generation Sequencing as a tool to study the etiology of plant virus diseases: the case study of the virome of a vineyard. BARD-sponsored workshop "Microarrays and NGS for the detection and identification of plant viruses". Beltville Agricultural Research Centre, USA. 17 – 19 November 17 – 19, 2011.
  Mrs Espach assisted with sample collection and processing as well as contributed to the data analysis. The presentation was delivered by Prof Burger.

- <u>Maree, H.J.</u>, **Nel[†], Y.**, Visser, M., Coetzee, B., Manicom, B., Burger, J.T., Rees, D.J.G. The study of plant virus disease etiology using next-generation sequencing technologies. 22nd International Conference on Virus and Other Graft Transmissible Diseases of Fruit Crops (ICVF). Rome, Italy. June 3 – 8, 2012.
  Mrs Espach assisted with sample collection and processing as well as contributed to the data analysis. The presentation was delivered by Dr Maree.

- **<u>Espach, Y.</u>**, Maree, H.J., Burger, J.T. The use of next-generation sequencing to identify novel mycoviruses in individual grapevine plants. 17th congress of the International Council for the study of Virus and Virus-like Diseases of the Grapevine (ICVG).Davis, California, USA. October 7 – 14, 2012.
  Poster summarizing the work in Chapters 3 and 4 presented by Mrs Espach.

- <u>Maree, H.J.</u>, **Espach, Y.**, Rees, D.J.G., Burger, J.T. A study of shiraz disease etiology using next-generation sequencing technology. 17th congress of the International Council for the study of Virus and Virus-like Diseases of the Grapevine (ICVG). Davis, California, USA. October 7 – 14, 2012.
  Mrs Espach assisted with sample collection and processing as well as contributed to the data analysis. The presentation was delivered by Dr Maree.

[†] Nel is the author's maiden name.

- **Espach, Y.**, Maree, H.J., <u>Burger, J.T.</u> The use of next-generation sequencing to identify novel mycoviruses in individual grapevine plants. COST Action FA1103 workshop "Endophytes in biotechnology and agriculture". Fondazione Edmund Mach, Italy. November 14 – 16, 2012.
  Poster summarizing the work in Chapters 3 and 4 presented by Prof Burger.

Local conference contributions

- <u>Maree, H.J.</u>, Coetzee, B., **Nel**[†]**, Y.**, Burger, J.T., Rees, D.J.G. Unravelling the complexity of grapevine viral diseases using next-generation sequencing. Agricultural Biotechnology International Conference (ABIC). Sandton, South Africa. September 6 – 9, 2011.
  Mrs Espach assisted with sample collection and processing as well as contributed to the data analysis. The poster was presented by Dr Maree.

- Coetzee, B., Maree, H.J., **Nel**[†]**, Y.**, Rees, D.J.G. and <u>Burger, J.T.</u> The use of Next-Generation sequencing in metagenomic studies of plant viruses. Africa Virology Conference. Cape Town, South Africa. November 29 – December 2, 2011.
  Mrs Espach assisted with sample collection and processing as well as contributed to the data analysis. The presentation was delivered by Prof Burger.

- <u>Maree, H.J.</u>, Coetzee, B., **Nel**[†]**, Y.**, Burger, J.T., Rees, D.J.G. Unravelling the complexity of grapevine viral diseases using next-generation sequencing. Africa Virology Conference. Cape Town, South Africa. November 29 – December 2, 2011.
  Mrs Espach assisted with sample collection and processing as well as contributed to the data analysis. The poster was presented by Dr Maree.

- **<u>Espach, Y.</u>**, Maree, H.J., Burger, J.T. The use of next-generation sequencing to identify novel mycoviruses in single grapevine plants. 34th South African Society for Enology and Viticulture (SASEV) Congress. Stellenbosch, South Africa. November 14 – 16, 2012.
  Presentation summarising the work in Chapters 3 and 4, presented by Mrs Espach.

- <u>Maree, H.J.</u>, **Espach, Y.**, Smyth, N., Rees, D.J.G., Burger, J.T. A study of shiraz disease etiology using next-generation sequencing technology. 34th South African Society for Enology and Viticulture (SASEV) Congress. Stellenbosch, South Africa. November 14 – 16, 2012.
  Mrs Espach assisted with sample collection and processing as well as contributed to the data analysis. The presentation was delivered by Dr Maree.

1.5 References

Alabi, O.J., Zheng, Y., Jagadeeswaran, G., Sunkar, R., Naidu, R.A., 2012. High-throughput sequence analysis of small RNAs in grapevine (*Vitis vinifera* L.) affected by grapevine leafroll disease. Mol. Plant Pathol. 13(9), 1060 – 1076.

[†] Nel is the author's maiden name.

Al Rwahnih, M., Daubert, S., Golino, D., Rowhani, A., 2009. Deep sequencing analysis of RNAs from grapevine showing Syrah decline symptoms reveals a multiple virus infection that includes a novel virus. Virology. 387(2), 395 – 401.

Al Rwahnih, M., Daubert, S., Úrbez-Torres, J.R., Cordero, F., Rowhani, A., 2011. Deep sequencing evidence from single grapevine plants reveals a virome dominated by mycoviruses. Arch. Virol. 156(3), 397 – 403.

Burger, P., Bouquet, A., Striem, M.J., 2009. Grape breeding. In: Jain, S.M., Priyadarshan, P.M. (eds), Breeding plantation tree crops: Tropical species. Springer Science, New York, 161 – 189.

Coetzee, B., Freeborough, M.J., Maree, H.J., Celton, J.M., Reese, D.J.G., Burger, J.T., 2010. Deep sequencing analysis of viruses infecting grapevines: Virome of a vineyard. Virology. 400(2), 157 – 163.

Giampetruzzi, A., Roumi, V., Roberto, R., Malossini, U., Yoshikawa, N., La Notte, P., Terlizzi, F., Credi, R., Saldarelli, P., 2012. A new grapevine virus discovered by deep sequencing of virus- and viroid-derived small RNAs in cv. *Pinot gris*. Virus Res. 163(1), 262 – 268.

Komar, V., Emmanuelle, V., Demangeat, G., Fuchs, M., 2007. Beneficial effect of selective virus elimination on the performance of *Vitis vinifera* cv. Chardonnay. Am. J. Enol. Vitic. 58(2), 202 – 210.

Martelli, G.P., 2003. Grapevine virology highlights: 2000 – 2003. Proceedings of the 14th Congress of the International Council for the Study of Virus and Virus-like Diseases of the Grapevine (ICVG), Locotoronto, Bari, Italy, September 12 – 17, 2003, 3 – 10.

Martelli, G.P., 2006. Infectious diseases and certification of grapevines. Proceedings of the Mediterranean Network on Grapevine Closteroviruses. Options Méditerr., Ser. B. 29, 47 – 64.

Martelli, G.P., 2009. Grapevine virology highlights: 2006 – 2009. Proceedings of the 16th Congress of the International Council for the Study of Virus and Virus-like Diseases of the Grapevine (ICVG), Dijon, France, August 31 – September 4, 2009, 15 – 23.

Martelli, G.P., 2012. Grapevine virology highlights: 2010 – 2012. Proceedings of the 17th Congress of the International Council for the Study of Virus and Virus-like Diseases of the Grapevine (ICVG), Davis, California, USA. October 7 – 14, 2012, 13 – 31.

Martelli, G.P., Boudon-Padieu, E., 2006. Directory of infectious diseases of grapevines and viroses and virus-like diseases of the grapevine: Bibliographic report 1998 - 2004. Options Méditerr. Ser. B: Stud. Res. 55, CIHEAM, 279.

Monette, P.L., James, D., 1990. Detection of two strains of grapevine virus A. Plant Dis. 74(11), 898 – 900.

Pantaleo, V., Saldarelli, P., Miozzi, L., Giampetruzzi, A., Gisel, A., Moxon, S., Dalmay, T., Bisztray, G., Burgyan, J., 2010. Deep sequencing analysis of viral short RNAs from an infected Pinot Noir grapevine. Virology. 408(1), 49 – 56.

Internet Resources:

South African Wine Industry Information and Systems (SAWIS) Macro-economic Impact of the Wine Industry on the South African Economy (also with reference to the Impacts on the Western Cape) – Final Report (09 December 2009): http://www.sawis.co.za/info/download/Macro_study_2009.pdf(accessed 10/12/2012)

36th Report of the South African wine industry statistics by South African Wine Industry Information and Systems (SAWIS): http://www.sawis.co.za/info/annualpublication.php (accessed 10/12/2012)

Statistical report on world vitiviniculture (2012) by International Organisation of Vine and Wine (OIV): http://www.oiv.int/oiv/info/enstatistiquessecteurvitivinicole#bilan (accessed 10/12/2012)

Chapter 2: An overview of grapevine infecting agents and detection techniques, with a focus on metagenomics and next-generation sequencing

2.1 Introduction

Grapevine is a deciduous woody crop that is cultivated in temperate regions across the world. The berries are used for raw consumption, or to produce wine, juice, brandy, vinegar, jam or raisins. The production of grapevine is therefore a key contributor to the global economy because of the many uses of its fruit. Grapevine is host to a large number of pathogens, including viruses, viroids, phytoplasmas, bacteria and fungi (Martelli and Boudon-Padieu, 2006). These pathogens negatively affect the quality, yield and productive life of the crop and are a constant threat to the industry. The study of grapevine, including the vast number of pathogens infecting the crop, is aided by the development of powerful technologies (Martelli and Boudon-Padieu, 2006).

The emergence of NGS has made a significant contribution to the study of viral and microbial populations as it is more cost-effective and less time consuming than traditional detection techniques (Beerenwinkel and Zagordi, 2011). One of its biggest advantages is the ability to detect new variants of viruses as well as completely novel viruses. Next-generation sequencing has been used to determine the virome of pooled grapevine samples (Al Rwahnih et al., 2011; Coetzee et al., 2010). Both these groups found mycoviruses to be strongly represented in the viromes of grapevine.

2.2 Grapevine pathogens and associated diseases

2.2.1 Fungal diseases

Grapevine is susceptible to infection by a large number of both pathogenic and endophytic fungi. Pathogenic fungus infections can lead to a number of different diseases which decrease the yield and vigour of crops. Some of the most significant fungal diseases in the South African context are powdery mildew (*Erysiphe necator*), downy mildew (*Plasmopara viticola*), Petri disease (*Phaeomoniella chlamydospora* and *Phaeoacremonium* species), black dead arm (*Botryosphaeria* spp.), phomopsis cane and leaf spot (*Phomopsis viticola*) and grey mould rot (*Botrytis cinerea*) (Burger and Deist, 2001; Fourie and Halleen, 2004). Conversely, endophytic fungi could either have no effect on the plant they inhabit, or be beneficial to the plant. Fungi are regarded as endophytic if they do not cause any visible symptoms in their host at that given time(Schulz and Boyle, 2005). Endophytic fungi can, however, become pathogenic at a later stage if the environmental conditions are altered (Schulz and Boyle, 2005). Endophytic fungi can inhibit the development of a disease caused by a pathogenic fungus by rapidly colonizing host tissue before the pathogen (Kortekamp, 1997; Musetti et al., 2006). Fungus endophytes identified from healthy

grapevine in South Africa include *Alternaria* spp., *Chaetomium* sp., *Cladosporium cladosporioides*, *Epicoccum nigrum*, *Fusarium* spp., *Gliocladium roseum*, *Nigrospora oryzae*, *Phoma* sp., *Phomopsis viticola*, *Pleospora herbarum*, *Sphaeropsis* sp., *Sporormiella minimoides*, *Trichoderma* sp. and *Verticillium* sp. (Mostert et al., 2000).

2.2.2 Viral diseases

Grapevine is susceptible to at least 63 different viruses, representing all the different genome types (single-stranded DNA (ssDNA), dsDNA, dsRNA, negative-sense ssRNA and positive-sense ssRNA) (Martelli, 2012). Not all of these viruses are associated with diseases in grapevine as some either cause negligible symptoms or are latent infections, causing no symptoms (Martelli, 2006). Five major virus associated diseases have been identified in grapevine, namely infectious degeneration, grapevine leafroll, rugose wood complex, graft incompatibility and fleck complex (Martelli and Boudon-Padieu, 2006). Viruses commonly detected in South African vineyards include grapevine leafroll-associated virus 3 (GLRaV-3) (family *Closteroviridae*), grapevine rupestris stem pitting-associated virus (GRSPaV), grapevine virus A (GVA), grapevine virus E (GVE) (family *Betaflexiviridae*) and grapevine fanleaf virus (GFLV) (family *Secoviridae*). The most prominent viral diseases in South Africa are GLD, SD, grapevine fanleaf disease and shiraz decline.

Grapevine leafroll Disease

Grapevine leafroll disease is present in all grape-growing countries, decreasing crop yields with 15-20% (Martelli and Boudon-Padieu, 2006). It is therefore a globally important disease and an economical threat to the industry. Grapevine leafroll disease is the most widespread disease affecting South African vineyards. Symptoms are most visible in autumn and differ between cultivars, but is generally more pronounced in red cultivars (Figure 2.1a-b) (Jooste et al., 2011). In red cultivars, reddening of the leaf surface occurs with the primary and secondary veins remaining green. Leaves also become brittle with the edges rolling downwards. Bunches ripen irregularly, but generally later, and have decreased sugar content and increased acidity. Symptoms are less distinct in white cultivars, with some white cultivars displaying no symptoms. In cultivars such as Chardonnay, which do display GLD symptoms, the leaves turn yellow instead of red, with the rest of the symptoms being similar to that of red cultivars (Jooste et al., 2011; Martelli and Boudon-Padieu, 2006). Although other viruses have also been associated with the disease, the most common causative agent is GLRaV-3 (Pietersen, 2004). Grapevine leafroll associated virus 3 is a single-stranded positive-sense RNA virus and is the type member of the genus *Ampelovirus*, family *Closteroviridae*. The virus is phloem-limited and is transmitted through vegetative propagation, grafting and by mealybug and soft scale vectors (*Planococcus ficus*, *Planococcus citri*, *Pseudococcus longispinus*, *Pseudococcus calceolariae*, *Pulvinaria vitis* and *Neopulvinaria innumerabilis*) (Martelli and Boudon-Padieu, 2006).

Shiraz disease

Plants affected with SD do not mature fully, remain unlignified and show delayed budding. The quality and number of bunches is affected and leaves that resemble GLD symptoms are shed later than in unaffected vines (Figure 2.1c-d). After the appearance of symptoms, the vines will degenerate and usually die within three to five years (Carstens, 1999). The disease is only observed in selected cultivars (Shiraz, Merlot, Malbec, Gamay, Tempranillo, Mourvèdre and Voignier) and is transmitted through grafting or mealybug vectors (Goszczynski, 2007a). Shiraz disease is found only in South Africa, but a similar disease, Australian shiraz disease, has been detected in Australian vineyards. Molecular variant group II of GVA has been associated with both these diseases (Goszczynski, 2007b; Habili, 2007), but the aetiology is still not clear and it is believed that multiple viruses could be involved (du Preez, 2005).

Grapevine fanleaf disease

Grapevine fanleaf disease is caused by GFLV, which belongs to the genus *Nepovirus*, family *Secoviridae*. It is one of the most devastating diseases of grapevine globally (Andret-Link et al., 2004) but, in South Africa, it is mostly restricted to the Breede River Valley in the Western Cape (Lamprecht et al., 2012b). The main symptoms include degeneration, yellowing of leaves and malformation of leaves, berries and shoots, leading to a decrease in the yield and quality (Figure 2.1e-g) (Lamprecht et al., 2012a; Martelli, 2006). The genome of GFLV consists of two segments of positive-sense, ssRNA, with some isolates of the virus having an associated satellite RNA (Martelli and Boudon-Padieu, 2006). The virus is transmitted by the nematode vector, *Xiphinema index*.

Shiraz decline

Although both shiraz disease and shiraz decline cause vines to degenerate and eventually die, the symptoms of the two diseases are distinct (Spreeth, 2005). Plants affected with shiraz decline have swollen graft joints with thickened bark above the graft union, deep cracks appear on the canes and the leaves redden prematurely (Figure 2.1h-j). Affected vines die within five to ten years. Although there are minor differences in the symptoms observed in the different regions, shiraz decline has been observed in French, Californian and South African vineyards (Al Rwahnih et al., 2009; Goszczynski, 2007a; Spreeth, 2005). Shiraz decline symptoms have only been observed in vines propagated from the French clone Syrah 99 in South Africa. The propagation of this clone has been discontinued which limits the prevalence of the disease (Goszczynski, 2010; Goszczynski, 2011; Spreeth, 2005). Three viruses have been associated with shiraz decline, namely GRSPaV, grapevine rupestris vein-feathering virus (GRVFV) and grapevine syrah virus 1 (GSyV-1) (Al Rwahnih et al., 2009; Lima et al., 2006). Of these, GRSPaV is the most prevalent. However, Goszczynski (2010) recently showed that GRSPaV is widely present in vineyards, but not necessarily associated with shiraz decline. Thus, the cause of the disease remains unknown.
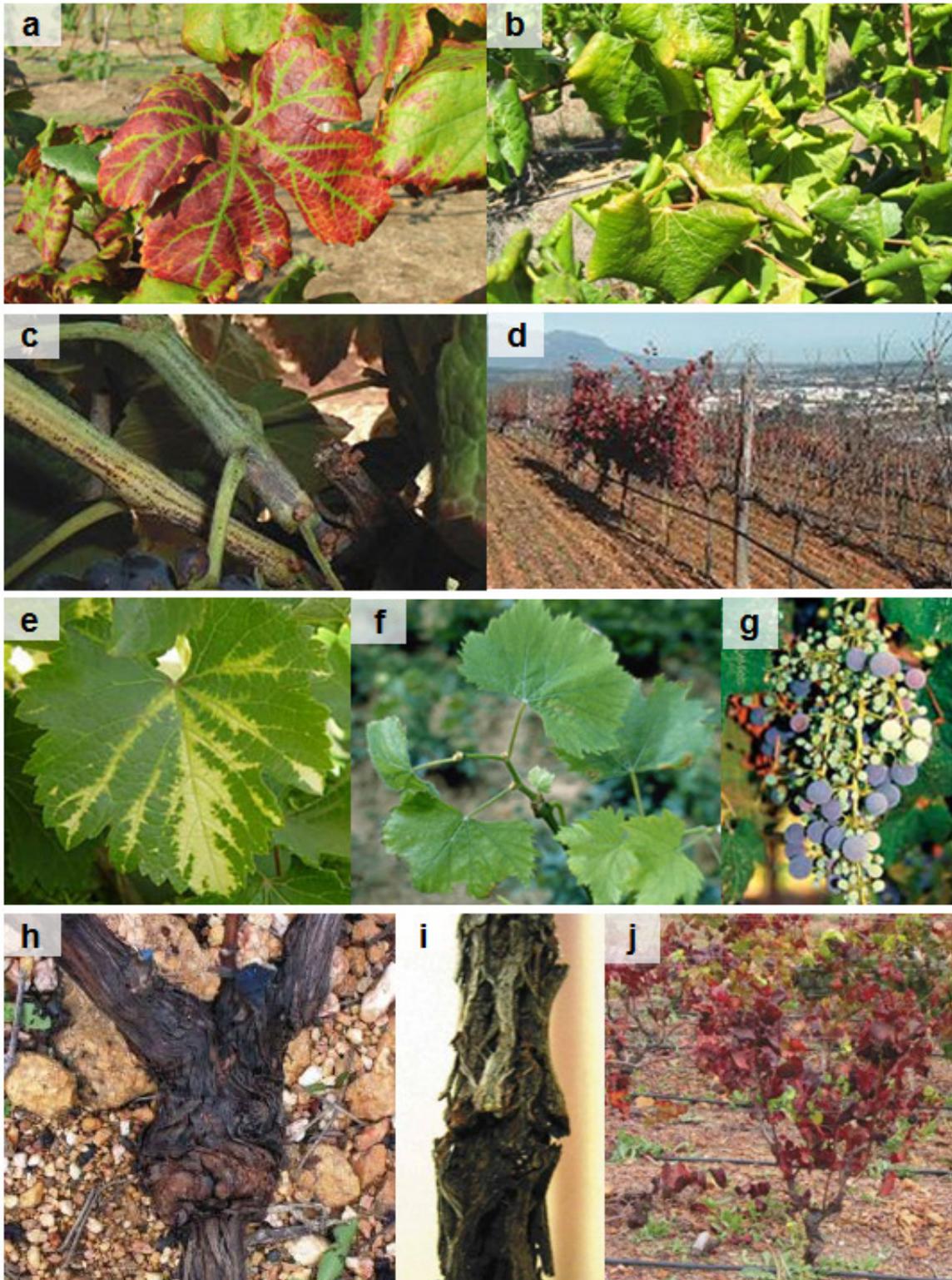
Figure 2.1: Symptoms displayed in grapevine diseases a) Grapevine leafroll disease symptoms in a red grapevine cultivar (Photo by E. Hellman). b) Grapevine leafroll disease symptoms in a white grapevine cultivar (Photo by H.J. Maree). c) Unlignified shoots in shiraz disease. d) Red leaves with delayed shedding in shiraz diseased vines. (Photos c-d from Goussard and Bakker (2006)). e) Yellowing of leaves in grapevine fanleaf disease (Photo by S. Jordan). f) Malformation of shoots and leaves, with leaves resembling a fan in grapevine fanleaf disease (Photo by W.M. Brown Jr.). g) Malformed berries in grapevine fanleaf disease (Photo by A. Schilder). h) Swollen graft joint and thickened bark in shiraz decline. i) Cracked cane in shiraz decline. j) Red discoloured leaves on a declining vine with shiraz decline. (Photos h-j from Spreeth (2005)).

2.3 Mycoviruses

Mycoviruses were recently detected in grapevine by two different research groups using NGS (Al Rwahnih et al., 2011; Coetzee et al., 2010). A significant number of mycoviral sequences were found to be present in grapevine belonging to the families *Chrysoviridae*, *Hypoviridae*, *Narnaviridae*, *Partitiviridae*, and *Totiviridae*. These were the first documented accounts of mycoviruses in grapevine and their effect on grapevine have not yet been determined. Al Rwahnih et al. (2011) predict that a complete grapevine virome will include a substantial number of mycoviruses.

True mycoviruses are distinguished from plant viruses that merely use fungi as vectors as they are able to replicate in fungi and cannot survive outside the fungus host cells (Buck, 1986; Tavantzis, 2008). A large number of mycoviruses have been identified to date, with more than 90 mycovirus species recognised by the International Committee for Taxonomy of Viruses (ICTV) (King et al., 2011). These mycoviruses are classified across 13 virus families, of which some families infect only fungi while others contain members that infect fungi, protozoa, plants or animals (Ghabrial and Suzuki, 2008). Approximately 20% of mycoviruses are still unassigned to either a genus or a family as this is difficult to do without adequate sequence or biological data (Pearson et al., 2009). Table 2.1 lists the number of identified mycoviruses belonging to each family (See Table S1 in supplementary data for a complete list of mycovirus species). As can be seen in the table, the majority of mycoviruses have dsRNA genomes. The genomes can be non-segmented or divided into two, four, 11 or 12 segments and are mostly encapsidated in isometric, non-enveloped particles that are 25-50 nm in diameter. Other particle morphologies have also been observed, while some mycoviruses like those belonging to the family *Endornaviridae*, are not encapsidated (Ghabrial and Suzuki, 2008; Pearson et al., 2009; Roossinck et al., 2011).

True mycoviruses infect all major taxonomic groups of fungi, namely Chytridiomycota, Zygomycota, Ascomycota and Basidiomycota, as well as Oomycota (Ghabrial and Suzuki, 2008; Pearson et al., 2009). The incidence of mycoviruses in fungus isolates has varied between studies from only a small percentage to as high as 85% of isolates containing mycoviruses, but most studies agree that mycoviruses are ubiquitous in fungi (Ghabrial and Suzuki, 2008; Pearson et al., 2009). Mycoviruses depend on their fungus hosts for transmission as they lack extracellular vectors and never leave the fungal cytoplasm. Vertical transmission through asexual spores is typical, while the ability to transmit via sexual spores depends on the mycovirus and host combination as sexual spore formation eliminates mycoviruses in some instances (Ghabrial and Suzuki, 2008). Horizontal transmission is only possible through anastomosis between genetically similar fungus species. Vegetative incompatibility groups limit the exchange of cytoplasm between unrelated fungi and anastomosis between incompatible fungi results in an apoptotic response. Vegetative incompatibility groups therefore limit the spread of mycoviruses, causing them to have very narrow fungus host ranges (McCabe et al., 1999; Nuss, 2005).

Table 2.1: List of mycovirus species adapted from King et al. (2011) and Pearson et al. (2009) including sequenced or partially sequenced mycoviruses from the NCBI database.

Family	Genus	Genome	Number of species
Alphaflexiviridae	*Botrexvirus*	ss(+)RNA	1
	Sclerodarnavirus		1
	Unassigned		1
Barnaviridae	*Barnavirus*		1
Chrysoviridae	*Chrysovirus*	dsRNA	12
Endornaviridae	*Endornavirus*	dsRNA	6
Gammaflexiviridae	*Mycoflexivirus*	ss(+)RNA	1
	Unassigned		1
Hypoviridae	*Hypovirus*	dsRNA	6
Megabirnaviridae	*Megabirnavirus*	dsRNA	1
Metaviridae	*Metavirus*	ss(+)RNA-RT	5
Narnaviridae	*Mitovirus*	ss(+)RNA	16
	Narnavirus	ss(+)RNA	3
	Unassigned	ss(+)RNA	2
Partitiviridae	*Partitivirus*	dsRNA	39
	Unassigned	dsRNA	1
Pseudoviridae	*Hemivirus*	ss(+)RNA-RT	4
	Pseudovirus	ss(+)RNA-RT	3
Reoviridae	*Mycoreovirus*	dsRNA	5
Totiviridae	*Totivirus*	dsRNA	22
	Victorivirus	dsRNA	10
Unassigned	*Rhizidiovirus*	dsDNA	1
	Unassigned	dsRNA	21
		ss(+)RNA	5
		ssDNA	1

Fungi have some defence mechanisms against mycovirus infection. RNA silencing, or quelling as it is called in the fungus world, is one such defence mechanism which is induced by dsRNA (Segers et al., 2007). The efficiency of RNA silencing differs between fungus species. In some species, such as the model organism *Neurospora crassa,* it appears to prevent mycovirus infection completely (Pearson et al., 2009), while other fungus species lack components that are essential for successful RNA silencing (Nakayashiki et al., 2006).

