Characterising the viral and microbial diversity in old and young grapevines

by

Kristin Oosthuizen

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

Supervisor: Prof. J.T. Burger

Co-supervisor: Dr. H.J. Maree

December 2017

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.

December 2017

Abstract

There is anecdotal evidence suggesting that old vines produce wines of higher quality than young vines. In South Africa, vines are generally regarded ‘old’ when they reach 35 years of age, while ‘young’ vines are less than ten years old. Grapevines are susceptible to a large spectrum of pathogens that have negative impacts on grape quality and yield. This crop is also colonised by diverse endophytic microorganisms that play an important role in plant growth, health and productivity. To date, limited molecular research has been performed to determine the complexity of the pathogenic and endophytic communities in old vines. This study aimed to characterise the viral and fungal profiles of old and young Pinotage grapevines, using next-generation sequencing in a metagenomics approach. To determine the viral diversity, double-stranded RNA was extracted from phloem to enrich for virus-specific nucleic acids, and sequenced on an Illumina platform. High-quality reads were assembled into contigs and classified through BLAST analysis against the NCBI database. Additionally, the reads were mapped to a database consisting of known grapevine virus and viroid genome sequences. Reverse-transcription PCR detection assays were performed to validate the presence of the identified viruses. The fungal communities were characterised by extracting total DNA from the vascular tissues of the cane, followed by amplification of the ITS2 region, and deep amplicon sequencing. The ITS2 sequences were clustered into operational taxonomic units at a 97% identity threshold and taxonomically classified through BLAST analysis against the UNITE database. Viruses of the families *Closteroviridae*, *Betaflexiviridae* and *Tymoviridae*, and four pospiviroids were detected. The virus community was more diverse in the old vines, with 31 and 16 virus variants detected in the old and young vines, respectively. This was expected, since old vines have been exposed to viral pathogens for a longer period. The economically important grapevine leafroll-associated virus 3 was the most abundant species present in the samples, consistent with previous surveys of vineyards in the Western Cape. Grapevine Syrah virus 1, and possibly grapevine rupestris vein feathering virus, was identified for the first time in South African grapevines, expanding the global distribution of the virus(es). The amplicon data revealed the presence of different filamentous and yeast-like fungal taxa commonly associated with grapevines, including species of *Alternaria*, *Aureobasidium*, *Cladosporium* and *Epicoccum*. Several pathogens of grapevine trunk diseases and postharvest rot, and endophytic species with biocontrol properties were detected. The young-vine sample group showed greater fungal diversity, as determined

by three alpha diversity metrics, although not statistically significant. It may be speculated that the fungal community of old vines is more accustomed to the environment, and therefore less diverse. No differences were observed between the old and young vines, with regards to the community composition. The data generated in this study has contributed to research on the complex viral and fungal communities inhabiting old vines.

Opsomming

Volgens wynkenners produseer ou wingerde wyne van hoër gehalte as jong wingerde. In Suid-Afrika word wingerde oor die algemeen as 'oud' beskou as hulle meer as 35 jaar oud is en ‘jonk’ as hulle minder as tien jaar oud is. Wingerdstokke is vatbaar vir veelvudige patogene wat die gehalte en opbrengs van die druiwe negatief beïnvloed. Hierdie gewas word ook beset deur diverse endofitiese microörganismes wat 'n belangrike rol in plantegroei, gesondheid en produktiwiteit speel. Tot op hede is min molekulêre navorsing uitgevoer om die kompleksiteit van die patogeniese en endofitiese gemeenskappe in ou wingerde te bepaal. Die doel van hierdie studie was om die virus- en swam-profiele van ou en jong Pinotage wingerdstokke te beskryf, deur gebruik te maak van volgende-generasie volgordebepalingstegnologie in 'n metagenomiese benadering. Die virus diversiteit is bepaal deur die suiwering van dubbelstring RNS vanuit floëem om vir virus-spesifieke nukleïensure te verryk, gevolg deur volgordebepaling met ‘n Illumina instrument. Hoë-kwailiteit volgorde-fragmente is saamgestel in langer konstrukte wat deur BLAST analise teen die NCBI databasis geklasifiseer is. Daarbenewens is die volgorde-fragmente vergelyk met 'n databasis bestaande uit genoomvolgordes van bekende wingerdvirusse en -viroïede. Tru-transkripsie amplifiseringsreaksies is uitgevoer om die teenwoordigheid van die geïdentifiseerde virusse te bevestig. Die swamgemeenskappe is beskryf deur die suiwering van DNS vanuit vaatweefsel, gevolg deur amplifisering van die ITS2 lokus, en Illumina amplikon volgordebapling. Die ITS2 volgorde-fragmente is in operasionele taksonomiese eenhede groepeer, gebaseer op 97% identiteit, en taksonomies geklassifiseer deur BLAST analise teen die UNITE databasis. Virusse van die families *Closteroviridae*, *Betaflexiviridae* en *Tymoviridae*, en vier pospiviroïede is geïdentifiseer. Die virusgemeenskap was meer divers in die ou wingerde; 31 en 16 virus variante is onderskeidelik in die ou en jong wingerde geïdentifiseer. Dit was nie onverwags nie, aangesien die ou wingerde vir 'n langer tydperk aan virale patogene blootgestel was. Die ekonomies-belangrike grapevine leafroll-associated virus 3 was die mees dominante spesie teenwoordig in die monsters, in ooreenstemming met vorige opnames van Wes-Kaapse wingerde. Grapevine Syrah virus 1, en moontlik grapevine rupestris vein feathering virus, is vir die eerste keer in Suid-Afrikaanse wingerdstokke geïdentifiseer. Die amplikon data het die teenwoordigheid van verskillende filament- en gisagtige swamme wat met wingerde geassosieer word aangedui, insluitende *Alternaria*, *Aureobasidium*, *Cladosporium* en *Epicoccum* spesies. Verskeie patogene

van wingerdstamsiektes en na-oesverrotting, en endofitiese spesies met biokontrole-eienskappe was teenwoordig in die monsters. Die jong-wingerd steekproefgroep het 'n groter swamdiversiteit getoon, soos bepaal deur drie alfa-diversiteitsmetrieë, alhoewel dit nie statisties betekenisvol was nie. Daar mag gespekuleer word dat die swamgemeenskap van ou wingerdstokke beter aangepas het by die omgewing, en daarom minder divers is. Geen verskille is tussen die ou en jong wingerde waargeneem met betrekking tot die gemeenskap samestelling nie. Die data wat in hierdie studie genereer is, het bygedra tot navorsing oor die komplekse virus en swam gemeenskappe wat in ou wingerde voorkom.

Acknowledgements

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

- My supervisor, Prof. J.T. Burger, for giving me the opportunity to be part of an exceptional research group, and for his guidance and support.
- My co-supervisor, Dr. H.J. Maree, for his mentorship and intellectual inputs.
- Beatrix Coetzee, for her friendship and constant assistance and encouragement throughout this study.
- Members of the Vitis laboratory, for the stimulating and uplifting working environment.
- My family and friends, for their moral support.
- Kanonkop wine estate, for permitting and assisting in sample collections.
- The Pinotage Association, for research funding.
- The Agricultural Research Council's Professional Development Programme, for personal financial assistance.

Table of contents

List of figures

List of tables

List of abbreviations

°C	degrees Celsius
µg	microgram(s)
µl	microlitre(s)
3’	three prime
5’	five prime
A	adenine
AGVd	Australian grapevine viroid
Avg. cov.	average depth of coverage
BDA	black dead arm
BLAST	Basic Local Alignment Search Tool
bp	base pair(s)
C	cytosine
cDNA	complimentary deoxyribonucleic acid
CP	coat protein
CPm	minor coat protein
CTAB	cetyltrimethylammonium bromide
cv.	cultivar
D	domain
DGGE	denaturing gradient gel electrophoresis
DNA	deoxyribonucleic acid
DNase	deoxyribonuclease
DOI	digital object identifier
dsRNA	double-stranded ribonucleic acid
E	exponent
Eds.	editors
EDTA	ethylenediaminetetraacetic acid
ELISA	enzyme-linked immunosorbent assay

E_{max}	maximum expected error
E-value	Expect value
F	forward primer
G	guanine
Gb	gigabase(s)
Gen. cov.	genome coverage
GFkV	Grapevine fleck virus
GHVd-like RNA	Grapevine hammerhead viroid-like ribonucleic acid
GLD	grapevine leafroll disease
GLRaV-2	Grapevine leafroll-associated virus 2
GLRaV-3	Grapevine leafroll-associated virus 3
GRGV	Grapevine Red Globe virus
GRSPaV	Grapevine rupestris stem pitting-associated virus
GRVFV	Grapevine rupestris vein feathering virus
GSyV-1	Grapevine Syrah virus 1
GVA	Grapevine virus A
GVB	Grapevine virus B
GVE	Grapevine virus E
GYSVd-1	Grapevine yellow speckle viroid 1
GYSVd-2	Grapevine yellow speckle viroid 2
HEL	helicase
HSVd	Hop stunt viroid
ICVG	International Council for the Study of Viruses and Virus-like Diseases of the Grapevine
ITS	internal transcribed spacer
kb	kilobase(s)
L.	Linnaeus system of binomial nomenclature
Len. frac.	length fraction
M	molar

Max.	maximum
mg	milligram(s)
Min.	minimum
ml	millilitre(s)
mM	millimolar
mRNA	messenger ribonucleic acid
NaCl	sodium chloride
NCBI	National Center for Biotechnology Information
ng	nanogram(s)
NGS	next-generation sequencing
nm	nanometre(s)
no.	number
nt	nucleotide(s)
O	old
OIV	International Organisation of Vine and Wine
OTA	ochratoxin A
OTU	operational taxonomic unit
PCR	polymerase chain reaction
PGPR	plant growth-promoting rhizobacterium
pH	potential of hydrogen
PhiX	*Enterobacteria phage phiX174*
p-value	calculated probability
PVP-10	polyvinylpyrrolidone-10
Q	Phred quality score
QIIME	Quantitative Insights Into Microbial Ecology
R	reverse primer
RBP	RNA-binding protein
RDP	Ribosomal Database Project

RM	read-mapping
RNA	ribonucleic acid
RNase	ribonuclease
RPKM	Reads Per Kilobase of transcripts per Million mapped reads
rpm	revolutions per minute
rRNA	ribosomal ribonucleic acid
RT-PCR	reverse transcription polymerase chain reaction
RWC	rugose wood complex
S	Svedberg units
SAWIS	South African Wine Industry Information and Systems
sp.	species (singular)
spp.	species (plural)
sRNAs	small ribonucleic acids
ssRNA	single-stranded ribonucleic acid
Suppl.	supplementary
T	thymine
T_a	annealing temperature
TAE	Tris-acetate-EDTA
Taq	*Thermus aquaticus*
TMV	Tobacco mosaic virus
T-RFLP	terminal restriction fragment length polymorphism
Tris-HCl	Tris-hydrochloride
V	(hyper)variable
v	version
VRR	Virus (or Viroid) Read Ratio
v/v	volume per volume
w/v	weight per volume
Y	young

Chapter 1: Introduction

1.1 General introduction

Grapevine (*Vitis vinifera* L.) is a commercially valuable agricultural crop that is cultivated on six continents, with a global vineyard surface area of roughly 7.5 million hectares (http://www.oiv.int/). In 2016, South Africa was ranked as having the 14th largest area under vines (wine and table grapes) and placed 7th in terms of global wine production. In the same year, 898 million litres of wine was produced locally, of which 47% was exported (http://www.sawis.co.za/). The wine industry contributes considerably to the South African economy; in 2013, the industry generated R36.1 billion, accounting for 1.2% of the national Gross Domestic Product, and provided employment for close to 300 000 people (http://www.sawis.co.za/).

The economic life of the average South African vineyard is approximately 20 years. This is primarily due to virus infection, but is also driven by the preferences of wine consumers. However, a number of vineyards have remained profitable beyond their life expectancy, despite the prolonged exposure to environmental stresses. Recent years have seen growing interest and investment in old vines, and the clonal propagation thereof (Heyns, 2013). While ‘old vine’ is a common description on wine labels, its definition is open to interpretation. According to Ms. Rosa Kruger, one of the directors of the Old Vine Project, vines in South Africa qualify for ‘old’ status when they reach 35 years of age (http://iamold.withtank.com/home/). Anecdotes from sensory panels indicate that old vines can produce exceptional, unique wines (Heyns, 2013). The consensus among connoisseurs is that these wines have more depth and complexity than wines produced from younger vines.

To date, limited experimental research has been performed to establish which factors are responsible for phenotypic differences between old and young vines, and how these differences potentially influence wine character. One hypothesis is that there is greater *terroir* expression from old vines. These old vines are more adapted to the soil type and climate of their specific environment, and will therefore show more pronounced regional characteristics, as reflected in the quality of the grapes (Easton, 2015). Viticulturists would specifically acknowledge old vines’ well-established root systems that can serve as a buffer in dry conditions, while molecular scientists may consider factors such as genetic variation and epigenetic modifications (Easton, 2015; Heyns, 2013). Another

potential contributing factor is the network of viruses and microbes interacting with the vine.

As a field-grown plant, grapevine is susceptible to a large spectrum of pathogens. More than 70 grapevine-infecting agents, including 65 viruses, have been described (Martelli, 2014). This number does not include fungal species that have been implicated in grapevine diseases. Although these pathogens cause various observable symptoms, it is their negative impacts on grape composition and yield that are of greatest concern to winemakers. Yet there are old vineyards that continue to produce wines of high quality, regardless of their pathogen status (Heyns, 2013). No efforts have been made to determine the extent of the pathogen populations in these old vines.

Not all organisms that inhabit grapevines are debilitating. This crop is also host to diverse endophytic fungal and bacterial communities that have important functions in promoting plant growth and health. The role of grape-associated microbes in wine fermentation is well documented (Barata *et al*., 2012). A biogeographical association for fungal and bacterial taxa of grape musts has been identified across different wine-producing zones, suggesting that there may a microbial contribution to regionally distinct wine characteristics (Bokulich *et al*., 2014; Pinto *et al*., 2015). This concept, referred to as 'microbial *terroir*', has yet to be established as a determining attribute of wine quality, but is further supported by regional associations among the grape microbiota, wine metabolite profiles and fermentation behaviour (Bokulich *et al*., 2016).

The grape microbiome is not dissociated from the rest of the grapevine niche; in fact, soil is considered a primary source of microbial diversity. It is plausible that differentially selected soil microbes, that endophytically colonise the roots and other vine tissues, shape the microbial assemblages of grapes, thereby indirectly influencing wine characteristics (Gilbert *et al*., 2014; Zarraonaindia *et al*., 2015). Like other aspects of *terroir*, it can be argued that the endospheres of old vines show more defined microbial regionalisation, hence the greater complexity of old-vine wines. However, this has not yet been investigated across old vineyards.

The focus of the current study is to identify potential differences in the viral and endophytic microbial communities of old and young vines, using next-generation sequencing (NGS). This technology has been successfully applied to sequence the total viral and microbial complement of grapevine (Al Rwahnih *et al*., 2009; Pinto *et al*., 2014). The information generated in this study could support viticultural practices;

identifying core microbes that are associated with old vines, and that contribute to the unique old-vine character of the final product could ultimately lead to the development of commercial vine probiotics, designed to improve the wine *terroir* for younger or imported clones. The pathogen data can support the development of more accurate diagnostic assays.

1.2 Aim and objectives

The study aimed to unravel the viral and microbial diversity in old and young grapevines using next-generation sequencing in a metagenomics approach. The following objectives were set out in order to achieve this aim:

- To sample genetically identical old and young *Vitis vinifera* vines from the same vineyard.
- To characterise the viral profiles of libraries prepared from double-stranded RNA-enriched samples, with NGS and bioinformatic analyses.
- To screen for the viruses identified in the NGS data with RT-PCR detection assays.
- To characterise the fungal communities by extracting total DNA and amplifying the internal transcribed spacer 2 region, followed by deep amplicon sequencing and bioinformatic analyses.
- To characterise the bacterial communities by extracting total DNA and amplifying the V3-V4 and V6-V8 regions of the 16S ribosomal RNA gene, followed by deep amplicon sequencing and bioinformatic analyses.

1.3 Chapter layout

This thesis is divided into five chapters: an introduction, literature review, two research chapters and a conclusion.

Each chapter is referenced separately.

Chapter 1: Introduction

A general introduction to the study and its significance, aims and objectives, an overview of the chapter layout, and the research outputs generated throughout the project, are provided.

Chapter 2: Literature review

An overview of literature pertaining to grapevine diseases and associated pathogens, grapevine endophytes, and the use of conventional molecular techniques and metagenomics to study viruses and microbial communities.

Chapter 3: The viral diversity in old and young grapevines

The viral diversity in four old and four young grapevines, as determined by NGS and bioinformatic analyses, is described. The use of RT-PCR detection assays to validate the presence of the viruses identified in the eight sequenced samples, as well as to screen 12 additional samples that had not been sequenced, is also discussed.

Chapter 4: The fungal diversity in old and young grapevines

The fungal diversity within and between the eight selected grapevine samples (see chapter 3), as determined by deep amplicon sequencing and bioinformatic analyses, is described.

Chapter 5: Conclusion

General concluding remarks, limitations and future prospects of the study.

1.4 Research outputs

The following publication and conference contributions were generated during the study:

Publication

Oosthuizen, K., Coetzee, B., Maree, H.J., Burger, J.T., 2016. First report of Grapevine Syrah virus 1 in South African grapevines. Plant Dis. 100(6), 1252.

This publication forms part of Chapter 3.

Conference contributions

Oosthuizen, K., Coetzee, B., Maree, H.J., Burger, J.T. Characterising the viromes of old and young Pinotage grapevines. Virology Africa Conference. Cape Town, South Africa. November 30 - December 3, 2015.

Poster summarising research performed in Chapter 3, presented by K. Oosthuizen.

Oosthuizen, K., Coetzee, B., Maree, H.J., Burger, J.T. Characterising the viral and fungal diversity in old and young Pinotage grapevines. 50th Anniversary Congress of the South African Society for Plant Pathology. Drakensberg, South Africa. January 15 - 19, 2017.

Presentation summarising research performed in Chapter 3 and preliminary results of Chapter 4, presented by K. Oosthuizen.

1.5 References

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

Barata, A., Malfeito-Ferreira, M., Loureiro, V., 2012. The microbial ecology of wine grape berries. Int. J. Food Microbiol. 153(3), 243-259.

Bokulich, N.A., Thorngate, J.H., Richardson, P.M., Mills, D.A., 2014. Microbial biogeography of wine grapes is conditioned by cultivar, vintage, and climate. Proc. Natl. Acad. Sci. U.S.A. 111(1), E139-E148.

Bokulich, N.A., Collins, T.S., Masarweh, C., Allen, G., Heymann, H., Ebeler, S.E., Mills, D.A., 2016. Associations among wine grape microbiome, metabolome, and fermentation behavior suggest microbial contribution to regional wine characteristics. MBio. 7(3), e00631-16. DOI: 10.1128/mBio.00631-16.

Easton, S., 2015. Strength in maturity? A look at old vines. The Drinks Business. [online] https://www.thedrinksbusiness.com/2015/05/strength-in-maturity-a-look-at-old-vines-wine/.

Gilbert, J.A., Van der Lelie, D., Zarraonaindia, I., 2014. Microbial *terroir* for wine grapes. Proc. Natl. Acad. Sci. U.S.A. 111(1), 5-6.

Heyns, E., 2013. Old vines, new opportunities? Wineland - Business & Marketing, Production. [online] http://www.wineland.co.za/old-vines-new-opportunities/.

Martelli, G.P., 2014. Directory of virus and virus-like diseases of the grapevine and their agents. J. Plant Pathol. 96(Suppl. 1), 1-136.

Pinto, C., Pinho, D., Sousa, S., Pinheiro, M., Egas, C., Gomes, A.C., 2014. Unravelling the diversity of grapevine microbiome. PLoS ONE. 9(1), e85622. DOI: 10.1371/journal.pone.0085622.

Pinto, C., Pinho, D., Cardoso, R., Custódio, V., Fernandes, J., Sousa, S., Pinheiro, M., Egas, C., Gomes, A.C., 2015. Wine fermentation microbiome: A landscape from different Portuguese wine appellations. Front. Microbiol. 6, 905. DOI: 10.3389/fmicb.2015.00905.

Zarraonaindia, I., Owens, S.M., Weisenhorn, P., West, K., Hampton-Marcell, J., Lax, S., Bokulich, N.A., Mills, D.A., Martin, G., Taghavi, S., Van der Lelie, D., Gilbert, J.A., 2015. The soil microbiome influences grapevine-associated microbiota. MBio. 6(2), e02527-14. DOI: 10.1128/mBio.02527-14.

Internet sources

Macro-economic impact of the wine industry on the South African economy (also with reference to the impacts on the Western Cape) - Final report. 2015. South African Wine Industry Information and Systems (SAWIS). [online] http://www.sawis.co.za/info/macro_study2014.php.

South African wine industry statistics - no. 41. 2016. South African Wine Industry Information and Systems (SAWIS). [online] http://www.sawis.co.za/info/annualpublication.php.

State of the vitiviniculture world market. 2016. International Organisation of Vine and Wine (OIV). [online] http://www.oiv.int/en/technical-standards-and-documents/statistical-analysis/state-of-vitiviniculture.

http://iamold.withtank.com/home/

Chapter 2: Literature review

2.1 Introduction

Grapevine is a woody perennial plant that is widely cultivated in temperate regions globally, including South Africa. It is an economically important crop, as its fruit is used for the production of table grapes, raisins, juice and wine. Grapevines are exposed to a large number of pests and pathogens. At least 70 intracellular infectious agents, exclusive of fungi, have been recorded from grapevine, the most recognised for any single crop (Martelli, 2014). These pathogens can be highly detrimental, compromising the plant's physiology, thereby causing severe losses and decreasing the overall productive lifespan of vineyards. Grapevines are also inhabited by diverse fungal and bacterial endophytes that have either neutral or beneficial effects on the host.

It is worth noting that several old vineyards showing disease symptoms have remained economically viable for the production of high-quality wines. To date, little molecular research has been done to determine the complexity of the viral, fungal and bacterial communities in old vines. Advances in genomic resources, particularly the development of next-generation sequencing (NGS) technologies and various bioinformatic tools, have significantly contributed to the field of phytopathology. Next-generation sequencing offers a cost-effective, culture-independent method to characterise the viral and microbial profiles of an environmental sample in an unbiased manner.

2.2 Grapevine diseases and associated pathogens

2.2.1 Viral diseases

At present, grapevine is susceptible to at least 65 different viruses (Martelli, 2014). There are five major viral disease complexes that affect grapevine; infectious degeneration, grapevine leafroll disease, rugose wood complex, graft incompatibility and fleck complex (Martelli, 2014). Additionally, Shiraz disease and Shiraz decline are discussed separately, as the placement of these diseases among the five complexes is ambiguous.

2.2.1.1 Infectious degeneration

The most prominent symptoms of infectious degeneration are those of fanleaf, one of the oldest viral diseases affecting grapevines globally (Martelli, 2014). In South Africa, the disease is mostly restricted to the Breede River Valley, Western Cape. Symptoms

include degeneration and malformation of the shoots and leaves (Figure 2.1A); the shoots show abnormal branching, double nodes and shortened internodes and the leaves are distorted and asymmetrical. Chlorotic mottling may accompany foliar abnormalities. Fanleaf disease also leads to malformation of the fruit, irregular ripening (Figure 2.1B), fewer and smaller bunches and decreased fruit quality. Yellow mosaic, induced by chromogenic virus strains, may occur on all vegetative tissues. Vein banding (Figure 2.1C) has also been observed in vineyards affected by infectious degeneration (Andret-Link *et al*., 2004; Martelli, 2014; Raski *et al*., 1983). Fanleaf disease is caused by *Grapevine fanleaf virus* of the genus *Nepovirus* in the family *Secoviridae* (Quacquarelli *et al*., 1976).

2.2.1.2 Grapevine leafroll disease

Among all the diseases affecting grapevines, grapevine leafroll disease (GLD) is the most prevalent and economically important, present in all grape-growing countries. Foliar symptoms of GLD differ between cultivars, but are usually more conspicuous in red cultivars. In red cultivars, premature reddening of the leaves occurs, while the primary and secondary veins remain green (Figure 2.1D). In some white cultivars, the leaves may display mild chlorotic mottling or yellowing (Figure 2.1E), while others may show no visual symptoms. In both red and white cultivars, the leaf margins usually roll downwards (Naidu *et al*., 2014). Other impacts of GLD include delayed ripening of the fruit, uneven fruit size, lower accumulation of sugars and anthocyanins, and increased titratable acidity (Naidu *et al*., 2014; Vega *et al*., 2011). Viruses of the family *Closteroviridae* are associated with the disease, and are collectively known as Grapevine leafroll-associated viruses (Boscia *et al*., 1995). Of these, *Grapevine leafroll-associated virus 3* of the genus *Ampelovirus* is recognised as the primary causal agent (Maree *et al*., 2013).

2.2.1.3 Rugose wood complex

The rugose wood complex (RWC) comprises four diseases; rupestris stem pitting, LN33 stem grooving, Kober stem grooving and corky bark (Martelli, 2014). The RWC occurs worldwide and is typically characterised by woody cylinder alterations, such as pit- and groove-like markings on either the scion or rootstock, or both (Figure 2.1F). The disease also disrupts the transport of nutrients and water through the vascular tissues and leads to vein necrosis, delayed budding and severe decline, with some vines dying within a few years (Bouyahia *et al*., 2005; Goheen, 1988; Martelli, 2014). Other symptoms

include prominent swelling at the bud union and, in certain cultivars, abnormal corky tissue production above the graft union, a disorder known as corky rugose wood (Bonavia *et al*., 1996). Several members of the genera *Vitivirus* and *Foveavirus* in the family *Betaflexiviridae* are associated with the RWC; including *Grapevine virus A*, *Grapevine virus B*, *Grapevine virus D* and *Grapevine rupestris stem pitting-associated virus* (Boscia *et al*., 2001).

2.2.1.4 Graft incompatibility

Viruses associated with graft incompatibility induce abnormalities in vines during the early stages of growth. Symptoms include stunting of young vines, short shoots, small leaves, downward rolling of the leaf margins, and prominent swelling and necrotic unions at scion-rootstock junctions (Figure 2.1G). In severe cases, declining vines may die within two years (Al Rwahnih *et al*., 2012; Golino *et al*., 2000; Martelli, 2014). A temporary form of incompatibility was observed in Italian vines, referred to as bushy stunt (Savino *et al*., 1991). *Grapevine leafroll-associated virus 2* of the genus *Closterovirus* has been associated with Kober 5BB incompatibility in Europe and, when in combination with the vitivirus, *Grapevine virus B*, young vine decline in Californian vines (Golino *et al*., 2000; Greif *et al*., 1995; Uyemoto *et al*., 2001).

2.2.1.5 Fleck complex

Prominent diseases of the fleck complex are grapevine fleck disease, asteroid mosaic and rupestris vein feathering. Associated viruses cause symptoms in *Vitis rupestris*, but occur mostly as latent infections in other species, with the exception of asteroid mosaic and rupestris vein feathering, which have been associated with symptoms in *Vitis vinifera* (Hewitt *et al*., 1972; Martelli, 2014). Fleck disease is typically characterised by localised clearing of the veinlets (Figure 2.1H), with severe flecking causing the leaves to become wrinkled and to curl upward. In vines affected by asteroid mosaic, the leaves display star-shaped chlorotic spots and appear asymmetrical and puckered along the veins. Other symptoms include vein banding, stunting and low fruit yield. Rupestris vein feathering induces mild asteroid mosaic-like symptoms and vein chlorosis (Constable & Rodoni, 2011; Martelli, 2014). Species of the genera *Maculavirus* and *Marafivirus*, family *Tymoviridae* are associated with diseases of the fleck complex, including *Grapevine fleck virus*, *Grapevine asteroid mosaic-associated virus* and *Grapevine rupestris vein feathering virus* (Boscia *et al*., 1991; Boscia *et al*., 1994; El Beaino *et al*., 2001).