Mycoviruses can lead to a wide range of phenotypes in fungus hosts, which further result in altered effects of mycovirus-infected fungi on plant hosts. The most studied and economically important phenotype is hypovirulence, where the presence of a mycovirus reduces the virulence of a pathogenic fungus (Nuss, 2005). This is achieved through decreasing mycotoxin production and fungal growth rate, inhibiting sporulation and reducing germination of spores (Pearson et al., 2009). A well-studied example of this is the cryphonectria parasitica hypovirus group that significantly reduce the incidence of chestnut blight in trees infected with the pathogenic fungus, *Cryphonectria parasitica* (Nuss, 2005). These viruses have been used successfully as biological control agents

against chestnut blight in Europe (Dawe and Nuss, 2001). Mycovirus infection can also cause hypervirulence, which is when mycovirus infection increases the virulence of a pathogenic fungus. Ahn and Lee (2001) identified a 6.0 kbp dsRNA fragment in *Nectria radicicola* which increased the virulence, sporulation, laccase activity and pigmentation of the fungus. Isolates that did not contain this dsRNA fragment were not virulent. Mycoviruses have been found to play a role in mutualistic relationships between endophytic fungi and their plant hosts. An example of this is the curvularia thermal tolerance virus (CThTV) which enables both the fungus host and a tropical panic grass to grow at high soil temperatures (Márquez et al., 2007). The presence of the virus therefore benefits both the fungus and the plant, leading to a three-way symbiosis. In spite of these noteworthy phenotypic effects, the majority of mycovirus infections are symptomless and persistent (Ghabrial and Suzuki, 2008).

Mycovirology is a new area of research when compared to plant and animal virology. The majority of mycovirus research is focussed on viruses infecting pathogenic or economically important fungi and very little is known about mycovirus populations in single hosts. It is also believed that a large number of mycoviruses are still to be discovered (Ghabrial and Suzuki, 2008).

Fungi do not naturally contain dsRNA, so the presence of dsRNA in fungi is indicative of mycovirus infection (Zabalgogeazcoa et al., 1998). Double-stranded RNA profiles in fungus isolates are therefore used to determine mycovirus incidence and variability. Mixed mycovirus infections are common occurrences, which complicates the detection of mycoviruses using dsRNA profiles. This is because of difficulties in distinguishing between individual viruses and the different segments of a single viral genome. Another difficulty is the fact that a number of fungi are considered unculturable. Better methods to detect and identify mycoviruses are therefore needed.

2.4 Virus detection techniques

Sensitive, specific and robust virus detection techniques are essential, both for the grapevine industry and for the study of grapevine. The most rudimental approach to pathogen identification is to visually observe the symptoms and to confirm the diagnosis using microscopy (Ward et al., 2004). Although this is a cheap and simple method, it requires specially trained and skilled individuals who are not always able to discriminate between similar symptoms and organisms. It is also not possible to visually diagnose virus infections in plants not yet displaying symptoms or that contain unknown pathogens. Serological and molecular techniques that detect viral proteins or nucleic acid molecules are therefore more sensitive and specific detection techniques. The most commonly used detection techniques in grapevine are ELISA and RT-PCR, and variations of these techniques.

2.4.1 ELISA

Enzyme-linked immunosorbent assay makes use of an antigen-specific antibody that binds to a viral antigen. Different forms of ELISA are used for different applications (Koenig and Paul, 1982). The double antigen sandwich ELISA (DAS-ELISA) is one of the more common serological techniques used for virus detection. In this method, a specific antibody is adsorbed to a polystyrene microtitre plate. After the test sample is added, virus antigens will bind to the adsorbed antibodies and excess components are washed off. A second, enzyme-labelled antibody is then added which also binds to the antigen. When the enzyme substrate is added, a change in colour will be observed in virus positive samples. This technique can also be used to quantify the virus titre by determining the intensity of the colour change (Clark and Adams, 1977; Ward et al., 2004). Although it has a high developmental cost, ELISA is cost-effective for high-throughput diagnostics and is a robust and sensitive technique (O'Donnell, 1999; Ward et al., 2004). The simplicity and effectiveness of ELISA also makes it the preferred diagnostic technique for known plant viral diseases (Ling et al., 2001).

2.4.2 RT-PCR

Reverse-transcription PCR is a technique where genome-specific primers are used to firstly make complimentary DNA (cDNA) from the viral RNA, and then to amplify a unique portion of the genome exponentially so that the product becomes visible after electrophoresis in stained gels. The presence of a certain molecular weight product is, however, not a conclusive result and further verification, like sequencing, is needed to confirm the presence of a virus (Schaad and Frederick, 2002). Reverse-transcription PCR is used for plant virus diagnostics as most plant viruses have RNA genomes. Detection techniques based on PCR are very sensitive and specific and have the potential to be multiplexed in order to detect multiple pathogens in a single reaction (Ward et al., 2004). Limitations of PCR are that accurate sequence information is needed for primer design and primers might not be able to detect all variants of a virus due to the rapid mutation rate of viruses (Coetzee et al., 2010). It is also not possible to accurately quantify the virus titre in a sample using RT-PCR or PCR alone (Ward et al., 2004).However, real-time RT-PCR, in which the fluorescence of intercalating dye is measured during each PCR cycle, can be used to quantify the virus titre in samples. The use of real-time RT-PCR also removes the need for post-reaction analysis of products in order to identify the virus present(Ward et al., 2004).

2.4.3 Novel virus discovery

The above-mentioned virus detection techniques require prior knowledge of the viruses to be detected and are not adequate for detecting unknown viruses. They also have the risk of not detecting unknown variants of known viruses. For example, the LC1/LC2 diagnostic primer pair that was traditionally used to detect known GLRaV-3 variants (Osman and Rowhani, 2006), were found to be unable to detect the newly identified variant group VI of GLRaV-3 (Bester, 2012).

Viruses do not have a conserved gene in common that can be exploited for the discovery of unknown viruses (Edwards and Rohwer, 2005). The high mutation rates of viruses, which includes the possibility of mutations in areas targeted by diagnostic primers, necessitates the development of diagnostic techniques that do not depend on viral sequences being known (Clem et al., 2007). Traditional techniques to detect novel or unknown viruses are laborious and time-consuming. One such approach is to isolate viral particles and observe them under an electron microscope (Kreuze et al., 2009). Another is the use of random primers to amplify unknown viral sequences and then to clone and sequence the resulting PCR products. Cloning can, however, introduce bias as some genomes or genome segments are notoriously difficult to clone (Mardis, 2008b). The fairly recent development of next-generation sequencing has provided a powerful alternative to these techniques. Although NGS is still too expensive to be used as a routine diagnostic tool, its ability to detect novel viruses still benefits the diagnostic environment. After novel viruses have been sequenced using NGS, standard PCR-based diagnostic techniques can be utilised as primers can be designed from the now known viral sequences.

2.5 Metagenomics and next-generation sequencing

2.5.1 Introduction to metagenomics and next-generation sequencing

Metagenomics refers to an approach where the entire community of organisms that inhabit a common environment is sampled and studied (Hugenholtz and Tyson, 2008). Some environments in which this approach has been taken include soil samples (Fierer et al., 2007; Kim et al., 2008), marine water (Bench et al., 2007; Breitbart et al., 2002; Williamson et al., 2008), human faeces (Breitbart et al., 2003; Zhang et al., 2005) and plant tissues (Adams et al., 2009; Al Rwahnih et al., 2009; Coetzee et al., 2010; Muthukumar et al., 2009; Roossinck et al., 2010). Metagenomics has provided scientists with the ability to determine the complexity of microbial populations in single environments (Hugenholtz and Tyson, 2008). The metagenomic workflow entails the extraction of nucleic acid from an environmental sample, shearing the genetic material into smaller fragments and cloning it into a vector, producing a library of clones to be sequenced (Hugenholtz and Tyson, 2008). The first metagenomic studies made use of traditional Sanger sequencing, which limited its use to cloned sequence fragments or fragments from which some sequence information is known that could be targeted with specific sequencing primers (Adams et al., 2009). With the advent of next-generation sequencing in 2005, the need for time-consuming cloning steps was circumvented. This is a drastic improvement on traditional capillary-based Sanger sequencing as it is more time and cost effective and has a significantly higher throughput (Adams et al., 2009; Margulies et al., 2005). Other terms that have been used to refer to NGS are high-throughput, massively parallel and deep sequencing.

For NGS, the entire genome or nucleic acid sample is sheared into small fragments. Adaptors are ligated to the sheared fragments and the fragments are immobilised on either a sequencing bead

or flow cell, depending on the platform used. The fragments are then clonally amplified using the adaptors as primers, producing a cluster of identical fragments on the surface of the bead or flow cell. The fragments are sequenced by pyrosequencing, by synthesis or by ligation (Ware et al., 2012). By using the adaptors to prime sequencing, an unbiased approach is ensured, which is essential if the technique is to be used for virus discovery. The majority of NGS studies detect novel viruses in their data sets, resulting in viral genome sequences being determined before the virus is identified or characterised (Koonin and Dolja, 2012).

Between 60 – 99% of metagenomic sequences generated are not homologous to known viruses and previous virus metagenomic studies have suggested that less than 1% of the existing virus pool has been identified. An NGS approach coupled to metagenomics has the potential to detect all viruses in an environmental sample, irrespective if they are novel or known, or whether their hosts are culturable or unculturable (Mokili et al., 2012). The technique is sensitive enough to identify low titre viruses, to distinguish different strains of a virus species from each other and to provide information regarding genomic variation (Al Rwahnih et al., 2009). The relative abundance of a virus in a community can also be inferred from the number of sequence reads as the reads are roughly proportional to population frequency (Mardis, 2008b).

There are three components to metagenomics, namely sample preparation, high-throughput sequencing and bioinformatic analysis (Mokili et al., 2012). Only a small amount of starting material is needed for sequencing, but it should be of high quality in order to obtain good quality data. By using RNA as starting material and producing cDNA with random primers, the RNA sequences of a large collection of pathogens can potentially be sequenced. This includes RNA viruses, viroids and actively replicating DNA viruses, as well as messenger RNA (mRNA) and ribosomal RNA (rRNA) from phytoplasmas, bacteria and fungi. Using RNA instead of DNA also avoids sequencing the host genome as only mRNA from active host genes will be included (Adams et al., 2009). Reverse transcription and amplification of RNA preparations prior to sequencing can result in coverage being variable along viral genomes (Yang et al., 2012).

With the exception of two newly identified DNA viruses, all known grapevine infecting viruses have either positive or negative ssRNA genomes (Martelli, 2012). When replicating, a complimentary RNA strand is synthesised on the single-stranded viral genome, producing a dsRNA intermediate. As mentioned earlier, the majority of mycoviruses also have dsRNA or ssRNA genomes. The use of dsRNA as starting material for metagenomic and NGS studies in grapevine will therefore serve as an enrichment step for virus-derived sequences.

2.5.2 Next-generation sequencing platforms

There are a number of NGS platforms commercially available, with the three most widely used platforms being the Illumina Genome Analyzer, the ABI SOLiD (Sequencing by Oligo Ligation Detection) and the Roche/454 FLX (Liu et al., 2012; Mardis, 2008a; Zhang et al., 2011). These

three are high-throughput analysers that produce a large amount of data per run. Smaller benchtop sequencers, known as personal genome machines (PGM), are also available that have quicker turnaround times, but generate less data (Liu et al., 2012). These are the MiSeq from Illumina, the Ion Torrent PGM from Life Technologies and the 454 GS Junior from Roche (Loman et al., 2012). The newest platforms, termed third-generation sequencers, do not require PCR amplification as a single DNA molecule is sequenced and the sequencing signal is detected in real time, which shortens the run time. The PacBio RS by Pacific Biosciences, which makes use of a SMRT cell (Single-molecule real-time), is a third-generation sequencer (Liu et al., 2012). Table 2.2 summarises and compares the main sequencing platforms and Figure 2.2 compares the workflows of the three dominant sequencing platforms.

Table 2.2: A comparison of different sequencing platforms adapted from Glenn (2011) and the respective websites[a].

Company	Instrument	Run time	Number of reads/run	Read length (bp)	Output
Conventional sequencing (Sanger)					
Life Technologies	3730xl DNA Analyzer	2 hrs	96	400-900	0.7-2 Mb
High throughput sequencers					
Roche/454	GS FLX Titanium XL+	23 hrs	1 M	1000	700 Mb
	GS FLX Titanium XLR70	10 hrs	1 M	600	450 Mb
Illumina	Genome Analyser IIx	14 d	640 M[b]	2x 150	95 Gb
	HiScanSQ	8.5 d	1.5 B[b]	2x 100	150 Gb
	HiSeq 1000/2000	8.5 d	3 B[b]	2x 100	300 Gb
	HiSeq 1500/2500	11 d[c]	6 B[b]	2x 100	600 Gb
Applied Biosystems	SOLiD 5500xl	7 d	2.8 B[b]	Mate pair: 60+60 Paired end: 75+35 Fragment: 75	180 Gb
Personal Genome Machines					
Roche/454	GS Junior	10 hrs	0.1 M	400	35 Mb
Illumina	MiSeq	39 hrs	34 M[b]	2x 250	8.5 Gb
Life Technologies	Ion torrent 314 chip	2.4 hrs	0.1 M	200	20 Mb
	Ion Torrent 316 chip	3 hrs	1 M	200	200 Mb
	Ion Torrent 318 chip	4.5 hrs	5 M	200	1 Gb
	Ion Proton I	4 hrs	80 M	200	10 Gb
Third-generation sequencer					
Pacific Biosciences	PacBio RS (SMRT)	2 hrs	0.01 M	2200	10 Mb

[a] Websites from which data was obtained:

Applied Biosystems: http://www.appliedbiosystems.com

Life Technologies: https://www.lifetechnologies.com/global/en/home.html

Roche: http://www.454.com

Illumina: http://www.illumina.com/technology/sequencing_technology.ilmn

Pacific Biosciences: http://www.pacificbiosciences.com

[b] Number of paired-end reads.

[c] Run time when using the dual flow cell.
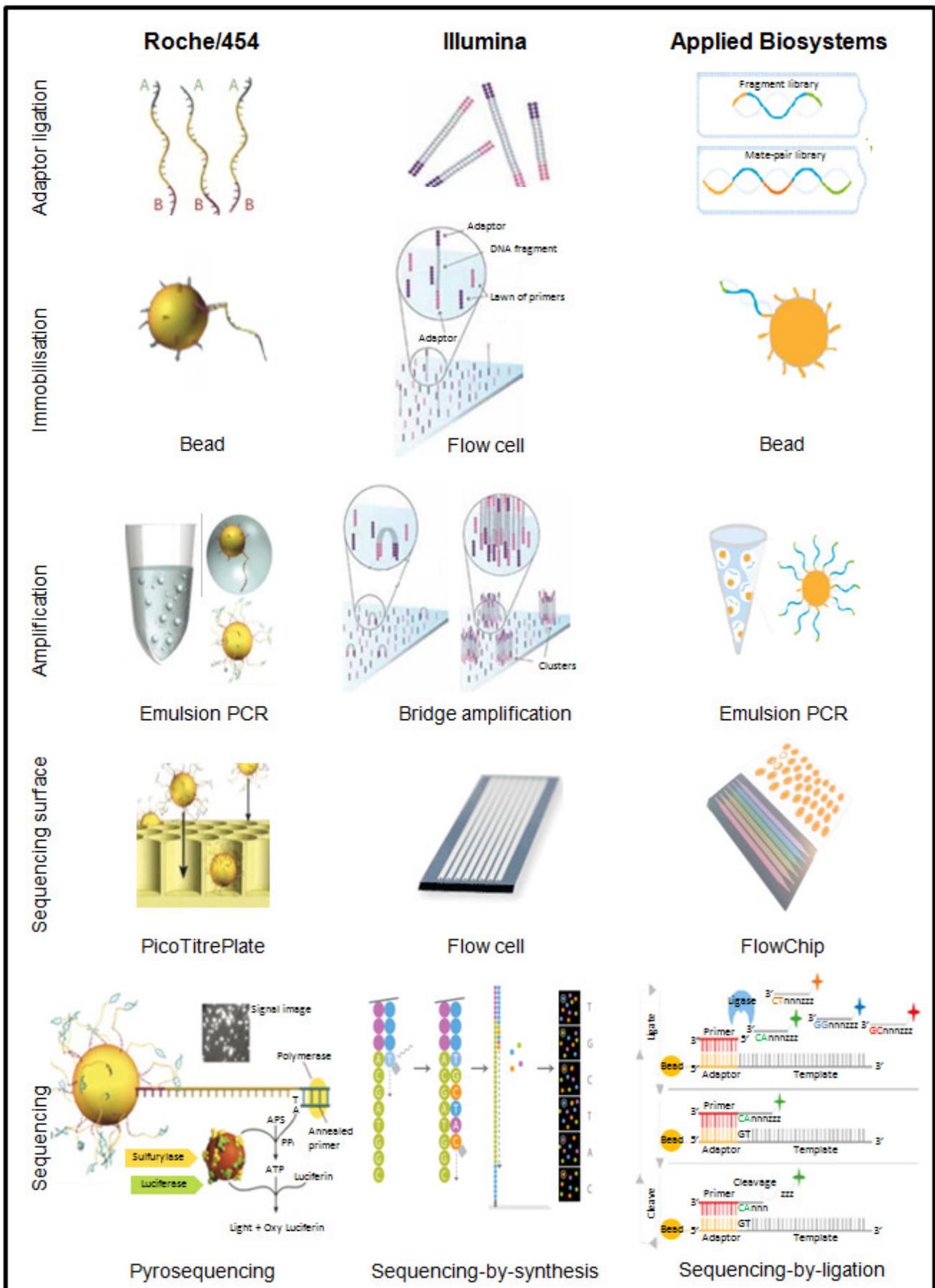
Figure 2.2: Schematic comparison of the three main sequencing chemistries namely Roche/454, Illumina and Applied Biosystems. Images obtained from the respective websites and from Mardis et al. (2008b).

Roche/454 was the first sequencer to be introduced commercially in 2004. The DNA to be sequenced is fragmented and ligated to adapters that are complimentary to adapters on agarose beads, causing them to associate with the beads. An individual single-stranded fragment associates with a bead and an aqueous micelle surrounded by oil is formed containing the bead, single fragment and PCR reactants. Emulsion PCR is used to produce approximately a million clonal copies of each fragment. The micelles are added into a microtitre plate in which each nucleotide solution (dATP, dGTP, dCTP, dTTP) is added sequentially, with an image being taken after each step. As a nucleotide is incorporated, pyrophosphate is released which leads to the production of light. The sequence is determined by measuring the amount of light that is released, which is proportional to the number of nucleotides incorporated. This is known as pyrosequencing (Liu et al., 2012; Mardis, 2008a; Mardis, 2008b; Mokili et al., 2012; Zhang et al., 2011). The Roche system produces the longest reads of all the current NGS platforms. The two latest Roche models on the market are the 454 Genome Sequencer FLX Titanium and the GS Junior.

The Solexa Genome Analyser, now known as the Illumina Genome Analyser, was introduced in 2006 and is the most widely used and adaptable NGS platform (Zhang et al., 2011). Sheared DNA fragments are ligated at both ends to platform-specific adapters and immobilised on a glass flow cell by associating with complimentary oligos in the flow cell. The flow cell contains eight lanes and each lane can contain a separate library. Bridge amplification is performed to produce clusters of one million clones of each single stranded fragment. The four 2',3'-dideoxynucleotide triphosphates (ddNTPs) are added simultaneously during sequencing. They are uniquely labelled and contain 3'-blocks so that a single base is incorporated per sequencing cycle. After each cycle, an image is taken and the fluorescent labels and 3'-blocks are removed. This continues for a user-defined number of cycles. The Illumina sequencing chemistry is known as sequencing-by-synthesis. After the sequencing run, poor quality reads are removed and the sequence of each cluster is determined by a base-calling algorithm (Liu et al., 2012; Mardis, 2008a; Mardis, 2008b; Mokili et al., 2012; Zhang et al., 2011). Currently available Illumina sequencers are the Genome Analyser IIx, the HiScanSQ, HiSeq and MiSeq Personal Sequencer.

The SOLiD platform was released by ABI in 2007. Platform-specific oligo adaptors are ligated to the ends of sheared DNA fragments to be sequenced when preparing either a fragment or a mate-paired library. The adaptor-ligated fragments are coupled to magnetic beads and emulsion PCR is performed to amplify the fragments. The beads are covalently bonded to a glass flow chip and two of these flow chips can be processed simultaneously per sequencing run. Up to eight different libraries can be added to a single flow chip (Mardis, 2008a). Fluorescently labelled semi-degenerate octamers, with two specific bases, are ligated by DNA ligase to the sequencing primer in the first sequencing cycle. Once the octamer is ligated, fluorescence is measured and bases six to eight, which contain the fluorescent label, are cleaved off. With each successive cycle, a new octamer is ligated to the remaining pentamer from the previous cycle. An extra quality check is

included because of the octamer labelling (Liu et al., 2012; Mardis, 2008a; Mardis, 2008b; Zhang et al., 2011). The SOLiD sequencing chemistry is known as sequencing-by-ligation and the latest model is the SOLiD 5500xl.

2.5.3 Bioinformatics associated with next-generation sequencing

Next-generation sequencing has made it possible to produce copious amounts of sequence data quickly and at a low cost. There is, however, a bottleneck when it comes to analysing and interpreting the data (Liu et al., 2012). Next-generation sequencing has a large computational requirement, both for storage of the masses of data produced and for sequence comparisons and analysis (Hugenholtz and Tyson, 2008; Zang et al., 2011). In order to do assemblies millions of comparisons and alignments must be performed in a realistic amount of time. Another bioinformatic difficulty is that the different NGS platforms produce different read lengths and use different formats for the data output. This means that universal bioinformatic tools cannot be used for all platforms (Zhang et al., 2011).

The basic bioinformatic workflow is to firstly trim low quality data from the raw short reads in order to continue with high quality data. Background reads from the host or other unwanted organisms can be filtered out and either *de novo* assemblies or mapping assemblies are performed with the remaining reads (Mokili et al., 2012). The assembled contiguous sequences (contigs) must then be characterised by aligning them to sequence databases like GenBank. It is not a necessity to assemble the reads before they are compared to databases (Edwards and Rohwer, 2005). It is generally best to compare translated query sequences to translated sequences in the database. Although this approach requires more computing power and time, it results in more hits as it is less specific and overcomes the problem of frameshifts due to incorrect base-calling (Edwards and Rohwer, 2005). A significantly large number of metagenomic reads or assembled contigs are not homologous to known sequences in databases. Methods to characterise these sequences are unfortunately still lacking (Mokili et al., 2012).

De novo assemblers, like the Velvet algorithm, divide the raw reads into fragments of a specified length (called the hash length). These fragments, called k-mers, are then used to construct de Bruijn graphs (Zerbino and Birney, 2008). This is the approach with the highest *de novo* assembling capability (Scholz et al., 2012). It uses only the overlaps between the different short reads to construct longer contigs. The k-mer based assembly reduces the assembly time, but requires a large amount of random-access memory, which is proportional both to the genome size being assembled and the amount of data (Scholz et al., 2012).

Mapping assemblers make use of a reference genome against which the short reads are mapped to produce a consensus sequence. The percentage genome coverage and depth of coverage can be obtained by doing a mapping assembly. Two different algorithms are used for mapping assemblies, either the spaced seed indexing or the Burrows-Wheeler Transform (BWT) based

algorithm (Schbath et al., 2012). The seed indexing algorithm is used by MAQ (Mapping and Assembly with Quality) and the BWT algorithm is used by Bowtie. The BWT index has a much smaller memory requirement than seed indexing and the alignment speed is significantly quicker (Trapnell and Salzberg, 2009).

Many of the available bioinformatic tools are open source with available support documentation (Trapnell and Salzberg, 2009). Most tools are Linux based and don not have graphical user interfaces, which limits their use. CLC Main Workbench (CLC Bio) is a commercial and integrated software package that is a user-friendly tool to view and analyse NGS data. It supports all the main NGS platforms and is available for Windows, Mac and Linux computers. Data quality can be determined, reads can be trimmed and both *de novo* and mapping assemblies can be performed. Data can also be easily imported and exported in a number of different formats. Supplementary Table S2 lists a number of available *de novo* and mapping assembly tools as well as quality assessment and visualisation tools.

2.5.4 Next-generation sequencing studies on grapevine diseases

The aetiologies of many grapevine diseases, like shiraz disease and shiraz decline, are still unclear (Coetzee et al., 2010). It is also believed that virus complexes may play a role in grapevine diseases, causing more severe symptoms than single infections (Prosser et al., 2007). This makes NGS the ideal tool for the study of grapevine virus diseases. A number of NGS studies have been done on grapevine since the advent of the technology (Alabi et al., 2012; Al Rwahnih et al., 2009; Al Rwahnih et al., 2011; Coetzee et al., 2010; Giampetruzzi et al., 2012; Pantaleo et al., 2010). Different sampling methods, starting material and NGS technology have been used. The following is an overview of some of these studies.

Al Rwahnih et al. (2009) sequenced both a total nucleic acid extract and a dsRNA extract from bark scrapings of a syrah decline vine and an asymptomatic vine from California, USA. The Roche 454 Genome Sequencer FLX was used to sequence the four samples. Duplicate reads were removed so that only one representative sequence remained, and the remaining reads were directly compared to the GenBank database, without prior assembly. Two viruses, GRSPaV and GRVFV were detected in both the syrah decline and the asymptomatic datasets. Additionally, grapevine leafroll-associated virus 9 (GLRaV-9), hop stunt viroid (HSVd), grapevine yellow speckle viroid (GYSVd) and australian grapevine viroid were also identified in the diseased dataset. A novel virus, GSyV-1, was also identified.

Coetzee et al. (2010) extracted dsRNA from phloem tissue of 44 randomly selected vines in a severely diseased South African vineyard (cv. Merlot). The dsRNA was pooled and sequenced as paired-end reads using the Illumina Genome Analyzer II. The Velvet *de novo* assembler was used to assemble the reads into scaffolds, after which the assembled scaffolds were subjected to BLAST searches. Mapping assemblies were then performed against identified viral reference

genomes using MAQ. Known grapevine infecting viruses were identified, including GLRaV-3, GRSPaV and GVA, as well as GVE, which had never been detected in South African vines prior to this study. This was also the first account of mycoviruses being present in grapevine. Mycoviruses belonging to the families *Chrysoviridae* and *Totiviridae* were detected.

Al Rwahnih et al. (2011) sequenced two dsRNA extracts from bark scrapings of two vines of unknown disease status using the 454 Genome Sequencer FLX platform. The two datasets were pooled and reads were assembled into larger contigs with the *de novo* 454 Newbler Assembler. Besides GRSPaV, 26 putative mycoviruses were identified belonging to the families *Chrysoviridae*, *Hypoviridae*, *Narnaviridae*, *Partitiviridae*, and *Totiviridae*. Fifteen primer sets were designed to target these mycoviral sequences and used to screen the original dsRNA as well as fungus cultures. All 15 of the mycoviruses for which were screened in the dsRNA were detected while only five were detected in fungus isolates. They identified *Penicillium* and *Botrytis* fungus species to be potential hosts for mycoviruses in grapevine.

2.6 Conclusion

A large number of viruses infect grapevine, making metagenomics and NGS the ideal approach to investigate the virus complexity and diversity in grapevine. In many of the previous NGS studies on grapevine, new grapevine infecting viruses were identified, including a number of different mycoviruses. This shows that, although a large number of viruses are known to infect grapevine, many more potentially remain unknown. Next-generation sequencing has revolutionised microbe, and specifically virus, discovery. A vast amount of data is generated through NGS projects, of which a significant proportion remains to be unknown sequences. Sequence databases are, however, expanding at a rapid pace, and with each new viral sequence identified, our understanding of virus complexity and evolution is enhanced.

2.7 References

Adams, I.P., Glover, R.H., Monger, W.A., Mumford, R., Jackeviciene, E., Navalinskiene, M., Samuitiene, M., Boonham, N., 2009. Next-generation sequencing and metagenomic analysis: a universal diagnostic tool in plant virology. Mol. Plant. Pathol. 10(4), 537 – 545.

Ahn, I.P., Lee, Y.H., 2001. A viral double-stranded RNA up-regulates the fungal virulence of *Nectria radicicola*. Mol. Plant Microbe Interact. 14(4), 496 – 507.

Alabi, O.J., Zheng, Y., Jagadeeswaran, G., Sunkar, R., Naidu, R.A., 2012. High-throughput sequence analysis of small RNAs in grapevine (*Vitis vinifera* L.) affected by grapevine leafroll disease. Mol. Plant Pathol. 13(9), 1060 – 1076.

Al Rwahnih, M., Daubert, S., Golino, D., Rowhani, A., 2009. Deep sequencing analysis of RNAs from grapevine showing Syrah decline symptoms reveals a multiple virus infection that includes a novel virus. Virology. 387(2), 395 – 401.

Al Rwahnih, M., Daubert, S., Úrbez-Torres, J.R., Cordero, F., Rowhani, A., 2011. Deep sequencing evidence from single grapevine plants reveals a virome dominated by mycoviruses. Arch. Virol. 156(3), 397 – 403.

Andret-Link, P., Laporte, C., Valat, L., Ritzenthaler, C., Demangeat, G., Vigne, E., Laval, V., Pfeiffer, P., Stussi-Garaud, C., Fuchs, M., 2004. Grapevine fanleaf virus: Still a major threat to the grapevine industry. J. Plant Pathol. 86(3), 183 – 195.

Beerenwinkel, N., Zagordi, O., 2011. Ultra-deep sequencing for the analysis of viral populations. Curr. Opin. Virol. 1(5), 413 – 418.

Bench, S.R., Hanson, T.E., Williamson, K.E., Ghosh, D., Radosovich, M., Wang, K., Eric, Wommack., 2007. Metagenomic characterization of Chesapeake Bay virioplankton. Appl. Environ. Microbiol. 73(23), 7629 – 7641.

Bester, R., 2012. Sequencing and detection of a new strain of grapevine leafroll-associated virus 3 in South Africa. Master of Science thesis. Stellenbosch: University of Stellenbosch.

Breitbart, M., Salamon, P., Andresen, B., Mahaffy, J.M., Segall, A.M., Mead, D., Azam, R., Rowher, F., 2002. Genomic analysis of uncultured marine viral communities. Proc. Natl. Acad. Sci. USA. 99(22), 14250 – 14255.

Breitbart, M., Hewson, I., Felts, B., Mahaffy, J.M., Nulton, J., Salamon, P., Rohwer, F., 2003. Metagenomic analyses of an uncultured viral community from human feces. J. Bacteriol. 185(20), 6220 – 6223.

Buck, K.W., 1986. Fungal Virology – An overview. In: Buck, K.W. (ed.), Fungal Virology, Boca Raton: CRC Press, 1 – 84.

Burger, J., Deist, J. (eds.), 2001. Wingerdbou in Suid Afrika.CTP Printers, Cape Town.

Carstens, R., 1999. Shiraz disease: Symptoms and management. Wynboer – A Technical Guide for Wine Producers. March 1999. http://www.wineland.co.za/index.php?option=com_zine&view=article&id=841:shiraz-disease-symptoms-and-management&q=shiraz+disease

Clark, M.F., Adams, A.N., 1977. Characteristics of the microplate method of enzyme-linked immunosorbent assay for the detection of plant viruses. J. Gen. Virol. 34(3), 475 – 483.

Clem, A.L., Sims, J., Telang, S., Eaton, J.W., Chesney, J., 2007. Virus detection and identification using random multiplex (RT)-PCR with 3'-locked random primers. Virol. J. 4(64), 65 DOI:10.1186/1743-422X-4-65

Coetzee, B., Freeborough, M.J., Maree, H.J., Celton, J.M., Reese, D.J.G., Burger, J.T. 2010. Deep sequencing analysis of viruses infecting grapevines: Virome of a vineyard. Virology. 400(2), 157 – 163.

Dawe, A.L., Nuss, D.L., 2001. Hypoviruses and chestnut blight: Exploiting viruses to understand and modulate fungal pathogenesis. Annu. Rev. Genet. 35(1), 1 – 29.

Du Preez, J., 2005. The construction of an infectious clone of Grapevine virus A (GVA).Master of Science thesis. Stellenbosch: University of Stellenbosch.

Edwards, R.A., Rohwer, F., 2005.Viral metagenomics. Nature Rev. Microbiol. 3(6), 504 – 510.

Fierer, N., Breitbart, M., Nulton, J., Salamon, P., Lozupone, C., Jones, R., Robeson, M., Edwards, R.A., Felts, B., Rayhawk, S., et al., 2007. Metagenomic and small subunit rRNA analyses reveal the genetic diversity of bacteria, archaea, fungi, and viruses in soil. Appl. Environ. Microbiol. 73(21), 7059 – 7066.

Fourie, P.H., Halleen, F., 2004. Occurrence of grapevine trunk disease pathogens in rootstock mother plants in South Africa. Australas. Plant Path. 33(2), 313 – 315.

Ghabrial, S.A., Suzuki, N., 2008. Fungal Viruses. In: Mahy, B.W.J., van Regenmortel, M.H.V. (eds.), Encyclopedia of virology, 3rd edition. Elserivier/Academic Press, London, 284 – 291.

Giampetruzzi, A., Roumi, V., Roberto, R., Malossini, U., Yoshikawa, N., La Notte, P., Terlizzi, F., Credi, R., Saldarelli, P., 2012. A new grapevine virus discovered by deep sequencing of virus- and viroid-derived small RNAs in cv. *Pinot gris*. Virus Res. 163(1), 262 – 268.

Glenn, T.C., 2011. Field guide to next-generation DNA sequencers. Mol. Ecol. Resour. 11(5), 759 – 769.

Goussard, P., Bakker, H., 2006. Characterisation of grapevines visually infected with Shiraz disease associated viruses. Wynboer – A Technical Guide for Wine Producers. December 2006. http://www.wynboer.co.za/recentarticles/200612shiraz.php3

Goszczynski, D.E., 2007a. Investigation of association of viruses of Viti- and Foveavirus genera with Shiraz disease and Shiraz (Syn. Syrah) decline in South Africa. Proceedings of the Syrah Vine Health Symposium, Davis, California, USA. November 7, 2007, 8 – 11.

Goszczynski, D.E., 2007b. Single-strand conformation polymorphism (SSCP), cloning and sequencing reveal a close association between related molecular variants of *Grapevine virus A* and Shiraz disease in South Africa. Plant Pathol. 56(5), 755 – 762.

Goszczynski, D.E., 2010. Rugose wood-associated viruses do not appear to be involved in Shiraz (Syrah) decline in South Africa. Arch. Virol. 155(9), 1463 – 1469.

Goszczynski, D.E., 2011. Aetiology of Syrah (Syn. Shiraz) decline remains a mystery after 18 years of study, and the confusion mounts. Wynboer – Pests & Diseases. November 2011. http://wineland.co.za/archive/index.php?option=com_zine&view=article&id=983:aetiology-of-syrah-syn-shiraz-decline-remains-a-mystery-after-18-years-of-study-and-the-confusion-mounts

Habili, N., 2007. The association of Grapevine Virus A with Australian Shiraz disease and its apparent spread in a commercial vineyard. Proceedings of the 15th Congress of the International Council for the Study of Virus and Virus-like Diseases of the Grapevine (ICVG), Stellenbosch, South Africa. April 3 – 7, 2006, 204 – 204.

Hugenholtz, P., Tyson, G.W., 2008. Microbiology: Metagenomics. Nature. 455(7212), 481 – 483.

Jooste, A.E.C., Pietersen, G., Burger, J.T., 2011. Distribution of grapevine leafroll associated virus-3 variants in South African vineyards. Eur. J. Plant. Pathol. 131(3), 371 – 381.

Kim, K.H., Chang, H.W., Nam, Y.D., Roh, S.W., Kim, M.S., Sung, Y., Jeon, C.O., Oh, H.M., Bae, J.W., 2008. Amplification of uncultured single-stranded DNA viruses from rice paddy soil. Appl. Environ. Microbiol. 74(19), 5975 – 5985.

King, A. M. Q., Lefkowitz, E. M., Adams, J., Carstens, E. B. (eds), 2011. Virus Taxonomy: Ninth Report of the International Committee on Taxonomy of Viruses (ICTV). San Diego, CA: Elsevier Academic.

Koenig, R., Paul, H.L., 1982. Variants of ELISA in plant virus diagnosis. J. Virol. Methods. 5(2), 113 – 125.

Koonin, E.V., Dolja, V.V., 2012. Expanding networks of RNA virus evolution. BMC Biol. 10, 54.

Kortekamp, A., 1997. *Epicoccum nigrum* Link: A biological control agent of *Plasmopara viticola* (Berk. et Curt.) Berl.et De Toni. Vitis. 36(4), 215 – 216.

Kreuze, J.F., Perez, A., Untiveros, M., Quispe, D., Fuentes, S., Barker, I., Simon, R., 2009. Complete viral genome sequence and discovery of novel viruses by deep sequencing of small RNAs: A generic method for diagnosis, discovery and sequencing of viruses. Virology. 388(1), 1 – 7.

Lamprecht, R.L., Maree, H.J., Stephan, D., Burger, J.T., 2012a. Complete nucleotide sequence of a South African isolate of grapevine fanleaf virus. Virus Genes. 45(2), 406 – 410.

Lamprecht, R.L., Maree, H.J., Stephan, D., Wetzel, T., Burger, J.T., 2012b. Molecular characterization of South African isolates of grapevine fanleaf virus and an associated new satellite RNA. Proceedings of the 17th Congress of the International Council for the Study of Virus and Virus-like Diseases of the Grapevine (ICVG), Davis, California, USA. October 7 – 14, 2012, 36 – 37.

Lima, M.F., Alkowni, R., Uyemoto, J.K., Golino, D., Osman, F., Rowhani, A., 2006. Molecular analysis of a California strain of *Rupestris stem pitting-associated virus* isolated from declining Syrah grapevines. Arch. Virol. 151(9), 1889 – 1894.

Ling, K.S., Zhu, H.Y., Petrovic, N., Gonsalves, D., 2001. Comparative effectiveness of ELISA and RT-PCR for detecting grapevine leafroll-associated closterovirus-3 in field samples. Am. J. Enol. Viticult. 52(1), 21 – 27.

Loman, N.J., Misra, R.V., Dallman, T.J., Constantinidou, C., Gharbia, S.E., Wain, J., Pallen, M.J., 2012. Performance comparison of benchtop high-throughput sequencing platforms. Nature Biotechnol. 30(5), 434 – 441.

Liu, L., Li, Y., Li, S., Hu, N., He, Y., Pong, R., Lin, D., Lu, L., Law, M., 2012. Comparison of next-generation sequencing systems. J. Biomed. Biotechnol. 2012, 1 – 11.

Mardis, E.R., 2008a. Next-generation DNA sequencing methods. Annu. Rev. Genomics Hum. Genet. 9, 387 – 402.

Mardis, E.R., 2008b. The impact of next-generation sequencing technology on genetics. Trends Genet. 24(3), 133 – 141.

Margulies, M., Egholm, M., Altman, W.E., Attiya, S., Bader, J.S., Bemben, L.A., Berka, J., Braverman, M.S., Chen, Y.J., Chen, Z., et al., 2005. Genome sequencing in microfabricated high-density picolitre reactors. Nature. 437(7057), 376 – 380.

Márquez, L.M., Redman, R.S., Rodriguez, R.J., Roossinck, M.J., 2007. A virus in a fungus in a plant: Three-way symbiosis required for thermal tolerance. Science. 315(5811), 513 – 515.

Martelli, G.P., 2006. Infectious diseases and certification of grapevines. Proceedings of the Mediterranean Network on Grapevine Closteroviruses. Options Méditerr., Ser. B: Stud. Res. 29, CIHEAM, 47 – 64.

Martelli, G.P., 2012. Grapevine virology highlights: 2010 – 2012. Proceedings of the 17th Congress of the International Council for the Study of Virus and Virus-like Diseases of the Grapevine (ICVG), Davis, California, USA. October 7 – 14, 2012, 13 – 31.

Martelli, G.P., Boudon-Padieu, E., 2006. Directory of infectious diseases of grapevines and viroses and virus-like diseases of the grapevine: Bibliographic report 1998 – 2004. Options Méditerr. Ser. B: Stud. Res. 55, CIHEAM, 279.

McCabe, P.M., Pfeiffer, P., Van Alfen, N.K., 1999. The influence of dsRNA viruses on the biology of plant pathogenic fungi. Trends Microbiol. 7(9), 377 – 381.

Mokili, J.L., Rohwer, F., Dutilh, B.E., 2012. Metagenomics and future perspectives in virus discovery. Curr. Opin. Virol. 2(1), 63 – 77.

Mostert, L., Crous, P.W., Petrini, O., 2000. Endophytic fungi associated with shoots and leaves of *Vitis vinifera*, with specific reference to the *Phomopsis viticola* complex. Sydowia. 52(1), 46 – 58.

Musetti, R., Vecchione, A., Stringher, L., Borselli, S., Zulini, L., Marzani, C., D'Ambrosio, M., Sanità di Toppi, L., Pertot, I., 2006. Inhibition of sporulation and ultrastructural alterations of grapevine downy mildew by the endophytic fungus *Alternaria alternata*. Phytopathology. 96(7), 689 – 698.

Muthukumar, V., Melcher, U., Pierce, M., Wiley, G.B., Roe, B.A., Palmer, M.W., Thapa, V., Ali, A., Ding, T., 2009. Non-cultivated plants of the Tallgrass Prairie Preserve of north-eastern Oklahoma frequently contain virus-like sequences in particulate fractions. Virus Res. 141(2), 169 – 173.

Nakayashiki, H., Kadotani, N., Mayama, S., 2006. Evolution and diversification of RNA silencing proteins in fungi. J. Mol. Evol. 63(1), 127 – 135.

Nuss, D.L. 2005. Hypovirulence: Mycoviruses at the fungal-plant interface. Nat. Rev. Microbiol. 3(8), 632 – 642.

O'Donnell, K.J., 1999. Plant pathogen diagnostics: Present status and future developments. Potato Res. 42(3–4), 437 – 447.

Osman, F., Rowhani, A., 2006. Application of a spotting sample preparation technique for the detection of pathogens in woody plants by RT-PCR and real-time PCR (TaqMan). J. Virol. Methods. 133(2), 130 – 136.

Pantaleo, V., Saldarelli, P., Miozzi, L., Giampetruzzi, A., Gisel, A., Moxon, S., Dalmay, T., Bisztray, G., Burgyan, J., 2010. Deep sequencing analysis of viral short RNAs from an infected Pinot Noir grapevine. Virology. 408(1), 49 – 56.

Pearson, M.N., Beever, R.E., Boine, B., Arthur, K., 2009. Mycoviruses of filamentous fungi and their relevance to plant pathology. Mol. Plant Pathol. 10(1), 115 – 128.

Pietersen, G., 2004. Spread of grapevine leafroll disease in South Africa – A difficult, but not insurmountable problem. Wynboer – A Technical Guide for Wine Producers. June 2004. http://www.wynboer.co.za/recentarticles/0406leaf.php3

Prosser, S.W., Goszczynski, D.E., Meng, B., 2007. Molecular analysis of double-stranded RNAs reveals complex infection of grapevines with multiple viruses. Virus Res. 124(1–2), 151 – 159.

Roossinck, M.J., Sabanadzovic, S., Okada, R., Valverde, R.A., 2011. The remarkable evolutionary history of endornaviruses. J. Gen. Virol. 92(11), 2674 – 2678.

Roossinck, M.J., Saha, P., Wiley, G.B., Quan, J., White, J.D., Lai, H., Chavarria, F., Shen, G., Roe, B.A., 2010. Ecogenomics: Using massively parallel pyrosequencing to understand virus ecology. Mol. Ecol. 19(1), 81-88.

Schaad, N.W., Frederick, R.D., 2002. Real-time PCR and its application for rapid plant disease diagnostics. Can. J. Plant Pathol. 24(3), 250 – 258.

Schbath, S., Martin, V., Zytnicki, M., Fayolle, J., Loux, V., Gibrat, J.F., 2012. Mapping reads on a genomic sequence: An algorithmic overview and a practical comparative analysis. J. Comput. Biol. 19(6), 796 – 813.

Scholz, M.B., Lo, C.C., Chain, P.S.G., 2012. Next generation sequencing and bioinformatic bottlenecks: The current state of metagenomic data analysis. Curr. Opin. Biotech. 23(1), 9 – 15.

Schulz, B., Boyle, C., 2005. The endophytic continuum. Mycol. Res. 109(6), 661 – 686.

Segers, G.C., Zhang, X., Deng, F., Sun, Q., Nuss, D.L., 2007. Evidence that RNA silencing functions as an antiviral defence mechanism in fungi. Proc. Nat. Acad. Sci. USA. 104(31), 12902 – 12906.

Spreeth, N., 2005. Shiraz decline. Wynboer – A Technical Guide for Wine Producers. February 2005. http://www.wineland.co.za/index.php?option=com_zine&view=article&id=454:shiraz-decline&q=shiraz+decline

Tavantzis, S., 2008. Partitiviruses of fungi. In: Mahy, B.W.J., van Regenmortel, M.H.V. (eds.), Encyclopedia of virology, 3rd edition. Elserivier/Academic Press, London. 63 – 68.

Trapnell, C., Salzberg, S.L., 2009. How to map billions of short reads onto genomes. Nature Biotechnol. 27(5), 455 – 457.

Ward, E., Foster, S.J., Fraaije, B.A., McCartney, H.A., 2004. Plant pathogen diagnostics: Immunological and nucleic acid-based approaches. Ann. Appl. Biol. 145(1), 1 – 16.

Ware, J.S., Roberts, A.M., Cook, S.A., 2012. Next generation sequencing for clinical diagnostics and personalised medicine: Implications for the new generation cardiologist. Heart. 98(4), 276 – 281.

Williamson, S.J., Rusch, D.B., Yooseph, S., Halpern, A.L., Heidelberg, K.B., Glass, J.I., Andrews-Pfannkoch, C., Fadrosh, D., Miller, C.S., Sutton, G., et al., 2008. The sorcerer II global ocean sampling expedition: Metagenomic characterization of viruses within aquatic microbial samples. PLoS One. 3(1), e1456 DOI:10.1371/journal.pone.0001456

Yang, X., Charlebois, P., Gnerre, S., Coole, M.G., Lennon, N.J., Levin, J.Z., Qu, J., Ryan, E.M., Zody, M.C., Henn, M.R., 2012. *De novo* assembly of highly diverse viral populations.BMC Genomics. 13, 475.

Zabalgogeazcoa, I., Benito, E.P., Ciudad, A.G., Criado, B.G., Eslava, O.P., 1998. Double-stranded RNA and virus-like particles in the grass endophyte *Epicloë festucae*. Mycol. Res. 102(8), 914 – 918.

Zhang, T., Breitbart, M., Lee, W.H., Run, J.Q., Wei, C.L., Soh, S.W.L., Hibberd, M.L., Liu, E.T., Rohwer, F., Ruan, Y., 2005. RNA viral community in human feces: Prevalence of plant pathogenic viruses. PLoS Biol. 4(1), e3 0108 – 0118 DOI:10.1371/journal.pbio.0040003

Zhang, J., Chiodini, R., Badr, A., Zhang, G., 2011. The impact of next-generation sequencing on genomics. J. Genet. Genomics. 38(3), 95 – 109.

Zerbino, D.R., Birney, E., 2008. Velvet: Algorithms for *de novo* short read assembly using de Bruijn graphs. Genome Res. 18(5), 821 – 829.

Chapter 3: Determination of the mycovirus complexity in five grapevine samples

3.1 Introduction

Grapevine is susceptible to a large number of infectious agents, of which fungi and viruses constitute the most devastating factors, leading to significant decreases in yield and vine vigour (Ferreira et al., 2004; Martelli and Boudon-Padieu, 2006). *Vitis vinifera* does not have natural resistance against these pathogens, which means that they are a constant threat to the grapevine industry (Fisher et al., 2004; Oliver and Fuchs, 2011; Yamamoto et al., 2000). Not all viruses and fungi are pathogenic as a number of cryptic viruses and endophytic fungi have been identified in grapevine (González and Tello, 2011; Martelli 2006; Mostert et al., 2000). Interactions between pathogens have been observed, where a virus only causes disease symptoms when associated with a second virus or where a virus alters the virulence of a fungus host. A holistic view of grapevine infectious agents, including pathogenic and cryptic agents, can therefore provide valuable information regarding the interactions between co-infecting organisms in grapevine.

In order to identify a population of organisms, like viruses, present in a host, a method capable of detecting a large number of organisms in an unbiased manner is required. Microarrays are able to detect a range of viruses simultaneously, but require sequence information of the viral genomes to be known (Boonham et al., 2007). Metagenomics, coupled to NGS, is another technique capable of detecting many viruses simultaneously, but it has the added advantage of not necessitating any prior knowledge of the viruses for detection (Adams et al., 2009). Next-generation sequencing makes use of universal adapters to sequence a library constructed from environmental samples, resulting in a dataset containing sequences of all the viruses present in the sample, including known viruses, novel variants of known viruses and completely novel viruses (Adams et al., 2009).

Recently, NGS was used to identify mycoviral sequences in grapevine for the first time (Al Rwahnih et al., 2011; Coetzee et al., 2010). Mycoviruses belonging to the families *Chrysoviridae*, *Hypoviridae*, *Narnaviridae*, *Partitiviridae*, and *Totiviridae* were identified. Coetzee et al. (2010) found a novel chrysovirus to be the second most prevalent virus in a pool of 44 vines, while Al Rwahnih et al. (2011) found that a single vine was infected with 26 different putative mycovirus groups. This is an indication that, in addition to the 64 known grapevine-infecting viruses (Martelli, 2012), grapevine hosts a significant number of mycoviruses.

Al Rwahnih et al. (2011) designed 15 primer sets from mycoviral contigs and detected all 15 of the putative mycoviral sequences in the original grapevine dsRNA used for NGS. They could, however, only detect five of the mycoviruses in fungus isolates (*Penicillium* and *Botrytis* sp.) using RT-PCR. A possible reason for the low detection rate in fungus isolates is that some fungi and fungus growth media can contain PCR inhibitors which may hinder the detection of mycoviruses in

fungi using PCR based techniques (Borman et al., 2008; Pancher et al., 2012). Another reason is that some of the mycoviruses that they identified could have unculturable hosts and they would thus not be able to detect the mycovirus in a cultured fungus isolates using RT-PCR.

Most mycoviruses do not cause any symptoms in their hosts, making it difficult to phenotypically distinguish between infected and uninfected fungi. Mycoviruses are most commonly identified in fungi through the isolation of dsRNA from fungal mycelium since the majority of mycoviruses have dsRNA genomes (Castillo et al., 2011). The number of mycoviruses with other genome organisations have increased rapidly, from only two ssRNA mycoviruses in the 7th ICTV report (van Regenmortel et al., 2000) to 28 in the 8th ICTV report, including one mycovirus with a dsDNA genome (Fauquet et al., 2005). The use of dsRNA profiles to detect mycovirus infection therefore excludes a small number of mycoviruses that do not have dsRNA or (positive or negative) ssRNA genomes. Viruses infecting unculturable fungi will also remain undetected unless a culture-independent detection technique is used. The unbiased nature of NGS provides a powerful alternative to dsRNA profiles when aiming to determine mycovirus populations.

The many uses of grapes make grapevine one of the most economically important fruit crops globally (Ferreira et al., 2004). This warrants research to be conducted on all grapevine infecting agents, be they pathogenic or cryptic. Although the majority of mycoviruses are cryptic, there are some species that attenuate the virulence of pathogenic fungi (Nuss, 2005). The investigation into mycoviruses infecting grapevine could therefore potentially lead to the identification of hypovirulent mycoviruses that can be exploited for biological control of grapevine pathogenic fungi.

In this chapter we describe the mining of five NGS datasets, generated by the Illumina HiScanSQ platform, for mycoviral contigs. We also describe the use of mycoviral contigs to develop diagnostic RT-PCR primers used to screen grapevine samples and fungus isolates for the detected mycoviruses.

3.2 Materials and Methods

3.2.1 Plant material

Canes were collected from 16 diseased *Vitis vinifera* plants from three different vineyards in the Western Cape, South Africa. Six samples were collected from ARC-Infruitec/Nietvoorbij in Stellenbosch (cv. Shiraz), six from Plaisir De Merle wine estate in Paarl (cv. Shiraz) and four from Fairview wine estate in Paarl (cv. Tempranillo). Ten of the samples displayed typical SD symptoms (sample numbers SD1, SD2, SD4, SD5, SD8, SD9, SD12, SD13, SD16 and SD17) and the remaining six samples displayed GLD symptoms (sample numbers SD3, SD6, SD11, SD14, SD15 and SD18). Sampling was done during autumn (April and May, 2011) when GLD and SD symptoms are most distinct. The basal ends of the canes were used for fungus isolation and the distal ends for dsRNA extraction.