2.2.1.6 Shiraz disease

Shiraz disease occurs only in South Africa; however, a similar disease, Australian Shiraz disease, has been reported in Australian grapevines (Goszczynski & Habili, 2012). The disease is characterised by reddening of the leaves and unlignified shoots (Figure 2.1I-J). The abnormal phloem development disrupts the transport of photosynthetic compounds to the grapes, affecting the quality and number of bunches and lowering the sugar content in the grapes. The vines are less vigorous and are unable to reach maturity. The most important symptom of the disease is the sudden degeneration of the plant; infected vines die within five years (Carstens, 1999; Goussard & Bakker, 2006). Grapevine virus A variant group II may have a possible association with Shiraz disease; however, the aetiology is unclear (Goszczynski *et al*., 2008; Goszczynski & Habili, 2012).

2.2.1.7 Shiraz decline

Although Shiraz decline is similar to Shiraz disease, in that affected vines decline and eventually die, the two diseases induce distinctive symptoms (Spreeth, 2005). Shiraz decline symptoms include swollen graft unions, thickening of the bark above graft unions, deep grooves and cracks on the canes (Figure 2.1K-L), premature reddening of the leaves, lower fruit yield and decreased vigour (Al Rwahnih *et al*., 2009; Spreeth, 2005). In South Africa, symptoms are only present in vines originating from the imported French Syrah clone 99, which has been discontinued to prevent disease spread (Spreeth, 2005). Grapevine rupestris stem pitting-associated virus, grapevine rupestris vein feathering virus and grapevine Syrah virus 1 have been detected in affected vines, though no association has been established and the aetiology remains unresolved (Al Rwahnih *et al*., 2009; Beuve *et al*., 2013; Goszczynski, 2010).

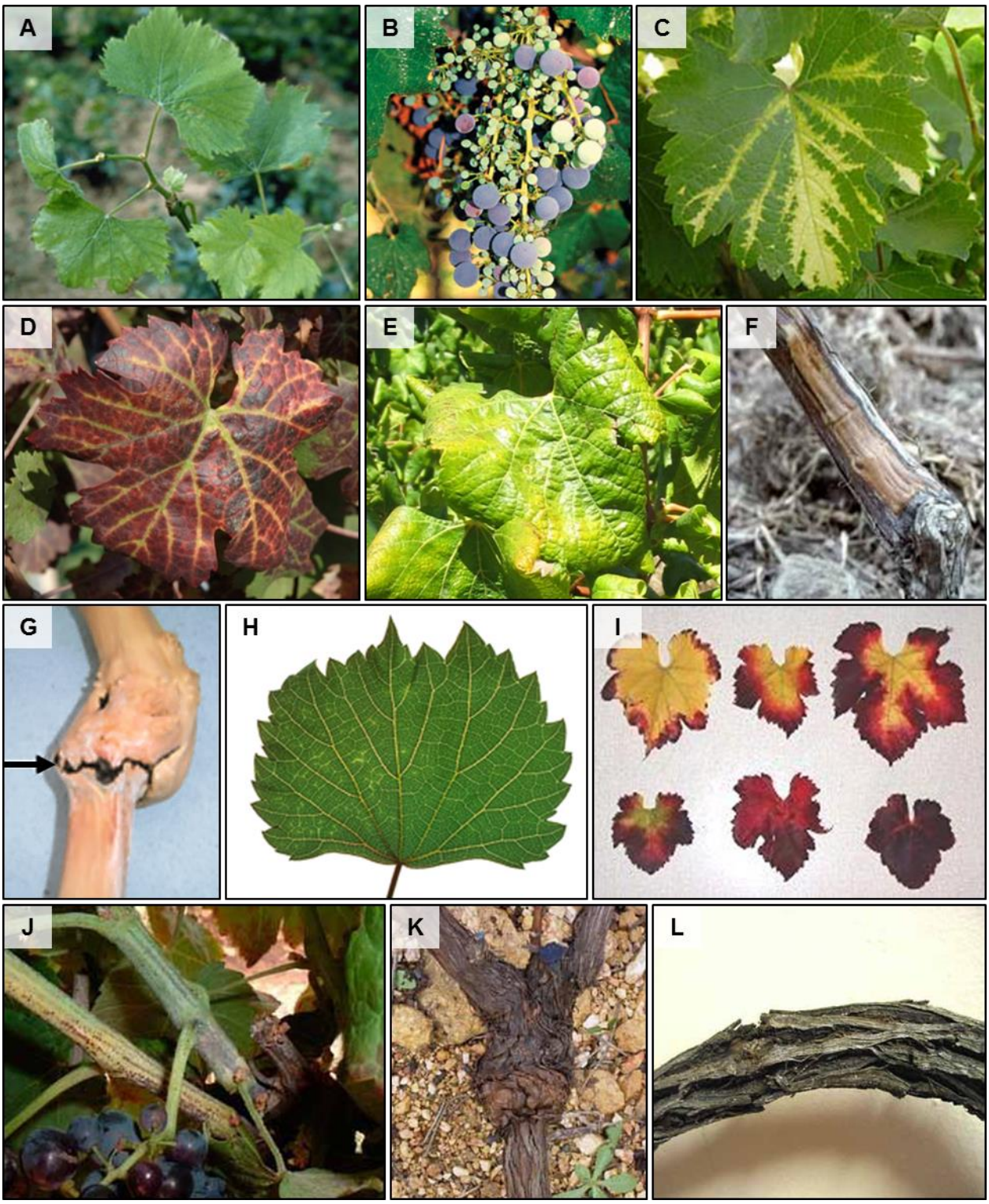

Figure 2.1: Symptoms of grapevine viral diseases. **Grapevine fanleaf disease** - A) Malformation of the leaves and shoots (Photo by W.M. Brown), B) Malformation and irregular ripening of the berries (Photo by A. Schilder) and C) Yellow vein banding (Photo by S. Jordan). **Grapevine leafroll disease** - D) Interveinal reddening of the leaves in a red cultivar) and E) Yellowing of the leaves in a white cultivar (Photos from Maree *et al*. (2013)). **Rugose wood complex** - F) Trunk displaying pit- and groove-like markings (Photo from Goussard (2013a)). **Graft incompatibility** - G) Necrotic union at the scion-rootstock junction, as indicated by the arrow (Photo from Al Rwahnih *et al*. (2012)). **Grapevine fleck disease** - H) Localised clearing of the veinlets (Photo courtesy of the University of California). **Shiraz disease** - I) Reddening of the leaves and J) Unlignified shoots (Photos from Goussard and Bakker (2006)). **Shiraz decline** - K) Swollen graft union with thickened bark (Photo from Spreeth (2005)) and L) Cane dispaying deep cracks and grooves (Photo from Goussard (2013b)).

2.2.2 Fungal diseases

Fungal pathogens of grapevine can cause economically important foliar, bunch and trunk diseases.

2.2.2.1 Foliar diseases

Powdery mildew affects the succulent tissues of grapevine, mostly the leaves. It is caused by *Erysiphe necator* and appears as a white-greyish powdery growth (Figure 2.2A) on the surface of diseased tissue (Gadoury *et al*., 2012). Young leaves may become distorted and stunted, while severely infected leaves dry out and drop prematurely. Inflorescences and berries are most vulnerable when young, with infection leading to desiccation and splitting of the berries (Figure 2.2B). Powdery mildew disrupts photosynthesis and lowers the sugar content of the fruit, resulting in decreased fruit and wine quality. The disease also causes overall decline in vine growth, vigour and yield (Gadoury *et al*., 2012; Wilcox, 2003).

Another disease affecting the green tissues is downy mildew, caused by the oomycete, *Plasmopara viticola* (Gessler *et al*., 2011). The most conspicuous symptom of this disease is the appearance of yellow, circular ‘oilspots’ on young leaves, that become dry and necrotic as they mature (Figure 2.2C). Under favourable conditions, a white downy growth will appear on the underside of infected leaves. On older leaves, the veinlets become resistant to infection, restricting the disease to small, angular spots that coalesce to form a mosaic-like pattern. When infected, young bunches turn brown (Figure 2.2D), decline and die. Mature berries, although symptomatic, do not support sporulation of the pathogen (Kennelly *et al*., 2007; Hewitt & Pearson, 1988).

2.2.2.2 Bunch diseases

A number of fungal species are associated with grapevine bunch rots, of which the most prominent is *Botrytis cinerea*, the causative agent of grey mould (McClellan *et al*., 1973). Grey mould is characterised by a grey fungal growth on the surface of ripening and mature berries, with infection spreading rapidly through contact. The pathogen invades clusters by two mechanisms; late-season infections which involve direct penetration of the berries through pores or wounds, or early-season infections that remain latent within the berries until the onset of ripening. At *véraison*, the pathogen will resume growth and progressively invade the entire cluster. First, the berries begin to rot

(Figure 2.2E), becoming soft and watery. Eventually rotted berries shrivel and drop off (McClellan *et al*., 1973; Moyer & Grove, 2011).

Other fungi that are responsible for grape rots include species of *Alternaria* (black mould) and *Colletotrichum* (ripe rot), *Greeneria uvicola* (bitter rot), *Elsinoe ampelina* (black spot) and *Guignardia bidwellii* (black rot), among many others. These fungi are mostly opportunistic pathogens or secondary invaders, whose symptoms are hard to differentiate by visual inspection (Steel *et al*., 2013). Bunch rots have negative impacts on the quality of grapes and wine. Compounds that have earthy, mushroom aromas and musty or bitter off-flavours have been isolated from diseased grapes, and wine. Certain saprophytic moulds also produce mycotoxins that contaminate the wine, further compromising the quality and introducing health risks to consumers (Steel *et al*., 2013). A well-studied example is ochratoxin A (OTA), which is produced by some *Aspergillus* species (Serra *et al*., 2005).

Trunk diseases

According to a review by Bertsch *et al*. (2013), the three most predominant grapevine trunk diseases are Eutypa dieback, esca and Botryosphaeria dieback.

The causal agent of Eutypa dieback is *Eutypa lata*, though several diatrypaceous species have also been associated with the disease (Carter, 1988; Trouillas *et al*., 2010). The pathogen penetrates through pruning wounds and colonises the vascular tissues of the trunk and cordons, developing brown wedge-shaped necrosis (Figure 2.2F). The shoots are stunted and have short internodes, while the leaves are small and cupped, with necrotic margins. Fruit infection causes berries to ripen unevenly and bunches to shrivel and drop off, leading to a decrease in fruit yield and wine quality. Infection restricts photosynthesis by degrading the tissues of the vascular system (Bertsch *et al*., 2013; Moller *et al*., 1974; Munkvold *et al*., 1993).

The esca complex comprises five syndromes; brown wood streaking, Petri disease, young esca (grapevine leaf stripe disease), esca (white rot) and esca proper (Surico *et al*., 2008; Surico, 2009). *Phaeomoniella chlamydospora*, *Phaeoacremonium aleophilum* and several wood-rotting basidiomycetes are the primary causal agents of the complex. Many other *Phaeoacremonium* species are also involved in the aetiology (Bertsch *et al*., 2013). Wood symptoms include dark streaking of the xylem and necrosis around the pith. Petri diseased vines show chlorotic leaf symptoms and reduced vigour and yield. Young esca is characterised by the appearance of spots, dispersed between the veins

or along the leaf margins that coalesce to form chlorotic, necrotic bands, a disorder known as ‘tiger stripes’ (Figure 2.2G). Infected berries display spots that have been described as ‘black measles’ (Figure 2.2H). Esca syndrome occurs in mature vines and induces white rot symptoms of the trunk and branches. Another symptom often observed in affected vines is apoplexy; a condition which involves the dieback of shoots, dropping of leaves and desiccation of bunches (Bertsch *et al*., 2013; Mugnai *et al*., 1999; Surico *et al*., 2008).

More than 20 Botryosphaeriaceae species are currently associated with the Botryosphaeria dieback complex. Three of these species, *Diplodia mutila*, *D*. *seriata* and *Neofusicoccum parvum*, are specifically associated with black dead arm (BDA) symptoms (Larignon *et al*., 2001; Lehoczky, 1974; Úrbez-Torres, 2011). The main indicator of BDA is wood necrosis of the trunk and arms. Foliar symptoms are very similar to that of young esca, but develop earlier in the season. Black dead arm also leads to brown wood streaking (Figure 2.2I) and apoplexy (Bertsch *et al*., 2013; Larignon *et al*., 2001; Lehoczky, 1974). The similarities in symptom expression have made it difficult to distinguish between BDA and young esca.

2.2.3 Bacterial diseases

Various economically important bacterial diseases have been identified in grapevine. Pierce’s disease is caused by *Xylella fastidiosa*, a bacterium that obstructs the movement of water through the xylem vessels (Hopkins, 1989; Wells *et al*., 1987). Symptoms of the disease include leaf scorch (Figure 2.3A), shrivelled berries, irregular cane lignification (‘green islands’) (Figure 2.3B), leaf separation from the petiole (‘matchsticks’), decreased vigour and eventually vine death (Goheen & Hopkins, 1988; Stevenson *et al*., 2005). The causative agent of bacterial blight is *Xylophilus ampelinus* (Willems *et al*., 1987). Diseased vines display dark streaks, cracks and cankers on the shoots, leaf spot and necrosis (Figure 2.3C) and overall decaying of the plant (Dreo *et al*., 2007). Bacterial inflorescence rot is caused by *Pseudomonas syringae* (pathovar *syringae*), with symptoms including angular leaf lesions (Figure 2.3D), longitudinal shoot lesions and rotting of inflorescences, leading to severe fruit losses (Whitelaw-Weckert *et al*., 2011). *Agrobacterium vitis* is a grapevine-associated pathogenic species that causes crown gall (Figure 2.3E). The bacterium is found systemically in vines and induces tumour formation at sites of wounding (Burr & Katz, 1984; Burr *et al*., 1998). The pathogen is disseminated through propagation material (Burr *et al*., 1998).

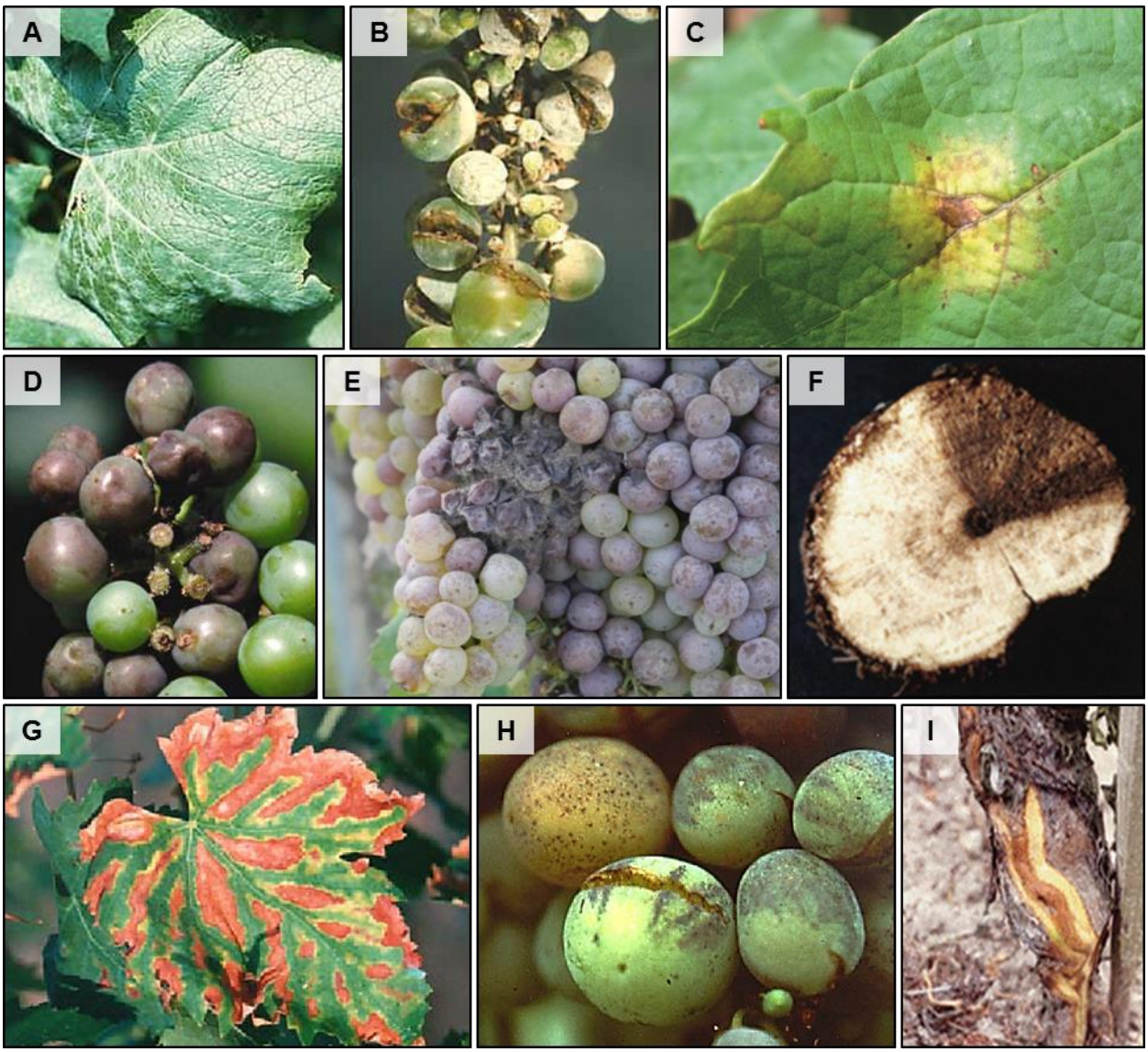

Figure 2.2: Symptoms of grapevine fungal diseases. **Powdery mildew** - A) White powdery fungal growth on the leaf surface and B) Desiccation and splitting of the berries (Photos by R. Pearson). **Downy mildew** - C) Yellow necrotic 'oilspot' (Photo from Gessler *et al*. (2011)) and D) Brown discolouration of the berries (Photo from Kennelly *et al*. (2005)). **Grey mould** - E) Grey growth, berry oozing and desiccation (Photo from Moyer and Grove (2011)). **Eutypa dieback** - F) Dark wedge-shaped necrosis of the wood (Photo from Bertsch *et al*. (2013)). **Esca** - G) Leaf displaying 'tiger stripes' and H) Berries displaying 'Black measles' (Photos from Mugnai *et al*. (1999)). **Botryosphaeria dieback** - I) Brown wood streaking (Photo from Bertsch *et al*. (2013)).

Grapevines are also susceptible to phytoplasmas, a group of obligate bacterial-like parasites with no cell wall that belong to the class Mollicutes, genus '*Candidatus* Phytoplasma' (Bertacinni, 2007). Phytoplasmas cause diseases of the grapevine yellows complex, including Australian grapevine yellows, flavescence dorée, bois noir and aster yellows. Aster yellows phytoplasma infection has been reported in South African vines, specifically in the Western Cape wine-producing regions (Engelbrecht *et al*., 2010). Affected vines display abnormal budding, premature discolouration, crackling and downward rolling of the leaves (Figure 2.3F), shortened internodes, incomplete

lignification of the shoots, and abortion of young bunches. Vines decline and eventually die (De Klerk & Carstens, 2016; Engelbrecht *et al*., 2010).

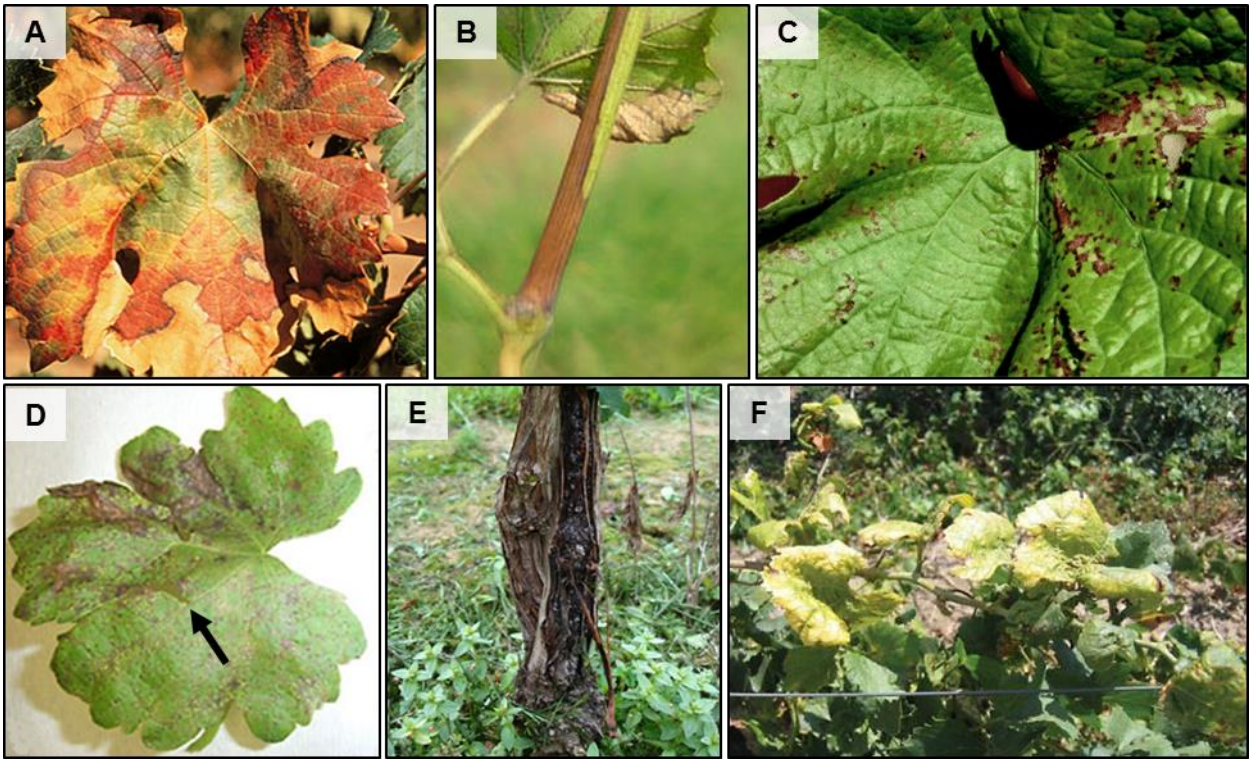

Figure 2.3: Symptoms of grapevine bacterial diseases. **Pierce's disease** - A) Leaf scorch (Photo from Goheen and Hopkins (1988)) and B) Irregular cane lignification, referred to as 'green islands' (Photo by T. Sutton). **Bacterial blight** - C) Leaf spot and necrosis (Photo from Dreo *et al*. (2007)). **Bacterial inflorescence rot** - D) Leaf displaying angular lesions, as indicated by the arrow (Photo from Whitelaw-Weckert *et al*. (2011)). **Crown gall** - E) Tumour growth on trunk (Photo by F. Westover). **Aster yellows** - F) Yellowing, crackling and downward rolling of the leaves in a white cultivar (Photo by J. Joubert).

2.3 Viroids

Viroids are subviral infectious agents of plants, consisting of non-encapsidated circular RNAs which do not code for any protein. Five distinct viroids infect grapevines globally, all of which belong to the family *Pospiviroidae* (Rezaian *et al*., 1991). Only two pospiviroids, *Grapevine yellow speckle viroid 1* and *Grapevine yellow speckle viroid 2* are pathogenic; associated with yellow speckle disease (Koltunow *et al*., 1989). This disease is characterised by chlorotic flecks dispersed over the surface, or localised along the primary veins of mature leaves, with the latter known as vein banding (Figure 2.1C). Concurrent infection with yellow speckle viroids and grapevine fanleaf virus is thought to induce vein banding symptoms (Szychowski *et al*., 1995).

2.4 Mycoviruses

Mycoviruses are viruses that infect fungi. Studies estimate that between 30 and 80% of fungal species, representing all major taxa of fungi, are inhabited by these viruses (Ghabrial & Suzuki, 2009). Most mycoviruses cause no apparent symptoms. However, studies have shown that some can induce a range of phenotypes in their fungal hosts. A well-studied example is hypovirulence, where the mycovirus attenuates the virulence of pathogenic fungi (Nuss, 2005). As a result, these viruses have been used as biocontrol agents of plant fungal diseases (Xie & Jiang, 2014). Some mycoviruses benefit their hosts by upregulating fungal virulence, while others have been implicated in mutualistic associations between fungal endophytes and plants (Ahn & Lee, 2001; Herrero *et al*., 2009). Mycoviruses are well represented in grapevine; up to six families have been detected (Al Rwahnih *et al*., 2011; Coetzee *et al*., 2010).

2.5 Endophytes

Plants are naturally colonised by a wide diversity of eukaryotic and prokaryotic endophytic microorganisms that do not cause apparent disease symptoms (Petrini, 1991; Sturz *et al*., 2000). Studies have shown that these asymptomatic organisms may enter into mutualistic relationships with their hosts, acquiring nutrients and protection from the host and in return, conferring enhanced ecological fitness (Carroll, 1988; Hallmann *et al*., 1997). Certain species have the ability to promote the plant's growth by producing phytohormones and other growth-stimulating substances, or increasing the uptake of minerals, such as nitrogen and phosphorus (Gasoni & De Gurfinkel, 1997; Malinowski & Belesky, 1999; Reis *et al*., 2000; Sturz *et al*., 2000). Another potential benefit is increased tolerance to abiotic stress (Malinowski & Belesky, 2000; Rodriguez *et al*., 2008). Some endophytes defend their hosts against fungal and bacterial pathogens, herbivores, or parasites by producing functional metabolites that have antimicrobial and other toxic properties (Tan & Zou, 2001). Others can act as inducers of host-related (systemic) resistance. For these reasons, endophytes have been applied in the biological control of plant diseases, insects and parasitic nematodes (Azevedo *et al*., 2000; Backman & Sikora, 2008; Hallmann & Sikora, 1996).

Grapevine fungal communities have been extensively studied. The following fungi have been isolated from the shoots and leaves of healthy South African vines: *Alternaria* spp., *Chaetomium* sp., *Cladosporium cladosporioides*, *Epicoccum nigrum*, *Gliocladium roseum*, *Nigrospora oryzae*, *Phoma* sp., *Pleospora herbarum*, *Sphaeropsis* sp.,

Sporormiella minimoides and *Verticillium* sp., among many other species that had lower isolation rates (Mostert *et al*., 2000). Other fungal endophytes that are commonly associated with grapevine include species of *Acremonium*, *Aureobasidium*, *Botryotinia*, *Fusarium*, *Giberella*, *Mucor*, *Nectria*, *Ophiostoma*, *Penicillium*, *Rhizopus* and *Trichoderma* (Casieri *et al*., 2009; González & Tello, 2011).

The use of endophytic fungi as biocontrol agents of grapevine diseases has also been explored. Studies have reported some level of antagonistic activity or induction of host resistance by the fungus, *Trichoderma harzianum* against the pathogens of grey mould, downy mildew, Eutypa dieback, Petri disease and black foot (Elad, 1994; Fourie *et al*., 2001; John *et al*., 2008; Palmieri *et al*., 2012). Other potential biocontrol agents of downy mildew include *Alternaria alternata*, *E. nigrum* and *Fusarium proliferatum* (Falk *et al*., 1996; Kortekamp, 1997; Musetti *et al*., 2006). *Verticillium lecanii* is described as an antagonist of powdery mildew due to its ability to penetrate and hyperparasitize *E. necator* spores (Heintz & Blaich, 1990). The yeast-like fungus, *Aureobasidium pullulans* is a well-known inhibitor of postharvest fungal pathogens, such as *B. cinerea*, *Aspergillus carbonarius* and *Aspergillus niger* (Dimakopoulou *et al*., 2008; Schena *et al*., 1999). Furthermore, this species is capable of reducing OTA contamination in must (De Felice *et al*., 2008).

The diversity of bacterial communities in grapevine is well documented. Species of the following genera have been endophytically isolated from different parts of grapevines, including the leaves and reproductive organs: *Bacillus*, *Curtobacterium*, *Enterobacter*, *Enterococcus*, *Erwinia*, *Ewingella*, *Paenibacillus*, *Pantoea*, *Pseudomonas*, *Rhodococcus*, *Staphylococcus* and *Streptomyces*, among others (Bulgari *et al*., 2009; Compant *et al*., 2011; West *et al*., 2010). Interestingly, Campisano *et al*. (2014) reported a rare inter-kingdom transfer event of the human acne-causing pathogen, *Propionibacterium acnes*, to grapevine. The bacterium, initially detected by NGS, was fluorescently visualised inside the bark, xylem and pith, likely having established itself as an obligate endophyte during the period of grapevine domestication. Other well-known human and animal pathogenic taxa, namely *Streptomyces*, *Roseomonas*, *Staphylococcus* and *Stenotophomonas*, have also been detected in the grapevine endosphere (Yousaf *et al*., 2014).