31

3.2.2 Fungus isolation and identification

Canes were surface sterilised by submersion in 70% ethanol (EtOH) for 30 seconds, followed by one minute in 1% NaClO and 30 seconds in 70% EtOH. The bark was removed from the sterilised canes and phloem tissue was plated on 39 g/l potato dextrose agar (PDA) containing 40mg/l streptomycin sulphate. Four Petri dishes were prepared per sample with each dish containing four tissue discs. The Petri dishes were incubated at 25°C for a week before morphologically different fungus colonies were subcultured onto fresh PDA plates. Fungus isolates were identified, as far as possible, to genus level through spore morphology and the frequencies with which each genus was isolated from the different disease states were determined.

3.2.3 Fungal dsRNA extractions

Double-stranded RNA extractions were performed on 24 fungus cultures, isolated from the five plants that were later sequenced (SD3, SD4, SD8, SD11 and SD12),using an adapted double cellulose affinity chromatography protocol (Valverde et al., 1990).Fungal mycelium was homogenised in 112 ml extraction buffer (9 ml 10x STE [1x STE (100 mM NaCl, 50 mM Tris-base, 1 mM Na_2EDTA pH 6.8)], 1.3% SDS, 48 mg Bentonite, 1 ml β-mercaptoethanol, 25 ml STE-saturated phenol, 25 ml chloroform) and shaken for 30 minutes at 150 rpm at room temperature. The homogenate was centrifuged (10 000 xg for 15 minutes at 20°C), the supernatant collected and 3 g cellulose (Sigma), 32 ml 99% EtOH (v/v) and 1x STE was added to a final volume of 200 ml. The mixture was shaken (at 150 rpm for 45 minutes at room temperature) and run through a chromatography column. The column was washed with 150 ml 1x STE/16% EtOH (v/v) and the dsRNA was eluted with 42 ml 1x STE. Half a gram of cellulose (Sigma) and 8 ml 99% EtOH (v/v) was added and the mixture was shaken at 150 rpm for 30 minutes at room temperature. The mixture was run through a second chromatography column, washed again with 50 ml 1x STE/16% EtOH (v/v) and the dsRNA finally eluted with 9 ml 1x STE. The dsRNA was precipitated overnight at -20°C after the addition of 900 µl 3M NaOAc (pH 5.5) and 25 ml 99% EtOH (v/v). The precipitated dsRNA was centrifuged at 16 000 xg for 60 minutes at 4°C, washed with 1 ml 70% EtOH (v/v) and resuspended in 50 µl distilled H_2O. The dsRNA was separated in 1% (w/v) TAE (40 mM Tris base, 0.1% glacial acetic acid, 1 mM Na_2EDTA pH 8.0) agarose gels and stained with ethidium bromide (EtBr). The dsRNA was excised from the gel and recovered using a Zymoclean™ Gel DNA Recovery Kit (Zymo Research).

3.2.4 Identification of fungal dsRNA

Two protocols were followed to identify the dsRNA extracted from fungus isolates. The first was an adaptation of the sequence-independent amplification of viral dsRNA genomes protocol described by Potgieter et al. (2009).The dsRNA isolated from fungi was denatured and a looped adapter was ligated to the ssRNA with T4 RNA Ligase 2 (New England BioLabs). Complimentary DNA was synthesised using avian myeloblastosis virus (AMV) reverse transcriptase (Thermo Scientific) and

a PCR was performed using KAPA Taq DNA polymerase (KAPA Biosystems) and a primer that binds to the looped adaptor. For the second protocol followed, random hexamer primers (Promega) were used in an RT-PCR to synthesise cDNA (AMV reverse transcriptase, Thermo Scientific) from the fungal dsRNA and to amplify the cDNA (KAPA Taq DNA polymerase, KAPA Biosystems) in order to clone and sequence the mycoviruses. Fungus isolates from which reasonable amounts of dsRNA were isolated (as judged by visible fragments on a gel) were classified through PCR amplification using ITS1 and ITS4 primers (White et al., 1990), and cloning and sequencing of the amplicons.

3.2.5 Double-stranded RNA extractions from grapevine and deep sequencing

Double-stranded RNA extractions were performed on 10 g of phloem tissue, ground to a fine powder in liquid nitrogen, from each of the 16 grapevine samples (Valverde et al., 1990) and visualised using 1% (w/v) TAE agarose gels stained with EtBr to determine the quality and quantity of the dsRNA. Five dsRNA samples were selected for sequencing based on quality, namely SD3, SD4, SD8, SD11 and SD12. Five individual RNA sequencing libraries were constructed using the ScriptSeq™ v2 RNA-Seq Library preparation kit (Epicentre) and the libraries were sequenced as single reads with an Illumina HiScanSQ instrument based at the Agricultural Research Council Biotechnology Platform (ARC-BP) in Pretoria.

3.2.6 Sequence analysis

Each of the five short read sequence datasets was processed individually. Sequencing adapters were trimmed off, as well as eight bases from the 5'-end due to unequal distribution of nucleotides in this region and 10 bases from the 3'-end as these bases had a Phred score[†] lower than Q20. Reads shorter than 20 nt were discarded from the sequence datasets, as were reads aligning to the grapevine chloroplast and mitochondria genomes.

The remaining short reads were assembled into contigs using both the *de novo* assembly algorithm, Velvet 1.1.04 (Zerbino and Birney, 2008), and the *de novo* assembly function of CLC Genomics Workbench 4.8 (CLC Bio). For the Velvet assemblies, a range of six hash lengths (21, 23, 25,27, 29 and 31) were used to construct the hash tables with the "velveth" executable. The only parameter set for the "velvetg" executable was a minimum contig length of 200 nt. The contigs produced by each of the six different hash lengths were pooled and larger scaffolds were constructed using the ContigExpress module of Vector NTI Advance™ 10.0 (Invitrogen™). The CLC contigs, Velvet scaffolds and unscaffolded Velvet contigs of each of the five samples were subjected to tBLASTx analysis against the non-redundant database of the NCBI (http://www.ncbi.nlm.nih.gov) through Blast2GO (Conesa et al., 2005). The E-value cut-off was set to 1.0E-6 so that high confidence matches could be reported. Figure 3.1 illustrates the bioinformatic workflow used to assemble and characterise the Illumina short read sequence data.

[†]A Phred score is a quality score that is assigned to each sequencing base call and is logarithmically related to the base-calling error probabilities. A base with a Phred score of Q20 has a 99% accuracy.
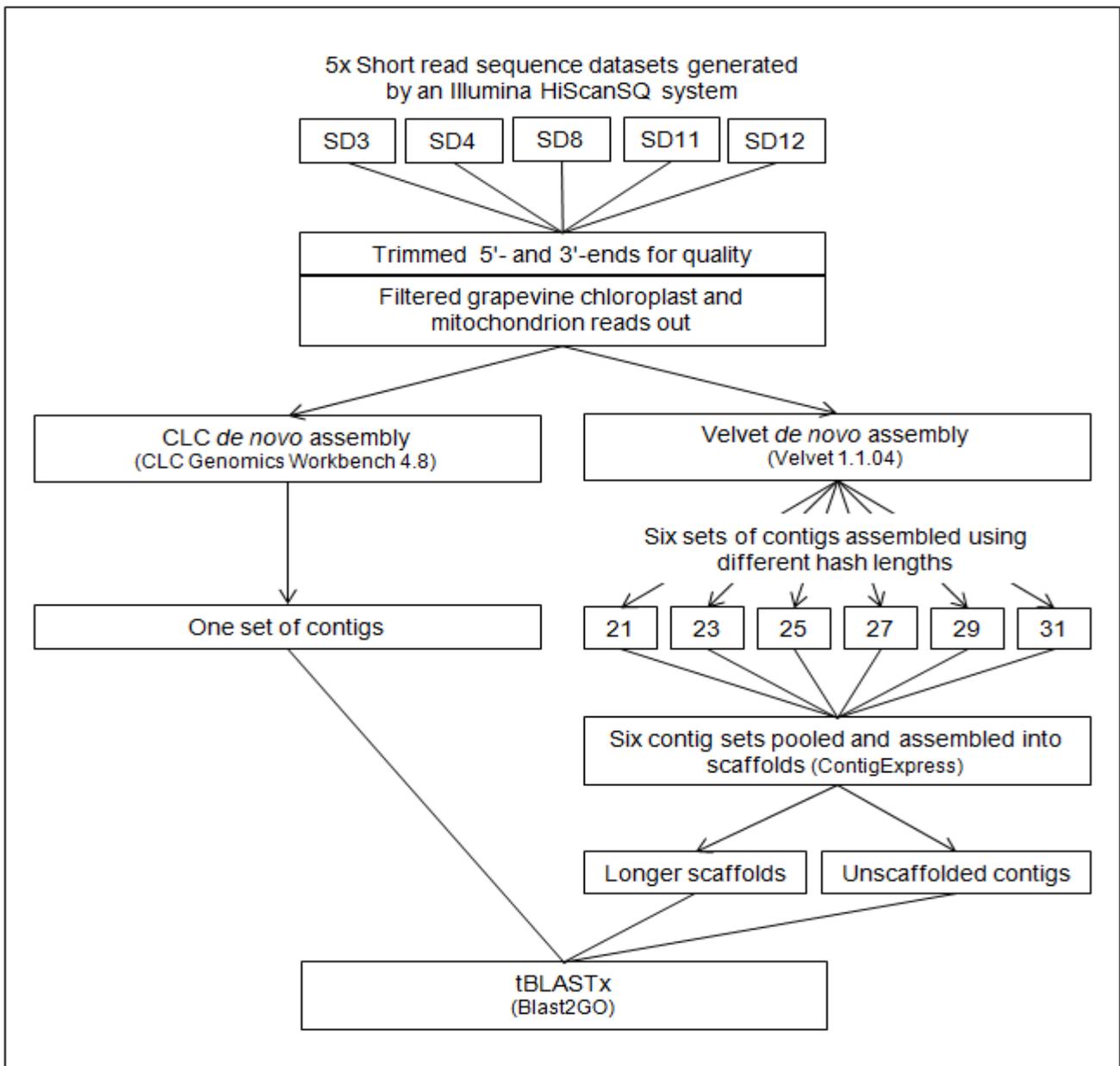
Figure 3.1: Diagram illustrating the bioinformatic workflow followed to assemble and identify the Illumina short read sequence data using CLC Genomics Workbench 4.8 and Velvet 1.1.04 for *de novo* assemblies and Blast2GO to characterise the sequences. This was done for each of the five sequence datasets.

3.2.7 Primer design

All the Velvet and CLC scaffolds and contigs aligning to the same mycoviral sequence were grouped together and consensus scaffolds constructed with the ContigExpress module of Vector NTI Advance™ 10.0 (Invitrogen™). Read mapping was performed against the consensus scaffolds using CLC Genomics Workbench 4.8 (CLC bio) in order to identify scaffolds with adequate coverage for primer design. Scaffolds that were less than 300 nt long or that had an average coverage of less than five were excluded from primer design. Twenty-two primer sets were designed using OligoExplorer 1.1.0 (Gene Link™), to areas on the scaffolds with deep coverage and a low number of mismatches, which was identified using the read-map graphs produced by CLC.

3.2.8 Screening for mycoviruses

Double-stranded RNA from the five grapevine samples was screened with the 22 mycovirus primer sets using RT-PCR. The dsRNA screened was from the same extraction that was sequenced. The dsRNA was converted to cDNA with AMV reverse transcriptase (Thermo Scientific) and KAPA Taq DNA polymerase (KAPA Biosystems) was used for the successive subsequent PCR. Newly extracted total RNA from the five grapevine samples was also screened with RT-PCR. The total RNA was extracted from phloem tissue using a modified cetyltrimethylammonium bromide(CTAB) extraction protocol (White et al., 2008). The GLRaV-3 diagnostic primer set, LR3.HRM4 (Bester et al., 2012), was used as a control for all screenings of plant material as all the plants were known to be infected with GLRaV-3.

Two approaches were used to screen the fungus isolates. Firstly, the dsRNA from fungi that contained dsRNA was screened with the 22 mycovirus primer sets using RT-PCR. Secondly, rapid-direct one tube PCR, in which the 22 mycovirus primer sets were used, was performed on all the fungus cultures(total of 24 cultures) isolated from the five grapevine plants (Osman et al., 2007). In this technique, approximately 150mg of fungal mycelia is homogenised using a tissue grinder in 1 ml of extraction buffer (15 mM Na_2CO_3, 34.9 mM $NaHCO_3$, pH 9.6; 2% PVP-40; 0.2% BSA; 0.05% Tween-20; 1% $Na_2S_2O_5$). Eight microlitres of this extraction solution is added to 50 µl GES buffer (0.1 M Glycine-NaOH, pH 9.0; 50 mM NaCl; 1 mM EDTA; 0.5% Triton X-100)and incubated at 95 °C for 10 minutes, followed by rapid cooling on ice for five minutes. Two microlitres of this crude nucleic acid extract is then used in a PCR (Final reagent concentrations: 1x KAPA Taq Buffer A (KAPA Biosystems); 0.625 µM forward primer; 0.625 µM reverse primer; 0.2 mM dNTPs; 5 mM DTT; 1x cresol/sucrose; 1 U KAPA Taq (KAPA Biosystems)). Fungus-specific primers targeting the conserved internal transcribed spacer region, ITS1 and ITS4 (White et al., 1990), were used as control primer set when fungal material was screened for mycoviruses.

The RT-PCR and PCR products were analysed using 1.5% (w/v) TAE agarose gels stained with EtBr. Amplicons of the expected size were excised from the gel and recovered using a Zymoclean™ Gel DNA Recovery Kit (Zymo Research) according to manufacturer's instructions. Recovered amplicons were cloned using the pGEM®-T Easy vector system (Promega Corporation) and transformed into JM109 chemically competent cells. Plasmid DNA was extracted using a GeneJET™ Plasmid Miniprep kit (Thermo Scientific) and the insert sequenced with either the SP6 or T7 primer at the Central Analytical Facility of Stellenbosch University.

3.3 Results

3.3.1 Fungus isolation and identification

Only GLD and SD symptoms were observed on sampled vines and no fungal disease symptoms were seen. Fungi were isolated on PDA dishes from 16 *Vitis vinifera* canes after surface

sterilisation and the removal of the bark. The isolates were subcultured individually in order to obtain pure cultures and observed under a light microscope to identify the fungus genera based on spore morphology. A total of 119 fungus isolates were obtained representing at least 12 different fungus genera. Fungus isolates which were sterile and that could not be identified by spore morphology were classified according to the mycelium colour (Table 3.1).

Table 3.1: Frequencies of the fungus genera isolated from phloem tissue of grapevine leafroll diseased and shiraz diseased grapevines.

	GLD samples	SD samples
Alternaria	0.50	0.80
Aspergillus	-	0.10
Aureobasidium	0.50	0.40
Chaetomium	-	0.10
Cladosporium	0.17	0.30
Epicoccum	-	0.20
Penicillium	0.17	0.60
Pestalotia	0.17	-
Phialostele-like	-	0.10
Sporobolomyces-like	-	0.50
Stemphylium	0.33	0.40
Trichoderma	-	0.10
Sterile brown mycelium	0.50	0.20
Sterile white mycelium	0.33	0.30
Sterile yellow mycelium	0.50	-

3.3.2 Fungal dsRNA extractions and identification of dsRNA

Double-stranded RNA extractions were performed on 24 fungus cultures isolated from five vines using cellulose affinity chromatography. Four of the isolates contained dsRNA fragments visible in 1% (w/v) TAE agarose gels (Figure 3.2) and these fungi were identified by ITS sequencing. Two of these isolates(8.3-3 and 8.4-1), identified to be *Penicillium digitatum*, contained two dsRNA bands each with sizes similar to the bipartite genome sizes of partitiviruses. The other two isolates (8.3-2 and 8.3-4) were identified as *Aspergillus ustus* and contained three dsRNA segments each, similar to the genome sizes of chrysoviruses. Isolate 8.3-4 also contained the two bands present in the *Penicillium digitatum* isolates. The mycoviral dsRNA isolated from fungus cultures could not be amplified and identified with either of the protocols followed (Potgieter el al. 2009 and random hexamer primer RT-PCR).

3.3.3 Grapevine dsRNA extractions and sequencing

Sixteen dsRNA extractions were performed on grapevine phloem tissue. The dsRNA was visualised using 1% (w/v) TAE agarose gels and five samples (SD3, SD4, SD8, SD11 and SD12)

were selected for sequencing based on dsRNA quantity and quality (Figure 3.3). Sequencing libraries were prepared from each of the samples and the samples were sequenced as single reads with an Illumina HiScanSQ instrument.
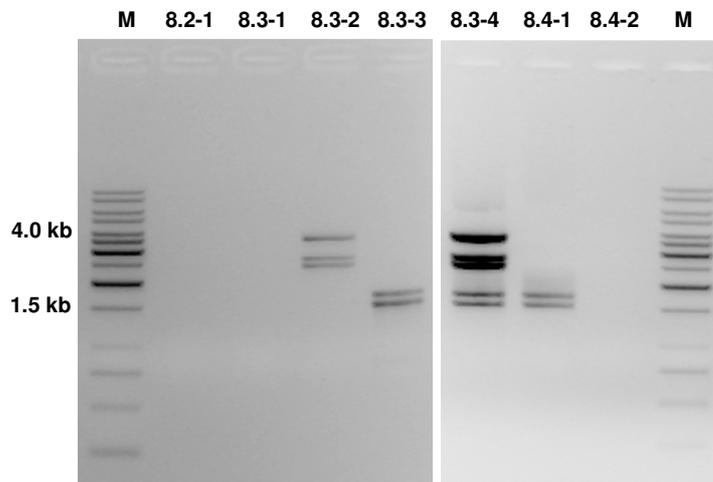
Figure 3.2: Electrophoretic separation of dsRNA extracted from fungus isolates in 1% (w/v) TAE agarose gels stained with ethidium bromide. Lanes marked M contain a GeneRuler™ 1 kb DNA Ladder (Thermo Scientific).

Figure 3.3: Double-stranded RNA extracted from the five grapevine samples that were selected to be sequenced. The samples were analysed using a 1% (w/v) TAE agarose gel stained with ethidium bromide. Lane M contained a GeneRuler™ 1 kb DNA Ladder (Thermo Scientific).
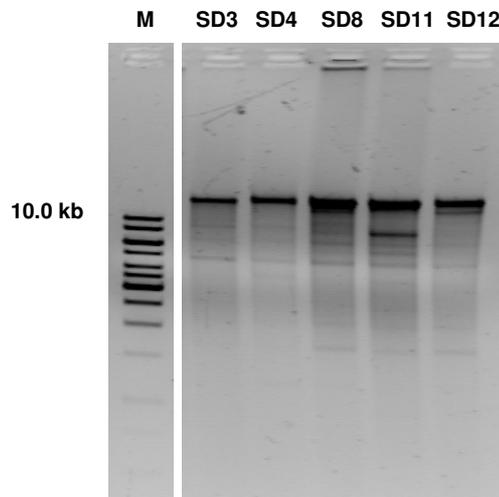
3.3.4 Sequence analysis

The Illumina HiScanSQ sequencer generated 97 nt-long single reads, with 3 540 256, 4 722 908, 3 788 524, 3 270 624 and 3 613 862 reads in each of the five sequence datasets respectively. The sequence datasets were trimmed and filtered for quality and *de novo* assemblies were performed using both the Velvet 1.1.04 algorithm (Zerbino and Birney, 2008) and CLC Genomics Workbench 4.8 (CLC bio). A total of 136 (SD3), 298 (SD4), 293 (SD8), 194 (SD11) and 305 (SD12) Velvet scaffolds and unscaffolded contigs were assembled and 190 (SD3), 376 (SD4), 370 (SD8), 315 (SD11) and 390 (SD12) CLC contigs were assembled for each of the samples. These contigs and

scaffolds were subjected to tBLASTx analysis using Blast2GO in order to be characterised. Table 3.2 shows the categories into which the sequences were classified and the number of Velvet scaffolds and unscaffolded contigs and the number of CLC contigs that were identified to fall into each category. Nine-hundred and sixty-nine sequences remained unidentified after tBLASTx analysis.

Two-hundred and fifty-eight sequences were identified to be mycovirus-like based on their homology with GenBank sequences. Twenty-six different putative mycovirus species, belonging to five different virus families, were identified on amino acid (aa) level(Table 3.3). The mycoviral sequences produced by the NGS data had aa sequence similarities of 40-100% with mycoviral sequences identified in the GenBank database. The majority of mycoviruses were less than 70% similar to known mycovirus species on aa level and are only distantly related to the mycoviruses present in the GenBank database. The highest aa identity was seen with the four genome segments of GaCV (average aa identity of 93%) and GaMV-1 (average aa identity of 87%). These were the only known mycovirus species that were detected in this study.

3.3.5 Primer design

Mycoviral consensus scaffolds were assembled from contigs aligning to each of the identified mycoviruses and used as reference sequences against which the short reads were mapped using CLC Genomics Workbench 4.8 (CLC bio). Scaffolds of deep coverage and few mismatches were identified and 22 primer sets were designed to target 13 of the most prominent mycoviruses identified. Table 3.4 contains a summary of these primer sets, including their sequences, the mycoviruses most closely related to their target viruses and the sizes of the target amplicons.

Table 3.2: Number of sequences *de novo* assembled by Velvet 1.1.04[a] and CLC Genomics Workbench 4.8 that were classified into each category after Blast2GO tBLASTx analysis.

| | GLD | | | | SD | | | | | |
| | SD3 | | SD11 | | SD4 | | SD8 | | SD12 | |
	Velvet	CLC	Velvet	CLC	Velvet	CLC	Velvet	CLC	Velvet	CLC
Vitis vinifera	77	93	111	157	82	96	96	121	112	135
Grapevine viruses[b]	19	29	64	83	58	67	22	41	14	42
Mycoviruses	10	12	4	3	70	60	28	22	24	25
Viroids	-	-	-	-	1	1	2	2	1	3
Other viruses	-	-	-	-	2	3	3	2	5	8
Fungus	2	2	3	4	5	3	6	7	7	4
Other	3	5	1	5	3	3	6	10	6	7
Unidentified	25	49	11	63	79	145	130	165	136	166

[a] A range of six hash lengths were used during Velvet assemblies and the resulting contigs were pooled and assembled into longer scaffolds. Sequences indicated as "Velvet" include the longer scaffolds and unscaffolded contigs.

[b] Known grapevine infecting viruses that were identified include GLRaV-3, GVA and GVE.

Table 3.3: List of mycovirus species most closely related to identified mycoviruses in each sample. The mycovirus species were identified with tBLASTx in Blast2GO.

Virus family	Mycovirus species	GLD		SD		
		SD3	SD11	SD4	SD8	SD12
Chrysoviridae	Aspergillus fumigatus chrysovirus	x		x		
	Cherry chlorotic rusty spot associated chrysovirus	x				
	Cryphonectria nitschkei chrysovirus 1	x		x		x
	Grapevine associated chrysovirus[a]	x		x		
	Helminthosporium victoriae 145S virus			x		
	Penicillium chrysogenum virus	x		x		
	Verticillium dahliae chrysovirus 1	x		x		
Endornaviridae	Chalara endornavirus 1		x		x	x
Narnaviridae	Ophiostoma novo-ulmi mitovirus 4		x			
	Phytophthora infestans RNA virus 4					x
	Thielaviopsis basicola mitovirus		x			
Partitiviridae	Botryotinia fuckeliana partitivirus 1	x				
	Discula destructiva virus 2				x	
	Flammulina velutipes isometric virus					x
	Helicobasidium mompa partitivirus V1-1				x	
	Ophiostoma quercus partitivirus					x
	Sclerotinia sclerotiorum partitivirus S				x	x
Totiviridae	Giardia canis virus				x	
	Ribes virus F				x	
	Tuber aestivum virus 1				x	x
	Ustilago maydis virus				x	x
Unassigned	Curvularia thermal tolerance virus			x		
	Diaporthe ambigua RNA virus 1					x
	Fusarium graminearum dsRNA mycovirus 4			x	x	x
	Grapevine associated mycovirus 1[a]				x	x
	M.lini MldsB3 mycovirus			x	x	x

[a] Known mycoviruses that were detected in this study.

3.3.6 Screening for mycoviruses

The 22 mycovirus primer sets were used to screen the five grapevine dsRNA extracts that were used for sequencing. Three of the primer sets, namely CeEV, Scl1 and Scl2, produced amplicons of the correct size and the amplicons were confirmed by sequencing. Sample SD8 tested positive for genome segments one and two of a partitivirus related to sclerotinia sclerotiorum partitivirus S and for an endornavirus related to CeEV-1. Genome segments one and two of the sclerotinia sclerotiorum partitivirus S-like mycovirus was also detected in sample SD12. Newly extracted total

RNA from the five grapevine plants was also screened using RT-PCR. Only a mycovirus related to CeEV-1 was detected in total RNA extracted from sample SD8.

Table 3.4: Primer sets targeting mycoviruses identified through next-generation sequencing data. These primer sets were used to screen plant and fungal material.

Primer set name	Sequence (5′ - 3')	Mycovirus most similar to target virus	Amplicon size (bp)
AFC1	ATCTCCCACTTTGAATGCTG AGAAGAGAAGTTACCAGACG	Aspergillus fumigatus chrysovirus, segment 1	391
AFC2	ATACGGCTTGACTCAGGCATAC AAGGATTGTGGTTGTGCGAC	Aspergillus fumigatus chrysovirus, segment 2	355
AFC3	CACACCTATCTCAACCAATC CCGCATCCTAAAAAGTCTG	Aspergillus fumigatus chrysovirus, segment 3	406
AFC4	TGAGGTCCGTAGGTACTG TGAGCAAGGGTGAAAGGG	Aspergillus fumigatus chrysovirus, segment 4	348
Bot1	AGTGAAACCTCGGACATC TTGAGCGTAAATGGAAGA	Botryotinia fuckeliana partitivirus 1, segment 1	277
CeEV	AGAAAGGCACTCAAACAC ATCATAACTACTTGCTGT	Chalara elegans endornavirus 1	671
Nits	TCCATTACACGCAGAAAC TGCCACAGGATTATGAGAAG	Cryphonectria nitschkei chrysovirus	292
Curv	TCGGCAATCTCCTCACAAG CACCCACGCTATGACGCA	Curvularia thermal tolerance virus	299
Fgram	GCTCTCGCAGGATGTCAAC TCGTAACCTGGGATTCAAGA	Fusarium graminearum dsRNA mycovirus 4	431
GaCV1	AAGTTCAAGACCGTGTTA GCAGATTGGTAGTGTTAGA	Grapevine associated chrysovirus 1	324
GaCV2	GTTTTGCCCACCTTGACATC TCGGAGTGGAGAAAGGATTG	Grapevine associated chrysovirus 2	1126
GaCV3	TCTGGCTCTTCATCATCAT ACGGTGGCTATCATCTCAG	Grapevine associated chrysovirus 3	1157
GaCV4	TGTCATCAGGTCGCTTTG TTAGGGGGAAGAGTTGATG	Grapevine associated chrysovirus 4	262
GaMV	CACGCTACCTGACGCACT CCAGGAGAAGGAAAACAGG	Grapevine associated mycovirus 1	175
PCV1	TCATCCCCCAGAAGTCAG TTGAACCACGACAGAATG	Penicillium chrysogenum virus, segment 1	249
PCV2	CTCCCAGTAGACAGGCAC AGGAGCAGCGAGAACAAT	Penicillium chrysogenum virus, segment 2	242
PCV4	TATCTGGTCACTCGGTTCG GGTGACTTTGGGATTGGG	Penicillium chrysogenum virus, segment 4	722
Scl1	ACACATCCAGTTCCAAAGTC CTATCATTCGCACTCGTCAA	Sclerotinia sclerotiorum partitivirus S, segment 1	602
Scl2	GTTTCTGGCGTTATCACT GAGAGTTGATTTGTCGTAG	Sclerotinia sclerotiorum partitivirus S, segment 2	604
Tube	ATAGGCGTCTTACAAACCAC GTAAACTCGGTGCTCAATG	Tuber aestivum virus 1	431
Ust	TCACATAAAATCATCTCATC ACCATAGAGAACTTGAACCT	Ustilago maydis virus H1	331
Vert	TATTACGATTGGACCTTCA ACCACCATATCCGAAGAG	Verticillium dahliae chrysovirus, segment 1	300

Double-stranded RNA, extracted from four fungus cultures isolated from grapevine sample SD8, was screened with RT-PCR. None of the primers produced amplicons of the correct size and we were not able to identify the dsRNA fragments present in the isolates. All 24 fungus isolates from the five grapevine plants were screened with rapid-direct one tube PCR. Genome segment one of a sclerotinia sclerotiorum partitivirus S-like mycovirus was detected in an *Alternaria* sp. isolate (Figure 3.4a) and genome segment two of a sclerotinia sclerotiorum partitivirus S-like mycovirus as well as a CeEV-1-like mycovirus was detected in a *Stemphylium* sp. FA-8J isolate (Figure 3.4b). Both the *Alternaria* sp. and the *Stemphylium* sp. were isolated from grapevine sample SD4. A CeEV-1-like mycovirus was also detected in an *Aureobasidium pullulans* isolate (Figure 3.4c) from grapevine sample SD11.
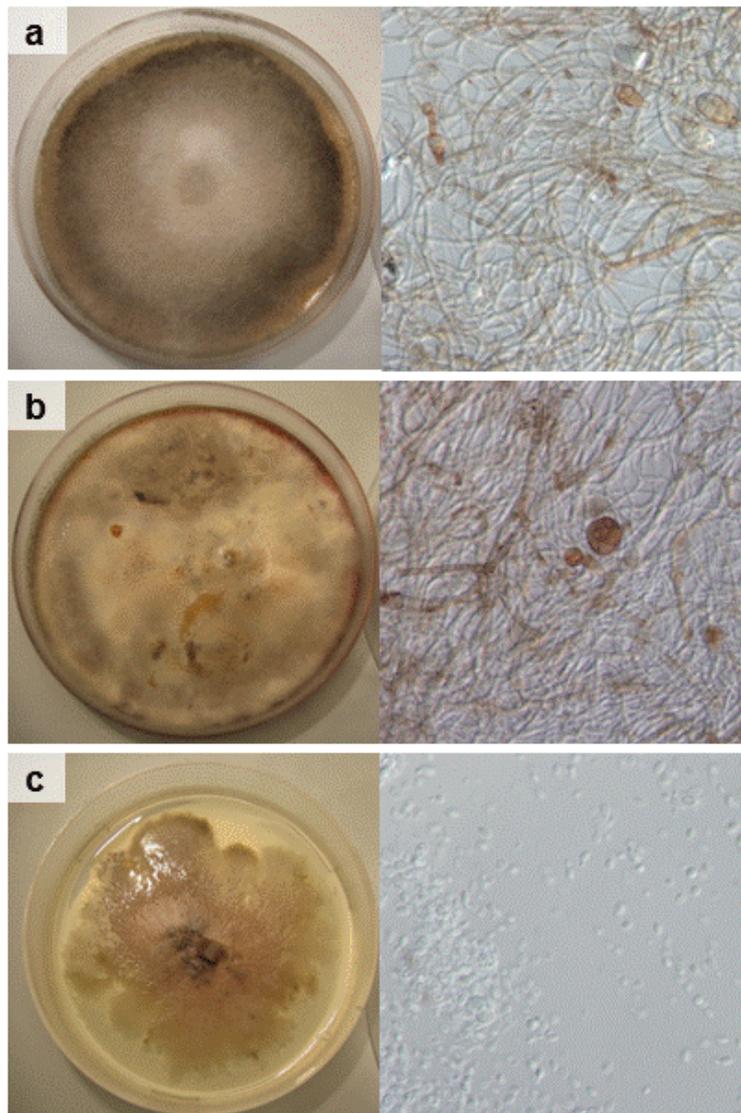
Figure 3.4: Images of fungal mycelium on the left and fungal spore morphology on the right of fungi in which mycoviruses were detected. a) *Alternaria* sp. in which genome segment one of a sclerotinia sclerotiorum partitivirus S-like mycovirus was detected. b) *Stemphylium* sp. FA-8J isolate in which genome segment two of a sclerotinia sclerotiorum partitivirus S-like mycovirus and a chalara elegans endornavirus 1-like mycovirus was detected. c) *Aureobasidium pullulans* isolate in which a chalara elegans endornavirus 1-like mycovirus was detected.

3.4 Discussion and Conclusion

A large number of putative mycoviruses were found to be present in phloem tissue of five GLD and SD *Vitis vinifera* cv. Shiraz vines. Contigs that were assembled from NGS data, generated by an Illumina HiScanSQ sequencing instrument, were subjected to tBLASTx analysis and 258sequences were identified that are related to 26 different known mycoviruses (Table 3.3). These viruses belong to five of the 13 different virus families that contain mycoviruses and as many as 12 putative species were identified in a single plant source. It can be concluded that a large diversity of mycoviruses are associated with grapevine phloem tissue. This finding is consistent with two previous studies that also found a considerable number of mycoviral sequences to be present in grapevine (Al Rwahnih et al., 2011; Coetzee et al., 2010).

The majority of mycoviral sequences identified had less than 70% aa sequence identity to known mycoviruses in the GenBank database. They could only be identified by tBLASTx against the NCBI non-redundant database and can be regarded as putative novel mycoviruses. Many virus NGS studies identify either previously uncharacterised viruses or previously characterised viruses not known to infect the specific host that was sequenced. It was therefore expected that most, if not all of the identified mycoviruses would be novel viruses.

Two known mycoviruses were detected in this study, namely GaCV and GaMV-1. The four genome segments of GaCV (Accessions GU108588, GU108589, GU108596 and GU10858 GU108597) and GaMV-1 (Accession GU108587) had the highest identity to viruses identified in this study, with average identities of 93% to GaCV and 87% to GaMV-1 on aa level. These mycoviruses were identified in a previous NGS study in grapevine that was conducted in California, USA (Al Rwahnih et al., 2011). The GaCV and GaMV-1 viruses detected in this study can be considered as different strains of the viruses identified by Al Rwahnih et al. (2011) since the sequences are significantly divergent from each other. Mycoviruses are believed to have originated at an early stage in the phylogeny of their hosts and to have co-evolved with their hosts (Ghabrial, 1998). This is an indication that these mycoviruses could have been associated with grapevine for many years.

Grapevine samples SD8 and SD12, both SD samples, contained the highest mycovirus diversity, with 11 and 12 different species identified through BLAST analysis in the samples respectively. These were also the only two grapevine samples in which mycoviruses were detected through RT-PCR. Both genome segments of a partitivirus related to sclerotinia sclerotiorum partitivirus S was detected in samples SD8 and in SD12 which supports the likelihood of these samples containing at least one partitivirus.

The fungus hosts for two of the mycoviruses could be determined after 24 fungus cultures, isolated from the five grapevine samples, were screened with the 22 mycovirus primer sets. The sclerotinia sclerotiorum partitivirus S-like mycovirus was detected in an *Alternaria* sp. isolate and in a

Stemphylium sp. (FA-8J isolate), both isolated from sample SD4. The CeEV-1-like mycovirus was detected in the *Stemphylium* sp. (FA-8J) isolated from SD4 and in *Aureobasidium pullulans* isolated from sample SD11. It is interesting to note that the sclerotinia sclerotiorum partitivirus S-like mycovirus was identified in both samples SD8 and SD12 through the NGS data and through RT-PCR of the grapevine dsRNA but not in any of the fungus cultures isolated from these samples. The virus was, however, identified in two fungus cultures isolated from sample SD4, in which the virus was not identified through the NGS data or through RT-PCR of the grapevine dsRNA. A similar observation is made with CeEV-1-like mycovirus. Fungus isolations do not select for all possible fungi present in plant material since fast growing fungi can overgrow the slower growing fungi. Isolations also do not select for unculturable fungi.

Generally, mycoviruses are not transmitted horizontally between different fungi and mycovirus host ranges are restricted to similar vegetative compatibility groups. There are only a few instances where horizontal transmission of mycoviruses between different fungus species, within the same genus (*Cryphonectria*, *Sclerotinia* and *Ophiostoma*),have been observed (Ikeda et al., 2005; Liu et al., 2003). Our findings that the same viral sequences were present in two different fungus genera are therefore significant. More research is needed to confirm this finding as the detected sequence in both cases was only ~600 nt long and not the complete viral genome.

Read mapping was performed against the contigs from which the primers were designed in order to ensure that they are designed in areas of deep coverage. Multiple primers were also designed to target the different genome segments of three chrysoviruses and one partitivirus to increase the chances of detection. In spite of these efforts, only three of the 22 mycovirus primer sets were successful in detecting the target mycoviruses in plant or fungus samples. No positive controls were available and the RT-PCR analysis could therefore not be optimised. The fungus PCRs were conducted on crude nucleic acid extracts which could still contain traces of growth media and fungal metabolites, compounds known to occasionally inhibit DNA polymerase (Pancher et al., 2012). We suspect that the dispersion of fungi and their associated mycoviruses through the vine play a more significant role in why we were unable to detect the viruses identified through NGS in fungi and total RNA isolated from different areas of the same vine. Additional plant material, from the five vines used in this study, could not be obtained to further investigate the possible role of dispersion as the vines were either removed or died due to their disease status.

Four fungus isolates from sample SD8 contained dsRNA that was visible on agarose gels. Mycoviruses with low virus titres would not have been observed on agarose gels and this could be why so few of the fungus isolates contained visible dsRNA. Double-stranded RNA in fungi is indicative of mycovirus infection as fungi do not contain natural dsRNA. Numerous attempts were made to amplify and sequence the dsRNA in order to identify the mycovirus(es), but none were successful. When observed on an agarose gel stained with EtBr, two bands between 1.5-2.0 kbp in length were found to migrate together. They were in the same size range as the bipartite genome

segments of partitiviruses, which are approximately 1.4-2.2 kbp in size (Rong et al., 2002). Three other bands were also observed to migrate together. One was between 3.4-4.0 kbp in size and the other two between 2.5-3.0 kbp. Although only three bands were observed, they were similar in size to the tetrapartite genome of chrysoviruses. The genomes of chrysoviruses are observed as three bands on agarose gels because of the small difference in size between RNA segments three and four (Jiang and Ghabrial, 2004). Mycoviral sequences related to three partitiviruses were identified in the NGS data of sample SD8, namely discula destructiva virus 2, helicobasidium mompa partitivirus V1-1 and sclerotinia sclerotiorum partitivirus S. The bipartite genome observed on the agarose gel is possibly one of these viruses. No viruses belonging to the family *Chrysoviridae* were identified in sample SD8 through the NGS data. This could be because the sample was not sequenced deep enough for the virus to be detected. Another possible reason is that the fungi were not isolated from the same plant material as was used for sequencing, although it was still from the same vine. Fungus dispersion through the vine could therefore influence the complexity of the mycovirus virome.

Many of the assembled contigs, a total of 969, did not have any similarity to sequences in the GenBank database and could not be identified. The dsRNA extractions prior to sequencing serve as an enrichment step for viral nucleic acids and it is therefore possible that additional novel viral sequences could be present among the unidentified contigs. The unidentified contigs should be analysed for the presence of open reading frames and protein domains as an initial step to identify them.

Bioinformatic analysis of NGS data generated in this study identified 26 different putative mycovirus species. Two of these mycoviruses are known to infect grapevine while the remainder are novel mycoviruses. We were able to validate the presence of two of the novel mycoviruses, a sclerotinia sclerotiorum partitivirus S-like and a CeEV-1-like mycovirus, back in grapevine and fungus isolates using RT-PCR and PCR, respectively. We suspect that the uneven distribution of mycoviruses through the plant material could have contributed to the low RT-PCR and PCR detection rates. Another possible reason for the low detection rate is poorly designed mycovirus primers with no positive controls to validate the primer sensitivity. An alternative technique is thus needed to validate the presence of the identified mycoviruses in grapevine. The fact that two known grapevine mycoviruses were detected and an additional two novel mycoviruses were successfully validated in grapevine is an indication that mycoviruses identified in this study are not sequencing or assembly artefacts.

3.5 References

Adams, I.P., Glover, R.H., Monger, W.A., Mumford, R., Jackeviciene, E., Navalinskiene, M., Samuitiene, M., Boonham, N., 2009. Next-generation sequencing and metagenomic analysis: a universal diagnostic tool in plant virology. Mol. Plant. Pathol. 10(4), 537 – 545.

Al Rwahnih, M., Daubert, S., Úrbez-Torres, J.R., Cordero, F., Rowhani, A., 2011. Deep sequencing evidence from single grapevine plants reveals a virome dominated by mycoviruses. Arch. Virol. 156(3), 397 – 403.

Bester, R., Maree, H.J., Burger, J.T., 2012. Real-time RT-PCR high-resolution melting curve analysis and multiplex RT-PCR to detect and differentiate grapevine leafroll-associated virus 3 variant groups I, II, III and VI. Virol. J. 9, 219 DOI:10.1186/1743-422X-9-219

Boonham, N., Tomlinson, J., Mumford, R., 2007. Microarrays for rapid identification of plant viruses. Annu. Rev. Phytopathol. 45, 307 – 328.

Borman, A.M., Linton, C.J., Miles, S.J., Johnson, E.M., 2008. Molecular identification of pathogenic fungi. J. Antimicrob. Chemother. 61(1), i7 – i12.

Castillo, A., Cottet, L., Castro, M., Sepúlveda, F., 2011. Rapid isolation of mycoviral double-stranded RNA from *Botrytis cinerea* and *Saccharomyces cerevisiae*. Virol. J. 8, 38 DOI:10.1186/1743-422X-8-38

Coetzee, B., Freeborough, M.J., Maree, H.J., Celton, J.M., Reese, D.J.G., Burger, J.T., 2010. Deep sequencing analysis of viruses infecting grapevines: Virome of a vineyard. Virology. 400(2), 157 – 163.

Conesa, A., Götz, S., García-Gómez, J.M., Terol, J., Talón, M., Robles, M., 2005. Blast2GO: a universal tool for annotation, visualization and analysis in functional genomics research. Bioinformatics. 21(18), 3674 – 3676.

Fauquet, C.M., Mayo, M.A., Maniloff, J., Desselberger, U., Ball, L.A. (eds.), 2005. Virus Taxonomy: Eighth Report of the International Committee on Taxonomy of Viruses. Elsevier Academic Press, San Diego.

Ferreira, R.B., Monteiro, S.S., Piçarra-Pereira, M.A., Teixeira, A.R., 2004. Engineering grapevine for increased resistance to fungal pathogens without compromising wine stability. Trends Biotechnol. 22(4), 168 – 173.

Fisher, B.M., Salakhutdinov, I., Akkurt, M., Eibach, R., Edwards, K.J., Töpfer, R., Zyprian, E.M., 2004. Quantitative trait locus analysis of fungal disease resistance factors on a molecular map of grapevine. Theor. Appl. Genet. 108(3), 501 – 515.

Ghabrial, S.A., 1998. Origin, adaptation and evolutionary pathways of fungal viruses. Virus Genes. 16(1), 119 – 131.

González, V., Tello, M.L., 2011. The endophytic mycota associated with *Vitis vinifera* in central Spain. Fungal Divers. 47(1), 29 – 42.

Ikeda, K.I., Nakamura, H., Arakawa, M., Masumoto, N., 2005. Dynamics of double-stranded RNA segments in a *Helicobasidium mompa* clone from a tulip tree plantation. FEMS Microbiol. Ecol. 51(2), 293 – 301.

Jiang, D., Ghabrial, S.A., 2004. Molecular characterization of *Penicillium chrysogenum virus*: Reconsideration of the taxonomy of the genus *Chrysovirus*. J. Gen. Virol. 85(7), 2111 – 2121.

Liu, Y.C., Linder-Basso, D., Hillman, B.I., Kaneko, S., Milgroom, M.G., 2003. Evidence for interspecies transmission of viruses in natural populations of filamentous fungi in the genus *Cryphonectria*. Mol. Ecol. 12(6), 1619 – 1628.

Martelli, G.P., 2006. Infectious diseases and certification of grapevines. Proceedings of the Mediterranean Network on Grapevine Closteroviruses. Options Méditerr., Ser. B: Stud. Res. 29, CIHEAM, 47 – 64.

Martelli, G.P., 2012. Grapevine virology highlights: 2010 – 2012. Proceedings of the 17th Congress of the International Council for the Study of Virus and Virus-like Diseases of the Grapevine (ICVG), Davis, California, USA. October 7 – 14, 2012, 13 – 31.

Martelli, G.P., Boudon-Padieu, E., 2006. Directory of infectious diseases of grapevines and viroses and virus-like diseases of the grapevine: Bibliographic report 1998 – 2004. Options Méditerr. Ser. B: Stud. Res. 55, CIHEAM, 279.

Mostert, L., Crous, P.W., Petrini, O., 2000. Endophytic fungi associated with shoots and leaves of *Vitis vinifera*, with specific reference to the *Phomopsis viticola* complex. Sydowia. 52(1), 46 – 58.

Nuss, D.L., 2005. Hypovirulence: Mycoviruses at the fungal-plant interface. Nat. Rev. Microbiol. 3(8), 632 – 642.

Oliver, J.E., Fuchs, M., 2011. Tolerance and resistance to viruses and their vectors in *Vitis* sp.: A virologist's perspective of the literature. Am. J. Enol. Vitic. 62(4), 438 – 451.

Osman, F., Leutenegger, C., Golino, D., Rowhani, A., 2007. Real-time RT-PCR (TaqMan®) assays for the detection of *Grapevine Leafroll associated viruses* 1-5 and 9. J. Virol. Methods. 141(1), 22 – 29.

Pancher, M., Ceol, M., Corneo, P.E., Longa, C.M.O., Yousaf, S., Pertot, I., Campisano, A., 2012. Fungal endophytic communities in grapevines (*Vitis vinifera* L.) respond to crop management. Appl. Environ. Microbiol. 78(12), 4308 – 4317.

Potgieter, A.C., Page, N.A., Liebenberg, J., Wright, I.M., Landt, O., van Dijk, A.A., 2009. Improved strategies for sequence-independent amplification and sequencing of viral double-stranded RNA genomes. J. Gen. Virol. 90(6), 1423 – 1432.

Rong, R., Rao, S., Scott, S.W., Carner, G.R., Tainter, F.H., 2002. Complete sequence of the genome of two dsRNA viruses from *Discula destructiva*. Virus Res. 90(1–2), 217 – 224.

Valverde, R., Nameth, S., Jordan, S., 1990. Analysis of double-stranded RNA for plant virus diagnosis. Plant Dis. 74(3), 255 – 258.

van Regenmortel, M.H.V., Fauquet, C.M., Bishop, D.H.L., Carstens, E.B., Estes, M.K., Lemon, S.M., Maniloff, J., Mayo, M.A., McGeoch, D.J., Pringle, C.R. et al. (eds.), 2000. Virus Taxonomy: Seventh Report of the International Committee on Taxonomy of Viruses. Academic Press, San Diego.

White, T.J., Bruns, T., Lee, S., Taylor, J., 1990. Amplification and direct sequencing of fungal ribosomal RNA genes for phylogenetics. In: Innis, M.A., Gelfand, D.H., Sninsky, J.J., White, T.J. (eds.), PCR protocols: a guide to methods and applications. Academic Press, New York, 315 – 322.

White, E.J., Venter, M., Hitten, N.F., Burger, J.T., 2008. Modified CTAB method improves robustness and versatility: The benchmark for plant RNA extraction. Biotechnol. J. 3(11), 1424 – 1428.

Yamamoto, T., Iketani, H., Ieki, H., Nishizawa, Y., Notsuka, K., Hibi, T., Hayashi, T., Matsuta, N., 2000. Transgenic grapevine plants expressing a rice chitinase with enhanced resistance to fungal pathogens. Plant Cell Rep. 19(7), 639 – 646.

Zerbino, D., Birney, E., 2008. Velvet: Algorithms for *de novo* short read assembly using de Bruijn graphs. Genome Res. 18(5), 821 – 829.

Chapter 4: Confirmation of mycovirus complexity through next-generation sequencing, with a focus on a novel endornavirus

4.1 Introduction

Next-generation sequencing is an indiscriminate approach that can produce high-quality sequence data of a sample, in the form of millions of short reads. The approach is more cost and time effective than traditional capillary sequencing for the amount of data produced (Suzuki et al., 2011). Three commercially available sequencing technologies, each using a different approach to clonal amplification and sequencing, currently dominate the NGS environment. They are Roche/454 (pyrosequencing), Illumina (sequencing-by-synthesis) and ABI (sequencing-by-ligation) (Radford et al., 2012; Zhang et al., 2011). Because of the different chemistries used by each platform, the quality of data produced by the platforms cannot be directly compared (Metzker, 2010).

The Roche/454 platform produces longer reads (up to 1000 nt) (http://454.com/downloads/GSFLXApplicationFlyer_FINALv2.pdf; December, 2012) while the Illumina and ABI (SOLiD) platforms produce short reads (100, 150 or 250 nt and 60 or 75 nt respectively) (http://www.illumina.com/systems/sequencing.ilmn; http://media.invitrogen.com.edgesuite.net/solid/pdf/CO18235-5500-Series-Spec-Sheet-F2.pdf; December, 2012). Alkan et al. (2011) suggests using a multiplatform approach when using NGS and *de novo* assemblies to determine novel genomes. According to Radford et al. (2012), the Illumina and SOLiD platforms have the best overall value when considering cost, accuracy and throughput and when high coverage depths are required.

Bester (2012) used NGS data, generated from 44 pooled vines using an Illumina Genome Analyser II, to assemble a draft genome sequence of a new GLRaV-3 strain. After confirming the genome sequence in a single vine using Sanger sequencing, it was found that the draft sequence had 98.9% nucleotide identity to the correct genome sequence. The genome sequence of a new isolate of bell pepper endornavirus (BPEV) has also been determined using NGS (Sela et al., 2012). A SOLiD 3 instrument was used to sequence small RNA extracted from *Capsicum annuum* and the sequence was assembled after the resulting reads were mapped to a BPEV reference sequence. Next-generation sequencing is thus a time and cost efficient approach to generate highly accurate complete genome sequences of novel viruses.

The family *Endornaviridae* is a new virus family (Carstens and Ball, 2009) that contains large viruses infecting plants, fungi, and oomycetes. The complete genomes of 12 endornaviruses have been sequenced previously; with a thirteenth endornavirus sequenced in this study (Espach et al.,

2012). Endornaviruses have dsRNA genomes which carry a single open reading frame (ORF). A discontinuity is present in the coding strand and some endornaviruses have a poly-C 3′-tail (Fukuhara et al., 2006). Endornaviruses are not encapsidated, they cause persistent infections, and horizontal transmissions are believed not to occur in plants or fungi (Fukuhara et al., 2006;Roossinck et al., 2011).An RNA-dependent RNA polymerase (RdRp) domain is the only conserved domain among all members, while other domains, like methyl-transferase, helicase, and glycosyltransferase, are present in some, but not all members. This suggests that members of the family *Endornaviridae* acquired different functional domains from various sources at different times, making the evolution of endornaviruses complex and dynamic (Roossinck et al., 2011).

Chapter 3 of this thesis describes the identification of 24 putative novel mycovirus species, one of which was an endornavirus, in five grapevines using NGS. Two additional mycoviruses, known to be present in grapevine, were also identified. Only two of the 26 mycoviruses identified could be verified in grapevine and fungus isolates using RT-PCR and PCR respectively. This chapter describes the use of two different NGS platforms, namely Illumina HiScanSQ and SOLiD 5500xl, to confirm the mycovirus complexity in grapevine. The complete genome assembly of a novel endornavirus, with the proposed name grapevine endophyte endornavirus (GEEV), from NGS data is also described.

4.2 Materials and Methods

4.2.1 Illumina sequencing

The RNA sequencing libraries, that were prepared as described in Chapter 3, section 3.2.5 from samples SD3, SD4, SD8, SD11 and SD12 (Chapter 3, section 3.2.1), were subjected to additional NGS using an Illumina HiScanSQ instrument in order to produce more sequence data than what was previously generated.

4.2.2 Applied Biosystems sequencing

Plant material from one GLD sample (SD3) and one SD sample (SD4) (refer to Chapter 3, section 3.2.1) was used. Double-stranded RNA was extracted from 10 g of leaf petioles from sample SD3 and 10 g of phloem tissue from sample SD4 using an adapted cellulose (Sigma) chromatography protocol from Valverde et al. (1990). The dsRNA from the two samples were pooled and an aliquot was analysed by electrophoresis in a 1% (w/v) TAE agarose gel. The SOLiD Total RNA-Seq kit (Life Technologies) was used to prepare a single whole transcriptome library which was sequenced as 75 nt fragments on one lane of a flowchip using a SOLiD 5500xl instrument (ABI) situated at the Central Analytical Facility at Stellenbosch University.