Bacterial endophytes have proved very effective in grapevine disease control. *Burkholderia phytofirmans* strain PsJN (previously named *Pseudomonas* sp.) is a plant growth-promoting rhizobacterium (PGPR) that endophytically colonises grapevine,

stimulating the growth of the plant whilst suppressing *B*. *cinerea* (Barka *et al*., 2002). Moreover, this PGPR has been shown to benefit grapevine by conferring enhanced tolerance to cold stress (Barka *et al*., 2006). *Pseudomonas fluorescens*, *Bacillus subtilis*, *Pantoea agglomerans* and *Acinetobacter lwoffi* strains have also been described as inhibitors of *B*. *cinerea*. The antagonism of these species is associated with differential induction of host-related defence mechanisms (Trotel-Aziz *et al*., 2008). Bulgari *et al*. (2011) identified bacterial endophytes in vines that have recovered from grapevine yellows phytoplasma infection. The authors suggest that phytoplasmas can alter the endophytic community of the host, selecting strains that are able to stimulate plant defence responses. However, the potential involvement of endophytes in the recovery phenomenon needs further investigation. All examples of bacterial strains with biocontrol properties, and their mechanisms of action against grapevine pathogens are reviewed by Compant *et al*. (2013).

The number of endophytic species with potential in the viticulture industry is steadily increasing. As there is still little known about the interactions among endophytes, and between these microbes and the host, it will be of great value to study such communities in greater depth.

2.6 Molecular virus detection methods

Grapevines have no natural resistance to viruses. It is therefore imperative to restrict the spread of viral diseases. Diagnostic services are in place to assist in virus certification schemes. Currently, most of these services rely on the use of sensitive and specific virus detection methods, such as serological assays and nucleic acid-based techniques. Two of the most frequently used molecular techniques for the detection of plant viruses are enzyme-linked immunosorbent assay (ELISA) and reverse transcription polymerase chain reaction (RT-PCR).

Enzyme-linked immunosorbent assay was first developed, and applied to plant viruses in the 1970s (Clark & Adams, 1977; Engvall & Perlmann, 1971). It is a plate-based method used to detect interactions between viral antigens and specific antibodies. Different formats of ELISA are available for various applications, and although they differ in the way antigens are captured or immobilised to the plate, and how the antigen-antibody complex is detected, the fundamental principle is similar (Koenig & Paul, 1982). In an ELISA, a test sample is incubated with specific antibodies conjugated to an enzyme, after which a substrate is added. Reaction of the substrate with the enzyme

that results in a change of colour indicates positive virus detection. This technique also has quantitative potential; the virus titre is proportional to the intensity of the colour reaction (Clark & Adams, 1977; Ward *et al*., 2004). This type of serological assay is a simple and economical approach for high-throughput diagnostics. Although ELISA remains one of the more established, robust virus detection methods, it has high development costs and is less sensitive than nucleic acid-based techniques (O'Donnell, 1999; Ward *et al*., 2004).

The most commonly used nucleic acid-based technique for virus detection is PCR. Detection of RNA viruses requires the application of RT-PCR, in which the RNA template is first converted to complementary DNA (cDNA). This variation is often used in plant virus diagnostics, as the majority of these viruses have RNA genomes (Ward *et al*., 2004). In order to detect several viruses or virus variants simultaneously, the PCR can be multiplexed, meaning multiple primer pairs with different pathogen targets are included in a single reaction. Multiplex assays are cost-effective, but can be difficult to develop (Henegariu *et al*., 1997; Ward *et al*., 2004). A more flexible technique for this purpose is microarray analysis; however, this technology is relatively expensive and is considered less sensitive than PCR (Boonham *et al*., 2007). Traditional PCR assays cannot accurately quantify virus titre. This difficulty was overcome with the development of real-time PCR assays, in which the amount of amplicon generated during each cycle is measured using fluorescent probes or intercalating dyes and a built-in fluorometer (Ward *et al*., 2004). All variations of PCR have the potential to be highly specific and sensitive, requiring only small amounts of nucleic acids.

Serological and nucleic acid-based methods, although effective, are mostly dependent on prior information about the virus and are unable to identify unknown viral agents, such as viruses infecting a new host, or previously uncharacterised viruses. Furthermore, considering the rapid rate at which RNA viruses mutate, divergent variants can evade detection. These limitations were overcome with the application of metagenomic shotgun NGS (Boonham *et al*., 2014).

2.7 Molecular methods in microbial community analyses

The first methods to describe microbial species relied on culturing and microscopy, with taxonomic classification based on morphology or selective staining. Initially, microorganisms cultured from a specific environment were thought to be the most abundant species in that environment. However, these species are not necessarily

numerically dominant within communities, but rather capable of multiplying rapidly on artificial growth media (Hugenholtz, 2002). One of the indicators of the discrepancy between culturable and *in situ* diversity was the 'great plate count anomaly', the observation that microscopic plate counts and viable cell counts can differ by "several orders of magnitude" (Staley & Konopka, 1985). With the discovery that many species cannot readily be grown in artificial media, it became evident that the microbial world is vastly unexplored (Handelsman, 2004). In fact, it is estimated that 99% of bacteria cannot be isolated by traditional culturing techniques, while only 5% of existing fungal species have been described (Hawksworth & Rossman, 1997, Hugenholtz, 2002). Indeed, studies using standard microbiological methods have considerably underestimated true microbial diversity.

Advances in molecular biology have spurred the development of culture-independent methods that allow more comprehensive studies of microbial ecology. The most commonly used approach is PCR-based microbial community profiling, also referred to as fingerprinting. Hybridisation methods, such as fluorescence *in situ* hybridisation and DNA microarray technology, have also been used; but these methods require extensive prior knowledge of the studied communities in order to design specific probes (Su *et al*., 2012).

Although there are multiple variations of community profiling, including the widely used denaturing gradient gel electrophoresis (DGGE) and terminal restriction fragment length polymorphism (T-RFLP), the underlying principle is the same. A typical workflow involves the extraction of total nucleic acids, followed by amplification of a microbial signature region (marker), and finally genotypic profiling of the PCR products (Nocker *et al*., 2007; Su *et al*., 2012). Selected markers should have both conserved and variable stretches, to provide a basis for universal primer design and taxonomic inference, respectively (Nocker *et al*., 2007). In principle, the region targeted should be universally homologous, capturing all microbes present in the sample, including the unculturable fraction (Rastogi & Sani, 2011). Amplified products are then subjected to genotypic profiling, during which heterogeneous fragments are separated by either gel or capillary electrophoresis. The trend is typically towards automation, as the detection of fluorescent signals provides increased sensitivity (Nocker *et al*., 2007). Depending on which variation is used, profiling may be performed with or without restriction enzyme digestion, and is based either on length or nucleotide sequence polymorphism (Nocker *et al*., 2007; Su *et al*., 2012).

In most cases, studies apply two or more culture-independent methods in combination with culturing techniques to minimise biases associated with using only one method (Su *et al*., 2012). These methods are mostly restricted to differentiating between members of communities and do not provide direct taxonomic identities of the species contributing to the signals. Profiling methods are therefore typically complemented with sequencing, where either individual signals or entire clone libraries are sequenced (Nocker *et al*., 2007; Rastogi & Sani, 2011). Community profiling also lacks the sensitivity to detect low-abundance taxa, a limitation that was overcome with the development of NGS, which allows analysis at a far greater depth.

2.8 Metagenomics

The term 'metagenomics' was first coined in 1998, and is described as the study of genetic material directly isolated from environmental samples (Handelsman *et al*., 1998; Handelsman, 2004). More specifically, it is a method to investigate the complexity of microbial communities that inhabit an environment through analysis of nucleotide sequence content (Hugenholtz & Tyson, 2008). Metagenomic studies may apply either a shotgun or targeted sequencing approach, depending on the type of environmental survey conducted. The first step in a metagenomics workflow is to extract total nucleic acids from an uncultured, unpurified sample. The genetic material is then randomly fragmented, or molecular markers are amplified. Traditionally, the fragments were cloned into vectors and the library of clones, Sanger sequenced (Cardenas & Tiedje, 2008; Hugenholtz & Tyson, 2008). However, the use of cloning strategies and capillary electrophoresis-based sequencing make this approach costly, time-consuming and labour-intensive.

2.8.1 Next-generation sequencing

The advent of NGS technology has revolutionised the field of metagenomics. Next-generation sequencing, also referred to as high-throughput, massively parallel or deep sequencing, has many advantages over Sanger sequencing: vast amounts of sequence data are produced in a single run, with a significantly lower cost per base; nucleic acids are sequenced directly, circumventing the need for cloning; and the use of universal adaptors, instead of specific primers allows researchers to sequence the combined genetic material of an environmental sample in an unbiased way, without requiring *a priori* biological or molecular knowledge. Furthermore, NGS can be used to estimate the relative abundance of viruses or microorganisms, as the sequence read count is

approximately proportional to the population frequency (Mardis, 2008a). This proves NGS to be a rapid, high-throughput and cost-effective tool in large-scale metagenomic studies (Hall, 2007; Mardis, 2008a).

Next-generation sequencing comprises three general steps, namely sample library preparation, clonal amplification (template preparation) and high-throughput sequencing (Mardis, 2008b; Metzker, 2010). Similar to traditional metagenomics, either a shotgun or targeted strategy is followed. During library preparation, the fragments or amplicons are ligated to adapters and immobilised on microbeads or a flow cell, depending on which NGS platform is utilised. The library is clonally amplified using the adaptors as primers, generating clusters of identical fragments. This step is a prerequisite for platforms that are not sensitive enough to detect the incorporation of single molecules during sequencing. Amplified templates are therefore necessary to enhance the signal-to-noise ratio (Mardis, 2013; Metzker, 2010). Finally, the clusters are subjected to massively parallel sequencing. So-called 'third-generation' platforms do not require a pre-sequencing amplification step and are capable of detecting single molecules, and capturing the sequencing signal in real time. These systems reduce bias and errors introduced by PCR and have a significantly shorter run time (Liu *et al*., 2012; Metzker, 2010). A number of platforms are currently available, each with inherent strengths and constraints. These platforms use different template preparation methods and sequencing signal-detecting chemistries (Mardis, 2013; Van Dijk *et al*., 2014). Current NGS studies tend to exploit Illumina, Ion Torrent or Pacific Biosciences sequencing technology (Mardis, 2013).

2.8.2 Viral metagenomics

No universal genes or highly conserved genomic regions exist across all viral families, and therefore a targeted approach like deep amplicon sequencing is not feasible (Edwards & Rohwer, 2005). Shotgun NGS has the potential to identify all virus species and genetic variants present in a sample, regardless whether they are known or novel, or if they occur at low titres (Al Rwahnih *et al*., 2009; Mokili *et al*., 2012). This approach has been extensively used to characterise the total viral complement, also referred to as the virome, of different environments. In plant virome studies, different nucleic acid preparations have served as starting material for library preparation, including total nucleic acids, double-stranded RNA (dsRNA), small RNAs (sRNAs) and virion-associated nucleic acids from partially or completely purified viral-like particles. The

advantages and limitations of each of these approaches are reviewed by Massart *et al*. (2014) and Roossinck *et al*. (2015).

The extraction of RNA, followed by cDNA synthesis using random primers enables direct sequencing of RNA viruses and viroids, as well as the transcripts of actively replicating DNA viruses (Adams *et al*., 2009). Sequencing cDNA, rather than genomic DNA, will also significantly reduce host background; only ribosomes and active host genes will be captured and sequenced (Adams *et al*., 2009). However, additional steps are usually necessary to further deplete host RNA, so as to enhance the signals of viral pathogens towards the level of detection (Kreuze *et al*., 2009). In RNA-Seq experiments, viruses can be detected after mRNA enrichment by ribo-depletion or poly-A selection; but sequencing poly-A selected RNA may not effectively capture non-adenylated viruses (Boonham *et al*., 2014; Visser *et al*., 2016). Adams *et al*. (2009) enriched for non-plant RNA by means of a subtractive hybridisation technique, in which cDNA from the infected plant is subtracted from that of an uninfected plant, thereby limiting the amount of sequencing required.

Another frequently used virus enrichment strategy is to purify dsRNA molecules that accumulate during the replication of both positive- and negative-sense single-stranded RNA (ssRNA) viruses (Al Rwahnih *et al*., 2009). Considering that endogenous plant RNAs do not readily form double-stranded structures, the presence of these molecules is indicative of viral infection (Dodds *et al*., 1984). This method has therefore proved very effective in minimising host contamination. Disadvantages of using dsRNA as starting material include limited detection of DNA viruses and the particular bias for dsRNA viruses (Roossinck *et al*., 2010). However, as most plant-infecting viruses have RNA genomes, dsRNA enrichment remains a popular approach for virome sequencing.

Since the pioneering work of Kreuze *et al*. (2009), various studies have sequenced sRNAs to characterise plant virus populations. This method relies on the mechanisms of RNA interference, where the plant's natural antiviral defence pathway generates sRNAs corresponding to invading viruses (Mlotshwa *et al*., 2008). This strategy has the potential to detect viruses of all genome types based on the virus-derived sRNAs, making it a universal method for virus enrichment and characterisation (Massart *et al*., 2014). Perhaps the most challenging aspect of sRNA profiling is the bioinformatic assembly of the short sequences; sRNAs are only 21 to 25 nucleotides in length. This is further complicated by infection of closely related virus strains (Boonham *et al*., 2014).

All of the strategies discussed have been successfully applied to the study of grapevine viromes (Al Rwahnih *et al*., 2009; Coetzee *et al*., 2010; Pantaleo *et al*., 2010). Most viruses and all viroids known to infect grapevine have ssRNA as their genetic material. Furthermore, the majority of mycoviruses that have been identified in grapevine have either ssRNA or dsRNA genomes. Presently, only two grapevine-infecting DNA viruses are recognised (Martelli, 2014), neither of which has been reported in South African grapevines.

2.8.3 Microbial metagenomics

Shotgun NGS may be used to characterise the microbiome of an environmental sample; however, the phylogenetic complexity of microbial communities and the generation of short reads that are “dissociated from their original species”, complicates genome assembly (Wooley *et al*., 2010). Therefore a targeted approach like deep amplicon sequencing can be considered more feasible. A typical workflow involves the extraction of total DNA and subsequent amplification of genomic regions that can be used for taxonomic classification. This is followed by amplicon library preparation and massively parallel sequencing.

The most frequently used target for fungi and bacteria is the DNA encoding for ribosomal operons, which are present across all organisms and in multiple copies in each cell, and contain both conserved domains and variable regions. These features make rRNA a universal and sensitive microbial marker (Ward *et al*., 2004). The predominant marker for the assessment of fungal diversity is the nuclear ribosomal internal transcribed spacer (ITS) region, recently designated as the universal fungal barcode, whereas the 16S rRNA gene is considered the gold standard for the study of bacterial ecology (Weisburg *et al*., 1991; Schoch *et al*., 2012).

One of the disadvantages of NGS, as opposed to Sanger sequencing, is the lower taxonomic resolution due to the shorter reads. Even with advances in NGS technologies, the current read length capabilities of most platforms do not permit full-length ITS and 16S rRNA gene sequencing. Considerably longer reads can be generated by the Pacific Biosystems technology, however at the expense of higher error rates. Instead, metabarcoding studies target the ITS1 or ITS2 subregion, and individual or adjacent hypervariable regions within the 16S rRNA gene (Cardenas & Tiedje, 2008; Lindahl *et al*., 2013). As a consequence, taxonomic resolution at the species level may be limited. Regardless, deep amplicon sequencing is an effective method to determine

and compare the microbial richness and composition of environmental samples. Unlike clone library Sanger sequencing and conventional community profiling methods, NGS technologies allow analysis at an unprecedented depth, and will therefore be able to detect many more rare taxa.

Deep amplicon sequencing has been widely applied in plant microbial ecology studies, mostly focussing on rhizosphere and phyllosphere communities, with particular interest in endophytes. Recent years have seen an increase in grapevine microbiome studies, in which regions of both the ITS and 16S rRNA gene were amplified (Bokulich *et al*., 2014; Bokulich *et al*., 2016; Perazzolli *et al*., 2014). Other studies have also sequenced the D1 and/or D2 region of the large subunit rRNA gene to characterise yeast profiles (Pinto *et al*., 2014; Pinto *et al*., 2015; Taylor *et al*., 2014).

2.9 Bioinformatics

Next-generation sequencing generates large quantities of data, which presents significant challenges for storage, management and analysis. To overcome these difficulties, substantial computational resources are needed. Several bioinformatic tools have been developed for this purpose, each with a specific application (Zhang *et al*., 2011).

2.9.1 Bioinformatics associated with viral metagenomics

The first step in the pipeline is to trim low-quality reads to improve the data for downstream analyses. Additionally, sequences that are derived from the host can be filtered from the data (Mokili *et al*., 2012). There are two main types of analyses performed on shotgun NGS data, namely *de novo* assembly and reference-based assembly, also referred to as read-mapping (Scholz *et al*., 2012). However, *de novo* assembly is regarded the main approach for virus metagenome reconstruction, and remains the only feasible method to identify sequences of an unknown virus, for which there is no existing reference genome (Massart *et al*., 2014).

One of the main dissimilarities between viruses and other biological entities is that there is a higher degree of sequence variation within virus populations (Roossinck *et al*., 2015). As a result, the stringency with which genome assemblies are performed is lower for viruses compared to other cellular organisms. Conversely, relaxed parameters may increase the probability of sequence chimerisation, i.e. the assembly of contigs that are comprised of multiple genetically distinct genomes (Roossinck *et al*., 2015). The most

efficient *de novo* assembly algorithms work by dividing the reads into smaller fragments of a defined length, called k-mers. The k-mers are subsequently used to build De Bruijn graphs (Figure 2.4), from which larger contiguous sequences, termed contigs, are constructed from the overlapping fragments. If paired-end sequencing was performed, the known insert length between the linked forward and reverse reads is used for scaffolding (Zerbino & Birney, 2008). The use of De Bruijn graphs shortens the time of assembly, but has large memory requirements (Scholz *et al*., 2012).

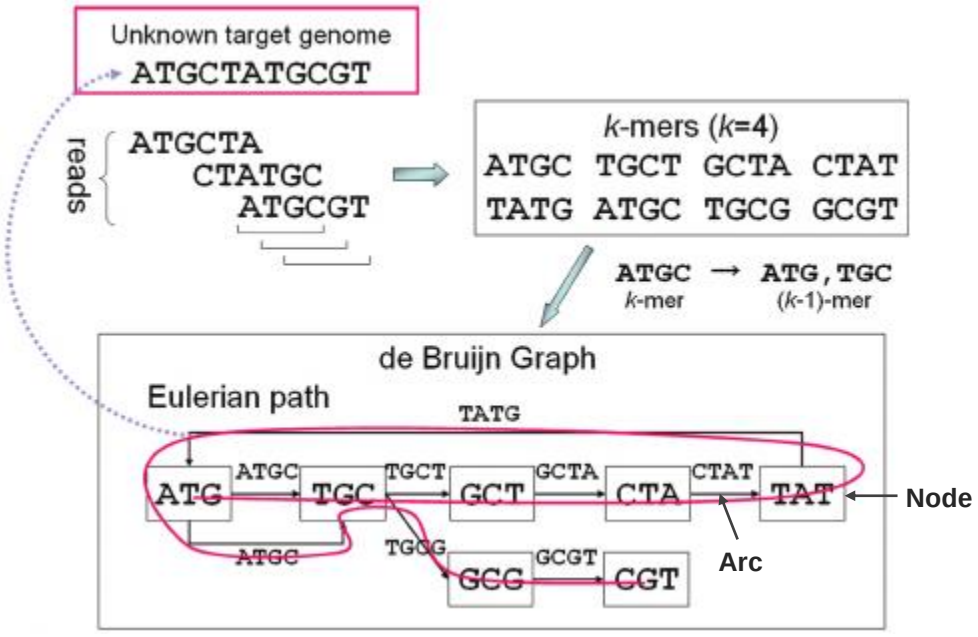

Figure 2.4: De Bruijn graph-based assembly. The sequence reads are first broken down into shorter fragments, termed k-mers, which are then linked in a De Bruijn graph. Overlapping k-mers are represented as nodes that are connected by arcs. A correct path through the nodes, called the Eulerian path, represents the sequence of the original reads or genome. Image adapted from Namiki *et al*. (2012).

Assembled contigs are usually classified by alignment-based similarity searches like BLAST against local or publicly available sequence databases (Altschul *et al*., 1990; Mokili *et al*., 2012). Notably, even when enriching for virus-specific nucleic acids and using sensitive searches like tBLASTx, a large portion of generated sequences has no significant similarity to any known nucleotide or amino acid sequences (Mokili *et al*., 2012; Roossinck *et al*., 2015). It is therefore likely that a number of viruses go undetected due to insufficient reference sequence data.

There are various open-source bioinformatic tools; including the widely used Velvet *de novo* assembler and Burrows-Wheeler Alignment tool (Li & Durbin, 2009; Zerbino & Birney, 2008). Most of these tools are Linux-based, requiring the user to become familiar with a command-line interface. The development of commercial software

packages with graphical user interfaces offers researchers a simplified approach to visualise and interpret NGS data. One such package is CLC Genomics Workbench (https://www.qiagenbioinformatics.com/); a platform that incorporates a versatile toolkit for quality control, *de novo* assembly and read-mapping, among other features.

2.9.2 Bioinformatics associated with microbial metagenomics

The first microbial ecology studies to make use of NGS discovered the so-called 'rare biosphere' (Sogin *et al*., 2006). However, the detection of these low-abundance populations was, in most cases, attributed to sequencing errors, which resulted in inflated diversity estimates. It is therefore imperative to apply extensive quality control measures in order to distinguish between sequencing errors and true biological diversity (Huse *et al*., 2010; Kunin *et al*., 2010).

The extent and type of sequencing errors is dependent on the chemistry underlining the technology. Until recently, the majority of microbiome studies were performed using pyrosequencing, which has high error rates in homopolymeric repeat regions (Balzer *et al*., 2011; Metzker, 2010). The current trend is towards the Illumina MiSeq platform, which offers higher coverage and lower costs (Bokulich *et al*., 2013; Kozich *et al*., 2013; Schmidt *et al*., 2013). This technology can be exploited to generate overlapping paired-end reads that can then merged to create larger consensus sequences, and to assist in error filtering by resolving mismatches in the alignment (Edgar & Flyvbjerg, 2015). Illumina reads may be filtered based on average Phred quality scores, an approach implemented in the Quantitative Insights Into Microbial Ecology (QIIME) software (Bokulich *et al*., 2013; Caporaso *et al*., 2010). Edgar and Flyvbjerg (2015) proposed using the expected number of errors in a more effective filtering strategy, calculated as the sum of the error probabilities as predicted by the individual quality scores.

The other key aspect of amplicon data analysis is to distinguish between chimeric and correct amplicon sequences. Chimeras are artefacts that are generated during PCR when partially extended strands act as primers to amplify related fragments (Schloss *et al*., 2011). It is estimated that approximately 5% of sequences in public repositories are anomalous, with the majority displaying chimeric patterns (Ashelford *et al*., 2005). In a subsequent study, chimeras were shown to represent up to 45% of sequences in 16S rRNA gene libraries (Ashelford *et al*., 2006). If undetected, these PCR artefacts may be falsely identified as novel taxa, leading to spurious inferences of microbial diversity

(Haas *et al*., 2011). A number of programs are available that can detect chimeras either *de novo* or by searching a chimera-free reference database.

After filtering, the sequences are clustered into operational taxonomic units (OTUs) at a pre-defined level of similarity, relevant to the study. It is generally accepted that sequences clustered at a 97% identity threshold correspond approximately to species; however precise species delimitation requires taxonomic expertise. The most efficient clustering method is implemented in the UPARSE pipeline, which has been optimised to reduce spurious OTUs (Edgar, 2013). Clustering is performed by identifying a set of OTU representative sequences (centroids) that share less than 97% pairwise identity, discarding chimeras, and comparing the input sequences to the existing OTUs. Assignment to an OTU is based on 97% or more sequence identity (Figure 2.6A-B). The more abundant input sequences are usually selected as OTU centroids, as these are more likely to reflect true biological sequences (Edgar, 2013).

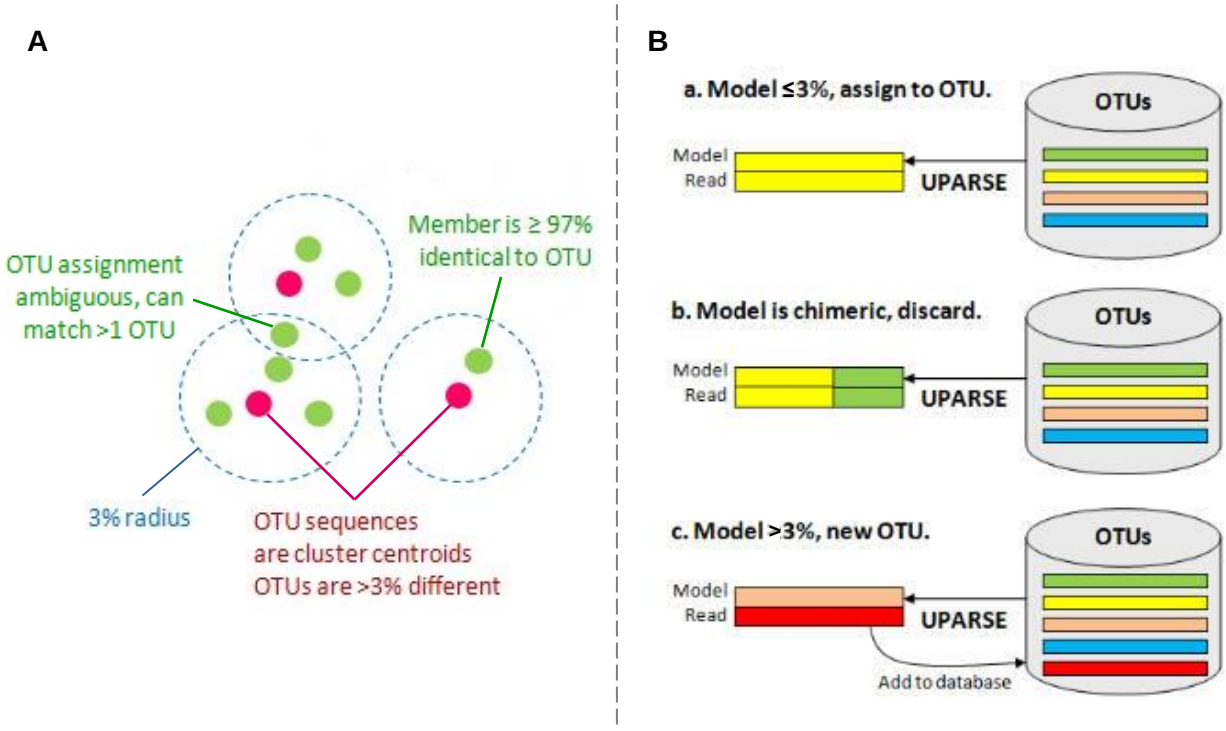

Figure 2.5: UPARSE OTU clustering. A) The UPARSE-OTU algorithm identifies OTU centroids that differ by more than 3%, thereby creating an OTU reference database. B) UPARSE-REF then compares all of the input sequences to the database, finding a maximum parsimony model for each sequence. The sequence is: a) assigned to an OTU, if the model is 97% or more identical to the OTU, b) discarded, if chimeric or c) added to the database as the centroid of a new OTU, if less than 97% identical to any existing OTU. Images adapted from http://www.drive5.com/usearch/manual/uparseotu_algo.html.

Taxonomic classification of OTUs is based on the closest match to a sequence in an annotated reference database, such as UNITE, the Ribosomal Database Project (RDP)

or Greengenes (Kõljalg *et al*., 2013; Cole *et al*., 2014; McDonald *et al*., 2012). The most widely used methods for taxonomic analysis are the RDP naïve Bayesian classifier and BLAST (Altschul *et al*., 1990; Wang *et al*., 2007). Additional classifiers are supported by the QIIME and Mothur software packages (Caporaso *et al*., 2010; Schloss *et al*., 2009). Perhaps the most significant constraint of taxonomic inference is the limited reference data, which could lead to high rates of false positive errors, particularly in datasets dominated by novel taxa.