4.2.3 Data analysis

FastQC (http://www.bioinformatics.babraham.ac.uk/projects/fastqc/) was used to identify low quality areas at the 5'- and 3'-terminal ends of the Illumina HiScanSQ short read datasets ("Illumina 2"). The fastx_trimmer and fastx_quality_filter commands of the FASTX-Toolkit (http://hannonlab.cshl.edu/fastx_toolkit/) were used to subsequently trim and filter the sequence datasets. The 5'-area with an imbalanced nucleotide composition and 3'-bases with Phred scores lower than Q20 were trimmed from the reads, and reads with average Phred scores lower than Q20, or in which more than 4% of bases in the read had a quality lower than Q20, were filtered out of the datasets. The P1 and P2 sequencing adaptors were removed from the reads generated by the SOLiD 5500xl instrument ("SOLiD") and the terminal ends of the reads were trimmed for quality using CLC Genomics Workbench 5.5 (CLC bio). Reads shorter than 30 nt were discarded from the five "Illumina 2" sequence datasets and from the "SOLiD" sequence dataset.

The reads from all six of the datasets were mapped against the *Vitis vinifera* genome (19 chromosomes), the *Vitis vinifera* mitochondrion and chloroplast genomes, and the genomes of viruses and viroids known to infect grapevine, using CLC Genomics Workbench 5.5 (CLC bio) (Supplementary Table S3 contains a list of sequences used as references for read mapping). The unmapped reads were used in *de novo* assemblies (CLC Genomics Workbench 5.5, CLC bio) to create contigs which were subjected to BLASTn and tBLASTx analysis against the NCBI non-redundant database using Blast2GO (Conesa et al., 2005).An E-value cut-off of 1.0E-6 was selected to ensure that high confidence matches were reported. The contigs were classified according to the sequences with the highest identity identified by BLAST searches.

4.2.4 The complete genome of a novel endornavirus assembled from next-generation sequence data

The high quality short read sequence dataset from sample SD8 (6 378 798 reads), generated by the Illumina HiScanSQ sequencing platform, was assembled into *de novo* contigs using CLC Genomics Workbench 4.9 (CLC bio). Assembled contigs were subjected to homology searches against the NCBI non-redundant database (BLASTn and tBLASTx) using Blast2GO (Conesa et al., 2005) and a 12 201 nt contig (contig 165) aligning to an endornavirus was identified. The short reads were mapped against contig 165 using the Bowtie 0.12.8 mapping assembler (Langmead et al., 2009) and assembly errors corrected manually. The low quality 5′- and 3′-ends of the contig were confirmed by sequencing amplicons generated by a poly(A)-tailing reaction on dsRNA from sample SD8. In brief, the 3′-ends of both the positive and negative strand were polyadenylated using *Saccharomyces cerevisiae* poly(A) polymerase (Affymetrix) and cDNA was synthesised using AMV reverse transcriptase (Thermo Scientific) and an oligo(dT) primer (5′-TACGATGGCTGCAGT$_{(17)}$-3′) (Meng et al., 2005). The 5′- and 3′-ends were amplified in a PCR using *Ex Taq* DNA polymerase (TaKaRa), two genome-specific primers and the oligo(dT) primer. A

reverse genome-specific primer (5′-TAGCCATCCAAAGCCTCAAG-3′; binding position to complete genome: 577) was used to amplify the 5′-end and a forward genome-specific primer(5′-GCCACAAGTTCCTCACCTG-3′; binding position to complete genome: 12 064) was used to amplify the 3′-end. The amplicons were cloned using a pGEM®-T Easy vector system (Promega) and seven clones were sequenced for each terminal end using Sanger sequencing.

The NCBI tools (http://www.ncbi.nlm.nih.gov/guide/all/#tools) were used to identify the ORF (ORF Finder) and to identify putative protein domains (Conserved Domain Search Service) (Marchler-Bauer et al., 2011) in the completed genome sequence. Read mapping, using the easyrun command of MAQ 0.7.1 (Li et al., 2008), was performed against the completed genome sequence to determine the depth of sequence and genome coverage. A phylogenetic tree was constructed using a multiple alignment of the RdRp aa sequences of the new endornavirus and 15 other endornaviruses for which sequence information are available. The GenBank accession numbers of the viruses used and their abbreviations are given in Supplementary Table S4. The RdRp aa sequence of pineapple mealybug wilt-associated virus 1 (PMWaV-1; family *Closteroviridae*, genus *Ampelovirus*) was used as an outgroup. The unweighted pair-group method with arithmetic mean (UPGMA) method, together with 1000 bootstrap replications was used in CLC Main Workbench 6.5 (CLC bio) to construct the phylogenetic tree.

4.3 Results

4.3.1 Illumina sequencing and pre-processing

RNA sequencing libraries, previously prepared from dsRNA extracted from five diseased grapevine samples, were sequenced as single reads with a length of 100 nt using an Illumina HiScanSQ instrument. The analysis generated 0.84 (SD3), 1.17 (SD4), 1.02 (SD8), 0.68 (SD11) and 1.00 (SD12) gigabases of data. FastQC (http://www.bioinformatics.babraham.ac.uk/projects/fastqc/) was used to view the composition and quality of the generated reads and to identify low quality areas to be trimmed. The fastx_trimmer command of FASTX-toolkit (http://hannonlab.cshl.edu/fastx_toolkit/) was used to trim low quality areas of the reads (Table 4.1) and the fastx_quality_filter was used to filter low quality reads out using a minimum quality score of Q20 (-q 20) and a minimum percentage of 96% of bases in a read to have a quality of Q20 or higher (-p 96) (Figure 4.1). After the sequence datasets were trimmed and filtered for quality, 56.3-66.9% of the reads in the original sequence datasets remained (Table 4.1).
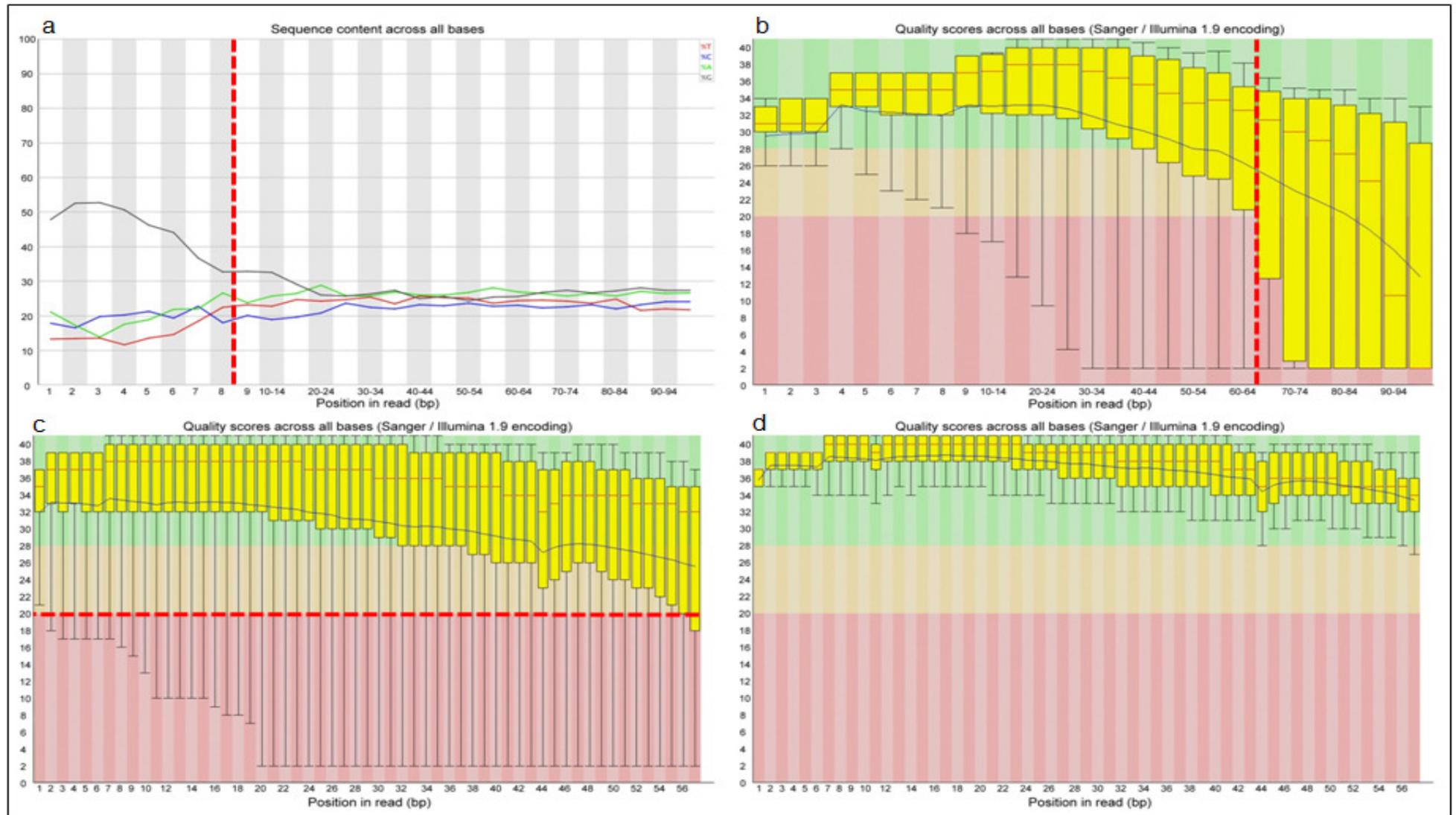
Figure 4.1: Graphs generated by FastQC depicting the quality of the SD8 sequence dataset. a) The percentage nucleotide composition per base. This was used to determine the 5′-end trim position. b) The average quality scores (Phred scores) at each nucleotide position. This was used to determine the 3′-end trim position. c) After the trimming, a minimum Phred score of Q20 was selected for filtering. d) The quality of the data after trimming and filtering. The red lines indicate the selected trimming and filtering thresholds.

4.3.2 Applied Biosystems sequencing and pre-processing

A pooled dsRNA sample, extracted from plants SD3 and SD4 (Figure 4.2), was sequenced as 75 nt fragments using a SOLiD 5500xl instrument. The analysis generated 142 496 471 reads, which translates to 10.69 gigabases of sequence data. Quality score information generated by the sequencing platform was used to identify 5′- and 3′-terminal positions to be trimmed and low quality reads were filtered out using CLC Genomics Workbench 5.5 (CLC bio). Ninety-seven percent of the reads were of high enough quality to be used in further analyses (Table 4.1).
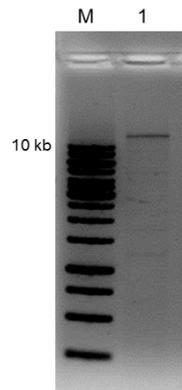
Figure 4.2: Lane 1 contained a pooled dsRNA sample, extracted from grapevine samples SD3 and SD4, separated in a 1% (w/v) TAE agarose gel stained with ethidium bromide. Lane M contained a GeneRuler™ 1 kb DNA Ladder (Thermo Scientific).

Table 4.1: Table indicating the number of sequence reads generated from each sample, read positions that were trimmed, read lengths and reads remaining after quality trimming and filtering.

Dataset[a]	Output (Gb)	Reads from sequencer	Trim	Read lengths	Reads after Q20 filter (%)	Reads after read mapping (%)
SD3	0.842	8 418 002	8-60	53	5 316 533 (63.2)	437 663 (8.2)
SD4	1.174	11 742 166	9-64	56	7 857 739 (66.9)	1 386 212 (17.6)
SD8	1.020	10 195 154	8-64	57	6 378 798 (62.6)	950 970 (14.9)
SD11	0.683	6 826 458	8-60	53	3 842 998 (56.3)	566 249 (14.7)
SD12	1.004	10 042 656	8-64	57	6 372 152 (63.5)	1 075 354 (16.9)
SD3/4	10.687	142 496 471	4-65	61	138 443 358 (97.2)	44 507 247 (32.1)

[a] Datasets SD3, SD4, SD8, SD11 and SD12 were generated by an Illumina HiScanSQ instrument ("Illumina 2") and dataset SD3/4 was generated by an Applied Biosystems SOLiD 5500xl instrument ("SOLiD").

4.3.3 Mapping assemblies

The reads from each of the six sequence datasets were mapped against the *Vitis vinifera* genome, the *Vitis vinifera* mitochondrion and chloroplast genomes, and against viruses and viroids known to infect grapevine. As shown in Table 4.2, the samples host a large number of known grapevine infecting viruses and viroids. Reads that mapped against the reference sequences were removed from the sequence datasets and the remaining reads are further referred to as the unmapped reads. Read mapping removed 91.8% (SD3), 82.4% (SD4), 85.1% (SD8), 85.3% (SD11), 83.1% (SD12) and 67.9% (SD3/4) of the reads from the respective sequence datasets (Table 4.1).

Table 4.2: Read mapping of the high quality reads against the grapevine genome, grapevine mitochondrion and chloroplast genomes and viruses and viroids known to infect grapevine. Mapped reads were removed from the sequence datasets. The number of reads that mapped against each reference and the percentage of the reference sequence that is covered, is shown. Where values are indicated in bold, the relevant virus is considered to be present in the sample.

| | | GLD | | | | SD | | | | | | GLD + SD | |
| | | SD3 | | SD11 | | SD4 | | SD8 | | SD12 | | Pooled SD3/SD4 | |
Reference sequence[a]	Reference length	Total read count	Genome coverage	Total read count	Genome coverage	Total read count	Genome coverage	Total read count	Genome coverage	Total read count	Genome coverage	Total read count	Genome coverage
Vitis vinifera chloroplast	160 928	562 683	97.8	510 501	98.0	126 877	93.9	99 451	91.0	92 565	91.4	9 421 778	99.9
Vitis vinifera mitochondrion	773 279	145 051	20.8	152 270	20.1	31 013	34.9	28 990	38.2	38 875	47.6	1 973 787	93.4
Vitis vinifera genome	303 085 820	2 466 244	0.5	2 300 017	0.6	161 796	0.5	2 219 231	0.6	113 964	0.4	10 376 233	28.2
GLRaV-3 isolate 621	18 498	**1 642 303**	**99.9**	**175 373**	**99.7**	**3 220 042**	**100.0**	40 292	92.0	66 053	76.2	**43 214 228**	**100.0**
GLRaV-3 isolate 623	18 498	30 790	52.0	1 863	20.2	**808 784**	**100.0**	**1 332 609**	**99.8**	**2 271 981**	**99.7**	**6 588 827**	**100.0**
GLRaV-3 isolate GP18	18 498	23 013	43.6	1 440	20.8	**430 553**	**99.7**	**1 244 751**	**99.7**	**1 979 180**	**99.9**	**4 695 020**	**100.0**
GLRaV-3 isolate PL-20	18 433	8 381	30.3	1 455	8.2	31 518	49.9	6 460	34.8	22 837	36.4	**675 870**	**97.8**
GLRaV-3 isolate GH11	18 671	271	25.6	**122 228**	**98.8**	**1 322 716**	**99.7**	2 879	84.9	609	67.8	**13 478 301**	**99.9**
GVA isolate BMo32-1	7 352	12	7.9	23	3.1	**84 410**	**99.9**	**179 712**	**99.9**	26 106	73.1	**797 864**	**100.0**
GVA isolate GTG11-1	7 351	-	-	22	1.4	83	5.7	1 522	16.5	7 205	21.7	931	72.6
GVA isolate GTR1-1	7 360	25	14.6	**11 194**	**97.9**	412	71.7	**259 960**	**99.9**	**418 908**	**100.0**	1 126	62.0
GVA isolate GTR1-2	7 351	1	0.7	287	4.4	479	19.9	4 051	24.8	29 819	47.4	20 707	62.5
GVA isolate P163-M5	7 471	-	-	2	1.1	3 867	37.6	4 933	50.1	**212 703**	**97.5**	19 288	84.5
GVA isolate PA3	7 351	-	-	17	1.6	71	6.7	1 788	16.3	8 119	25.9	1 484	59.6
GVB isolate 94_971	7 599	-	-	1	0.4	2	1.1	-	-	6 425	1.7	185	38.7
GVE isolate SA94	7 568	11	7.7	-	-	**247 636**	**100.0**	492	67.6	7	4.6	**2 428 794**	**100.0**
GRSPaV-1	8 744	1	0.6	-	-	5	1.9	46	15.7	5	1.6	189	36.8
GFLV RNA 1	7 342	-	-	1	0.7	-	-	-	-	-	-	124	37.5
GFLV RNA 2	3 774	1	0.9	-	-	-	-	-	-	-	-	61	35.6
GFLV satellite RNA	1 114	-	-	-	-	-	-	-	-	-	-	19	27.6
HSVd isolate H1	297	8	64.0	5	49.8	**381**	**98.7**	**146**	**98.3**	363	99.7	**72 358**	**100.0**
HSVd isolate HSVd 2.7	297	13	63.3	4	64.0	**660**	**100.0**	**286**	**96.0**	624	99.0	**72 677**	**100.0**
GYSVd-1	366	27	79.0	14	71.3	**119**	**97.3**	25	47.0	76	89.9	**36 769**	**100.0**
GYSVd-2	363	**35**	**100.0**	25	93.4	102	100.0	131	100.0	181	100.0	59 482	100.0
GVd clone Cari-1	368	-	-	4	43.8	-	-	24	64.4	64	66.6	2	13.0
GVd clone Cari-2	364	-	-	3	31.6	1	11.3	**39**	**90.4**	**116**	**97.5**	2	15.1
GVd clone Syrah	361	-	-	-	-	-	-	10	33.2	13	49.0	5	36.8

[a] GLRaV-3 Grapevine leafroll-associated virus 3; GVA Grapevine virus A; GVB Grapevine virus B; GVE Grapevine virus E; GRSPaV-1 Grapevine rupestris stem pitting-associated virus 1; GFLV Grapevine fanleaf virus; HSVd Hop stunt viroid; GYSVd Grapevine yellow speckle viroid; GVd Grapevine viroid.

4.3.4 De novo contig assemblies and analyses

The unmapped short reads were assembled into contigs with the *de novo* assembly function of CLC Genomics Workbench 5.5 (CLC bio). The reads were assembled into 76 (SD3), 305 (SD4), 273 (SD8), 99 (SD11), 216 (SD12) and 1 499 (SD3/4) contigs. The characteristics of the *de novo* assemblies and the contigs that were assembled are summarised in Table 4.3. Only a small percentage of the original sequence datasets were assembled into contigs. The assembled contigs were subjected to BLASTn and tBLASTx analyses against the NCBI non-redundant database and classified according to the database sequences to which they had the highest identity (Table 4.4). A large number of contigs remained unidentified after BLASTn and tBLASTx analyses (1 889 and 1431 contigs respectively). The percentage of unmapped reads that could not be assembled into contigs or that were assembled into contigs that could not be identified were 5.9% (SD3), 7.6% (SD4), 11.8% (SD8), 13.9% (SD11) and 5.9% (SD12) for the "Illumina 2" sequence dataset and 30.9% of the "SOLiD" sequence dataset.

Table 4.3: Characteristics of the *de novo* assemblies and the contigs that were assembled.

Sample	Contigs assembled	Reads assembled (%)	N50[a]	Max contig length (average)	Total contig length	% of original dataset
SD3	76	126 824 (29.0)	386	3 948 (414)	31 487	1.5
SD4	305	880 292 (63.5)	444	5 386 (441)	134 448	7.5
SD8	273	260 495 (27.4)	509	12 048 (461)	125 757	2.6
SD11	99	531 122 (93.8)	375	3 402 (397)	39 326	7.8
SD12	216	233 814 (21.7)	740	12 167 (546)	117 837	2.3
SD3/4	1 499	10 133 248 (22.8)	327	3 674 (352)	528 034	7.1

[a] N50 is a weighted median contig length in a set of sequences. It is a statistical measure used to indicate the contig length (N50) at which 50% of all bases in the sequence set are incorporated into contigs with lengths equal to or larger than N50.

Table 4.4: Number of *de novo* assembled contigs that were classified into each category after Blast2GO tBLASTx and BLASTn analyses.

	GLD				SD						Diseased	
	SD3		SD11		SD4		SD8		SD12		SD3/SD4	
	tx	n	tx	n	tx	n	tx	n	tx	n	tx	n
Vitis vinifera	9	9	15	15	17	17	23	22	22	21	89	91
Grapevine viruses	-	-	-	-	10	9	2	2	6	6	1	5
Mycoviruses	21	16	3	1	110	37	43	19	44	15	358	130
Bacteriophage	2	2	2	2	-	1	-	1	-	1	-	-
Other viruses	1	0	11	10	10	8	12	2	8	-	26	5
Fungus	2	1	1	-	7	3	6	4	6	5	52	10
Other	7	7	45	49	21	21	23	20	10	9	12	3
Unidentified	34	41	22	22	130	209	164	203	120	159	961	1255

tx tBLASTx; n BLASTn

4.3.5 Mycovirus hits

More contigs could be classified as mycoviral contigs using tBLASTx than using BLASTn (Table 4.4). After tBLASTx analysis, 0.49% of the "SOLiD" unmapped data and 0.09% (SD3), 1.83%

(SD4), 2.45% (SD8), 0.01% (SD11) and 10.12% (SD12) of the "Illumina 2" unmapped data aligned to mycoviral sequences in the NCBI non-redundant database. Fifty-one putative mycovirus species, belonging to at least seven different virus families, were identified using tBLASTx analysis (Table 4.5). Twenty-two of the putative species have been identified in five of the samples during a previous sequencing run with an Illumina HiScanSQ sequencing instrument ("Illumina 1") (Chapter 3). The "SOLiD" sequence data set, generated from a SD3/SD4 pooled sample, contained the highest diversity of mycoviruses, with 42 putative mycovirus species of which 22 were not identified in the individual SD3 or SD4 "Illumina 2" sequence datasets.

Table 4.5: List of mycovirus species most closely related to identified mycoviruses. The mycovirus species were identified with tBLASTx against the NCBI non-redundant database using Blast2GO. The number of contigs aligning to each virus is indicated with the average amino acid identity indicated in brackets.

Mycovirus family and species	"Illumina 2"					"SOLiD"
	GLD		SD			GLD+SD
	SD3	SD11	SD4	SD8	SD12	SD3/SD4
Chrysoviridae						
Aspergillus fumigatus chrysovirus[a]	2(71)		18(65)			27(66)
Cherry chlorotic rusty spot associated chrysovirus[a]	1(66)					
Cryphonectria nitschkei chrysovirus 1[a]			3(63)		1(73)	5(66)
Grapevine associated chrysovirus[a]	5(89)		16(87)	1(58)	5(83)	47(89)
Helminthosporium victoriae 145S virus[a]	1(51)		8(76)			10(77)
Magnaporthe oryzae chrysovirus 1			1(71)			3(62)
Penicillium chrysogenum virus[a]			15(67)			21(69)
Raphanus sativas chrysovirus 1	1(74)					
Uncultured chrysovirus			6(76)	2(66)		24(73)
Verticillium dahliae chrysovirus 1[a]	1(53)		7(64)		1(54)	4(63)
Endornaviridae						
Chalara endornavirus 1[a]		2(58)	2(64)	1(41)	1(41)	4(55)
Gammaflexiviridae						
Grapevine associated gammaflexiviridae 1						1(66)
Hypoviridae						
Valsa ceratosperma hypovirus 1						1(61)
Narnaviridae						
Ophiostoma mitovirus 1a	1(62)					
Ophiostoma mitovirus 1b	1(69)					
Ophiostoma novo-ulmi mitovirus 4[a]		1(84)				
Phytophthora infestans RNA virus 3						1(58)
Phytophthora infestans RNA virus 4[a]					2(72)	
Thielaviopsis basicola mitovirus[a]	1(77)					
Partitiviridae						
Amasya cherry disease associated partitivirus						2(65)
Aspergillus fumigatus partitivirus 1						1(91)
Botryotinia fuckeliana partitivirus 1[a]	2(89)					1(94)
Cherry chlorotic rusty spot associated partitivirus						2(84)
Flammulina velutipes isometric virus[a]				1(60)		1(65)
Fusarium poae virus 1			1(68)			1(84)
Helicobasidium mompa partitivirus V1 1[a]						1(68)
Ophiostoma partitivirus 1	1(93)					
Pleurotus ostreatus virus 1				1(57)		2(69)
Rhizoctonia solani virus						1(68)
Rosellinia necatrix partitivirus 1						6(76)
Rosellinia necatrix partitivirus 4						3(61)
Sclerotinia sclerotiorum partitivirus S[a]				3(67)	2(64)	1(77)
Uncultured partitivirus	4(92)		4(69)	4(82)	3(71)	12(74)

Table 4.5 continued: List of mycovirus species most closely related to identified mycoviruses in each sample. The mycovirus species were identified with tBLASTx in Blast2GO. The number of contigs aligning to each virus is indicated with the average amino acid identity indicated in brackets.

Mycovirus family and species	"Illumina 2"					"SOLiD"
	GLD		SD			GLD+SD
	SD3	SD11	SD4	SD8	SD12	SD3/SD4
Totiviridae						
Aspergillus mycovirus 1816			2(68)			5(74)
Giardia canis virus[a]				2(57)	1(59)	1(55)
Saccharomyces cerevisiae virus La						1(58)
Tolypocladium cylindrosporum virus 2						1(54)
Uncultured totivirus			5(87)	1(92)	2(87)	18(71)
Ustilago maydis virus[a]				1(53)		2(55)
Xanthophyllomyces dendrorhous virus L1A						1(63)
Xanthophyllomyces dendrorhous virus L1b			1(70)	11(70)	7(70)	33(70)
Xanthophyllomyces dendrorhous virus L2						1(54)
Unassigned						
Alternaria alternata virus 1						10(99)
Botrytis porri RNA virus 1			2(65)		1(63)	45(65)
Curvularia thermal tolerance virus[a]					1(62)	4(66)
Diaporthe ambigua RNA virus 1[a]				2(69)	2(59)	
Fusarium graminearum dsRNA mycovirus 2						1(52)
Fusarium graminearum dsRNA mycovirus 4[a]			5(67)	2(69)	7(68)	13(74)
Grapevine associated mycovirus 1[a]			2(88)	6(91)	2(90)	5(78)
M.lini MldsB3 mycovirus[a]			7(61)	2(51)	2(48)	29(57)
Uncultured dsRNA virus			5(77)	3(90)	4(76)	6(80)
Total number of putative mycoviruses	**12**	**2**	**19**	**16**	**17**	**42**
Total number of mycoviral contigs	**21**	**3**	**110**	**43**	**44**	**358**

[a] Mycovirus species that have been detected in a previous Illumina HiScanSQ sequencing run.