The final step in microbial data analysis is to determine the sample diversity. Both QIIME and Mothur incorporate multiple diversity metrics. QIIME also generates graphs of rarefaction curves, which are useful to determine whether the sequencing depth is adequate to capture the complete microbial community within the samples (Kuczynski *et al*., 2011). Rarefaction analysis is performed by random subsampling, without replacement, at different intervals. The estimated diversity (OTU richness) is then plotted as a function of the simulated sequencing depth (Kuczynski *et al*., 2011). Species richness counts can be validly compared only once the curve reaches an asymptote (Gotelli & Colwell, 2001).

2.10 Conclusion

Grapevine is susceptible to a large spectrum of pathogens that compromise the plant's health, reducing the yield and quality of the fruit and influencing the character of the wine. This plant is also host to diverse endophytic microbial communities that confer enhanced ecological fitness by providing increased protection against debilitating organisms and promoting plant growth. In fact, there is a growing list of species with potential in the viticulture industry, specifically as biocontrol agents, some of which are already commercially available. In South Africa, several vineyards older than 35 years have remained economically productive, despite the prolonged exposure to biotic and abiotic stresses. Limited research has been performed to investigate the pathogen and endophyte populations in these old vines. Next-generation sequencing presents an unprecedented approach to characterise the virome and microbiome of grapevine, without the need for prior sequence knowledge. The data generated in this research could not only support plant protection and crop improvement initiatives, but also contribute to the study of the natural environment of grapes and wine *terroir*, as influenced by the grapevine microbiome.

2.11 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.

Al Rwahnih, M., Daubert, S., Golino, D., Rowhani, A., 2009. Deep sequencing analysis of RNAs from a 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.

Al Rwahnih, M., Rowhani, A., Smith, R.J., Uyemoto, J.K., Sudarshana, M.R., 2012. Grapevine necrotic union, a newly recognized disease of unknown etiology in grapevine grafted on 110 Richter rootstock in California. J. Plant Pathol. 94(1), 149-156.

Altschul, S.F., Gish, W., Miller, W., Myers, E.W., Lipman, D.J., 1990. Basic Local Alignment Search Tool. J. Mol. Biol. 215(3), 403-410.

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.

Ashelford, K.E., Chuzhanova, N.A., Fry, J.C., Jones, A.J., Weightman, A.J., 2005. At least 1 in 20 16S rRNA sequence records currently held in public repositories is estimated to contain substantial anomalies. Appl. Environ. Microbiol. 71(12), 7724-7736.

Ashelford, K.E., Chuzhanova, N.A., Fry, J.C., Jones, A.J., Weightman, A.J., 2006. New screening software shows that most recent large 16S rRNA gene clone libraries contain chimeras. Appl. Environ. Microbiol. 72(9), 5734-5741.

Azevedo, J.L., Maccheroni, W., Pereira, J.O., De Araújo, W.L., 2000. Endophytic microorganisms: A review on insect control and recent advances on tropical plants. Electron. J. Biotechnol. 3(1), 40-65.

Backman, P.A., Sikora, R.A., 2008. Endophytes: An emerging tool for biological control. Biol. Control. 46(1), 1-3.

Balzer, S., Malde, K., Jonassen, I., 2011. Systematic exploration of error sources in pyrosequencing flowgram data. Bioinformatics. 27(13), i304-i309.

Barka, E.A., Gognies, S., Nowak, J., Audran, J., Belarbi, A., 2002. Inhibitory effect of endophyte bacteria on *Botrytis cinerea* and its influence to promote the grapevine growth. Biol. Control. 24(2), 135-142.

Barka, E.A., Nowak, J., Clément, C., 2006. Enhancement of chilling resistance of inoculated grapevine plantlets with a plant growth-promoting rhizobacterium, *Burkholderia phytofirmans* strain PsJN. Appl. Environ. Microbiol. 72(11), 7246-7252.

Bertacinni, A., 2007. Phytoplasmas: Diversity, taxonomy, and epidemiology. Front. Biosci. 12(2), 673-698.

Bertsch, C., Ramírez-Suero, M., Magnin-Robert, M., Larignon, P., Chong, J., Abou-Mansour, E., Spagnolo, A., Clément, C., Fontaine, F., 2013. Grapevine trunk diseases: Complex and still poorly understood. Plant Pathol. 62(2), 243-265.

Beuve, M., Moury, B., Spilmont, A., Sempé-Ignatovic, L., Hemmer, C., Lemaire, O., 2013. Viral sanitary status of declining grapevine Syrah clones and genetic diversity of Grapevine rupestris stem pitting-associated virus. Eur. J. Plant Pathol. 135(2), 439-452.

Bokulich, N.A., Subramanian, S., Faith, J.J., Gevers, D., Gordon, J.I., Knight, R., Mills, D.A., Caporaso, J.G., 2013. Quality-filtering vastly improves diversity estimates from Illumina amplicon sequencing. Nat. Methods. 10(1), 57-59.

Bokulich, N.A., Thorngate, J.H., Richardson, P.M., Mills, D.A., 2014. Microbial biogeography of wine grapes is conditioned by cultivar, vintage, and climate. Proc. Natl. Acad. Sci. U.S.A. 111(1), E139-E148.

Bokulich, N.A., Collins, T.S., Masarweh, C., Allen, G., Heymann, H., Ebeler, S.E., Mills, D.A., 2016. Associations among wine grape microbiome, metabolome, and fermentation behavior suggest microbial contribution to regional wine characteristics. MBio. 7(3), e00631-16. DOI: 10.1128/mBio.00631-16.

Bonavia, M., Digiaro, M., Boscia, D., Boari, A., Bottalico, G., Savino, V., Martelli, G.P., 1996. Studies on corky rugose wood of grapevine and on the diagnosis of Grapevine virus B. Vitis. 35(1), 53-58.

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

Boonham, N., Kreuze, J., Winter, S., Van der Vlugt, R., Bergervoet, J., Tomlinson, J., Mumford, R., 2014. Methods in virus diagnostics: From ELISA to next generation sequencing. Virus Res. 186, 20-31.

Boscia, D., Martelli, G.P., Savino, V., Castellano, M.A., 1991. Identification of the agent of grapevine fleck disease. Vitis. 30(2), 97-105.

Boscia, D., Sabanadzovic, S., Savino, V., Kyriakopoulou, P.E., Martelli, G.P., Lafortezza, R., 1994. A non-mechanically transmissible virus associated with asteroid mosaic of the grapevine. Vitis. 33(2), 101-102.

Boscia, D., Greif, C., Gugerli, P., Martelli, G.P., Walter, B., Gonsalves, D., 1995. Nomenclature of grapevine leafroll-associated putative closteroviruses. Vitis. 34(3), 171-175.

Boscia, D., Digiaro, M., Safi, M., Garau, R., Zhou, Z., Minafra, A., Abou Ghanem-Sabanadzovic, N., Bottalico, G., Potere, O., 2001. Production of monoclonal antibodies to Grapevine virus D and contribution to the study of its aetiological role in grapevine diseases. Vitis. 40(2), 69-74.

Bouyahia, H., Boscia, D., Savino, V., La Notte, P., Pirolo, C., Castellano, M.A., Minafra, A., Martelli, G.P., 2005. Grapevine rupestris stem pitting-associated virus is linked with grapevine vein necrosis. Vitis. 44(3), 133-137.

Bulgari, D., Casati, P., Brusetti, L., Quaglino, F., Brasca, M., Daffonchio, D., Bianco, P.A., 2009. Endophytic bacterial diversity in grapevine (*Vitis vinifera* L.) leaves described by 16S rRNA gene sequence analysis and length heterogeneity-PCR. J. Microbiol. 47(4), 393-401.

Bulgari, D., Casati, P., Crepaldi, P., Daffonchio, D., Quaglino, F., Brusetti, L., Bianco, P.A., 2011. Restructuring of endophytic bacterial communities in grapevine yellows-diseases and recovered *Vitis vinifera* L. plants. Appl. Environ. Microbiol. 77(14), 5018-5022.

Burr, T.J., Katz, B.H., 1984. Grapevine cuttings as potential sites of survival and means of dissemination of *Agrobacterium tumefaciens*. Plant Dis. 68(11), 976-978.

Burr, T.J., Bazzi, C., Süle, S., Otten, L., 1998. Crown gall of grape - Biology of *Agrobacterium vitis* and the development of disease control strategies. Plant Dis. 87(12), 1288-1297.

Campisano, A., Ometto, L., Compant, S., Pancher, M., Antonielli, L., Yousaf, S., Varotto, C., Anfora, G., Pertot, I., Sessitsch, A., Rota-Stabelli, O., 2014. Interkingdom transfer of the acne-causing agent, *Propionibacterium acnes*, from human to grapevine. Mol. Biol. Evol. 31(5), 1059-1065.

Caporaso, J.G., Kuczynski, J., Stombaugh, J., Bittinger, K., Bushman, F.D., Costello, E.K., Fierer, N., Peña, A.G., Goodrich, J.K., Gordon, J.I., Huttley, G.A., Kelley, S.T., Knights, D., Koenig, J.E., Ley, R.E., Lozupone, C.A., McDonald, D., Muegge, B.D., Pirrung, M., Reeder, J., Sevinsky, J.R., Turnbaugh, P.J., Walter, W.A., Widmann, J., Yatsunenko, T., Zaneveld, J., Knight, R., 2010. QIIME allows analysis of high-throughput community sequencing data. Nat. Methods. 7(5), 335-336.

Cardenas, E., Tiedje, J.M., 2008. New tools for discovering and characterizing microbial diversity. Curr. Opin. Biotechnol. 19(6), 544-549.

Carroll, G., 1988. Fungal endophytes in stems and leaves: From latent pathogen to mutualistic symbiont. Ecology. 69(1), 2-9.

Carstens, R., 1999. Shiraz disease: Symptoms and management. Wineland - Production. [online] http://www.wineland.co.za/shiraz-disease-symptoms-and-management/.

Carter, M.V., 1988. Eutypa dieback. In: Pearson, R.C., Goheen, A.C. (Eds.). Compendium of grape diseases, 32-34. APS Press, St Paul, MN, USA.

Casieri, L., Hofstetter, V., Viret, O., Gindro, K., 2009. Fungal communities living in the wood of different cultivars of young *Vitis vinifera* plants. Phytopathol. Mediterr. 48(1), 73-83.

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.

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

Cole, J.R., Wang, Q., Fish, J.A., Chai, B., McGarrell, D.M., Sun, Y., Brown, C.T., Porras-Alfaro, A., Kuske, C.R., Tiedje, J.M., 2014. Ribosomal Database Project: Data and tools for high throughput rRNA analysis. Nucleic Acids Res. 42(Database issue), D633-D642.

Compant, S., Mitter, B., Colli-Mull, J.G., Gangl, H., Sessitsch, A., 2011. Endophytes of grapevine flower, berries, and seeds: Identification of cultivable bacteria, comparison with other plant parts, and visualization of niches of colonization. Microb. Ecol. 62(1), 188-197.

Compant, S., Brader, G., Muzammil, S., Sessitsch, A., Lebrihi, A., Mathieu, F., 2013. Use of beneficial bacteria and their secondary metabolites to control grapevine pathogen diseases. Biocontrol. 58(4), 435-455.

Constable, F., Rodoni, B., 2011. Grapevine fleck and associated viruses. Grape and Wine Research and Development Corporation, Fact sheet, July 2011.

De Felice, D.V., Solfrizzo, M., De Curtis, F., Lima, G., Visconti, A., Castoria, R., 2008. Strains of *Aureobasidium pullulans* can lower ochratoxin A contamination in wine grapes. Phytopathology. 98(12), 1261-1270.

De Klerk, A., Carstens, R., 2016. Aster yellows and leafhoppers. Wineland - Viticulture research, Winetech Technical. [online] http://www.wineland.co.za/aster-yellows-and-leafhoppers/.

Dimakopoulou, M., Tjamos, S.E., Antoniou, P.P., Pietri, A.P., Battilani, P., Avramidis, N., Markakis, E.A., Tjamos, E.C., 2008. Phyllosphere grapevine yeast *Aureobasidium pullulans* reduces *Aspergillus carbonarius* (sour rot) incidence in wine-producing vineyards in Greece. Biol. Control. 46(2), 158-165.

Dodds, J.A., Morris, T.J., Jordan, R.L., 1984. Plant viral double-stranded RNA. Annu. Rev. Phytopathol. 22, 151-168.

Dreo, T., Gruden, K., Manceau, C., Janse, J.D., Ravnikar, M., 2007. Development of a real-time PCR-based method for detection of *Xylophilus ampelinus*. Plant Pathol. 56(1), 9-16.

Edgar, R.C., 2013. UPARSE: Highly accurate OTU sequences from microbial amplicon reads. Nat. Methods. 10(10), 996-998.

Edgar, R.C., Flyvbjerg, H., 2015. Error filtering, pair assembly and error correction for next-generation sequencing reads. Bioinformatics. 31(21), 3476-3482.

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

Elad, Y., 1994. Biological control of grape grey mould by *Trichoderma harzianum*. Crop Prot. 13(1), 35-38.

Engelbrecht, M., Joubert, J., Burger, J.T., 2010. First report of aster yellows phytoplasma in grapevines in South Africa. Plant Dis. 94(3), 373.

Engvall, E., Perlmann, P., 1971. Enzyme-linked immunosorbent assay (ELISA) - Quantitative assay of immunoglobulin G. Immunochemistry. 8(9), 871-874.

El Beaino, T., Sabanadzovic, S., Digiaro, M., Abou Ghanem-Sabanadzovic, N., Rowhani, A., Kyriakopoulou, P.E., Martelli, G.P., 2001. Molecular detection of Grapevine fleck virus-like viruses. Vitis. 40(2), 65-68.

Falk, S.P., Pearson, R.C., Gadoury, D.M., Seem, R.C., Sztejnberg, A., 1996. *Fusarium proliferatum* as a biocontrol agent against grape downy mildew. Phytopathology. 86(10), 1010-1017.

Fourie, P.H., Halleen, F., Van der Vyver, J., Schreuder, W., 2001. Effect of *Trichoderma* treatments on the occurrence of decline pathogens in the roots and rootstocks of nursery grapevines. Phytopathol. Mediterr. 40(3), 473-478.

Gadoury, D.M., Cadle-Davidson, L., Wilcox, W.F., Dry, I.B., Seem, R.C., Milgroom, M.G., 2012. Grapevine powdery mildew (*Erysiphe necator*): A fascinating system for the study of the biology, ecology and epidemiology of an obligate biotroph. Mol. Plant Pathol. 13(1), 1-16.

Gasoni, L., De Gurfinkel, B.S., 1997. The endophyte *Cladorrhinum foecundissimum* in cotton roots: Phosphorus uptake and host growth. Mycol. Res. 101(7), 867-870.

Gessler, C., Pertot, I., Perazzolli, M., 2011. *Plasmopara viticola*: A review of knowledge on downy mildew of grapevine and effective disease management. Phytopathol. Mediterr. 50(1), 3-44.

Ghabrial, S.A., Suzuki, N., 2009. Viruses of plant pathogenic fungi. Annu. Rev. Phytopathol. 47, 353-384.

Goheen, A.C., 1988. Rupestris stem pitting. In: Pearson R.C., Goheen A.C. (Eds). Compendium of grape diseases, 52-53. APS Press, St Paul, MN, USA.

Goheen, A.C., Hopkins, D.L., 1988. Pierce's disease. In: Pearson, R.C., Goheen, A.C. (Eds.). Compendium of grape diseases, 44-45. APS Press, St Paul, MN, USA.

Golino, D.A, Sim, S., Rowhani, A., 2000. Identification of the latent viruses associated with young vine decline in California. Extended Abstracts 13th Meeting of ICVG, Adelaide, Australia, 85-86.

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

Goszczynski, D.E., Du Preez, J., Burger, J.T., 2008. Molecular divergence of Grapevine virus A (GVA) variants associated with Shiraz disease in South Africa. Virus Res. 138(1-2), 105-110.

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., Habili, N., 2012. Grapevine virus A variants of group II associated with Shiraz disease in South Africa are present in plants affected by Australian Shiraz disease, and have also been detected in the USA. Plant Pathol. 61(1), 205-214.

Gotelli, N.J., Colwell, R.K., 2001. Quantifying biodiversity: Procedures and pitfalls in the measurement and comparison of species richness. Ecol. Lett. 4(4), 379-391.

Goussard, P., Bakker, H., 2006. Characterisation of grapevines visually infected with Shiraz disease associated viruses. Wineland - Production. [online] http://www.wineland.co.za/characterisation-of-grapevines-visually-infected-with-shiraz-disease-associated-viruses/.

Goussard, P., 2013a. A guide to grapevine abnormalities in South Africa: Virus and virus-like diseases - Stem grooving and corky bark (rugose wood complex) (Part 4.3). Wineland - Production. [online] http://www.wineland.co.za/a-guide-to-grapevine-abnormalities-in-south-africa-virus-and-virus-like-diseases-stem-grooving-and-corky-bark-rugose-wood-complex-part-4-3/.

Goussard, P., 2013b. A guide to grapevine abnormalities in South Africa: Virus and virus-like diseases - Shiraz decline (Part 4.5). Wineland - Production. [online] http://www.wineland.co.za/a-guide-to-grapevine-abnormalities-in-south-africa-virus-and-virus-like-diseases-shiraz-decline-part-4-5/.

Greif, C., Garau, R., Boscia, D., Prota, V.A., Fiori, M., Bass, P., Walter, B., Prota, U., 1995. The relationship of grapevine leafroll-associated closterovirus 2 with a graft incompatibility condition of grapevines. Phytopathol. Mediterr. 34(3), 167-173.

Haas, B.J., Gevers, D., Earl, A.M., Feldgarden, M., Ward, D.V., Giannoukos, G., Ciulla, D., Tabbaa, D., Highlander, S.K., Sodergren, E., Methé, B., DeSantis, T.Z., The Human Microbiome Consortium, Petrosino, J.F., Knight, R., Birren, B.W., Chimeric 16S rRNA sequence formation and detection in Sanger and 454-pyrosequenced PCR amplicons. Genome Res. 21(3), 494-504.

Hall, N., 2007. Advanced sequencing technologies and their wider impact in microbiology. J. Exp. Biol. 210(9), 1518-1525.

Hallmann, J., Sikora, R.A., 1996. Toxicity of fungal endophyte secondary metabolites to plant parasitic nematodes and soil-borne plant pathogenic fungi. Eur. J. Plant Pathol. 102(2), 155-162.

Hallmann, J., Quadt-Hallmann, A., Mahaffee, W.F., Kloepper, J.W., 1997. Bacterial endophytes in agricultural crops. Can. J. Microbiol. 43(10), 895-914.

Handelsman, J., Rondon, M.R., Brady, S.F., Clardy, J., Goodman, R.M., 1998. Molecular biological access to the chemistry of unknown soil microbes: A new frontier for natural products. Chem. Biol. 5(10), R245-R249.

Handelsman, J., 2004. Metagenomics: Application of genomics to uncultured microorganisms. Microbiol. Mol. Biol. Rev. 68(4), 669-685.

Hawksworth, D.L., Rossman, A.Y., 1997. Where are all the undescribed fungi? Phytopathology. 87(9), 888-891.

Heintz, C., Blaich, R., 1990. *Verticillium lecanii* as a hyperparasite of grapevine powdery mildew (*Uncinula necator*). Vitis. 29(4), 229-232.

Henegariu, O., Heerema, N.A., Dlouchy, S.R., Vance, G.H., Vogt, P.H., 1997. Multiplex PCR: Critical parameters and step-by-step protocol. BioTechniques. 23(3), 504-511.

Herrero, N., Márquez, S.S., Zabalgogeazcoa, I., 2009. Mycoviruses are common among different species of endophytic fungi of grasses. Arch. Virol. 154(2), 327-330.

Hewitt, W.B., Goheen, A.C., Cory, L., Luhn, C., 1972. Grapevine fleck disease, latent in many varieties, is transmitted by graft inoculation. Ann. Phytopathol. 43-47.

Hewitt, W.B., Pearson, R.C., 1988. Downy mildew. In: Pearson, R.C., Goheen, A.C. (Eds.). Compendium of grape diseases, 11-13. APS Press, St Paul, MN, USA.

Hopkins, D.L., 1989. *Xylella fastidiosa*: Xylem-limited bacterial pathogen of plants. Annu. Rev. Phytopathol. 27, 271-290.

Hugenholtz, P., 2002. Exploring prokaryotic diversity in the genomic era. Genome Biol. 3(2), reviews0003.1-reviews0003.8.

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

Huse, S.M., Welch, D.M., Morrison, H.G., Sogin, M.L., 2010. Ironing out the wrinkles in the rare biosphere through improved OTU clustering. Environ. Microbiol. 12(7), 1889-1898.

John, S., Wicks, T.J., Hunt, J.S., Scott, E.S., 2008. Colonisation of grapevine wood by *Trichoderma harzianum* and *Eutypa lata*. Aust. J. Grape Wine Res. 14(1), 18-24.

Kennelly, M.M., Gadoury, D.M., Wilcox, W.F., Magarey, P.A., Seem, R.C., 2005. Seasonal development of ontogenic resistance to downy mildew in grape berries and rachises. Phytopathology. 95(12), 1445-1452.

Kennelly, M.M., Gadoury, D.M., Wilcox, W.F., Magarey, P.A., Seem, R.C., 2007. Primary infection, lesion productivity, and survival of sporangia in the grapevine downy mildew pathogen *Plasmopara viticola*. Phytopathology. 97(4), 512-522.

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

Kõljalg, U., Nilsson, R.H., Abarenkov, K., Tedersoo, L., Taylor, A.F.S., Bahram, M., Bates, S.T., Bruns, T.D., Bengtsson-Palme, J., Callaghan, T.M., Douglas, B., Drenkhan, T., Eberhardt, U., Dueñas, M., Grebenc, T., Griffith, G.W., Hartmann, M., Kirk, P.M., Kohout, P., Larsson, E., Lindahl, B.D., Lücking, R., Martín, M.P., Matheny, P.B., Nguyen, N.H., Niskanen, T., Oja, J., Peay, K.G., Peintner, U., Peterson, M., Põldmaa, K., Saag, L., Saar, I., Schüßler, A., Scott, J.A., Senés, C., Smith, M.E., Suija, A., Taylor, D.L., Telleria, M.T., Weiss, M., Larsson, K., 2013. Towards a unified paradigm for sequence-based identification of fungi. Mol. Ecol. 22(1), 5271-5277.

Koltunow, A.M., Krake, L.R., Johnson, S.D., Rezaian, M.A., 1989. Two related viroids cause grapevine yellow speckle disease independently. J. Gen. Virol. 70(12), 3411-3419.

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.

Kozich, J.J., Westcott, S.L., Baxter, N.T., Highlander, S.K., Schloss, P.D., 2013. Development of a dual-index sequencing strategy and curation pipeline for analyzing amplicon sequence data on the MiSeq Illumina sequencing platform. Appl. Environ. Microbiol. 79(17), 5112-5120.

Kreuze, J.F., Perez, A., Untiveros, M., Quispe, D., Fuentes, S., Barker, I., Simon, R., 2009. Complete viral genome sequence 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.

Kuczynski, J., Stombaugh, J., Walters, W.A., González, A., Caporaso, J.G., Knight, R., 2011. Using QIIME to analyze 16S rRNA gene sequences from microbial communities. Curr. Protoc. Bioinformatics. 36, 10.7.1-10.7.2.

Kunin, V., Engelbrekston, A., Ochman, H., Hugenholtz, P., 2010. Wrinkles in the rare biosphere: Pyrosequencing errors can lead to artificial inflation of diversity estimates. Environ. Microbial. 12(1), 118-123.

Larignon, P., Fulchic, R., Cere, L., Dubos, B., 2001. Observation on black dead arm in French vineyards. Phytopathol. Mediterr. 40(3), 336-342.

Lehoczky, J., 1974. Black dead-arm disease of grapevine caused by *Botryosphaeria stevensii* infection. Acta Phytopathol. Acad. Sci. Hung. 9, 319-327.

Li, H., Durbin, R., 2009. Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics. 25(14), 1754-1760.

Lindahl, B.D., Nilsson, R.H., Tedersoo, L., Abarenkov, K., Carlsen, T., Kjøller, R., Kõljalg, U., Pennanen, T., Rosendahl, S., Stenlid, J., Kauserud, H., 2013. Fungal community analysis by high-throughput sequencing of amplified makers - A user's guide. New Phytol. 199(1), 288-299.

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, 251364. DOI: 10.1155/2012/251364.

Malinowski, D.P., Belesky, D.P., 1999. *Neotyphodium coenophialum*-endophyte infection affects the ability of tall fescue to use sparingly available phosphorus. J. Plant Nutr. 22(4-5), 835-853.

Malinowski, D.P., Belesky, D.P., 2000. Adaptations of endophyte-infected cool-season grasses to environmental stresses: Mechanisms of drought and mineral stress tolerance. Crop Sci. 40(4), 923-940.

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

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

Mardis, E.R., 2013. Next-generation sequencing platforms. Annu. Rev. Anal. Chem. 6, 287-303.

Maree, H.J., Almeida, R.P.P., Bester, R., Chooi, K.M., Cohen, D., Dolja, V.V., Fuchs, M.F., Golino, D.A., Jooste, A.E.C., Martelli, G.P., Naidu, R.A., Rowhani, A., Saldarelli, P., Burger, J.T., 2013. Grapevine leafroll-associated virus 3. Front. Microbiol. 4, 82. DOI: 10.3389/fmicb.2013.00082.

Martelli, G.P., 2014. Directory of virus and virus-like diseases of the grapevine and their agents. J. Plant Pathol. 96(Suppl. 1), 1-136.

Massart, S., Olmos, A., Jijakli, H., Candresse, T., 2014. Current impact and future directions of high throughput sequencing in plant virus diagnostics. Virus Res. 188, 90-96.

McClellan, W.D., Hewitt, W.B., La Vine, P., Kissler, J., 1973. Early Botrytis rot of grapes and its control. Am. J. Enol. Vitic. 24(1), 27-30.

McDonald, D., Price, M.N., Goodrich, J., Nawrocki, P., DeSantis, T.Z., Probst, A., Andersen, G.L., Knight, R., Hugenholtz, P., 2012. An improved Greengenes taxonomy with explicit ranks for ecological and evolutionary analyses of bacteria and archaea. ISME J. 6(3), 610-618.

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

Mlotshwa, S., Pruss, G.J., Vance, V., 2008. Small RNAs in viral infection and host defense. Trends Plant Sci. 13(7), 375-382.

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

Moller, W.J., Kasimatis, A.N., Kissler, J.J., 1974. A dying arm disease of grape in California. Plant Dis. Rep. 58, 869-871.

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, 46-58.

Moyer, M.M., Grove, G.G., 2011. Botrytis bunch rot in commercial Washington grape production: Biology and disease management. Washington State University Extension, FS046E.

Mugnai, L., Graniti, A., Surico, G., 1999. Esca (black measles) and brown wood-streaking: Two old and elusive diseases of grapevines. Plant Dis. 83(5), 404-418.

Munkvold, G.P., Duthie, J.A., Marois, J.J., 1994. Reductions in yield and vegetative growth of grapevines due to Eutypa dieback. Phytopathology. 84(2), 186-192.

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

Naidu, R., Rowhani, A., Fuchs, M., Golino, D., Martelli, G.P., 2014. Grapevine leafroll: A complex viral disease affecting a high-value fruit crop. Plant Dis. 98(9), 1172-1185.

Namiki, T., Hachiya, T., Tanaka, H., Sakakibara, Y., 2012. MetaVelvet: An extension of Velvet assembler to *de novo* metagenome assembly from short sequence reads. Nucleic Acids Res. 40(20), e155. DOI: 10.1093/nar/gks678.

Nocker, A., Burr, M., Camper, A.K., 2007. Genotypic microbial community profiling: A critical technical review. Microb. Ecol. 54(2), 276-289.

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

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

Palmieri, M.C., Perazzolli, M., Matafora, V., Moretto, M., Bachi, A., Pertot, I., 2012. Proteomic analysis of grapevine resistance induced by *Trichoderma harzianum* T39 reveals specific defence pathways activated against downy mildew. J. Exp. Bot. 63(17), 6237-6251.

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.

Perazzolli, M., Antonielli, L., Storari, M., Puopolo, G., Pancher, M., Giovannini, O., Pindo, M., Pertot, I., 2014. Resilience of the natural phyllosphere microbiota of the grapevine to chemical and biological pesticides. Appl. Environ. Microbiol. 80(12), 3585-3596.

Petrini, O., 1991. Fungal endophytes of tree leaves. In: Andrews, J.H., Hirano, S.S. (Eds.). Microbial ecology of leaves, 179-197. Springer, New York, NY, USA.

Pinto, C., Pinho, D., Sousa, S., Pinheiro, M., Egas, C., Gomes, A.C., 2014. Unravelling the diversity of grapevine microbiome. PLoS ONE. 9(1), e85622. DOI: 10.1371/journal.pone.0085622.

Pinto, C., Pinho, D., Cardoso, R., Custódio, V., Fernandes, J., Sousa, S., Pinheiro, M., Egas, C., Gomes, A.C., 2015. Wine fermentation microbiome: A landscape from different Portuguese wine appellations. Front. Microbiol. 6, 905. DOI: 10.3389/fmicb.2015.00905.

Quacquarelli, A., Gallitelli, D., Savino, V., Martelli, G.P., 1976. Properties of Grapevine fanleaf virus. J. Gen. Virol. 32(3), 349-360.

Raski, D.J., Goheen, A.C., Lider, L.A., Meredith, C.P., 1983. Strategies against Grapevine fanleaf virus and its nematode vector. Plant Dis. 67(3), 335-339.

Rastogi, G., Sani, R.K., 2011. Molecular techniques to assess microbial community structure, function, and dynamics in the environment. In: Ahmad, I., Ahmad, F., Pichtel, J. (Eds.). Microbes and microbial technology, 29-57. Springer, New York, NY, USA.

Reis, V.M., Baldani, J.I., Baldani, V.L.D., Dobereiner, J., 2000. Biological dinitrogen fixation in *Gramineae* and palm trees. Crit. Rev. Plant Sci. 19(3), 227-247.

Rezaian, M.A., Koltunow, A.M., Krake, L.R., Skene, K.G., 1991. Grapevine viroids. Proceedings 10th Meeting of ICVG, Volos, Greece, 297.

Rodriguez, R.J., Henson, J., Van Volkenburgh, E., Hoy, M., Wright, L., Beckwith, F., Kim, Y., Redman, R.S., 2008. Stress tolerance in plants via habitat-adapted symbiosis. ISME J. 2(4), 404-416.

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

Roossinck, M.J., Martin, D.P., Roumagnac, P., 2015. Plant virus metagenomics: Advances in virus discovery. Phytopathology. 105(6), 717-727.

Savino, V., Di Terlizzi, B., Rivieccio, S., Di Silvio, F., 1991. Presence in clonal rootstocks of a graft-transmissible factor that induces stunting and bushy growth in European grapevines. Proceedings 10th Meeting of ICVG, Volos, Greece, 202-210.

Schena, L., Ippolito, A., Zahavi, T., Cohen, L., Nigro, F., Droby, S., 1999. Genetic diversity and biocontrol activity of *Aureobasidium pullulans* isolates against postharvest rots. Postharvest Biol. Technol. 17(3), 189-199.

Schloss, P.D., Westcott, S.L., Ryabin, T., Hall, J.R., Hartmann, M., Hollister, E.B., Lesniewski, R.A., Oakley, B.B., Parks, D.H., Robinson, C.J., Sahl, J.W., Stres, B., Thallinger, G.G., Van Horn, D.J., Weber, C.F., 2009. Introducing Mothur: Open-source, platform-independent, community-supported software for describing and comparing microbial communities. Appl. Environ. Microbiol. 75(23), 7537-7541.

Schloss, P.D., Gevers, D., Westcott, S.L., 2011. Reducing the effects of PCR amplification and sequencing artifacts on 16S rRNA-based studies. PLoS ONE. 6(12), e27310. DOI: 10.1371/journal.pone.0027310.

Schmidt, P., Bálint, M., Greshake, B., Bandow, C., Römbke, J., Schmitt, I., 2013. Illumina metabarcoding of a soil fungal community. Soil Biol. Biochem. 65, 128-132.

Schoch, C.L., Seifert, K.A., Huhndorf, S., Robert, V., Spouge, J.L., Levesque, C.A., Chen, W., Fungal Barcoding Consortium, 2012. Nuclear ribosomal internal transcribed spacer (ITS) region as a universal DNA barcode marker for fungi. Proc. Natl. Acad. Sci. U.S.A. 109(16), 6241-6246.

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

Serra, R., Braga, A., Venâncio, A., 2005. Mycotoxin-producing and other fungi isolated from grapes for wine production, with particular emphasis on ochratoxin A. Res. Microbiol. 156(4), 515-521.

Sogin, M.L., Morrison, H.G., Huber, J.A., Welch, D.M., Huse, S.M., Neal, P.R., Arrieta, J.M., Herndl, G.J., 2006. Microbial diversity in the deep sea and the underexplored rare biosphere. Proc. Natl. Acad. Sci. U.S.A. 103(32), 12115-12120.

Spreeth, N., 2005. Shiraz decline. Wineland - Business & Marketing. [online] http://www.wineland.co.za/shiraz-decline/.

Staley, J.T., Konopka, A., 1985. Measurement of *in situ* activities of nonphotosynthetic microorganisms in aquatic and terrestrial habitats. Annu. Rev. Microbiol. 39, 321-346.

Steel, C.C., Blackman, J.W., Schmidtke, L.M., 2013. Grapevine bunch rots: Impact on wine composition, quality, and potential procedures for the removal of wine faults. J. Agric. Food Chem. 61(22), 5188-5206.

Stevenson, J.F., Matthews, M.A., Rost, T.L., 2005. The developmental anatomy of Pierce's disease symptoms in grapevines: Green islands and matchsticks. Plant Dis. 89(6), 543-548.

Sturz, A.V., Christie, B.R., Nowak, J., 2000. Bacterial endophytes: Potential role in developing sustainable systems of crop production. Crit. Rev. Plant Sci. 19(1), 1-30.

Su, C., Lei, L., Duan, Y., Zhang, K., Yang, J., 2012. Culture-independent methods for studying environmental microorganisms: Methods, application, and perspective. Appl. Microbiol. Biotechnol. 93(3), 993-1003.

Surico, G., Mugnai, L., Marchi, G., 2008. The esca disease complex. In: Ciancio, A., Mukerji, K.G. (Eds.). Integrated management of diseases caused by fungi, phytoplasma and bacteria, 119-136. Springer Science+Business Media, Dordrecht, Netherlands.

Surico, G., 2009. Towards a redefinition of the diseases within the esca complex of grapevine. Phytopathol. Mediterr. 48(1), 5-10.

Szychowski, J.A., McKenry, M.V., Walker, M.A., Wolpert, J.A., Credi, R., Semancik, J.S., 1995. The vein-banding disease syndrome: A synergistic reaction between grapevine viroids and fanleaf virus. Vitis 34(4), 229-232.

Tan, R.X., Zou, W.X., 2001. Endophytes: A rich source of functional metabolites. Nat. Prod. Rep. 18(4), 448-459.

Taylor, M.W., Tsai, P., Anfang, N., Ross, H.A., Goddard, M.R., 2014. Pyrosequencing reveals regional differences in fruit-associated fungal communities. Environ. Microbiol. 16(9), 2848-2858.

Trotel-Aziz, P., Couderchet, M., Biagianti, S., Aziz, A., 2008. Characterization of new bacterial biocontrol agents *Acinetobacter*, *Bacillus*, *Pantoea* and *Pseudomonas* spp. mediating grapevine resistance against *Botrytis cinerea*. Environ. Exp. Bot. 64(1), 21-32.

Trouillas, F., Úrbez-Torres, J.R., Gubler, W.D., 2010. Diversity of diatrypaceous fungi associated with grapevine canker diseases in California. Mycologia. 102(2), 319-336.

Úrbez-Torres, J.R., 2011. The status of Botryosphaeriaceae species infecting grapevines. Phytopathol. Mediterr. 50(Suppl.), S5-S45.

Uyemoto, J.K., Rowhani, A., Luvisi, D., Krag, C.R., 2001. New closterovirus in Redglobe grape causes decline of grafted plants. Calif. Agric. 55(4), 28-31.

Van Dijk, E.L., Auger, H., Jaszczyszyn, Y., Thermes, C., 2014. Ten years of next-generation sequencing technology. Trends Genet. 30(9), 418-426.

Vega, A., Gutiérrez, R.A., Peña-Neira, A., Cramer, G.R., Arce-Johnson, P., 2011. Compatible GLRaV-3 viral infections affect berry ripening decreasing sugar accumulation and anthocyanin biosynthesis in *Vitis vinifera*. Plant Mol. Biol. 77(3), 261-274.

Visser, M., Bester, R., Burger, J.T., Maree, H.J., 2016. Next-generation sequencing for virus detection: Covering all the bases. Virol. J. 13, 85. DOI: 10.1186/s12985-016-0539-x.

Wang, Q., Garrity, G.M., Tiedje, J.M., Cole, J.R., 2007. Naïve Bayesian classifier for rapid assignment of rRNA sequences into the new bacterial taxonomy. Appl. Environ. Microbiol. 73(16), 5261-5267.

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.

Weisburg, W.G., Barns, S.M., Pelletier, D.A., Lane, D.J., 1991. 16S ribosomal DNA amplification for phylogenetic study. J. Bacteriol. 173(2), 697-703.

Wells, J.M., Raju, B.C., Hung, H., Weisburg, W.G., Mandelco-Paul, L., Brenner, D.J., 1987. *Xylella fastidiosa* gen. nov., sp. nov.: Gram-negative, xylem-limited, fastidious plant bacteria related to *Xanthomonas* spp. Int. J. Syst. Bacteriol. 37(2), 136-143.

West, E.R., Cother, E.J., Steel, C.C., Ash, G.J., 2010. The characterization and diversity of bacterial endophytes of grapevine. Can. J. Microbiol. 56(3), 209-216.

Whitelaw-Weckert, M.A., Whitelaw, E.S., Rogiers, S.Y., Quirk, L., Clark, A.C., Huang, C.X., 2011. Bacterial inflorescence rot of grapevine caused by *Pseudomonas syringae* pv. *syringae*. Plant Pathol. 60(2), 325-337.

Wilcox, W., 2003. Grapevine powdery mildew. Cornell University Cooperative Extension, Disease identification sheet no. 102GFSG-D2.

Willems, A., Gillis, M., Kersters, K., Van den Broecke, L., De Ley, J., 1987. Transfer of *Xanthomonas ampelina* Panagopoulos 1969 to a new genus, *Xylophilus* gen. nov., as *Xylophilus ampelinus* (Panagopoulos 1969) comb. nov. Int. J. Syst. Bacteriol. 37(4), 422-430.

Wooley, J.C., Godzik, A., Friedberg. I., 2010. A primer on metagenomics. PLoS Comput. Biol. 6(2), e1000667. DOI: 10.1371/journal.pcbi.1000667.

Xie, J., Jiang, D., 2014. New insights into mycoviruses and exploration for the biological control of crop fungal diseases. Annu. Rev. Phytopathol. 52, 45-68.

Yousaf, S., Bulgari, D., Bergna, A., Pancher, M., Quaglino, F., Casati, P., Campisano, A., 2014. Pyrosequencing detects human and animal pathogenic taxa in the grapevine endosphere. Front. Microbiol. 5, 327. DOI: 10.3389/fmicb.2014.00327.

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

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

Internet sources

https://www.qiagenbioinformatics.com/

http://www.drive5.com/usearch/manual/uparseotu_algo.html.

Chapter 3: The viral diversity in old and young grapevines

3.1 Introduction

Grapevine is susceptible to a large number of pathogens. At present, grapevine is known to host at least 65 viruses and five viroids, the most recognised for any woody agricultural crop (Martelli, 2014). These viruses can be highly detrimental, posing a serious threat to the global wine industry by affecting the quality and yield of the fruit. The aetiologies of prominent grapevine diseases are not well understood and are complicated by multiple virus infections (Prosser *et al*., 2007). Therefore, research on the diversity of viral pathogens in grapevines is necessary.

Conventional methods for virus detection, such as nucleic acid-based techniques and serological assays, are unable to identify unknown viruses that may be associated with disease symptoms. Moreover, considering the rapid and error-prone nature of RNA virus replication, not all variants are detected using these conventional assays. Next-generation sequencing (NGS), used in a metagenomics approach, allows researchers to sequence the viral component of environmental samples in an unbiased manner (Adams *et al*., 2009). Next-generation sequencing has been applied in several grapevine disease studies, using different templates for library preparation. This has facilitated the discovery of a number of novel grapevine viruses.

Al Rwahnih *et al*. (2009) sequenced both total RNA and double-stranded RNA (dsRNA) library preparations from Californian vines displaying decline symptoms. The vines were revealed to support a mixed infection with seven different RNA genomes, which included a novel marafivirus in the family *Tymoviridae*. The scope of metagenomics is further reflected in studies that characterised the virome of an entire vineyard and the mycovirome of a single vine (Al Rwahnih *et al*., 2011; Coetzee *et al*., 2010). Coetzee *et al*. (2010) performed a survey of over 40 pooled samples of randomly selected vines from a South African vineyard displaying viral disease symptoms, and identified four grapevine viruses and two mycovirus families; *Chrysoviridae* and *Totiviridae*. The following year, Al Rwahnih *et al*. (2011) revealed the grapevine virome to be dominated by mycoviruses. Species of the families *Chrysoviridae*, *Hypoviridae*, *Narnaviridae*, *Partitiviridae* and *Totiviridae* were detected, with three mycoviruses associated with the pathogenic fungus, *Botrytis cinerea*.

Another study sequenced small RNAs from vines showing symptoms of vein clearing and vine decline, and reported the first DNA virus in grapevine. This virus was present in six diseased cultivars across the Midwestern United States and was taxonomically classified in the family *Caulimoviridae* (Zhang *et al*., 2011). A second DNA virus, of the family *Geminiviridae* was detected in Californian vineyards showing symptoms of red blotch disease. The virus was revealed to be identical to Grapevine Cabernet Franc-associated virus, which had previously been detected in declining grapevines in New York, and was renamed *Grapevine red blotch-associated virus* (Al Rwahnih *et al*., 2013; Krenz *et al*., 2012). Two novel *Betaflexiviridae* species were also described in recent grapevine virome studies, including a fifth vitivirus, *Grapevine virus F*, and a new trichovirus, *Grapevine Pinot gris virus* (Al Rwahnih *et at*., 2012; Giampetruzzi *et al*., 2012). Next-generation sequencing enables the efficient detection of known viruses, viruses occurring at low titres, as well as unknown viruses and divergent variants, making it a valuable tool to unravel the complexity of grapevine diseases. Furthermore, the generated data can be used to improve current virus detection assays.

A number of old vineyards have remained profitable, despite the presence of several economically important viruses. Vines are regarded ‘old’ if they are more than 35 years old and ‘young’ if they are less than ten years old. These old vineyards were established before virus certification schemes were in place. To date, little research has been done to determine the extent of the viral populations in old grapevines. In this study, the viromes of both old and young Pinotage vines, that displayed grapevine leafroll disease (GLD) symptoms, were characterised using a metagenomic NGS approach.

3.2 Materials and methods

3.2.1 Plant material

Ten old and ten young randomly selected grapevines (*Vitis vinifera* cv. Pinotage) were identified at a wine estate in the Stellenbosch region of the Western Cape. The vineyard historically had a high incidence of GLD. The vines are managed as bush vines, maximising the area for leaf, trunk and deep root development. In this vineyard, several old vines displaying poor vigour had been replaced with younger vines of the same clone. Prior to propagation, these vines were subjected to heat therapy to reduce the risk of virus transmission. At the time of sampling, the old vines were 40 years old, while the young vines were approximately seven years old. Canes were sampled during the

dormant season (July, 2014) before annual pruning of the vineyard. Phloem tissue was obtained by scraping canes with a surface-sterilised knife, and stored at -80°C.

3.2.2 Double-stranded RNA extraction and next-generation sequencing

Double-stranded RNA was extracted to enrich for virus-specific nucleic acids. The extractions were performed on approximately 12 grams of phloem scrapings, pulverised in liquid nitrogen. A cellulose affinity chromatography extraction method, with a combination of MN 2100 (Macherey-Nagel) and medium-fibre (Sigma-Aldrich) cellulose, was used (Burger & Maree, 2015). The dsRNA extracts were treated with RQ1 RNase-free DNase (Promega) and RNase T1 (Thermo Scientific) and the integrity and quantity was evaluated by gel electrophoresis (1% [w/v] agarose-TAE). Double-stranded RNA samples from four old vines, KP1.27O, KP3.29O, KP4.25O and KP4.30O, and four young vines, KP1.29Y, KP3.20Y, KP5.37Y and KP7.30Y were selected for NGS, based on yield. Sequencing libraries (~400nt insert size) were prepared using the TruSeq RNA sample preparation kit. A protocol adapted for dsRNA as input RNA was followed (Burger & Maree, 2015). The libraries were sequenced in two separate paired-end runs (2x250nt and 2x125nt) on the Illumina HiSeq 2500 platform at the Agricultural Research Council's Biotechnology Platform in Pretoria.

3.2.3 Pre-processing of sequence data

The quality of the datasets generated for each sample was assessed using FastQC v0.11.3 (http://www.bioinformatics.babraham.ac.uk/projects/fastqc/). Trimmomatic v0.33 was used to trim the reads for low-quality bases and to identify and remove adapters (Bolger *et al*., 2014). The trimming parameters used included: ILLUMINACLIP:TruSeq3-PE-2.fa:2:30:10; HEADCROP:10 and TRAILING:20. In short, the specified Illumina adapter sequences used in the TruSeq RNA library preparation were removed, the first 10 bases were trimmed off from the 5'-end of the reads and the bases displaying a Phred quality score below Q20, denoting 99% base-calling accuracy, were removed from the 3'-end of the reads. Figure 3.1 illustrates the bioinformatic workflow that was followed to analyse the NGS data.

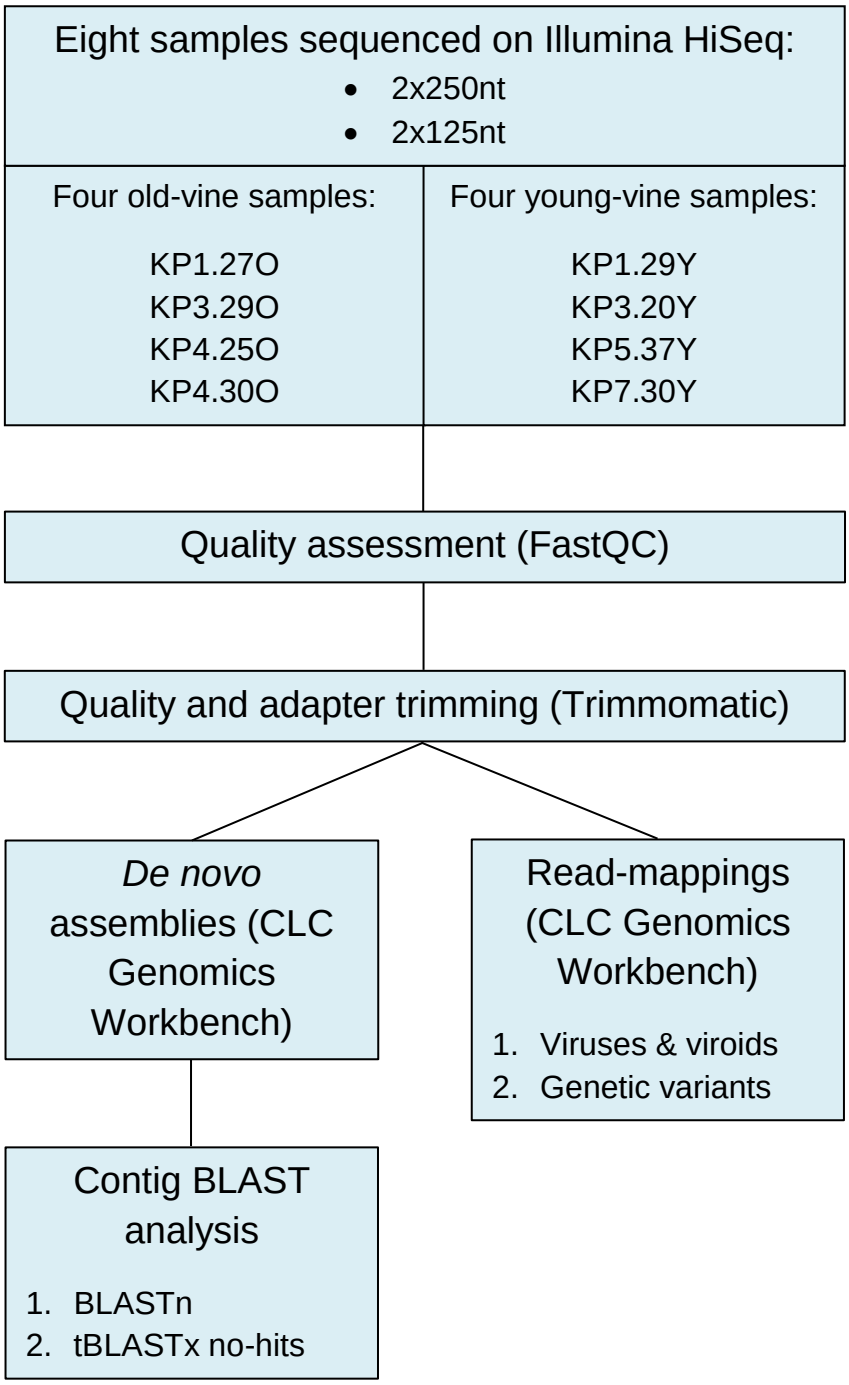

Figure 3.1: Bioinformatic workflow followed to analyse the NGS datasets generated by the two separate paired-end Illumina sequencing runs (2x250nt and 2x125nt). Quality assessment and trimming was performed individually for each sequencing run and sample. The trimmed reads, generated by the two separate runs were then combined for each sample for downstream analyses.

3.2.4 *De novo* assembly and read-mapping analyses

The quality-trimmed datasets were imported into CLC Genomics Workbench v8.0.3 (https://www.qiagenbioinformatics.com/) as paired-end FASTQ files. *De novo* assemblies of pooled reads (2x250nt and 2x125nt) were performed using CLC's assembly algorithm, which works by employing De Bruijn graphs. This involved two successive steps. First, the reads were assembled into contig sequences, without

retaining information about the reads used to build each contig. Second, the reads were mapped using the contig sequences as reference. The resulting contig coverage serves as an indication of assembly accuracy. It is worth noting that the alignment of a read to a contig does not necessarily mean that the information from this read was used in the assembly of the contig, as the mapping is a separate part of the algorithm.

De novo assembly parameters included: default word size (automatically calculated); bubble size of 50nt; minimum contig length of 200nt and scaffolding of contigs. Reads were mapped to the contigs using the following parameters: optimised read alignment settings (mismatch cost of 2; linear gap cost; insertion and deletion costs of 3; length fraction of 0.5 and similarity fraction of 0.8); global alignment of reads and auto-detection of paired distances. Contigs were updated based on the mapped reads. Consequently, erroneous contigs or regions within contigs where no reads aligned were omitted or trimmed. Generated contigs were identified through command-line BLASTn analysis (BLAST+ v2.2.31) against the National Center for Biotechnology Information (NCBI) non-redundant nucleotide database, using default parameters (Altschul *et al*., 1990; Camacho *et al*., 2009). The contigs were classified based on the match with the highest bit score. Contigs that could not be identified with BLASTn were further subjected to tBLASTx searches.

To obtain a more accurate indication of the abundance of virus and viroid populations in each vine, the number of reads was normalised based on the concept of RPKM (Reads Per Kilobase of transcripts per Million mapped reads), as applied to RNA-Seq transcriptome experiments (Mortazavi *et al*., 2008). This was necessary to account for the varying sequencing depth across the eight samples, as well as differences in the sizes of grapevine virus genomes. Furthermore, considering the possibility of multiple virus variant infections, normalising against the total concatenated contig length for each virus will skew the virus abundance estimation. This is due to the fact that contigs assemble separately for different variants. Therefore, a better normalising factor is the species reference genome length. In this study, a new method for abundance estimation was implemented; the virus and viroid read distribution was expressed as the number of reads that mapped to the contigs of a given virus species or family, per kilobase reference sequence length, per million reads mapped to the total assembled contigs, hereafter referred to as the Virus (or Viroid) Read Ratio (VRR). The VRR is calculated as follows: read count [contigs of species or family] / reference genome length * read count [total assembled contigs] * 1E+03 * 1E+06.

Additionally, the pooled quality-trimmed reads (2x250nt and 2x125nt) were mapped to a database of known grapevine-infecting virus and viroid sequences (Supplementary Table S1), to verify the BLAST results. The database comprises 87 complete or near-complete genome sequences representing 60 different viruses. The database included sequences of all available genomic segments in the case of multipartite viruses. Read-mapping parameters included: no masking of reference sequences; optimised read alignment settings; global alignment and auto-detection of paired distances. Non-specific matches were mapped randomly, i.e. if a read pair mapped to multiple positions with an equal score, the pair was mapped indiscriminately. A virus was classified as present in a sample, based on an arbitrary threshold of 50% reference genome coverage. As there is no definitive read-mapping threshold for positive virus detection, RT-PCR detection assays were performed to confirm the bioinformatic results (section 3.2.5).

Where applicable, the reads were further mapped to representative virus genomes of different genetic variants (Table 3.1). The same parameters were used, except the fraction of the total alignment length that had to match the reference sequence was set at a stringency of 0.9 and non-specific matches were ignored. This parameter will exclude a read pair from the final mapping if it aligns at multiple positions, thereby enabling the unique mapping of reads, so as to distinguish between variants. To estimate, from the *de novo* assembled contigs, the number of variants present in each sample, the concatenated contig length for each virus species was divided by the reference genome length of that virus species. This estimation is hereafter referred to as the length fraction, not to be confused with CLC's length fraction read alignment parameter.

Table 3.1: Representative genome sequences of commonly recognised genetic variants of prevalent grapevine viruses. The corresponding GenBank accession numbers are indicated. Selection of variant groups to include in the mapping analysis was based on that of recent grapevine virus diversity studies.

Virus & variant group identifier	Isolate/strain	Accession	Reference
Grapevine leafroll-associated virus 2			
93/955	93/955	NC_007448	
OR1	OR1	FJ436234	
BD	BD	DQ286725	Poojari *et al*., 2013
RG	RG	AF314061	
SG	SG	KF220376	
Grapevine leafroll-associated virus 3			
Group I	621	GQ352631	
Group II	GP18	EU259806	
Group III	PL-20	GQ352633	Maree *et al*., 2015
Group VI	GH30	JQ655296	
Group VII	GH24	KM058745	
Grapevine rupestris stem pitting-associated virus			
Group 1	SY	AY368590	
Group 2a	SG1	AY881626	Alabi *et al*., 2010
Group 2b	1	NC_001948	
Group 3	BS	AY881627	
Grapevine virus A			
Group I	GTG11-1	DQ855084	
Group II	GTR1-2	DQ855086	Alabi *et al*., 2014
Group III	GTR1-1	DQ787959	
Group IV	PA3	AF007415	
Grapevine virus B			
94/971	94/971	EF583906	Hu *et al*., 2014
QMWH	QMWH	KF700375	
Grapevine virus E			
TvAQ7	TvAQ7	NC_011106	Alabi *et al*., 2013
SA94	SA94	GU903012	

3.2.5 Reverse transcription PCR screening

In order to validate the presence of grapevine viruses detected in the NGS data of the eight sequenced samples, primers were selected for each virus (Table 3.2) and used in reverse transcription (RT-) PCR detection assays. Additionally, all primer pairs were also used to screen the other 12 samples that had not been sequenced.

Table 3.2: Primers selected to screen the 20 samples for viruses detected in the NGS data. The virus, and gene or domain of interest of each virus is indicated.