4.3.6 The assembly of the complete genome of a novel endornavirus

CLC Genomics Workbench 4.9 (CLC bio) was used to assemble *de novo* contigs from trimmed and filtered reads from the SD8 "Illumina 2" sequence dataset. A contig of 12 201 nt was assembled that had the highest nucleotide identity to CeEV-1 (GenBank accession number GQ494150). The nucleotide identity was however low (44%), which suggested a novel endornavirus. An assembly error, which caused a premature stop codon in the ORF, was manually corrected after mapping the reads against contig 165, using Bowtie 0.12.8 (Langmead et al., 2009) to determine the position of the error (Figure 4.3). The sequence of the 5′- and 3′-ends of the viral genome were confirmed and the exact terminal nucleotides determined by Sanger sequencing the amplicons of the poly(A) tailing reactions. The corrected contig 165 was longer than the complete genome as determined after the Sanger sequencing, with a 23 nt longer 5′-end and a 20 nt longer 3′-end than the complete genome.

The complete genome of the virus, for which we propose the name grapevine endophyte endornavirus (GEEV), is 12 154 bp long and the sequence has been deposited in the GenBank database under accession number JX678977. The genome contains a single, large ORF stretching from nucleotide positions 12-12 095 on the positive strand. The ORF encodes a putative polyprotein of 4 027 aa with a predicted molecular mass of 452.7 kDa. Predicted protein domains

include a viral helicase (superfamily: cl15862) near the middle of the polyprotein and an RdRp (superfamily: cl03049) near the C-terminus (Figure 4.4). A total of 146 438 reads, representing 2.3% of the SD8 sequencing dataset, were mapped against the genome of GEEV using MAQ 0.7.1 (Li et al., 2008). The average depth of sequence was 687 nt and 99.9% of the genome was covered (Figure 4.5). Only 12 bp in the 5′-end and 1 bp in the 3′-end were not covered.
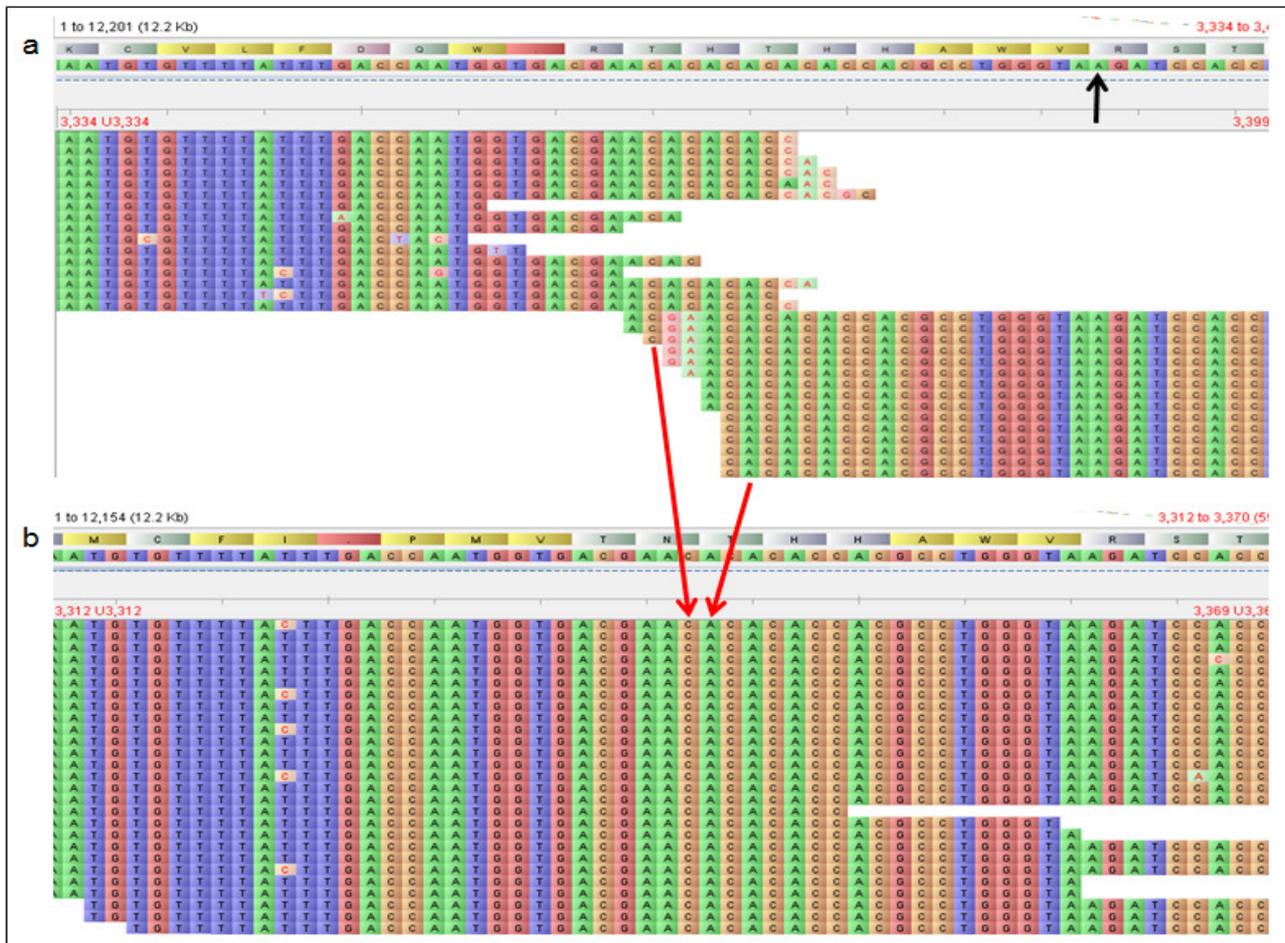
Figure 4.3: The high quality Illumina HiScanSQ sequencing reads from sample SD8 mapped against the uncorrected contig 165 (a) and against the corrected contig 165 (b) using Bowtie 0.12.8 (Langmead et al., 2009). The black arrow indicates the position of a premature stop codon in the uncorrected contig 165. The red arrows indicate an insertion in the AC-repeat region which was removed to correct the draft sequence.
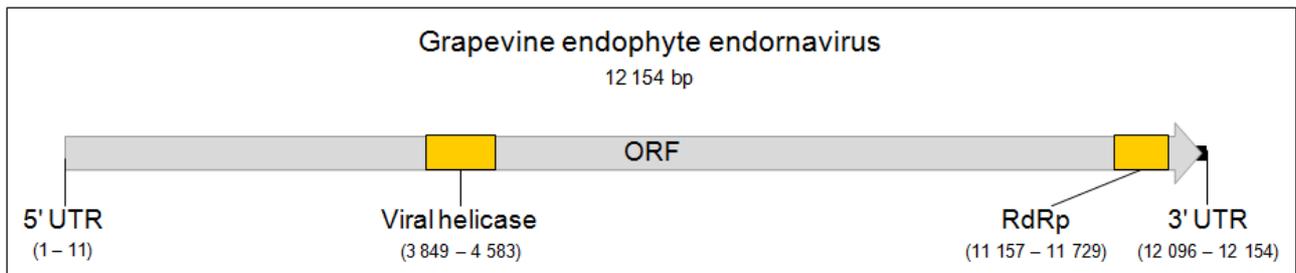
Figure 4.4: Schematic representation of the genome organisation of grapevine endophyte endornavirus (drawn to scale). The nucleotide positions are indicated in brackets underneath the genome features. UTR Untranslated region; ORF Open reading frame; RdRp RNA dependent RNA polymerase.
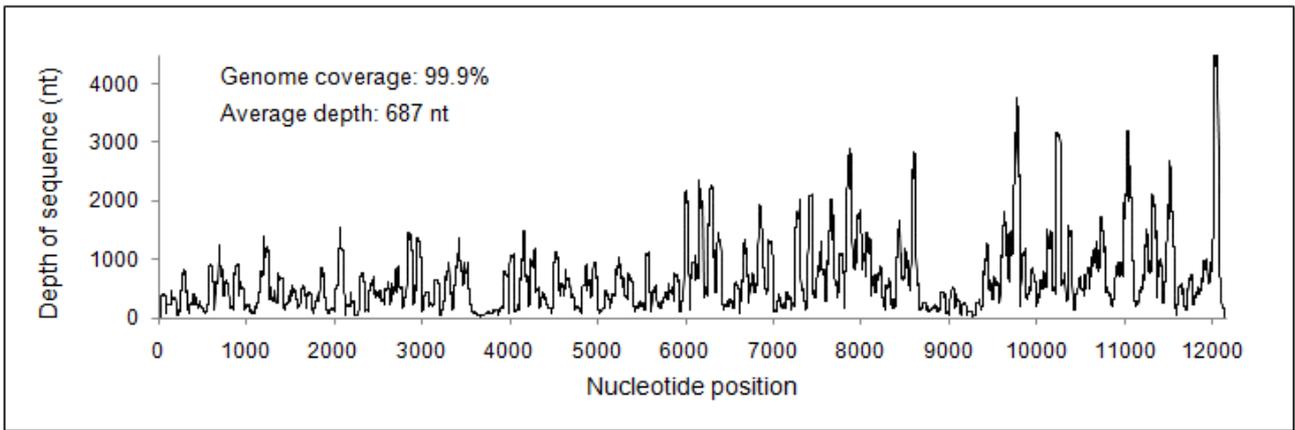
Figure 4.5: Graphical representation of the genome coverage and depth of sequencing of grapevine endophyte endornavirus (GEEV). The short read data of sample SD8 were mapped against the complete genome of GEEV using the MAQ 0.7.1 mapping assembler (Li et al., 2008).
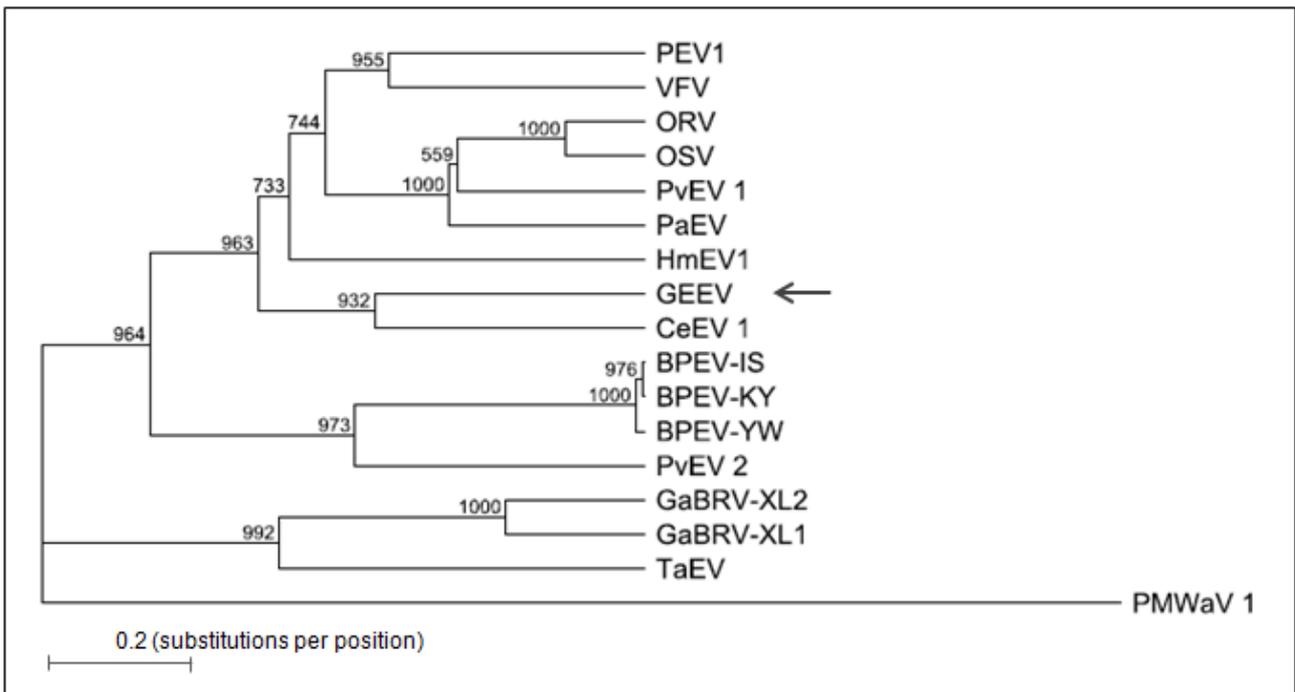
Figure 4.6: Phylogenetic tree of endornavirus RdRp aa sequences constructed with the UPMGA method using 1000 bootstrap replicates. PMWaV-1 was used as an outgroup. Bootstrap values are indicated on the branch nodes. GenBank accession numbers and full virus names are provided in Supplementary Table S4.

The RdRp aa sequences of GEEV and 15 other endornaviruses were used to construct a phylogenetic tree in which GEEV formed a clade with CeEV-1 (Figure 4.6). The topology of the tree does not coincide with the virus hosts as some mycoviruses are more closely related to plant viruses than to other mycoviruses.

4.4 Discussion and Conclusion

Twenty-six putative mycovirus species were bioinformatically identified in five grapevine samples in an earlier Illumina HiScanSQ sequencing run ("Illumina 1") (Described in Chapter 3). Primers designed from mycoviral sequences were able to detect two mycoviruses in grapevine samples

and fungus isolates, namely a CeEV-1-like and a sclerotinia sclerotiorum partitivirus S-like mycovirus. Although most of the mycoviruses could not be verified experimentally, the large amount of mycoviral sequences generated (37% of the viral contigs) suggests that these sequences are not sequencing or assembly artefacts. Two additional sequencing runs ("Illumina 2" and "SOLiD"), using different sequencing chemistries, were therefore employed to validate the mycovirus findings of the earlier sequencing run.

Twenty-two of the 26 mycoviruses initially detected could be detected again based on homology searches (Table 4.5). The viral contigs were classified according to mycoviral sequences in the GenBank database against which they aligned. Since the contigs did not have 100% sequence identity to mycoviruses in the database, and only partial genome segments were represented in the contigs, it is possible that different areas in the genomes of novel mycoviruses could align to different known mycoviruses within a virus family. This could explain why four of the mycoviruses detected in the first sequencing run were not detected in the subsequent sequencing runs.

Twenty-nine mycovirus species not detected in the "Illumina 1" sequence datasets were detected in the subsequent "Illumina 2" and "SOLiD" sequence datasets, including mycoviruses from two new families, namely *Gammaflexiviridae* and *Hypoviridae*. The "Illumina 2" and "SOLiD" sequence datasets had both a greater volume and better quality compared to the "Illumina 1" sequence datasets. The majority of contigs assembled from the "Illumina 1" sequence dataset aligned to the *Vitis vinifera* genome and other viruses known to infect grapevine. Reads aligning to these sequences were removed during analyses of the "Illumina 2" and "SOLiD" sequence datasets to simplify the datasets. Other factors that could have led to the detection of more mycoviruses include a different approach to assemble contigs, upgraded software versions and a larger NCBI database (the analyses were performed approximately a year apart).

The average aa identity to known mycoviruses ranged between 52-99%. Viruses with aa identities lower than 55% represent different genera of the same virus family, 55-65% aa identity represents species of the same genus and aa identity higher than 65% represent different strains of the same species (Fauquet et al., 2005). Six viruses could therefore be identified on family level, 14 on genus level and 31 represented variants of known mycovirus species (Table 4.5).

A large diversity of mycoviral sequences was detected in the samples, which is consistent with the findings of the "Illumina 1" sequence dataset. However, as complete genomes were not assembled from the NGS data, it is possible that the actual mycovirus complexity is lower than what it appears as contigs originating from different areas of the same viral genome can hit to different known mycoviruses during blast analysis. The SD samples contained more mycoviruses than the GLD samples. However, we believe this to be an interesting coincidence and do not propose that mycoviruses play a role in the aetiology of SD. Most mycoviral sequences identified in this study aligned to cryptic mycoviruses that are not known to cause symptoms in their hosts. Shiraz disease

symptoms are more severe than GLD symptoms and the SD vines that were sampled were physiologically more compromised than the GLD vines. The higher virus load in these vines could therefore be due to a decreased defence against fungus and virus pathogens.

Significantly more mycoviral sequences were detected in the sample sequenced using the SOLiD 5500xl sequencing platform, with more than double the number of mycoviruses identified compared to samples sequenced with the HiScanSQ platform. The "SOLiD" sequence dataset contained ten times more reads than the largest "Illumina 2" sequence dataset (Table 4.1) and was thus able to detect viruses with lower titres which could not be detected in the "Illumina 2" sequence datasets. The higher number of reads did not improve the *de novo* contig assemblies of the "SOLiD" short reads. Instead of leading to the assembly of longer contigs as would be expected, a significantly larger number of contigs were assembled that were shorter than the contigs assembled from the smaller "Illumina 2" sequence datasets. The N50 value of 327 indicates that 50% of the more than 618 million "SOLiD" bases that were assembled were assembled into contigs shorter than 327 nt. The N50 value, maximum contig length and average contig length of the "SOLiD" contigs were shorter than the same values for the contigs assembled from the "Illumina 2" sequence data (Table 4.3).

After trimming and filtering, 56.3-66.9% of the "Illumina 2" dataset and 97.2% of the "SOLiD" sequence datasets remained (Table 4.1). This difference between the two sequencing platforms is not necessarily an indication that the SOLiD 5500xl platform produced higher quality sequence data as different tools and parameters were used to trim and filter the sequence reads (FASTX-Toolkit and CLC Genomics Workbench 5.5 for the "Illumina 2" and "SOLiD" data respectively). However, the sequencing-by-ligation strategy used by the SOLiD platform leads to each position in the read being sequenced twice, which generally increases the accuracy of this approach (Radford et al., 2012). The "SOLiD" dataset contained the highest percentage of reads that could not be assembled into contigs or that were assembled into contigs that could not be identified (30.9%). This is consistent with other comparative studies that found the SOLiD sequencing platform to produce the highest fraction of unusable data (Shen et al., 2008; Suzuki et al., 2011; Walter et al., 2009). Possible reasons for this are the inclusion of sequencing adapters in the reads, the presence of multiple templates on a single bead during emulsion PCR or fluorescence detection of closely located neighbouring beads (Suzuki et al., 2011).

The complete genome of a novel virus related to CeEV-1 was assembled from the "Illumina 2" sequence dataset (contig 165). We named this virus grapevine endophyte endornavirus (GEEV). The genome contains a single large ORF which encodes a polyprotein with a viral helicase and an RdRp domain. Most endornavirus genomes have a discontinuity (nick) in the positive strand near the 5'-end (Fukuhara et al., 2006). We were unable to determine the position of the nick (if any) from the sequencing data. The size and structure of the genome and sequence similarity substantiates that GEEV represents the genome of a member of the family *Endornaviridae*.

As previously shown, the accuracy of genomes assembled from NGS data is high. The quality of sequences in the 5′- and 3′-ends of assembled contigs are, however, lower due to a decreased depth of sequencing in these areas. Only the 5'- and 3'-terminal ends of the genome of GEEV were thus confirmed with Sanger sequencing. In addition to contig 165, three other large contigs representing the genome of GEEV were assembled. Contigs 165 and 137 were assembled from samples SD8 and SD12 respectively using CLC Genomics Workbench 4.9 (CLC bio) and contigs 3 and 2 were assembled from samples SD8 and SD12 respectively using CLC Genomics Workbench 5.5 (CLC bio). Pairwise comparisons between the completed genome of GEEV and these contigs showed similarities of 99.13-100.00% (data not shown). The majority of differences were at the ends of the contigs as all of the contigs were shorter than the genome of GEEV, except contig 165 which was longer. The average depth of sequencing for the contigs were 686 nt (contig 165), 2 843 nt (contig 137), 684 nt (contig 3) and 2 798 (contig 3) which further increases confidence in the accuracy of the assembled genome sequence of GEEV.

As expected from the BLAST analysis, GEEV clustered with CeEV-1 in a phylogenetic tree comprised of the RdRp aa sequences of 15 endornaviruses. The endornaviruses did not cluster together according to their fungus, plant and oomycete hosts or according to protein domains present. Endornaviruses are believed to have an interesting evolutionary history, with protein domains not found consistently in all members and similar protein domains being acquired independently by the different viruses (Roossinck et al., 2011).

We were able to show that NGS of grapevine dsRNA generates a large amount of high quality read data that can be used to determine the virus complexity in individual grapevines. The read data was generated using two different sequencing technologies, namely an Illumina HiScanSQ and an ABI SOLiD 5500xl sequencing instrument. In addition to a number of viruses and viroids known to infect grapevine, 51 putative mycovirus species were identified in the five grapevines analysed. Twenty-two of these putative mycovirus species were previously identified in the samples using an earlier NGS run, which demonstrates that the results are reproducible. Twenty-nine additional mycovirus species were also identified. By assembling the complete genome of one of the identified mycoviruses, we demonstrate that NGS is a viable approach, which is both time and cost effective, to identify mycoviruses in grapevine. The preliminary mycoviral sequence data generated in this study can be used to determine the full genome sequences of the remaining identified mycoviruses, after which further investigations can be made regarding the fungus hosts and phenotypic effects of these newly identified viruses in grapevine.

4.5 References

Alkan, C., Sajjadian, S., Eichler, E.E., 2011. Limitations of next-generation genome sequence assembly. Nat. Methods. 8(1), 61 – 65.

Bester, R., 2012. Sequencing and detection of a new strain of grapevine leafroll-associated virus 3 in South Africa. Master of Science thesis. Stellenbosch: University of Stellenbosch.

Carstens, E.B., Ball, L.A., 2009. Ratification vote on taxonomic proposals to the International Committee on Taxonomy of Viruses. Arch. Virol. 154(7), 1181 – 1188.

Conesa, A., Götz, S., García-Gómez, J.M., Terol, J., Talón, M., Robles, M., 2005. Blast2GO: a universal tool for annotation, visualization and analysis in functional genomics research. Bioinformatics. 21(18), 3674 – 3676.

Espach, Y., Maree, H.J., Burger, J.T., 2012. Complete genome of a novel endornavirus assembled from next-generation sequence data. J. Virol. 86(23), 13142.

Fauquet, C.M., Mayo, M.A., Maniloff, J., Desselberger, U., Ball, L.A. (eds.), 2005. Virus Taxonomy: Eighth Report of the International Committee on Taxonomy of Viruses. Elsevier Academic Press, San Diego.

Fukuhara, T., Koga, R., Aoki, N., Yuki, C., Yamamoto, N., Oyama, N., Udagawa, T., Horiuchi, H., Miyazaki, S., Higashi, Y. et al., 2006. The wide distribution of endornaviruses, large double-stranded RNA replicons with plasmid-like properties. Arch. Virol. 151(5), 995 – 1002.

Langmead, B., Trapnell, C., Pop, M., Salzberg, S.L., 2009. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. Genome Biol. 10(3), R25.

Li, H., Ruan, J., Durbin, R., 2008. Mapping short DNA sequencing reads and calling variants using mapping quality scores. Genome Res. 18(11), 1851 – 1858.

Marcher-Bauer, A., Lu, S., Anderson, J.B., Chitsaz, F., Derbyshire, M.K., DeWeese-Scott, C., Fong, J.H., Geer, L.Y., Geer, R.C., Gonzales, N.R. et al. 2010. CDD: A conserved domain database for the functional annotation of proteins. Nucl. Acids Res. 39(1), 225 – 229.

Meng, B., Li, C., Goszczynski, D.E., Gonsalves, D., 2005. Genome sequences and structures of two biologically distinct strains of Grapevine leafroll-associated virus 2 and sequence analysis. Virus Genes. 31(1), 31 – 41.

Metzker, M.L., 2010. Sequencing technologies – The next generation. Nat. Rev. Genet. 11(1), 31 – 46.

Radford, A.D., Chapman, D., Dixon, L., Chantrey, J., Darby, A.C., Hall, N., 2012. Application of next-generation sequencing technologies in virology. J. Gen. Virol. 93(9), 1853 – 1868.

Roossinck, M.J., Sabanadzovic, S., Okada, R., Valverde, R.A., 2011. The remarkable evolutionary history of endornaviruses. J. Gen. Virol. 92(11), 2674 – 2678.

Sela, N., Luria, N., Dombrovsky, A., 2012. Genome assembly of bell pepper *Endornavirus* from small RNA. J. Virol. 86(14), 7721.

Shen, Y., Sarin, S., Liu, Y., Hobert, O., Pe'er, I., 2008. Comparing platforms for *C. elegans* mutant identification using high-throughput whole-genome sequencing. PLoS One. 3(12), e4012 DOI: 10.1371/journal.pone.0004012

Suzuki, S., Ono, N., Furusawa, C., Ying, B.W., Yomo, T., 2011. Comparison of sequence reads obtained from three next-generation sequencing platforms. PLoS One. 6(5), e19534 DOI:10.1371/journal.pone.0019534

Valverde, R., Nameth, S., Jordan, S., 1990. Analysis of double-stranded RNA for plant virus diagnosis. Plant Dis. 74(3), 255 – 258.

Walter, N.A., Bottomly, D., Laderas, T., Mooney, M.A., Darakjian, P., Searles, R.P., Harrington, C.A., McWeeny, S.K., Hitzemann, R., Buck, K.J., 2009. High throughput sequencing in mice: a platform comparison identifies a preponderance of cryptic SNPs. BMC Genomics. 10, 379 DOI:10.1186/1471-2164-10-379

Zhang, J., Chiodini, R., Badr, A., Zhang, G., 2011. The impact of next-generation sequencing on genomics. J. Genet. Genomics. 38(3), 95 – 109.

Internet resources

Applied Biosystems: http://www.appliedbiosystems.com/absite/us/en/home.html

Illumina: http://www.illumina.com/

Roche/454: http://www.454.com/

Chapter 5: Conclusion

The development of PCR (Mullis et al., 1986) and Sanger sequencing (Sanger et al., 1977) revolutionised molecular biology and provided scientists with the ability to amplify and sequence genetic material with relative ease. Limitations of these techniques are that they are not time and cost effective when employed for large scale experiments since prior knowledge or cloning of the genome is needed for amplification or sequencing (Radford et al., 2012). The development of the first NGS platform in 2005 was another revolutionary advancement in molecular biology that made large scale sequencing experiments attainable to a wider range of laboratories (Margulies et al., 2005). Metagenomics and NGS have provided us with the ability to produce copious amounts of sequence data, in which multiple organisms are represented, in a single run. The sensitivity and unbiased nature of NGS of dsRNA metagenomic samples ensures that all viruses are detected, including viruses with low titres, novel viruses and cryptic viruses.