Target	Gene/ Domain	Sequence (5' - 3')	Amplicon size (bp)	T_a (°C)	Reference
GLRaV-2	CPm	F: TATGAGTTCCAACACAAGCGTGC R: ACACCGTGCTTAGTACCTCC	682	58	M. Engelbrecht, unpublished
GLRaV-3	HEL	F: GGGRACGGARAAGTGTTACC R: TCCAAYTGGGTCATRCACAA	144	53	Bester *et al*., 2014
GRSPaV	CP	F: ACTTTCAAAGACGGTGGACATGAG R: AGCCATAGCTTGTCTGAGCACTTG	523	54	M. Engelbrecht, unpublished
GVA	RBP	F: AGGTCCACGTTTGCTAAG R: CATCGTCTGAGGTTTCTACTA	238	56	MacKenzie, 1997
GVB	CP	F: TGACCTTCGTAACTGATGCT R: GCTGTGAAGACGTTCTTAGCAC	493	57	M. Engelbrecht, unpublished
GVE	CP	F: ATGATTTGATGCTCAGTCACAGG R: GGGTTCTTATGGCCTGCTTA	217	58	Jooste *et al*., 2015
GFkV	CP	F: CCGTCTGCTGACCAGCCTG R: CGACGCAGGCGTCATTGCG	520	55	Glasa *et al*., 2011
GFkV	CP	F: TGACCAGCCTGCTGTCTCTA R: TGGACAGGGAGGTGTAGGAG	179	56	Gambino & Gribaudo, 2006
GRVFV[a]	CP	F: CATCCCCTTCCAGTGGGTTG R: ACGCTGACCATGCCACGAA	441	53	This study
GSyV-1[a]	CP	F: CCAATGGGTCGCACTTGTTG R: ACTTCATGGTGGTGCCGGTG	374	56	Glasa *et al*., 2015

[a]Viruses have not previously been reported in South African grapevines and primers need to be verified. GLRaV-2 = Grapevine leafroll-associated virus 2; GLRaV-3 = Grapevine leafroll-associated virus 3; GRSPaV = Grapevine rupestris stem pitting-associated virus; GVA = Grapevine virus A; GVB = Grapevine virus B; GVE = Grapevine virus E; GFkV = Grapevine fleck virus; GRVFV = Grapevine rupestris vein feathering virus; GSyV-1 = Grapevine Syrah virus 1. CPm = minor coat protein; HEL = helicase; CP = coat protein; RBP = RNA-binding protein.

Total RNA was extracted from roughly 2 grams of pulverised phloem scrapings using a modified cetyltrimethylammonium bromide (CTAB) extraction method (White *et al*., 2008). The integrity, and concentration and quality of the RNA were assessed by gel electrophoresis (1% [w/v] agarose-TAE) and spectrophotometry (NanoDrop 2000, Thermo Scientific), respectively. For each sample, 20µl complementary DNA (cDNA) was synthesised from approximately 1µg of total RNA primed with 150ng Random Primers (Promega) using Maxima Reverse Transcriptase and RiboLock RNase Inhibitor (Thermo Scientific) as per manufacturer's instructions. The primer annealing reaction was incubated for 5 minutes at 65°C, followed by 2 minutes on ice and cDNA synthesis was carried out for 30 minutes at 50°C after a 10-minute incubation step at 25°C. Thereafter, 25µl amplification reactions were performed with 2.5µl cDNA using KAPA *Taq* DNA Polymerase (Kapa Biosystems), as instructed by the manufacturer. Cycling conditions included a 5-minute initial denaturation step at 94°C; 35 cycles of denaturation at 94°C for 30 seconds, annealing at the appropriate temperature (table 3.1) for 30 seconds and extension at 72°C for 20 to 40 seconds, depending on the

expected amplicon size; and a final extension step at 72°C for 7 minutes to complete the reaction. The PCR products were separated and visualised on 1.5% [w/v] agarose-TAE gels, stained with ethidium bromide. Amplicons generated for viruses not previously reported in South African grapevines were purified (Zymoclean Gel DNA Recovery Kit, Zymo Research) and Sanger sequenced for verification.

3.3 Results and discussion

3.3.1 Double-stranded RNA extraction

Double-stranded RNA extractions were performed on all 20 collected samples. Only seven old- and four young-vine samples yielded dsRNA adequate for NGS. Several optimisation strategies were implemented, including using different combinations of cellulose and varying amounts of pulverised plant material. The insufficient dsRNA yields of the remaining samples may be ascribed to low viral load in these plants. The four highest dsRNA-yielding old-vine samples were selected for sequencing, along with the four young-vine samples.

3.3.2 Pre-processing of sequence data

Initially, samples were sequenced in one paired-end (2x250nt) run on the Illumina HiSeq instrument, requesting 2 gigabases of sequence data per sample. Less than 0.4 gigabases of data were generated for six of the eight samples. The quality rapidly declined towards the end of the reads, an error pattern typically observed in Illumina data, as a result of phasing and pre-phasing during the sequencing process. The mean quality of the raw 2x250nt data dropped below a Phred score of Q20 after roughly 170 bases in the forward reads (Figure 3.2A), and 150 bases in the reverse reads. Significant 3’ adapter contamination was observed, indicating adapter read-through. Adapter read-through occurs when the fragments to be sequenced are shorter than the read length itself, due to suboptimal library size selection.

To compensate for the insufficient data, the libraries were re-sequenced in a second paired-end (2x125nt) run. The second run generated between 0.4 and 1.9 gigabases of data. The number of raw read pairs generated for both sequencing runs combined ranged between 233 980 and 7 709 522. The variation in the data output across the samples may be the result of inaccuracies in library quantification prior to pooling. As anticipated, the raw 2x125nt datasets were of higher quality (Figure 3.2B), and the adapter content was considerably lower, because a shorter read length was sequenced.

Unevenly distributed nucleotides were also observed at the 5'-end of the reads for both sequencing runs.

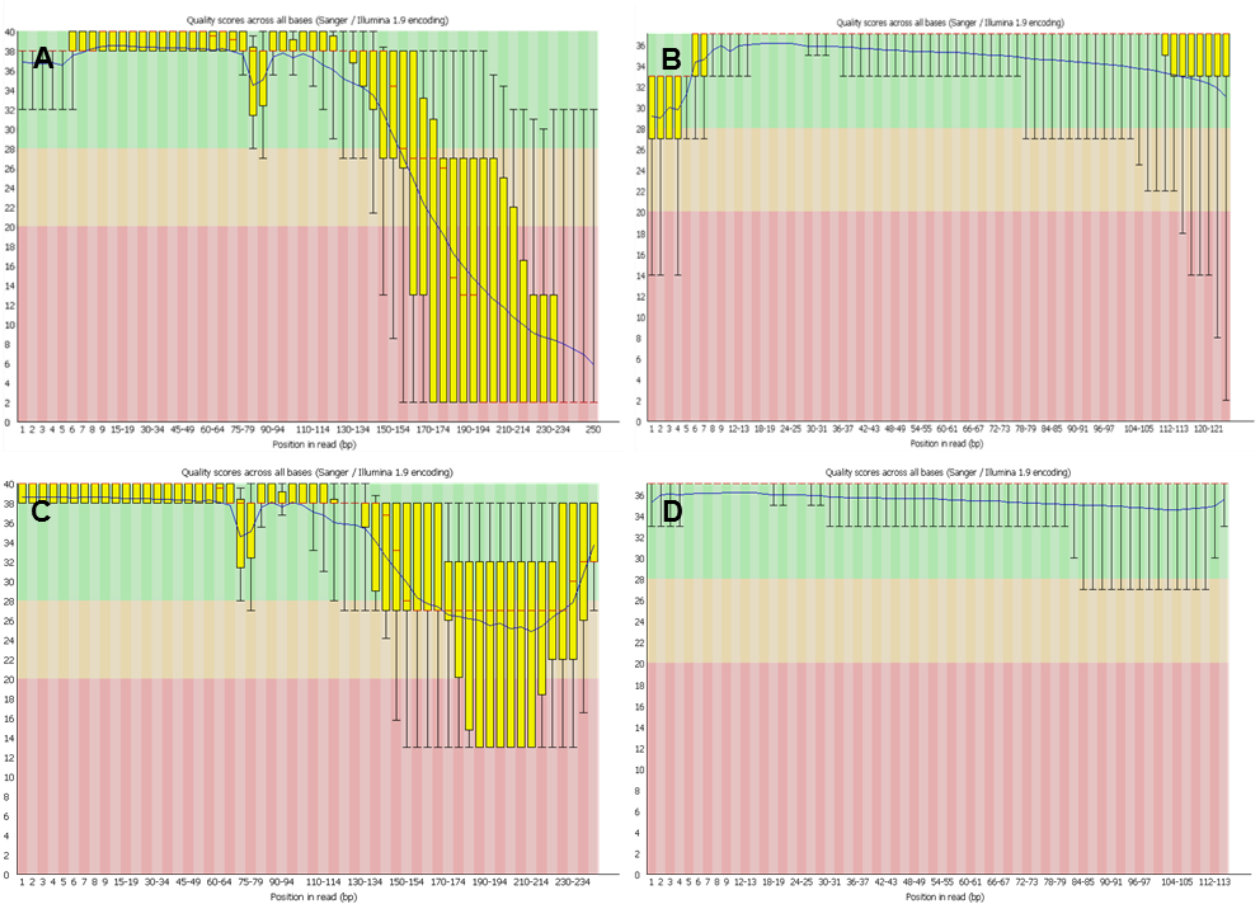

Figure 3.2: FastQC graphs illustrating the quality of the data before and after trimming. A) Raw 2x250nt, B) Raw 2x125nt, C) Trimmed 2x250nt and D) Trimmed 2x125nt data. The quality scores are shown on the y-axis. The blue line represents the mean quality and the red central line indicates the median value. The yellow boxes represent the inter-quartile range from 25 to 75%, and the upper and lower whiskers mark the 10th and 90th percentiles, respectively.

After the individual datasets were trimmed for adapters and low-quality bases, 28 to 80%, and 69 to 94% of read pairs were retained, for the first and second sequencing run, respectively. After trimming the first 10 bases with the HEADCROP parameter, the uneven nucleotide distribution was no longer evident. RNA-Seq libraries, such as those produced by the TruSeq kit, inherit intrinsic biases in the nucleotide composition at the start of the reads. This is caused by random hexamer priming for first-strand cDNA synthesis during library preparation (Hansen *et al*., 2010). As Figures 3.2C and 3.2D illustrate, trimming improved the overall quality of the data for downstream analyses; however, the quality of the 2x250nt datasets was less favourable. Table 3.3 summarises the raw data output per sample and the number of read pairs after trimming. The low quality of the raw data, the adapter content and the resulting loss of a

substantial amount of read pairs after trimming, specifically of that generated by the 2x250nt run, suggest substandard library preparation and sequencing. No correlation was observed between inadequate data output and low dsRNA yield.

Table 3.3: Sequence data (gigabases[a]) and trimming results. The number of raw read pairs and the number and percentage of read pairs remaining after trimming are indicated.

Sample	Raw output Gb	Raw read pairs	Trimmed Gb	Trimmed read pairs (%)
2x250nt				
KP1.27O	1.31	2 618 818	1.05	2 108 544 (80.52)
KP3.29O	0.36	722 818	0.10	206 594 (28.58)
KP4.25O	0.19	378 418	0.06	126 443 (33.41)
KP4.30O	0.14	286 364	0.05	90 937 (31.76)
KP1.29Y	0.12	233 980	0.04	74 772 (31.96)
KP3.20Y	0.13	259 571	0.05	93 141 (35.88)
KP5.37Y	0.15	293 693	0.06	120 569 (41.05)
KP7.30Y	1.31	2 617 415	0.49	984 467 (37.61)
2x125nt				
KP1.27O	1.93	7 709 522	1.81	7 250 803 (94.05)
KP3.29O	1.14	4 542 770	0.86	3 447 122 (75.88)
KP4.25O	0.63	2 536 629	0.50	2 003 413 (78.98)
KP4.30O	0.50	2 002 121	0.36	1 438 098 (71.83)
KP1.29Y	0.42	1 681 123	0.32	1 275 756 (75.89)
KP3.20Y	0.44	1 770 697	0.35	1 388 262 (78.40)
KP5.37Y	0.54	2 148 300	0.43	1 737 860 (80.89)
KP7.30Y	1.50	5 992 799	1.03	4 158 691 (69.39)

[a]gigabases (Gb) = 2 * read pairs * read length / 1E+09.

3.3.3 *De novo* assembly and read-mapping analyses

Trimmed reads were assembled into contigs using a minimum contig length of 200nt. A short minimum contig length was chosen to avoid excluding sequences originating from viroids. Contigs were updated based on the mapping of the reads, a CLC feature developed to filter out erroneous contigs. As a result, fewer and shorter contigs may remain than initially assembled. Between 1 021 and 2 572 contigs were generated for all eight samples, with an N50 value between 349 and 871nt. The contig lengths ranged from 85 to 26 660nt. The percentage of matched reads (reads that mapped to contigs) ranged between 75 and 96% (Table 3.4). A number of contigs had a low average coverage, henceforth referred to as average depth of coverage. Contigs originating from low-titre viruses and viruses with small genomes will typically have low depth of

coverage. In order to prevent the loss of such sequences, no coverage cut-off was selected and all of the updated contigs were subjected to BLAST analysis.

Table 3.4: *De novo* assembly output statistics for the pooled quality-trimmed 2x250nt and 2x125nt datasets. The contig features and the number and percentage of reads that mapped to the assembled contigs are listed for each sample.

Sample	Contig measurements						Matched reads (%)
			Length (nt)				
	Count	Total bases	Min.	N50[a]	Average	Max.	
KP1.27O	1 409	943 782	90	871	670	26 660	17 926 355 (95.77)
KP3.29O	1 298	689 717	134	624	531	14 905	6 810 956 (93.21)
KP4.25O	1 021	543 206	85	612	532	18 655	3 732 641 (87.63)
KP4.30O	1 454	724 705	115	550	498	10 181	2 771 084 (90.62)
KP1.29Y	1 278	542 802	170	457	425	18 570	2 360 012 (87.37)
KP3.20Y	1 487	661 414	150	496	445	5 386	2 355 371 (79.50)
KP5.37Y	2 572	960 498	117	349	373	16 343	2 806 504 (75.51)
KP7.30Y	1 917	864 136	115	411	451	15 581	9 152 092 (89.00)

[a]N50 is a statistical measure used to assess assembly quality and is comparable to the median of contig lengths, but with more weight given to longer contigs. It can be defined as the length for which all contigs of that length or longer comprise at least 50% of bases in the sequence dataset.

As determined by BLAST analysis, the greatest percentage (57 to 95%) of reads across the samples originated from grapevine chromosomes, chloroplasts and mitochondria. Although endogenous dsRNA could account for some of the host contamination, the extent thereof would indicate that the DNase and RNase treatment did not completely digest host DNA and single-stranded RNA. Regardless, dsRNA extraction is an effective strategy for enrichment of viral nucleic acids, with the second largest fraction of the data identified as grapevine virus- and viroid-related sequences. Between 0.08 and 1.2% of reads in each sample represent fungal, bacterial and other species.

BLASTn searches revealed the presence of multiple viruses known to infect local vineyards (Coetzee *et al*., 2010). These results are summarised in Table 3.5, with the abundance of the pathogens expressed as the VRR. To support the BLAST results, the reads were mapped to the genomes of 60 different grapevine-infecting viruses and viroids. The percentage reference genome coverage was used as an indication of the presence of a virus; the greater the coverage, the more confidence there is in positive identification. A virus was regarded present in the sample if the reads mapped to more than 50% of the reference genome, as indicated in Table 3.5. The read-mappings correlated with the BLASTn results for the known grapevine-associated viruses.

Table 3.5: The distribution of the reads accounting for contigs that were classified as viruses and viroids with BLASTn, expressed as the VRR[a]. Species that had more than 50% genome coverage in the read-mapping analyses (RM) were considered present, as indicated.

Virus or viroid		Sample															
		KP1.27O		KP3.29O		KP4.25O		KP4.30O		KP1.29Y		KP3.20Y		KP5.37Y		KP7.30Y	
Family	Species	VRR	RM	VRR	RM	VRR	RM	VRR	RM	VRR	RM	VRR	RM	VRR	RM	VRR	RM
Closteroviridae	GLRaV-2	-	-	-	-	-	-	436.18	✓	0.26		-	-	-	-	177.55	✓
	GLRaV-3	10 519.04	✓	9 995.31	✓	14 307.83	✓	11 199.70	✓	10 580.91	✓	2 016.15	✓	1 671.61	✓	9 170.57	✓
Betaflexiviridae	GRSPaV	278.33	✓	121.55	✓	154.21	✓	169.04	✓	-	-	115.22	✓	386.80	✓	73.00	✓
	GVA	77.33	✓	40.60	✓	76.04	✓	75.90	✓	53.37	✓	-	-	62.40	✓	-	-
	GVB	159.30	✓	135.40	✓	172.01	✓	180.46	✓	-	-	-	-	-	-	165.80	✓
	GVE	254.97	✓	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Tymoviridae	GFkV	64.54	✓	166.89	✓	164.17	✓	47.85	✓	-	-	127.08	✓	437.24	✓	144.86	✓
	GRGV[b]	-	-	0.45		0.61		-	-	-	-	-	-	4.13		-	-
	GRVFV	25.42	✓	23.85	✓	27.33	✓	23.40	✓	-	-	-	-	17.93	✓	-	-
	GSyV-1	6.97	✓	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Virgaviridae	TMV	-	-	-	-	-	-	-	-	-	-	0.27		-	-	-	-

Virus or viroid		Sample															
		KP1.27O		KP3.29O		KP4.25O		KP4.30O		KP1.29Y		KP3.20Y		KP5.37Y		KP7.30Y	
Family	Species	VRR	RM	VRR	RM	VRR	RM	VRR	RM	VRR	RM	VRR	RM	VRR	RM	VRR	RM
Pospiviroidae	AGVd	32.50	✓	112.21	✓	45.01	✓	31.29	✓	401.91	✓	-	-	421.01	✓	37.61	✓
	GYSVd-1	56.24	✓	74.21	✓	27.08	✓	106.49	✓	292.90	✓	306.24	✓	1 224.71	✓	45.97	✓
	GYSVd-2	16.90	✓	23.46	✓	20.66	✓	49.71	✓	547.46	✓	686.55	✓	708.70	✓	-	-
	HSVd	8.13	✓	14.10	✓	12.42	✓	5.97	✓	32.27	✓	46.39	✓	95.57	✓	9.41	✓
Unclassified	GHVd-like RNA	-	-	10.18		-	-	-	-	-	-	-	-	-	-	-	-
Narnaviridae	Mycovirus	3.88		2.39		11.61		-	-	-	-	-	-	5.28		8.26	
Partitiviridae	Mycovirus	-	-	-	-	-	-	-	-	-	-	-	-	3.73		-	-
Totiviridae	Mycovirus	0.27		-	-	-	-	-	-	-	-	-	-	0.51		-	-

[a]VRR (Virus (or Viroid) Read Ratio) = read count [contigs of species or family] / reference genome length * read count [total assembled contigs] * 1E+03 * 1E+06. The reference genome length of each species is indicated in Supplementary Table S1. [b]Due to the lack of a complete genome sequence at the time of this study, the average genome size of the genus, 7.5kb, was used. Mycoviruses are classified at the family level, because of the low confidence in species classification. For the mycovirus families, the average genome size of the family was used as reference length: *Narnaviridae* = 2.7kb; *Partitiviridae* (2 genome segments) = 4.2kb; *Totiviridae* = 5.6kb. GLRaV-2 = Grapevine leafroll-associated virus 2; GLRaV-3 = Grapevine leafroll-associated virus 3; GRSPaV = Grapevine rupestris stem pitting-associated virus; GVA = Grapevine virus A; GVB = Grapevine virus B; GVE = Grapevine virus E; GFkV = Grapevine fleck virus; GRGV = Grapevine Red Globe virus; GRVFV = Grapevine rupestris vein feathering virus; GSyV-1 = Grapevine Syrah virus 1; TMV = Tobacco mosaic virus; AGVd = Australian grapevine viroid; GYSVd-1 = Grapevine yellow speckle viroid 1; GYSVd-2 = Grapevine yellow speckle viroid 2; HSVd = Hop stunt viroid; GHVd-like RNA = Grapevine hammerhead viroid-like RNA.

Grapevine leafroll-associated virus 3 (GLRaV-3) was detected with the highest abundance in every sample. GLRaV-3 had a VRR ranging between 1 671 and 14 308 across the eight samples, orders of magnitude more than any other virus. The prevalence of GLRaV-3 was expected, as the vines presented typical GLD symptoms earlier in the season. Although several closteroviruses are associated with the disease, GLRaV-3 is considered the primary causal agent (Maree *et al*., 2013). A number of contigs aligned to GLRaV-3 with either a low query coverage or nucleotide identity, suggesting a divergent variant or putative recombinant virus, although further investigation is necessary to confirm this. This may also be due to erroneous assembly that resulted in chimeric contigs. Grapevine leafroll-associated virus 2 (GLRaV-2) was identified in KP4.30O and KP7.30Y, verified with the read-mappings. In a third sample, KP1.29Y, a single 216nt contig aligned to GLRaV-2. The presence of GLRaV-2 in this sample was disregarded, as only five read pairs mapped to the reference sequence. Apart from GLRaV-2 and GLRaV-3, no other members of the family *Closteroviridae* were identified.

The family *Betaflexiviridae* was well represented in all of the samples. Grapevine rupestris stem pitting-associated virus (GRSPaV) was detected in seven samples. Two samples, KP1.27O and KP5.37Y, had a VRR of approximately 278 and 387, respectively, while the other five samples had VRRs below 170. Three vitiviruses were identified; grapevine virus A (GVA), grapevine virus B (GVB) and grapevine virus E (GVE). GVA and GVB were identified in all four old-vine samples, with VRRs ranging from 40 to 78 for GVA, and 135 to 181 for GVB. Two young vines were infected with GVA and one with GVB. GVE was identified only in KP1.27O, with a VRR of approximately 255.

The third family of viruses detected was *Tymoviridae*. Thus far, only two genera in this family have been identified in grapevines; *Maculavirus* and *Marafivirus*. Grapevine fleck virus (GFkV) was detected in seven samples. KP5.37Y had a VRR of 437; nearly three times that of the other six samples. Field spread of this *Maculavirus* reference strain has previously been reported in South African vineyards (Engelbrecht & Kasdorf, 1990). Additionally, several contigs suggested the presence of two marafiviruses that have not been reported in South African grapevines before, namely grapevine rupestris vein feathering virus (GRVFV) and grapevine Syrah virus 1 (GSyV-1). BLASTn results revealed five samples, of which four were from old vines, to be infected with a GRVFV-like species. Analysis of all the GRVFV-like contigs revealed low sequence identity, with

more than half of the contigs showing less than 85% nucleotide identity to the marafivirus. In some cases, low query coverage was observed, decreasing the confidence of the BLAST hit. It is conceivable that a divergent variant of GRVFV had been detected.

GSyV-1 was identified in only one sample, KP1.27O, with a VRR of approximately 7. In this sample, four contigs aligned to the virus with 84 to 90% sequence identity. As this may be the first detection of GSyV-1 in South African grapevines, additional assemblies were performed, generating larger GSyV-1 contigs. A 6 425nt draft genome was assembled, spanning 99% of the reference sequence (NC_012484). The draft sequence shared approximately 85% nucleotide and 95% amino acid identities with eight complete GSyV-1 genome sequences available on GenBank. The presence of the virus was validated with RT-PCR.

A second maculavirus, grapevine Red Globe virus (GRGV) was identified in three samples, KP3.29O, KP4.25O and KP5.37Y. In total, five short contigs were associated with the virus, with four aligning to two partial sequences encoding a methyltransferase domain and the other aligning to a partial RNA-dependent RNA polymerase sequence. These contigs, 244 to 385nt in length, shared 81 to 87% sequence identity with GRGV, though BLAST analysis revealed comparable similarity with corresponding sequences of GRVFV and, to a lesser extent, that of other *Tymoviridae* members. The presence of GRGV was not validated with read-mapping; however, it is important to note that the GRGV reference is represented by a partial sequence, as no complete genome sequence was available on GenBank at the time of this study. The lack of a full-length reference genome may have resulted in false negative detection.

There is a possibility that these contigs may originate from GRVFV or a closely related virus, but due to sequencing and/or assembly errors, and the short length of the sequences, the contigs show similarity to GRGV. This is plausible considering that the assemblies failed to generate contigs that span other parts of the 2 006bp GRGV sequence (AF521977). Results showed that the three samples were also infected with GFkV and a virus resembling GRVFV. Considering the mixed infection of tymoviruses, and potentially multiple variants within these samples, providing conditions for genetic exchange, the possibility of recombination cannot be excluded. Currently, limited sequence data is available for GRGV and GRVFV, and future studies investigating the genetic diversity and occurrence of recombination are would be of interest.

A single 288nt contig, with a total read count of four, aligned to a tobacco mosaic virus sequence with 92% identity. The virus was excluded from the data, as the read-mapping analysis showed only two read pairs aligned to the reference genome. Considering the poor genome coverage, it is possible that the virus was present at a very low titre at the time of extraction or was introduced as a contaminant during sample or library preparation.

Viroids of the family *Pospiviroidae* were present in each sample. These included Australian grapevine viroid (AGVd), hop stunt viroid (HSVd), grapevine yellow speckle viroid 1 (GYSVd-1) and grapevine yellow speckle viroid 2 (GYSVd-2). The viroids were generally present at higher VRRs in the young-vine samples (Table 3.5). In KP3.29O, a contig aligned to an unclassified grapevine hammerhead viroid-like (GHVd) RNA sequence with 100% coverage and 98% nucleotide identity. This circular viroid-like RNA was discovered in grapevine using a computational algorithm named progressive filtering of overlapping small RNAs. It encodes an active hammerhead ribozyme that enables the viroid to self-cleave (Wu *et al*., 2012). Hammerhead ribozymes have not been identified in pospiviroid sequences and it is speculated that cleavage is instead mediated by an RNase III-like enzyme (Flores *et al*., 2009). Currently, the International Council for the Study of Viruses and Virus-like Diseases of the Grapevine (ICVG) does not recognise GHVd as one of the viroids that infect grapevine, citing the suspected, but unproven viroidal nature of the organism as the reason for its omission (Martelli, 2014). This clarifies why it is referred to as a viroid-like RNA entity, rather than a viroid. The available sequence was therefore not included in the virus reference database for the read-mapping analyses.

Apart from viruses that primarily infect grapevine, several mycovirus species of the families *Totiviridae*, *Partitiviridae* and *Narnaviridae* were detected. Of these, *Narnaviridae* was the best represented family, specifically the genus *Mitovirus*. Interestingly, mitochondrial mitovirus-like sequences have been found to be widespread in plant genomes, appearing to be the result of horizontal transfer from fungi (Bruenn *et al*., 2015). This may aid in plant defences against fungal pathogens (Shackelton & Holmes, 2008). The presence of these mycovirus families could not be validated with read-mapping. This is due to the limited number of mycovirus sequences that were included as references in the read-mapping analyses. Only accessions of grapevine-specific mycoviruses were included, of which all but one, are partial genomic sequences.

Figure 3.3 displays the viral diversity between the old and young vines. Bioinformatic results revealed a more complex viral diversity in the old vines; between six and eight grapevine-associated viruses were detected in the four old-vine samples, while two to five viruses infected the young vines. This may be due to recent virus infections, i.e. the viruses have not had time to accumulate. The collective virus and viroid component in two of the young-vine samples (KP3.20Y and KP5.37Y) had VRRs ranging from 3 297 to 5 000, compared to more than 9 800 in the other six samples. Several viruses of the families *Closteroviridae*, *Betaflexiviridae* and *Tymoviridae* were detected. Viroid infection was more uniform, with all four pospiviroids present in six of the eight samples.

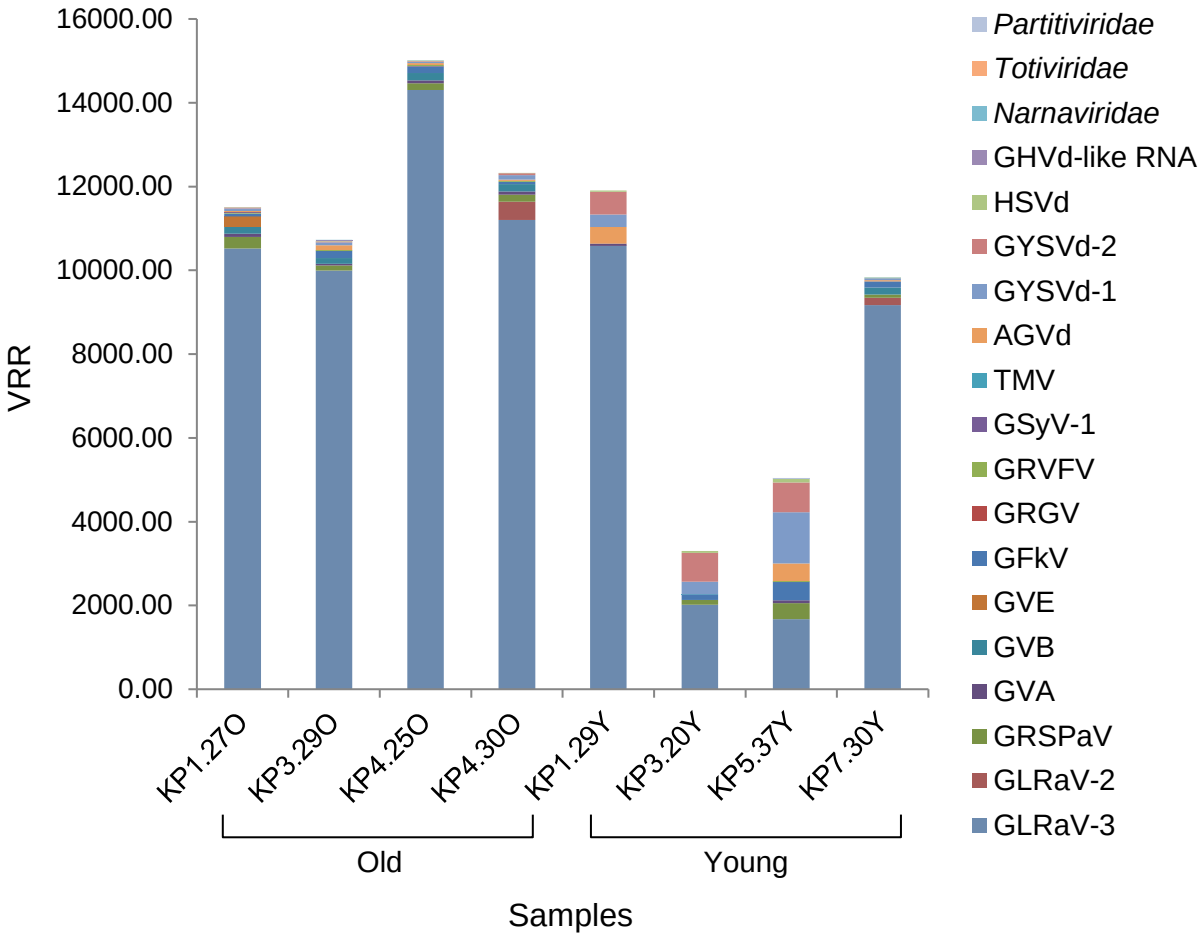

Figure 3.3: The diversity of grapevine viruses and viroids in the old- and young-vine samples. The relative abundance of the pathogens is expressed as the VRR. VRR (Virus (or Viroid) Read Ratio) = read count [contigs of species or family] / reference genome length * read count [total assembled contigs] * 1E+03 * 1E+06. GLRaV-3 = Grapevine leafroll-associated virus 3; GLRaV-2 = Grapevine leafroll-associated virus 2; GRSPaV = Grapevine rupestris stem pitting-associated virus; GVA = Grapevine virus A; GVB = Grapevine virus B; GVE = Grapevine virus E; GFkV = Grapevine fleck virus; GRGV = Grapevine Red Globe virus; GRVFV = Grapevine rupestris vein feathering virus; GSyV-1 = Grapevine Syrah virus 1; TMV = Tobacco mosaic virus; AGVd = Australian grapevine viroid; GYSVd-1 = Grapevine yellow speckle viroid 1; GYSVd-2 = Grapevine yellow speckle viroid 2; HSVd = Hop stunt viroid; GHVd-like RNA = Grapevine hammerhead viroid-like RNA.

Approximately 7 to 17% of contigs in all of the samples, representing 0.03 to 0.4% of assembled reads, could not be classified through nucleotide BLAST analysis. The unidentified contigs were additionally subjected to tBLASTx to search for amino acid identity with available sequences on GenBank. As these contigs could not be identified based on nucleotide similarity, lowering the confidence in species classification, the contigs were grouped in the respective virus families. Table 3.6 displays the distribution of the reads, expressed as the VRR.

BLAST analysis revealed 15 contigs with similarity to *Tymoviridae* species in four samples. Eleven contigs aligned to GRVFV with 53 to 97% amino acid identity, while the remaining four had similarity with sequences of GFkV, GSyV-1 and two marafiviruses not associated with grapevine, namely citrus sudden death-associated virus and oat blue dwarf virus. These results support the presence of a possible divergent variant in *Tymoviridae* in these vines.

Five mycovirus families were identified in the eight samples, namely *Chrysoviridae*, *Endornaviridae*, and the previously identified *Narnaviridae*, *Partitiviridae* and *Totiviridae*, consistent with the results of previous grapevine metagenomic studies (Al Rwahnih *et al*., 2011; Coetzee *et al*., 2010; Espach, 2013). The family *Narnaviridae*, identified in seven samples, had the largest VRR in six samples, ranging between 2 and 72. These sequences had little amino acid similarity to known mycovirus species. Of the 60 contigs, only five were more than 70% identical to mycoviruses, while 28 showed less than 50% identity. This may suggest that these sequences represent putative novel mycoviruses. Further investigation of these mycoviruses and their biological impact would be of interest.

Other sequences related to viruses that do not infect plants and regarded as contaminants, were also detected. These included mammalian viruses belonging to the families *Picobirnaviridae* and *Picornaviridae*, among others. The source of the contamination is unknown, but it may have been introduced during sample processing, library preparation or sequencing. Seventeen contigs did not align to any available GenBank nucleotide or protein sequences. The inability to classify these sequences may be ascribed to assembly errors.

Table 3.6: The distribution of the reads accounting for contigs that were classified in grapevine virus families with tBLASTx, expressed as the VRR[a].

Virus family	Sample							
	KP1.27O	KP3.29O	KP4.25O	KP4.30O	KP1.29Y	KP3.20Y	KP5.37Y	KP7.30Y
Closteroviridae	-	-	-	-	5.77	0.76	-	0.55
Tymoviridae	-	2.04	6.80	5.01	-	-	2.29	-
Chrysoviridae	0.02	-	-	-	-	-	-	0.16
Endornaviridae	-	-	-	-	0.32	-	-	0.06
Narnaviridae	2.25	5.22	72.43	22.99	2.04	2.83	-	23.39
Partitiviridae	-	0.59	2.74	-	-	0.81	7.80	0.62
Totiviridae	0.04	-		-	-	-	-	-

[a]VRR (Virus (or Viroid) Read Ratio) = read count [contigs of species or family] / reference genome length * read count [total assembled contigs] * 1E+03 * 1E+06. The average genome size of each virus family was used as reference length, except for *Closteroviridae*, as only one species, grapevine leafroll-associated virus 3 was identified, of which the reference length is indicated in Supplementary Table S1. Average genome sizes of virus families: *Tymoviridae* = 6.7kb; *Chrysoviridae* (4 genome segments) = 12.3kb; *Endornaviridae* = 15.8kb; *Narnaviridae* = 2.7kb; *Partitiviridae* (2 genome segments) = 4.2kb; *Totiviridae* = 5.6kb.

In order to determine which genetic variants of the most prevalent viruses were present, the reads were mapped to representative genome sequences of different variant groups (Table 3.1). Non-specific matches were ignored, permitting reads to be mapped uniquely. The confidence in positive detection increases with genome coverage. Variants with more than 80% coverage were regarded present with a high level of confidence, while those with genome coverage between 50 and 80% were suspected to be present. Table 3.7 displays the percentage genome coverage and average depth of coverage of each reference sequence, in the eight samples.

GLRaV-3 variant group II was found to be the most prevalent across the samples, with a significantly higher average depth of coverage than any other variant, except for groups I and III in KP1.27O (Table 3.7). This is consistent with the findings of a previous South African virome study, and past surveys of grapevine-infecting viruses in the wine-producing regions of the Western Cape (Coetzee *et al*., 2010; Jooste *et al*., 2011; Jooste *et al*., 2015). Mixed infections between variants of GRSPaV group 2a, 2b and 3 were common among old and young vines. GVA variants were well represented in the old vines, with group I present in all four old-, but no young-vine samples. Group II was identified in only one sample, KP1.27O. It has been suggested that group II variants are associated with Shiraz disease in South African vineyards and with Australian Shiraz disease in Australian grapevines, specifically the cultivars Shiraz and Merlot, although

the aetiology remains unresolved (Goszczynski *et al*., 2008; Goszczynski & Habili, 2012).

The number of virus variants, as estimated by the length fraction, corresponds approximately to that detected by read-mapping. This shows that contigs are assembled separately for different variants present. However, due to sequence similarity between variants and the way the assembly algorithm works, multiple contigs may be generated for each variant. Therefore, the number and total length of contigs cannot be used to give an indication of virus species abundance. In this study, these factors were accounted for by calculating the ratio of the total concatenated contig length to the reference genome length, for each species, as a variant number indicator.

Based on the 50% genome coverage cut-off, 31 virus variants were detected in the old and 16 in the young vines. This was expected, as the old vines have had exposure to viral pathogens over a longer period. Considering that the young vines underwent heat therapy before propagation and that these were exposed to the natural virus inoculum for a much shorter period, it is surprising that vines accumulate multiple viruses in a relatively short time. The young vines have also had less time to be infected with multiple genetic variants.

Several co-infecting viruses are typically associated with grapevine diseases. Viral complexes are a common occurrence in South African vineyards and can result in increased disease severity (Jooste *et al*., 2015; Prosser *et al*., 2007). Different virus variants may lead to differences in disease symptom expression. It is therefore necessary to characterise not only the virus, but also the variant complexity. Mixed infections can complicate reliable virus and variant detection through conventional methods. In this study, NGS is shown to be a reliable tool to investigate the virus and variant diversity in grapevines in an unbiased manner.

Table 3.7: The percentage genome coverage[a] (% Gen. cov.), average depth of coverage[b] (Avg. cov.) and length fraction[c] (Len. frac.) of the respective virus variants in the eight samples. Yellow and blue shading is used to indicate variants with 50 to 80%, and greater than 80% genome coverage, respectively.

Virus	Variant group	Sample															
		KP1.27O		KP3.29O		KP4.25O		KP4.30O		KP1.29Y		KP3.20Y		KP5.37Y		KP7.30Y	
		% Gen. cov. (Avg. cov.)	Len. frac.	% Gen. cov. (Avg. cov.)	Len. frac.	% Gen. cov. (Avg. cov.)	Len. frac.	% Gen. cov. (Avg. cov.)	Len. frac.	% Gen. cov. (Avg. cov.)	Len. frac.	% Gen. cov. (Avg. cov.)	Len. frac.	% Gen. cov. (Avg. cov.)	Len. frac.	% Gen. cov. (Avg. cov.)	Len. frac.
GLRaV-2	93/955	0.77 (0.23)		0.55 (0.10)		0.28 (0.06)		3.07 (0.20)		0.28 (0.04)		0.42 (0.03)		0.39 (0.06)		3.23 (2.35)	
	OR1	1.42 (2.11)		0.47 (0.96)		1.14 (0.79)		99.90 (140.26)		1.70 (0.38)		0.44 (0.44)		1.29 (0.94)		100 (1 535.22)	
	BD	1.30 (3.33)		0.91 (1.26)		0.85 (0.64)		1.01 (2.13)	1.00	0.67 (0.43)		0.82 (0.45)		0.57 (1.25)		1.60 (47.60)	0.97
	RG	1.52 (1.22)		0.60 (0.24)		0.66 (0.18)		0.34 (0.19)		0.29 (0.09)		0.29 (0.14)		0.64 (0.30)		0.96 (1.98)	
	SG	0.33 (0.00)		0.14 (0.00)		-		0.91 (0.09)		-		-		0.10 (0.00)		2.71 (0.24)	
GLRaV-3	Group I	99.99 (4 657.12)		100 (921.77)		16.23 (1.60)		100 (840.82)		18.13 (1.86)		49.19 (42.29)		5.85 (0.60)		55.07 (112.75)	
	Group II	100 (5 348.57)		100 (5 927.59)		100 (6 328.44)		100 (1 774.43)		100 (2 963.10)		98.51 (1 310.43)		100 (2 750.72)		100 (12 235.69)	
	Group III	100 (4 423.69)	3.67	14.40 (1.55)	2.50	6.29 (0.18)	1.02	13.55 (7.62)	2.42	8.35 (0.48)	1.04	6.72 (0.17)	1.41	2.49 (0.03)	0.97	10.86 (0.66)	1.78
	Group VI	13.94 (3.56)		4.20 (0.29)		0.69 (0.01)		6.06 (0.44)		0.12 (0.00)		0.95 (0.01)		0.52 (0.01)		3.70 (0.18)	
	Group VII	18.89 (32.48)		11.16 (1.25)		0.62 (0.01)		11.81 (2.01)		0.49 (0.01)		8.51 (0.35)		0.22 (0.00)		9.98 (0.71)	
GRSPaV	Group 1	2.95 (0.14)		1.10 (0.09)		0.37 (0.05)		6.15 (0.20)		0.38 (0.04)		0.35 (0.02)		0.84 (0.06)		4.28 (0.48)	
	Group 2a	99.81 (86.18)		99.59 (24.62)		27.59 (7.18)		83.46 (15.06)		-		23.67 (2.66)		96.36 (42.53)		76.29 (24.15)	
	Group 2b	50.99 (76.25)	2.35	38.58 (10.93)	1.70	28.34 (8.94)	0.98	86.60 (27.58)	1.27	0.47 (0.03)		26.33 (3.80)	1.03	61.18 (22.11)	1.59	88.19 (61.62)	0.90
	Group 3	77.88 (415.43)		72.66 (60.08)		76.19 (50.27)		67.40 (10.80)		4.47 (0.07)		74.28 (25.09)		75.80 (68.34)		57.20 (15.15)	

Virus	Variant group	Sample															
		KP1.27O		KP3.29O		KP4.25O		KP4.30O		KP1.29Y		KP3.20Y		KP5.37Y		KP7.30Y	
		% Gen. cov. (Avg. cov.)	Len. frac.	% Gen. cov. (Avg. cov.)	Len. frac.	% Gen. cov. (Avg. cov.)	Len. frac.	% Gen. cov. (Avg. cov.)	Len. frac.	% Gen. cov. (Avg. cov.)	Len. frac.	% Gen. cov. (Avg. cov.)	Len. frac.	% Gen. cov. (Avg. cov.)	Len. frac.	% Gen. cov. (Avg. cov.)	Len. frac.
GVA	Group I	99.88 (117.14)	1.94	99.07 (13.45)	2.49	99.52 (13.28)	1.88	98.97 (14.31)	1.84	-	0.99	-		-	0.73	-	
	Group II	85.96 (34.12)		1.56 (0.03)		-		0.86 (0.01)		-		-		-		2.24 (0.07)	
	Group III	11.59 (3.35)		99.96 (26.81)		99.99 (22.97)		99.20 (14.84)		98.64 (14.95)		0.94 (0.02)		58.67 (14.90)		0.95 (0.06)	
	Group IV	41.61 (6.36)		6.14 (0.18)		0.34 (0.01)		3.66 (0.04)		0.33 (0.00)		0.33 (0.00)		3.54 (0.11)		1.62 (0.02)	
GVB	94/971	99.53 (295.67)	0.97	98.58 (71.84)	1.85	92.47 (50.64)	1.00	93.45 (41.60)	1.00	-		0.29 (0.00)		0.34 (0.00)		91.42 (117.25)	1.08
	QMWH	35.91 (17.09)		31.42 (2.88)		21.77 (2.37)		15.32 (1.84)		0.50 (0.02)		0.49 (0.02)		0.57 (0.03)		38.28 (5.09)	
GVE	TvAQ7	1.26 (0.02)	0.97	0.62 (0.01)		0.46 (0.01)		-		-		-		0.33 (0.00)		0.22 (0.00)	
	SA94	100 (601.41)		2.43 (0.03)		1.84 (0.03)		0.37 (0.00)		3.45 (0.04)		0.34 (0.00)		0.22 (0.00)		1.33 (0.01)	

[a]Percentage genome coverage refers to the percentage of the reference genome length covered by the mapped reads. [b]Average depth of coverage, as determined by CLC's read-mapping algorithm, is the sum of the bases of the aligned portion of all the mapped reads divided by the reference genome length. [c]Length fraction denotes the ratio of the total concatenated contig length to the reference genome length, for each virus species. The reference genome length of each virus species is indicated in Supplementary Table S1. GLRaV-2 = Grapevine leafroll-associated virus 2; GLRaV-3 = Grapevine leafroll-associated virus 3; GRSPaV = Grapevine rupestris stem pitting-associated virus; GVA = Grapevine virus A; GVB = Grapevine virus B; GVE = Grapevine virus E.

3.3.4 Reverse transcription PCR screening

Total RNA was successfully extracted from the 20 samples, with a mean RNA concentration and standard deviation of 562 ± 171ng/µl, and an average 260/280 and 260/230 absorbance ratio of 2.13 and 2.02, respectively. Reverse transcription PCR detection assays were performed to validate the presence of viruses detected in the NGS data (not including viroids and mycoviruses). All the collected samples were screened. Table 3.8 indicates the viruses detected in the 20 (eight NGS and 12 additional) samples. The most prevalent viruses identified in both old and young vines are GLRaV-3, GFkV and GRSPaV. There were 65 positive virus detections in the old-vine samples compared to 34 in the young-vine samples.

Results confirmed the presence of the identified viruses in all eight samples, except GVA in KP1.29Y and KP5.37Y, and GRVFV in KP5.37Y. The false negative GVA RT-PCR results in KP1.29Y and KP5.37Y can be attributed to a 3' nucleotide mismatch of the forward primer with genetic variants of GVA group III, the only group detected in these two samples. The mismatch was confirmed by the read-mapping analysis. Two sets of primers were selected to screen for GFkV, both targeting the coat protein. One pair yielded positive GFkV detection in all 20 samples. This contradicts the NGS result for KP1.29Y, in which GFkV was not detected with *de novo* assemblies or read-mappings. The other primer pair detected 16 samples positive for GFkV, excluding KP1.29Y. Due to the inconsistent detection of this virus, the amplicons obtained by both primer pairs were Sanger sequenced. BLAST analysis revealed 94 to 98% nucleotide, and 94 to 100% amino acid identities with corresponding GFkV sequences. The inability to identify GFkV in the NGS data for sample KP1.29Y remains unexplained. A speculative possibility is low concentrations of the virus in, and/or the low sequencing depth obtained for sample KP1.29Y.

Primers were designed to amplify a 441bp fragment of the coat protein of GRVFV. Nine old-vine samples and none of the young-vine samples tested positive for GRVFV. To validate the efficiency of the primers, the amplicon was directly Sanger sequenced. The sequence shared 81 to 85% nucleotide, and 96 to 99% amino acid identities with the available GRVFV coat protein sequences. However, the amplicon sequence had a number of ambiguous base calls, roughly 8 to 9%, that can be attributed to mixed

variant infection, or the quasispecies[1] nature of plant viruses. GRVFV-associated reads were also identified in KP5.37Y, but the virus could not be detected with RT-PCR. Poor binding specificity of the reverse primer is suspected, though this could not be validated due to a lack of sequence data, generated for this sample that is specific to that region of the genome. This result would be the first account of GRVFV in South African grapevines; however, until its complete genome is characterised and the genetic diversity of the virus is studied in greater depth, the virus is tentatively regarded as a GRVFV-like species.

Table 3.8: Semi-quantitative RT-PCR results of all 20 samples. Sequenced samples are indicated in bold. Double and single plus signs represent strong and weak amplification, respectively.

Sample	Virus									
	GLRaV-2	GLRaV-3	GRSPaV	GVA	GVB	GVE	GFkV[a]	GFkV[b]	GRVFV	GSyV-1
KP1.27O		++	++	++	++	++	++	++	++	++
KP2.33O		++	++				++	++		
KP2.34O		++					++	++	++	
KP3.28O		++	++				++	++	++	
KP3.29O		++	++	++	++		++	++	++	
KP3.30O		++	++	++		+	++	++	++	
KP3.34O		++	++	++	+		++	++	++	
KP4.25O		++	++	++	++		++	++	++	
KP4.30O	++	++	++	++	+		++	++	++	
KP4.31O		++	++	++	++		++	++	++	
KP1.29Y		++						++		
KP1.32Y		++		++				++		
KP3.20Y		++	++				++	++		
KP3.31Y		+						++		
KP4.37Y		++						++		
KP5.34Y		++					++	++		
KP5.37Y		++	++				++	++		
KP6.27Y		++	++				++	++		
KP6.32Y		++	++				++	++		
KP7.30Y	++	++	++		+		++	++		

[a]Glasa *et al*. (2011) and [b]Gambino & Gribaudo (2006) primer pairs. GLRaV-2 = Grapevine leafroll-associated virus 2; GLRaV-3 = Grapevine leafroll-associated virus 3; GRSPaV = Grapevine rupestris stem pitting-associated virus; GVA = Grapevine virus A; GVB = Grapevine virus B; GVE = Grapevine virus E; GFkV = Grapevine fleck virus; GRVFV = Grapevine rupestris vein feathering virus; GSyV-1 = Grapevine Syrah virus 1.

[1]A quasispecies is defined as a cloud of viruses from a single origin that are related through mutation. The evolution of quasispecies, relating to RNA viruses, is influenced by high mutations rates as a result of short replication times and the error-prone nature of RNA-dependent RNA polymerases.

Several contigs with similarity to the marafivirus, GSyV-1, and spanning 99% of its genome, were identified in sample KP1.27O. To verify the presence of this virus, RT-PCR was performed in triplicate and the expected fragments were cloned into the pGEM-T Easy vector (Promega) and Sanger sequenced. BLASTn analysis confirmed the sequences (KT898045-47) to be GSyV-1. These sequences shared 89 to 97% nucleotide, and 98 to 100% amino acid identities with 55 GSyV-1 coat protein sequences available on GenBank (Oosthuizen *et al*., 2016). To determine the prevalence of the virus, 302 samples from a number of cultivars, previously collected from vineyards across the Western Cape (Jooste *et al*., 2015), were screened for GSyV-1. Only one additional Merlot vine tested positive. This is the first reported detection of GSyV-1 in South African grapevines (Oosthuizen *et al*., 2016).

Future studies on the incidence and biological impact of the identified tymoviruses would be valuable. GFkV infections have been reported across the world, while the other members of *Tymoviridae* have a more limited distribution (Martelli, 2014). According to the directory of virus and virus-like diseases of the grapevine and their agents, GFkV and GRVFV are associated with the grapevine fleck complex, causing fleck and vein feathering symptoms in *Vitis rupestris* and occurring mostly as latent or semi-latent infections in *Vitis vinifera* (Martelli, 2014). Although GSyV-1 was previously isolated from declining Syrah vines, there is no known association between the virus and grapevine disease symptoms (Al Rwahnih *et al*., 2009; Glasa *et al*., 2015).

3.4 Conclusion

In this study, the viromes of four old and four young Pinotage grapevines, that displayed symptoms of GLD earlier in the season, were sequenced. Several viruses of the families *Closteroviridae*, *Betaflexiviridae* and *Tymoviridae* were detected, along with four grapevine-associated viroids. The virus community was more diverse in the old vines with six to eight different viruses, exclusive of mycoviruses, detected in the four old-vines samples. The economically important GLRaV-3, specifically variants of group II, was found to be the most prevalent virus, present in all eight sequenced samples with the highest relative abundance. Additionally, GLRaV-3 was detected in the other 12 vines that had not been sequenced. This is consistent with the results of previous studies focussed on vines affected by GLD in the Western Cape. GSyV-1, and probably GRVFV, was identified for the first time in South African grapevines, expanding the global distribution of the virus(es). The results obtained in this study would suggest that

the old vines have remained viable for the production of high-quality wines, despite the presence of a number of viral pathogens.

This study has once again demonstrated the value of metagenomic NGS studies to investigate grapevine viruses and viral diseases. Reference-mappings may serve as a bioinformatic virus detection tool, but should be used in conjunction with *de novo* assemblies and alignment-based similarity searches. Furthermore, it is important to consider the differences in grapevine virus genome size, the varying sequencing depth across the samples and the number of infecting virus variants. To account for these factors, new methods to normalise the NGS data and estimate the number of virus variants, namely the VRR and length fraction, were implemented.

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.

Alabi, O.J., Martin, R.R., Naidu, R.A., 2010. Sequence diversity, population genetics and potential recombination events in Grapevine rupestris stem pitting-associated virus in Pacific North-West vineyards. J. Gen. Virol. 91(1), 265-276.

Alabi, O.J., Poojari, S., Sarver, K., Martin, R.R., Naidu, R.A., 2013. Complete genome sequence analysis of an American isolate of Grapevine virus E. Virus Genes. 46(3), 563-566.

Alabi, O.J., Al Rwahnih, M., Mekuria, T.A., Naidu, R.A., 2014. Genetic diversity of Grapevine virus A in Washington and California vineyards. Phytopathology. 104(5), 548-560.

Al Rwahnih, M., Daubert, S., Golino, D., Rowhani, A., 2009. Deep sequencing analysis of RNAs from a 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.

Al Rwahnih, M., Sudarshana, M.R., Uyemoto, J.K., Rowhani, A., 2012. Complete genome sequence of a novel vitivirus isolated from grapevine. J. Virol. 86(17), 9545.

Al Rwahnih, M., Dave, A., Anderson, M.M., Rowhani, A., Uyemoto, J.K., Sudarshana, M.R., 2013. Association of a DNA virus in grapevines affected by red blotch disease in California. Phytopathology. 103(10), 1069-1076.

Altschul, S.F., Gish, W., Miller, W., Myers, E.W., Lipman, D.J., 1990. Basic Local Alignment Search Tool. J. Mol. Biol. 215(3), 403-410.

Bester, R., Pepler, P.T., Burger, J.T., Maree, H.J., 2014. Relative quantitation goes viral: An RT-qPCR assay for a grapevine virus. J. Virol. Methods. 210, 67-75.

Bolger, A.M., Lohse, M., Usadel, B., 2014. Trimmomatic: A flexible trimmer for Illumina sequence data. Bioinformatics. 30(15), 2114-2120.

Bruenn, J.A., Warner, B.E., Yerramsetty, P., 2015. Widespread mitovirus sequences in plant genomes. Peer J. 3, e876. DOI: 10.7717/peerj.876.

Burger, J.T., Maree, H.J., 2015. Metagenomic next-generation sequencing of viruses infecting grapevines. In: Lacomme, C. (Ed.). Plant Pathology. Methods in Molecular Biology, 1302: 315-330. Humana Press, New York, NY, USA.

Camacho, C., Coulouris, G., Avagyan, V., Ma, N., Papadopoulos, J., Bealer, K., Madden, T.L., 2009. BLAST+: Architecture and applications. BMC Bioinformatics. 10, 421. DOI: 10.1186/1471-2105-10-421.

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

Engelbrecht, D.J. Kasdorf, G.G.F., 1990. Field spread of corky bark, fleck, leafroll and Shiraz decline diseases and associated viruses in South African grapevines. Phytophylactica. 22(3), 347-354.

Espach, Y., 2013. The determination of mycoviral sequences in grapevine using next-generation sequencing. MSc thesis, Stellenbosch University.

Flores, R., Gas, M., Molina-Serrano, D., Nohales, M., Carbonell, A., Gago, S., De la Peña, M., Daròs, J., 2009. Viroid Replication: Rolling-circles, enzymes and ribozymes. Viruses. 1(2), 317-334.

Gambino, G., Gribaudo, I., 2006. Simultaneous detection of nine grapevine viruses by multiplex reverse transcription-polymerase chain reaction with coamplification of a plant RNA as internal control. Phytopathology. 96(11), 1223-1229.

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.

Glasa, M., Predajňa, L., Komínek, P., 2011. Grapevine fleck virus isolates split into two distinct molecular groups. J. Phytopathol. 159(11-12), 805-807.

Glasa, M., Predajňa, L., Šoltys, K., Sabanadzovic, S., Olmos, A., 2015. Detection and molecular characterisation of Grapevine Syrah virus 1 isolates from central Europe. Virus Genes. 51(1), 112-121.

Goszczynski, D.E., Du Preez, J., Burger, J.T., 2008. Molecular divergence of Grapevine virus A (GVA) variants associated with Shiraz disease in South Africa. Virus Res. 138(1-2), 105-110.

Goszczynski, D.E., Habili, N., 2012. Grapevine virus A variants of group II associated with Shiraz disease in South Africa are present in plants affected by Australian Shiraz disease, and have also been detected in the USA. Plant Pathol. 61(1), 205-214.