The majority of mycoviruses do not cause symptoms in their hosts, and therefore the presence of mycoviruses in grapevine remained unknown until a recent study, using NGS, identified mycoviruses related to the families *Chrysoviridae* and *Totiviridae* in a pooled grapevine sample (Coetzee et al., 2010). The chrysovirus related sequences were the second most abundant viral sequences generated by this study. A subsequent NGS study also detected a considerable number of mycoviral sequences in grapevine, with as many as 26 putative mycovirus species reported in two vines (Al Rwahnih et al., 2011). In the present study, NGS was used to show that, in addition to a number of known grapevine infecting viruses, the grapevine samples analysed host a large diversity of mycoviruses.

Five diseased grapevine samples (*Vitis vinifera* cv. Shiraz) were sequenced using an Illumina HiScanSQ NGS instrument ("Illumina 1"). The short reads were assembled and classified according to BLAST results. Twenty-six putative mycovirus species were identified, of which two (GaCV and GaMV-1) have previously been identified in grapevine (Al Rwahnih et al., 2011) and the rest were novel viruses. Primers, designed to target 13 of the identified mycoviruses, were used to screen dsRNA from the five grapevine samples and fungal mycelium from endophytic fungi isolated from the five grapevine samples. Two of the mycoviruses related to sclerotinia sclerotiorum partitivirus S and CeEV-1 could be detected in grapevine dsRNA and in fungus isolates. There were some inconsistencies regarding the samples in which these viruses were identified bioinformatically after sequencing and the samples in which they were identified using RT-PCR (grapevine dsRNA) and rapid direct one tube PCR (fungus isolates). We believe that this is caused by the dispersion of the fungus hosts in grapevine which leads to an uneven distribution of the mycoviruses in the vine.

Additionally, two putative mycoviruses were detected in fungus isolates after dsRNA extractions from fungal mycelium. The dsRNA segments visible in TAE agarose gels were similar in size to the

genome sizes of chrysoviruses and partitiviruses. None of the primers, of which 13 targeted chrysoviruses and three targeted partitiviruses, produced amplicons of the correct size when the fungal dsRNA was screened using RT-PCR. The fungi were isolated from the same plant but not the exact material that was used for sequencing, resulting in the possibility that genome sequences of mycoviruses infecting these fungi were not present in the sequence data used for primer design.

As only two of the 26 putative mycovirus species identified could be verified using PCR-based techniques, two additional sequencing runs using platforms based on different chemistries were employed to further investigate the presence and diversity of mycoviruses in the five grapevine samples. The five cDNA sequencing libraries were sequenced for a second time using an Illumina HiScanSQ instrument, and newly extracted dsRNA from two of the samples was pooled and sequenced using the SOLiD 5500xl system from ABI. More reads and better quality data was produced in these runs than during the "Illumina 1" sequencing run. Twenty-two of the 26 mycovirus species identified in the "Illumina 1" sequence dataset were again identified in the "Illumina 2" and "SOLiD" sequence datasets, as well as 29 additional mycovirus species not identified in "Illumina 1". The aa similarities to known mycovirus species ranged between 52-99%, so a number of the identified mycoviral sequences represented novel mycoviruses. When considering aa identities to known mycovirus species, six viruses could be identified on family level, 14 on genus level and 31 represented variants of known mycovirus species.

Up to 19 different mycovirus species from at least four different virus families were identified in a single grapevine plant. This demonstrates that individual grapevines potentially host a large diversity of mycoviruses. However, only partial mycoviral genomes were assembled from the NGS data and it is possible that areas of a single mycoviral genome could hit to different known mycovirus species during BLAST analyses. The actual mycovirus diversity could therefore be lower than what it currently appears to be. The role of the identified mycoviruses in vine health status has not yet been determined. Although more mycoviral sequences were identified in SD vines than in GLD vines, it does not indicate an association between mycoviruses and the development of SD. We believe the higher number of mycoviruses in SD vines is caused by the compromised physiology of theses vines due to more severe symptoms, leading to a higher susceptibility to fungus and virus infection.

The complete genome of a novel endornavirus, for which we propose the name grapevine endophyte endornavirus (GEEV), was assembled from the "Illumina 2" sequence data of sample SD8. The genome is 12 154 bp long and contains a single large ORF which encodes a polyprotein with putative RdRp and viral helicase domains. This is the first complete genome of a mycovirus identified in grapevine. The virus was detected using rapid-direct one tube PCR in *Stemphylium* sp. and *Aureobasidium pullulans* isolated from grapevine phloem tissue. This is interesting since mycoviruses have not been found to be naturally associated with two different fungi from different genera in the past. Further research is needed to verify and investigate this phenomenon.

Additional future research will determine the complete genomes of other novel mycoviruses identified in this study as well as further characterise these genomes. This will provide a better understanding of the true mycovirus diversity in grapevine samples analysed. Parallel sequencing studies of DNA enriched for fungal sequences and dsRNA enriched for viral sequences from a single vine will provide a more complete representation of possible fungus hosts for identified mycoviruses and will circumvent the problem of unculturable fungus hosts. The role of mycoviruses in fungi infecting grapevine also needs to be determined. Although the majority of mycoviruses are cryptic, the possibility of identifying a hypovirulent mycovirus that can be exploited for the biological control of a grapevine pathogenic fungus, justifies further research on mycoviruses in this economically important crop.

References

Al Rwahnih, M., Daubert, S., Úrbez-Torres, J.R., Cordero, F., Rowhani, A., 2011. Deep sequencing evidence from single grapevine plants reveals a virome dominated by mycoviruses. Arch. Virol. 156(3), 397 – 403.

Coetzee, B., Freeborough, M.J., Maree, H.J., Celton, J.M., Reese, D.J.G., Burger, J.T., 2010. Deep sequencing analysis of viruses infecting grapevines: Virome of a vineyard. Virology. 400(2), 157 – 163.

Margulies, M., Egholm, M., Altman, W.E., Attiya, S., Bader, J.S., Bemben, L.A., Berka, J., Braverman, M.S., Chen, Y.J., Chen, Z., et al., 2005. Genome sequencing in microfabricated high-density picolitre reactors. Nature. 437(7057), 376 – 380.

Mullis, K., Faloona, F., Scharf, S., Saiki, R., Horn, G., Erlich, H., 1986. Specific enzymatic amplification of DNA in vitro: the polymerase chain reaction. Cold Spring Harb. Symp. Quant. Biol. 51(1), 263 – 273.

Radford, A.D., Chapman, D., Dixon, L., Chantrey, J., Darby, A.C., Hall, N., 2012. Application of next-generation sequencing technologies in virology. J. Gen. Virol. 93(9), 1853 – 1868.

Sanger, F., Nicklen, S., Coulson, A.R., 1977.DNA sequencing with chain-terminating inhibitors. Proc. Natl. Acad. Sci. USA. 74(12), 5463 – 6467.

Supplementary data

Table S1: List of mycovirus species adapted from King et al. (2011) and Pearson et al. (2009), including sequenced or partially sequenced mycoviruses from the NCBI database. Italicised viruses are recognized by the International Committee on Taxonomy of Viruses (ICTV).

Family	Genus	Genome	Species
Alphaflexiviridae	*Botrexvirus*	ss(+)RNA	*Botrytis virus X*
	Sclerodarnavirus		*Sclerotinia sclerotiorum debilitation-associated RNA virus*
	Unassigned		Grapevine-associated alphaflexiviridae 1
Barnaviridae	*Barnavirus*		*Mushroom bacilliform virus*
Chrysoviridae	*Chrysovirus*	dsRNA	*Agaricus bisporus virus 1*
			Aspergillus fumigatus chrysovirus
			Cherry chlorotic rusty spot associated chrysovirus
			Cryphonectria nitschkei chrysovirus 1
			Grapevine associated chrysovirus
			Helminthosporium victoriae 145S virus
			Magnaporthe oryzae chrysovirus 1
			Penicillium brevicompactum virus
			Penicillium chrysogenum virus
			Penicillium cyaneo-fulvum virus
			Raphanus sativas chrysovirus 1
			Verticillium dahliae chrysovirus 1
Endornaviridae	*Endornavirus*	dsRNA	Chalara endornavirus 1
			Grapevine endophyte endornavirus
			Gremmeniella abietina type B RNA virus XL1
			Helicobasidium mompa endornavirus 1
			Phytophthora endornavirus 1
			Tuber aestivum endornavirus
Gammaflexiviridae	*Mycoflexivirus*	ss(+)RNA	*Botrytis virus F*
	Unassigned		Grapevine-associated gammafelxiviridae 1
Hypoviridae	*Hypovirus*	dsRNA	*Cryphonectria hypovirus 1*
			Cryphonectria hypovirus 2
			Cryphonectria hypovirus 3
			Cryphonectria hypovirus 4
			Grapevine-associated hypovirus 1
			Valsa ceratosperma hypovirus 1
Megabirnaviridae	*Megabirnavirus*	dsRNA	*Rosellinia necatrix megabirnavirus 1*
Metaviridae	*Metavirus*	ss(+)RNA-RT	*Cladosporium fulvum T-1 virus*
			Fusarium oxysporum Skippy virus
			Saccharomyces cerevisiae Ty3 virus
			Schizosaccharomyces pombe Tf1 virus
			Schizosaccharomyces pombe Tf2 virus
Narnaviridae	*Mitovirus*	ss(+)RNA	Botrytis cinerea debilitation-related virus
			Cryphonectria mitovirus 1
			Gremmeniella mitovirus S1
			Helicobasidium mompa mitovirus 1-18
			Ophiostoma mitovirus 1a

68

Family	Genus	Genome	Species
Metaviridae	*Metavirus*	ss(+)RNA-RT	Ophiostoma mitovirus 1b
			Ophiostoma mitovirus 3a
			Ophiostoma mitovirus 4
			Ophiostoma mitovirus 5
			Ophiostoma mitovirus 6
			Ophiostoma mtiovirus 1a
			Ophiostoma mtiovirus 1b
			Ophiostoma mtiovirus 2
			Ophiostoma mtiovirus 3b
			Ophiostoma novo-ulmi mitovirus 4-Ld
			Thielaviopsis basicola mitovirus
	Narnavirus	ss(+)RNA	Grapevine-associated narnavirus 1
			Saccharomyces 20S RNA narnavirus
			Saccharomyces 23S RNA narnavirus
	Unassigned	ss(+)RNA	Phytophthora infestans RNA virus 4
			Rhizoctonia virus M2
Partitiviridae	*Partitivirus*	dsRNA	*Agaricus bisporus virus 4*
			Amasya cherry disease associated partitivirus
			Aspergillus foetidus virus S
			Aspergillus fumigatus partitivirus 1
			Aspergillus niger virus S
			Aspergillus ochraceous virus
			Atkinsonella hypoxylon virus
			Botryotinia fuckeliana partitivirus 1
			Ceratocystis polonica
			Ceratocystis resinifera virus 1
			Cherry chlorotic rusty spot associated partitivirus
			Diplocarpon rosae virus
			Discula destructiva virus 1
			Discula destructiva virus 2
			Flammulina velutipes isometric virus
			Fusarium poae virus 1
			Fusarium solani virus 1
			Gaeumannomyces abietina RNA virus T1-A
			Gaeumannomyces graminis virus 019/6-A
			Gaeumannomyces graminis virus T1-A
			Grapevine-associated partitivirus 1
			Grapevine-associated partitivirus 2
			Gremmeniella abietina RNA virus MS1
			Helicobasidium mompa virus
			Heterobasidion annosum virus
			Heterobasidion RNA virus 1
			Ophiostoma partitivirus 1
			Ophiostoma quercus partitivirus
			Penicillium stoloniferum virus F
			Penicillium stoloniferum virus S
			Phialophora radicicola virus 2-2-A

Family	Genus	Genome	Species
Partitiviridae	*Partitivirus*	dsRNA	*Pleurotus ostreatus virus 1*
			Rhizoctonia solani virus 717
			Rosellinia necatrix partitivirus 2
			Rosellinia necatrix partitivirus 3
			Rosellinia necatrix partitivirus 4
			Rosellinia necatrix partitivirus 5
			Rosellinia necatrix virus 1
			Sclerotinia sclerotiorum partitivirus S
	Unassigned	dsRNA	Fusarium graminearum dsRNA mycovirus 2
Pseudoviridae	*Hemivirus*	ss(+)RNA-RT	*Candida albicans Tca2 virus*
			Candida albicans Tca5 virus
			Saccharomyces cerevisiae Ty5 virus
			Saccharomyces paradoxus Ty5 virus
	Pseudovirus	ss(+)RNA-RT	*Saccharomyces cerevisiae Ty1 virus*
			Saccharomyces cerevisiae Ty2 virus
			Saccharomyces cerevisiae Ty4 virus
Reoviridae	*Mycoreovirus*	dsRNA	*Mycoreovirus 1*
			Mycoreovirus 2
			Mycoreovirus 3
			Rosellinia anti-rot virus
			Roselinnia necatrix mycoreovirus-3
Totiviridae	*Totivirus*	dsRNA	*Aspergillus foetidus virus 3*
			Aspergillus mycovirus 1816
			Aspergillus niger virus 3
			Botryotinia fuckeliana totivirus 1
			Chalara elegans RNA Virus 2
			Gaeumannomyces graminis virus 87-1-H
			Giardia canis virus
			Grapevine-associated totivirus 1
			Grapevine-associated totivirus 2
			Grapevine-associated totivirus 3
			Mycogone perniciosa virus
			Ophiostoma minus totivirus
			Ribes virus F RdRp gene
			Saccharomyces cerevisiae virus L-A
			Saccharomyces cerevisiae virus L-A(L1)
			Saccharomyces cerevisiae virus L-BC (La)
			Tolypocladium cylindrosporum virus 2
			Tuber aestivum virus 1
			Ustilago maydis virus H1
			Xanthophyllomyces dendrorhous virus L1A
			Xanthophyllomyces dendrorhous virus L1b
			Xanthophyllomyces dendrorhous virus L2
	Victorivirus	dsRNA	*Chalara elegans RNA Virus 1*
			Coniothyrium minitans RNA virus
			Epichloe festucae virus 1
			Gremmeniella abietina RNA virus L1

Family	Genus	Genome	Species
Totiviridae	*Victorivirus*	dsRNA	*Helicobasidium mompa totivirus 1-17*
			Helminthosporium victoriae virus 190S
			Magnaporthe oryzae virus 1
			Sphaeropsis sapinea RNA virus 1
			Sphaeropsis sapinea RNA virus 2
			Tolypocladium cylindrosporum virus 1
Unassigned	*Rhizidiovirus*	dsDNA	*Rhizidiomyces virus*
	Unassigned		Allomyces arbuscula virus
			Alternaria alternata virus 1
			Aspergillus foetidus virus F
			Botrytis cinerea virus CVg25
			Botrytis porri RNA virus 1
			Colletotrichum lindemunthianum virus
			Curvularia thermal tolerance virus
			Fusarium graminearum dsRNA mycovirus 1
			Fusarium graminearum dsRNA mycovirus 3
			Fusarium graminearum dsRNA mycovirus 4
			Fusarium graminearum virus DK21
			Gaeumannomyces graminis virus 45/101-C
			Grapevine associated mycovirus 1
			Grapevine associated mycovirus 2
			Grapevine associated mycovirus 3
			Helminthosporium maydis virus
			LaFrance isometric virus
			Lentinus edodes virus
			Perconia circinata virus
			Rosellinia necatrix megabirnavirus 1/W779
			Sclerotinia sclerotiorum dsRNA mycovirus-L
		ss(+)RNA	Botrytis cinerea virus F
			Diaporthe ambigua RNA virus 1
			Diaporthe RNA virus
			Oyster mushroom spherical virus
			Phytophthora infestans RNA virus 1
		ssDNA	Sclerotinia sclerotiorum hypovirulence-associated DNA virus 1

Table S2: List of bioinformatic tools available, which sequencing platforms each is compatible with and website addresses for more information. Table was adapted from Zang et al. (2011).

Tool	Platform[a]	Website address
***De novo* assembly**		
Abyss	I	http://www.bcgsc.ca/platform/bioinfo/software/abyss
ALLPATHS	I	http://www.broadinstitute.org/science/programs/genome-biology/crd
AMOScmp	R	http://sourceforge.net/projects/amos/files/
ARACHNE	R	http://www.broadinstitute.org/science/programs/genome-biology/crd
CAP3	R	http://pbil.univ-lyon1.fr/cap3.php
Consensus/SeqCons	R	http://www.seqan.de/downloads/projects.html
Curtain	I/R/A	http://code.google.com/p/curtain/
Edena	I	http://www.genomic.ch/edena
Euler-SR	I/R	http://euler-assembler.ucsd.edu/portal/?q=team
FuzzyPath	I/R	ftp://ftp.sanger.ac.uk/pub/zn1/fuzzypath/fuzzypath_v3.0.tgz
IDBA	I	http://www.cs.hku.hk/walse/idba/
MIRA/MIRA3	I/R	http://chevreux.org/projects_mira.html
Newbler	R	roche-applied-science.com/
Phrap	I/R	http://www.phrap.org/consed/consed.html#howToGet
RGA	I	http://rga.cgrb.oregonstate.edu/
QSRA	I	http://qsra.cgrb.oregonstate.edu/
SHARCGS	I	http://sharcgs.molgen.mpg.de/
SHORTY	A	http://www.cs.sunysb.edu/wskiena/shorty/
SOAPdenovo	I	http://soap.genomics.org.cn
SOPRA	I/A	http://www.physics.rutgers.edu/%7Eanirvans/SOPRA/
SR-ASM	R	http://bioserver.cs.put.poznan.pl/sr-asm-short-reads-assembly-algorithm
SSAKE	I/R	http://www.bcgsc.ca/platform/bioinfo/software/ssake
Taipan	I	http://sourceforge.net/projects/taipan/files/
VCAKE	I/R	http://sourceforge.net/projects/vcake
Velvet	I/R/A	http://www.ebi.ac.uk/%7Ezerbino/velvet
Mapping assembly		
BFAST	I/A	http://sourceforge.net/apps/mediawiki/bfast/index.php?title=Main_Page
Bowtie	I/R/A	http://bowtie-bio.sourceforge.net
BWA	I/A	http://bio-bwa.sourceforge.net/bwa.shtml
CoronaLite	A	http://solidsoftwaretools.com/gf/project/corona/
CABOG	R/A	http://wgs-assembler.sf.net
ELAND/ELAND2	I/A	http://www.illumina.com/
EULER	I	http://euler-assembler.ucsd.edu/portal/
Exonerate	R	http://www.ebi.ac.uk/wguy/exonerate
EMBF	I	http://www.biomedcentral.com/1471-2105/10?issue=S1
GenomeMapper	I	http://1001genomes.org/downloads/genomemapper.html
GMAP	I	http://www.gene.com/share/gmap
gnumap	I	http://dna.cs.byu.edu/gnumap/
ICON	I	http://icorn.sourceforge.net/
Karma	I/A	http://www.sph.umich.edu/csg/pha/karma/
LAST	I	http://last.cbrc.jp/
LOCAS	I	http://www-ab.informatik.uni-tuebingen.de/software/locas
Mapreads	A	http://solidsoftwaretools.com/gf/project/mapreads/
MAQ	I/A	http://maq.sourceforge.net
MOM	I	http://mom.csbc.vcu.edu/
Mosaik	I/R/A	http://bioinformatics.bc.edu/marthlab/Mosaik

Tool	Platform[a]	Website address
Mapping assembly		
mrFAST/mrsFAST	I	http://mrfast.sourceforge.net/
MUMer	A	http://mummer.sourceforge.net/
Nexalign	I	http://genome.gsc.riken.jp/osc/english/dataresource/
Novocraft	I	http://www.novocraft.com/
PerM	I/A	http://code.google.com/p/perm/
RazerS	I/A	http://www.seqan.de/projects/razers.html
RMAP	I	http://rulai.cshl.edu/rmap
segemehl	I/R	http://www.bioinf.uni-leipzig.de/Software/segemehl/
SeqCons	R	http://www.seqan.de/projects/seqcons.html
SeqMap	I	http://biogibbs.stanford.edu/*jiangh/SeqMap/
SHRiMP	I/R/A	http://compbio.cs.toronto.edu/shrimp
Slider/SliderII	I	http://www.bcgsc.ca/platform/bioinfo/software/slider
SOCS	A	http://solidsoftwaretools.com/gf/project/socs/
SOAP/SOAP2	I/A	http://soap.genomics.org.cn
SSAHA/SSAHA2	I/R	http://www.sanger.ac.uk/Software/analysis/SSAHA2
Stampy	I	http://www.well.ox.ac.uk/wmarting/
SXOligoSearch	I	http://synasite.mgrc.com.my:8080/sxog/NewSXOligoSearch.php
SHORE	I	http://1001genomes.org/downloads/shore.html
Vmatch	I	http://www.vmatch.de/
Quality assessment		
FastQC	I/A	http://www.bioinformatics.bbsrc.ac.uk/projects/fastqc/
PIQA	I	http://bioinfo.uh.edu/PIQA
ShortRead	I/R	http://bioconductor.org/packages/2.6/bioc/html/ShortRead.html
TileQC	I	http://www.science.oregonstate.edu/wdolanp/tileqc
Visualization tool		
Artemis/ACT	I/R	http://www.sanger.ac.uk/resources/software/artemis/
Consed	I/R	http://www.genome.washington.edu/consed/consed.html
EagleView	I/R	http://bioinformatics.bc.edu/marthlab/EagleView
Gambit	I/R	http://bioinformatics.bc.edu/marthlab/Gambit
Hawkeye	I/R	http://amos.sourceforge.net/hawkeye
IGV	I	http://www.broadinstitute.org/igv/?q¼home
LookSeq	I/R	http://lookseq.sourceforge.net
MagicViewer	I	http://bioinformatics.zj.cn/magicviewer/
MapView	I	http://evolution.sysu.edu.cn/mapview/
NGSView	I/A	http://ngsview.sourceforge.net
Savant	I/R	http://compbio.cs.toronto.edu/savant/
Tablet	I/R	http://bioinf.scri.ac.uk/tablet
XMatchView	I/R	http://www.bcgsc.ca/platform/bioinfo/software/xmatchview
Yenta	I	http://genome.wustl.edu/tools/cancer-genomics

[a] I Illumina; R Roche; A Applied Biosystems

Table S3: Organism names and accession numbers of sequences used as reference sequences for mapping assemblies.

Organism	Accession number
Vitis vinifera mitochondrion genome	NC_012119
Vitis vinifera chloroplast genome	NC_007957
Vitis vinifera chromosome 1	NC_012007
Vitis vinifera chromosome 2	NC_012008
Vitis vinifera chromosome 3	NC_012009
Vitis vinifera chromosome 4	NC_012010
Vitis vinifera chromosome 5	NC_012011
Vitis vinifera chromosome 6	NC_012012
Vitis vinifera chromosome 7	NC_012013
Vitis vinifera chromosome 8	NC_012014
Vitis vinifera chromosome 9	NC_012015
Vitis vinifera chromosome 10	NC_012016
Vitis vinifera chromosome 11	NC_012017
Vitis vinifera chromosome 12	NC_012018
Vitis vinifera chromosome 13	NC_012019
Vitis vinifera chromosome 14	NC_012020
Vitis vinifera chromosome 15	NC_012021
Vitis vinifera chromosome 16	NC_012022
Vitis vinifera chromosome 17	NC_012023
Vitis vinifera chromosome 18	NC_012024
Vitis vinifera chromosome 19	NC_012025
Grapevine viroid clone Cari2	AF462155
Grapevine viroid clone Syrah	AF462154
Grapevine fanleaf virus RNA 1	NC_003615
Grapevine fanleaf virus RNA 2	NC_003623
Grapevine fanleaf virus satellite RNA	NC_003203
Grapevine leafroll-associated virus-3 isolate 621	GQ352631
Grapevine leafroll-associated virus-3 isolate 623	GQ352632
Grapevine leafroll-associated virus-3 isolate GH11	JQ655295
Grapevine leafroll-associated virus-3 isolate GP18	EU259806
Grapevine leafroll-associated virus-3 isolate PL-20	GQ352633
Grapevine viroid clone Cari1	AF462156
Grapevine virus A isolate BMo32-1	DQ855087
Grapevine virus A isolate GTG11-1	DQ855084
Grapevine virus A isolate GTR1-1	DQ787959
Grapevine virus A isolate GTR1-2	DQ855086
Grapevine virus A isolate P163-M5	DQ855082
Grapevine virus A isolate PA3	AF007415
Grapevine virus B isolate 94_971	EF583906
Grapevine virus E isolate SA94	GU903012
Grapevine yellow speckle viroid 1	AF059712
Grapevine yellow speckle viroid 2	FJ597941
Hop stunt viroid H1	HM357802
Hop stunt viroid HSVd 2.7	GU825977
Rupestris stem pitting associated virus-1	NC_001948

Table S4: Viruses and amino acid sequences used to construct the endornavirus phylogenetic tree.

Virus	Host	Acronym	GenBank accession
Bell pepper endornavirus-IS	Plant	BPEV-IS	AFM10600
Bell pepper endornavirus-KY	Plant	BPEV-KS	BAK52155
Bell pepper endornavirus-YW	Plant	BPEV-YW	YP_004765011
Chalara elegans endornavirus 1[a]	Fungus	CeEV-1	ADN43901
Grapevine endophyte endornavirus	Fungus	GEEV	AFV91541
Gremmeniella abietina type B RNA virus XL1	Fungus	GaBRV-XL1	ABD73305
Gremmeniella abietina type B RNA virus XL2	Fungus	GaBRV-XL2	ABD73306
Helicobasidium mompa endornavirus 1	Fungus	HmEV1	YP_003280846
Oryza rufipogon endornavirus	Plant	ORV	YP_438202
Oryza sativa endornavirus	Plant	OSV	YP_438200
Persia americana endornavirus	Plant	PaEV	YP_005086952
Phaseolus vulgaris endornavirus 1[b]	Plant	PvEV-1	BAM68539
Phaseolus vulgaris endornavirus 2[b]	Plant	PvEV-2	BAM68540
Phytophthora endornavirus 1	Oomycete	PEV1	YP_241110
Tuber aestivum endornavirus	Fungus	TaEV	YP_004123950
Vicia faba endornavirus	Plant	VFV	YP_438201
Pineapple mealybug wilt-associated virus 1	Plant	PMWaV-1	YP_001642337

[a] Only a partial coding sequence is available for this virus.

[b] Complete coding sequences are available for these viruses, but not a complete genome sequence.