Hansen, K.D., Brenner, S.E., Dudoit, S., 2010. Biases in Illumina transcriptome sequencing caused by random hexamer priming. Nucleic Acids Res. 38(12), e131. DOI: 10.1093/nar/gkq224.

Hu, G.J., Dong, Y.F., Zhang, Z.P., Fan, X.D., Ren, F., Zhu, H.J., 2014. Complete nucleotide sequence of a new isolate of Grapevine virus B from China. J. Plant Pathol. 96(2), 403-406.

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.

Jooste, A.E.C., Molenaar, N., Maree, H.J., Bester, R., Morey, L., De Koker, W.C., Burger, J.T., 2015. Identification and distribution of multiple virus infections in grapevine leafroll diseased vineyards. Eur. J. Plant Pathol. 142(2), 363-375.

Krenz, B., Thompson, J.R., Fuchs, M., Perry, K.L., 2012. Complete genome sequence of a new circular DNA virus from grapevine. J. Virol. 86(14), 7715.

MacKenzie, D.J., 1997. A standard protocol for the detection of viruses and viroids using a reverse transcription-polymerase chain reaction technique. Document CPHBT-RT-PCR1.00. Canadian Food Inspection Agency.

Maree, H.J., Almeida, R.P.P., Bester, R., Chooi, K.M., Cohen, D., Dolja, V.V., Fuchs, M.F., Golino, D.A., Jooste, A.E.C., Martelli, G.P., Naidu, R.A., Rowhani, A., Saldarelli, P., Burger, J.T., 2013. Grapevine leafroll-associated virus 3. Front. Microbiol. 4, 82. DOI: 10.3389/fmicb.2013.00082.

Maree, H.J., Pirie, M.D., Oosthuizen, K., Bester, R., Rees, D.J.G., Burger, J.T., 2015. Phylogenomic analysis reveals deep divergence and recombination in an economically important grapevine virus. PLoS ONE. 10(5), e0126819. DOI: 10.1371/journal.pone.0126819.

Martelli, G.P., 2014. Directory of virus and virus-like diseases of the grapevine and their agents. J. Plant Pathol. 96(Suppl. 1), 1-136.

Mortazavi, A., Williams, B.A., McCue, K., Schaeffer, L., Wold, B., 2008. Mapping and quantifying mammalian transcriptomes by RNA-Seq. Nat. Methods. 5(7), 621-628.

Oosthuizen, K., Coetzee, B., Maree, H.J., Burger, J.T., 2016. First report of Grapevine Syrah virus 1 in South African grapevines. Plant Dis. 100(6), 1252.

Poojari, S., Alabi, O.J., Naidu, R.A., 2013. Molecular characterization and impacts of a strain of Grapevine leafroll-associated virus 2 causing asymptomatic infection in a wine grape cultivar. Virol. J. 10, 324. DOI: 10.1186/1743-422X-10-324.

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.

Shackelton, L.A., Holmes, E.C., 2008. The role of alternative genetic codes in viral evolution and emergence. J. Theor. Biol. 254(1), 128-134.

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

Wu, Q., Wang, Y., Cao, M., Pantaleo, V., Burgyan, J., Li, W., Ding, S., 2012. Homology-independent discovery of replicating pathogenic circular RNAs by deep sequencing and a new computational algorithm. Proc. Natl. Acad. Sci. U.S.A. 109(10), 3938-3943.

Zhang, Y., Singh, K., Kaur, R., Qui, W., 2011. Association of a novel DNA virus with the grapevine vein-clearing and vine decline syndrome. Phytopathology. 101(9), 1081-1090.

Internet sources

http://www.bioinformatics.babraham.ac.uk/projects/fastqc/

https://www.qiagenbioinformatics.com/

Chapter 4: The fungal diversity in old and young grapevines

4.1 Introduction

Grapevine has an intricate microbiome that plays an essential role in plant health, growth and productivity through the complex interactions between the different microbes (Barata *et al*., 2012). Fungi and bacteria colonise different parts of the crop, either internally as endophytes, or externally as epiphytes. However, the interactions between these organisms and the host are not well understood. Grape-associated microbes are undoubtedly active participants in the initial wine fermentation stages (Barata *et al*., 2012). Endophytes colonising other vine tissues may play an important role in shaping the microbial composition of grapes, thereby indirectly influencing the wine quality, as assessed by sensory panels.

The role of plant microbes, apart from pathogens, has been mostly overlooked. In the past, studies of microbial communities were limited by biases associated with culture-dependent techniques, and the low resolution of conventional molecular methods. The development of next-generation sequencing (NGS) has significantly facilitated microbial ecology research. This technology provides adequate sequencing depth for the detection of rare taxa, allowing a more comprehensive analysis of the microbiome. Deep amplicon sequencing, whereby taxonomically informative markers are amplified from total DNA and subjected to NGS, is a common targeted metagenomics approach used in environmental surveys.

The grapevine microbiome, particularly the microbial consortia of the soil, rhizosphere and above-ground tissues, has received considerable attention in recent years. Pinto *et al*. (2014) surveyed the fungal, yeast and bacterial diversity in both healthy and diseased leaves along the vegetative cycle of the crop, revealing a highly diverse and dynamic microbiome, with a negative influence from chemical treatments on the balance between pathogens and their endophytic antagonists. Another study evaluated the effects of both chemical and biological treatment on the foliar microbiota, but observed negligible changes in community structure (Perazzolli *et al*., 2014). Viticultural practices were, however, shown to shape the bacterial assemblages of the shoots and rhizosphere (Campisano *et al*., 2014; Vega-Avila *et al*., 2015). A survey of the microbial communities associated with the soil, rhizosphere, roots and phyllosphere concluded that soil serves as a primary source of colonisation. The authors described local regionalisation of soil bacterial taxa (Zarraonaindia *et al*., 2015).

The microbial landscape of the grapes and various stages of wine fermentation have also been extensively studied. Bokulich *et al*. (2014) studied the fungal and bacterial consortia of grape musts and observed distinct microbial biogeography across major Californian grape-growing regions, with some influences from the cultivar, vintage, and climate. Since then, similar regional patterns have been reported across different viticultural zones in Portugal, Italy and New Zealand, and between neighbouring vineyards in South Africa (Marzano *et al*., 2016; Pinto *et al*., 2015; Setati *et al*., 2015; Taylor *et al*., 2014). Bokulich *et al*. (2016) repeated their previous sequencing strategy on samples collected at various steps during the winemaking process and identified a correlation between the grape microbiota and chemical composition of finished wines. Indeed, NGS has proven to be a valuable tool to investigate microbial dynamics in grapevine.

Several vineyards have managed to remain profitable beyond their life expectancy (~20 years). Moreover, there is anecdotal evidence suggesting that old vines produce wines of higher quality than young vines. As previously stated, vines qualify for 'old' status when they reach 35 years of age, whereas 'young' vines are less than ten years old. To date, little research has been performed to characterise the endophytic fungal communities, referred to as the mycobiome, in old vines. In this study, the fungal diversity in old and young Pinotage vines was assessed using NGS in a targeted metagenomics approach.

4.2 Materials and methods

4.2.1 Plant material and DNA extraction

The four old and four young Pinotage vines selected for sequencing, as described in Chapter 3 (section 3.2.2) were sampled for this study. For the purpose of characterising the endophytic fungal communities in the vascular tissues, cane material was used for nucleic acid extraction. The canes were surface sterilised by submersion as follows: 45 seconds in 70% ethanol, 1 minute in 1% sodium hypochlorite, 45 seconds in 70% ethanol and 2x 45 seconds in sterile reverse osmosis water. The bark was removed and cane cuttings were reduced to powder in liquid nitrogen. For each sample, total DNA was extracted from approximately 100 milligrams of pulverised cane material using a robust cetyltrimethylammonium bromide (CTAB) method (2% [w/v] CTAB, 2.5% [w/v] PVP-10, 100mM Tris-HCl pH8, 1.4M NaCl, 20mM EDTA pH8 and 3% [v/v] β-mercaptoethanol). This extraction procedure (unpublished, Supplementary Protocol S1)

included treatment with RNase A (Thermo Scientific), three chloroform-isoamyl alcohol (24:1) extraction steps and precipitation with isopropanol. The DNA concentration and quality were measured with the NanoDrop 2000 Spectrophotometer (Thermo Scientific), and the integrity was assessed by gel electrophoresis (1% [w/v] agarose-TAE).

4.2.2 Amplification and next-generation sequencing

Primers were selected to amplify the nuclear ribosomal internal transcribed spacer 2 (ITS2) region: ITS3F (5'-GCATCGATGAAGAACGCAGC-3') and ITS4R (5'-TCCTCCGC TTATTGATATGC-3') (White *et al*., 1990). The locus-specific primers were ligated to Illumina adapter overhang sequences (Nextera transposase adapters) that are compatible with Illumina index and sequencing adapters, as follows: forward overhang (5'-TCGTCGGCAGCGTCAGATGTGTATAAGAGACAG-[ITS3F]) and reverse overhang (5'-GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAG-[ITS4R]). Amplification was carried out in 50µl reactions with 120ng template DNA using Phusion High-Fidelity DNA Polymerase (New England BioLabs) as per manufacturer's instructions. Cycling conditions consisted of an initial denaturation step at 98°C for 30 seconds, followed by 35 cycles of denaturation at 98°C for 10 seconds, annealing at 54°C for 30 seconds and extension at 72°C for 30 seconds; and a final extension step at 72°C for 10 minutes. Amplicons were visualised on 0.8% [w/v] agarose-TAE gels, and the expected fragments excised, and purified with the Zymoclean Gel DNA Recovery Kit (Zymo Research). The DNA was eluted in 20 µl PCR-grade water. The quality and concentration of the purified fragments were determined by the NanoDrop 2000 Spectrophotometer and the Qubit 2.0 Fluorometer (Thermo Scientific), respectively.

Libraries were prepared and sequenced according to the standard Illumina protocol for amplicon sequencing (https://support.illumina.com/downloads/16s_metagenomic_sequencing_library_preparation.html). This involved the addition of dual indices and sequencing adapters to both ends of the amplicon targets using the Nextera XT Index kit, and subsequent quantification, normalisation and pooling. The libraries were sequenced in a paired-end (2x300nt) run on the Illumina MiSeq platform at the Agricultural Research Council's Biotechnology Platform in Pretoria. The run included a 5% PhiX spike-in as an internal control for the low-diversity libraries.

4.2.3 Amplicon data analyses

The dataset generated for each sample was assessed for quality using FastQC v0.11.5 (http://www.bioinformatics.babraham.ac.uk/projects/fastqc/). The Nextera transposase

adapters, appended to the primer sequences, were removed and reads smaller than 50nt discarded, using the ILLUMINACLIP:NexteraPE-PE.fa:2:30:10 and MINLEN:50 parameters in Trimmomatic v0.36 (Bolger *et al*., 2014). The following bioinformatic data analysis pipeline (Figure 4.1) incorporates tools from UPARSE and the Quantitative Insights Into Microbial Ecology (QIIME) software (Caporaso *et al*., 2010; Edgar, 2013). Default parameters were used, unless otherwise stated.

4.2.3.1 Data analysis using UPARSE

The initial data processing steps were carried out following the UPARSE pipeline (http://drive5.com/useach/manual/uparse_pipeline.html), as implemented in USEARCH v8.1.1756 (Edgar, 2010; Edgar, 2013). Overlapping read pairs were merged, and subsequently filtered for quality, as recommended by Edgar and Flyvbjerg (2015). Reads were allowed a minimum overlap of 50nt and merged sequences smaller than 200nt were considered uninformative and discarded. Quality filtering was performed at a maximum expected error cut-off (E_{max}) of 1.0, so that the most probable number of errors in the sequences is zero. To enable accurate cross-sample comparative analysis, the filtered sequences from the eight samples were pooled and, to reduce the computational need, 'dereplicated' based on full-length matching, a process whereby identical sequences are collapsed and assigned size annotations. The resulting unique sequences were sorted in order of decreasing abundance and sequences occurring exactly once, i.e. singletons, were removed. Considering that singletons typically represent PCR and sequencing artefacts, this approach will effectively decrease the error rate (Edgar, 2013).

As the inclusion of the flanking conserved ribosomal RNA (rRNA) genes in sequence similarity-based analyses can lead to misleading results, the variable ITS2 region was extracted with the ITSx v1.0.11 software (Bengtsson-Palme *et al*., 2013). During this step, HMMER3 compares the input sequences to hidden Markov models computed from large alignments within the fungal kingdom (Eddy, 2011). The region is extracted based on the predicted positions of the rRNA genes in the sequence. The ITS2 sequences were clustered into operational taxonomic units (OTUs) with UPARSE-OTU at a canonical 97% identity threshold (Edgar, 2013). This algorithm performs *de novo* clustering and chimera filtering concurrently, and produces a set of OTU representative sequences (section 2.9.2). *De novo* clustering methods were shown to outperform reference-based methods (Westcott & Schloss, 2015). Additionally, the OTU sequences were subjected to chimera filtering against the UNITE UCHIME v7.0 ITS2 reference

dataset (https://unite.ut.ee/repository.php), using the UCHIME algorithm (Edgar *et al*., 2011; Nilsson *et al*., 2015). This dataset has an estimated 99.5% chimera detection rate. Sequences that passed the quality filtering step before dereplication were mapped to the chimera-free OTUs at a 97% identity threshold to generate an OTU map.

4.2.3.2 Data analysis using QIIME

Taxonomic and diversity analyses were performed in QIIME v1.9.1 (Caporaso *et al*., 2010). Standard scripts and parameters are available online (http://qiime.org/scripts/). The OTUs were taxonomically classified using the BLAST method, as implemented in QIIME (Altschul *et al*., 1990; Caporaso *et al*., 2010). The input sequences were searched against the dynamic set of the UNITE v7.0 database (https://unite.ut.ee/repository.php), pre-formatted for QIIME (Kõljalg *et al*., 2013). A maximum E-value of 1E-10 was used to record an OTU assignment. The OTU map generated with UPARSE, and the taxonomically annotated OTUs were used to construct an OTU table, a matrix wherein the sequence counts and taxonomic identities are indicated per OTU per sample. Low-confidence OTUs, resulting from carryover contamination between MiSeq runs, were filtered at an abundance threshold of 0.005%, as suggested by Bokulich *et al*. (2013) for Illumina datasets without a mock community control.

Certain diversity measures are sensitive to differences in sampling effort (Magurran, 2004). To normalise the data, the OTU table was rarefied to an even sample depth, based on the sequence count of the sample with the lowest number of sequences so that no sample would be omitted. As the variability of the ITS region presents difficulties for phylogenetic inference, non-phylogenetic diversity metrics were applied (Lindahl *et al*., 2013).

To measure the alpha diversity, multiple rarefaction analysis, i.e. iterative (10x) subsampling at different depth intervals, was performed and the Chao1 richness estimator, observed OTUs and Shannon diversity index computed for each rarefied OTU table (Chao, 1984; Shannon, 1948). The collated Chao1 results served as input to plot rarefaction curves describing the OTU richness for the old- and young-vine sample groups. Generating rarefaction curves was also necessary to determine whether the sequencing depth was sufficient to uncover the total fungal diversity (Kuczynski *et al*., 2011). To test for significant differences between the sample groups, non-parametric two-sample t-tests were performed for each alpha diversity metric, using the default number of Monte Carlo permutations to calculate the p-values. The OTUs were further

grouped into phylotypes according to their taxonomic identities, ranging from the phylum to species level, and the relative sequence abundance of the identified taxa evaluated per sample.

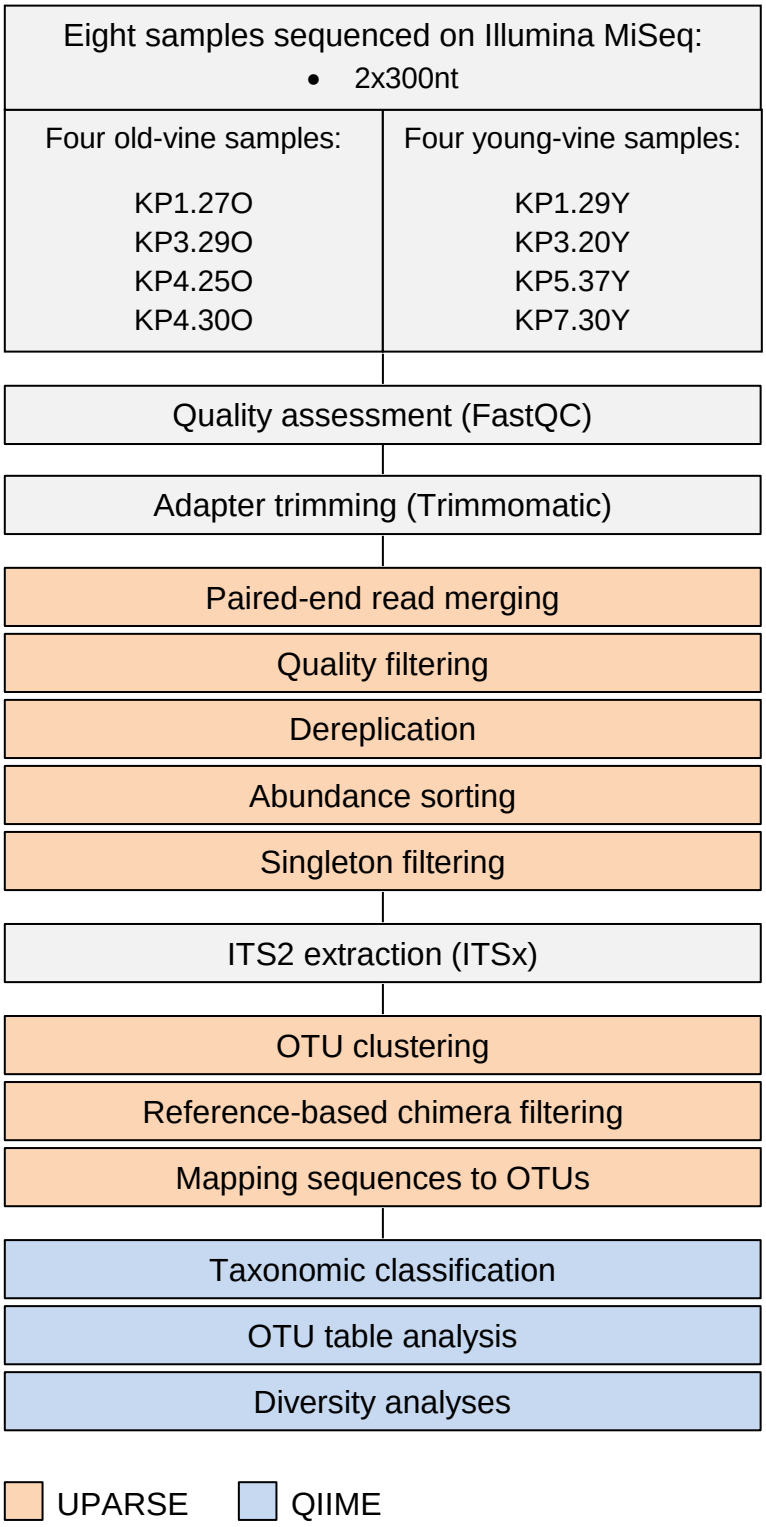

Figure 4.1: Bioinformatic workflow followed to analyse the NGS data generated by the paired-end Illumina sequencing run (2x300nt). Quality assessment, adapter trimming, merging and filtering were performed individually for each sample. The filtered sequences from each sample were pooled for downstream analyses.

4.3 Results and Discussion

4.3.1 DNA extraction

In order to optimise the protocol, DNA was extracted from three vascular tissue type combinations of cane material, after surface sterilisation and bark removal: 1) phloem, 2) xylem and pith and 3) phloem, xylem and pith. The DNA obtained from the phloem scrapings was of suboptimal purity, with a 260/230 absorbance ratio of 1.29, while the xylem-pith extracts had the lowest concentration, measuring at 47ng/µl. An increase in absorbance at 230nm may indicate contamination by salts used during the robust extraction method, or plant organic compounds. Polyphenols and polysaccharides are particularly abundant in tissues of woody plant species like grapevine, and are difficult to separate from nucleic acids (Lodhi *et al*., 1994). These co-purified contaminants may interfere with enzymatic manipulations of DNA.

The phloem-xylem-pith combination achieved an adequate balance between DNA purity and quantity, for subsequent DNA amplification, with a 260/230 absorbance ratio of 1.83 and a concentration of 225ng/µl. Total DNA was successfully extracted from the eight collected samples, using phloem-xylem-pith cuttings as starting material, with a mean DNA concentration and standard deviation of 142 ± 37ng/µl, and an average 260/280 and 260/230 absorbance ratio of 2.05 and 1.98, respectively.

4.3.2 Amplification of the internal transcribed spacer 2

While the conserved small and large subunit rRNA genes are suitable for sequence alignment and phylogenetic inference, the ITS region is regarded the universal barcode for fungal identification (Schoch *et al*., 2012). The ITS region is composed of the variable spacers, ITS1 and ITS2, separated by the highly conserved 5.8S rRNA gene. At present, the Illumina MiSeq offers 2x300nt read lengths, and is therefore incapable of sequencing the full-length fungal ITS region (~550bp), whilst retaining an overlap between the paired reads (Bálint *et al*., 2014). Furthermore, the presence of a highly conserved segment, in this case the intercalary 5.8S gene, facilitates chimera formation during PCR, and reduces the accuracy of species identification (Bengtsson-Palme *et al*., 2013; Fonesca *et al*., 2012; Nilsson *et al*., 2015). Considering these limitations, it is advised that amplification be restricted to either the ITS1 or ITS2 subregion, both of which offer high taxonomic resolution (Lindahl *et al*., 2013; Nilsson *et al*., 2008). The regions yield comparable results with regards to community composition (Blaalid *et al*.,

2013). However, the ITS2 is better represented in public sequence databases and less variable in length (Lindahl *et al*., 2013; Nilsson *et al*., 2009).

In this study, the ITS2 was amplified to characterise the endophytic fungal communities in grapevine, consistent with recent grapevine microbiome studies (Pinto *et al*., 2014; Pinto *et al*., 2015). Despite efforts to optimise the reaction and cycling conditions, multiple fragments were observed due to the variability of the ITS2 region and non-specific amplification. This was not unforeseen, considering that the template DNA was extracted from environmental samples. Fragments of the expected size (~400bp including overhang adapters) were excised, and purified using a spin column-based gel extraction kit. The recovered DNA had a mean concentration of 65ng/µl, with a standard deviation of 20ng/µl. The 260/280 absorbance ratios ranged between 1.81 and 1.91, indicative of high-quality DNA. The 260/230 ratios, though regarded adequate for sequencing (>1.5), were less favourable. Absorbance at 230nm may be attributed to traces of salts in the purified DNA.

4.3.3 Amplicon data analyses

The number of read pairs generated ranged between 360 436 and 822 991 across the eight samples. The variation in the amount of sequence data may be the result of technical errors in library preparation, but could also be ascribed to impurities derived from gel extraction that may have influenced accurate quantification of the purified DNA. In future studies, quantitative PCR can be used to prevent such discrepancies. Figures 4.2A and 4.2B illustrate the quality of the forward and reverse reads, respectively. As is typically observed in Illumina data, the quality of the reads declined towards the 3'end. The reverse reads were of lower quality, with the mean quality falling below a Phred score of Q20 after roughly 220 bases, compared to 290 bases in the forward reads. Nextera Transposase adapters, that had been appended to the primer sequences, were detected in the data; however, at a negligible level. Regardless, these non-biological sequences were removed to improve the accuracy of downstream analyses. Between 95.7 and 99.8% of read pairs were retained. Error-prone reads are abundant in high-throughput sequencing data and inflate diversity estimates (Huse *et al*., 2010; Kunin *et al*., 2010). To minimise spurious inferences of fungal diversity, extensive quality control measures were implemented, as discussed hereafter.

4.3.3.1 Data analysis using UPARSE

Paired reads were merged prior to filtering, so as to exploit the re-calculated quality scores in the overlapping region, which serve as an improved prediction of true error rates. This step was performed in USEARCH, as other read-merging programs were shown to have high rates of false positive alignments, and calculate incorrect 'posterior' quality scores according to Bayes' theorem (Edgar & Flyvbjerg, 2015). Approximately 98.4% of read pairs were merged, with a mean length of 345nt across the eight samples. A total of 82 sequences were discarded based on length. There is a visible increase in quality in the middle of the sequences where the reads overlap (Figure 4.2C), illustrating the value of read-merging as a pre-filtering step.

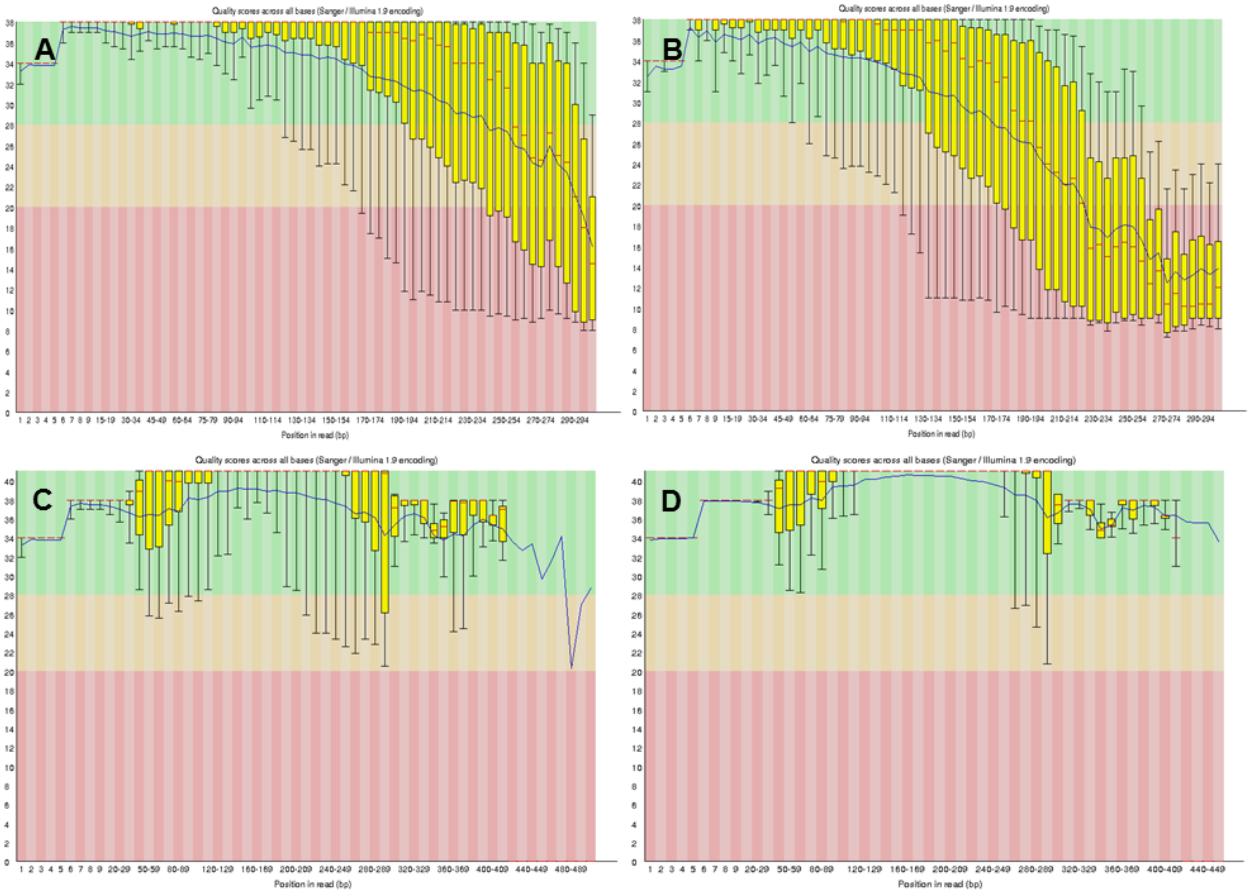

Figure 4.2: FastQC graphs illustrating the quality of the 2x300nt data. A) Forward reads, B) Reverse reads, C) Merged sequences and D) Filtered sequences. Note the difference in scaling on the x-axis. The quality scores are shown on the y-axis. The blue line represents the mean quality and the red central line indicates the median value. The yellow boxes represent the inter-quartile range from 25 to 75%, and the upper and lower whiskers mark the 10th and 90th percentiles, respectively.

After filtering for quality, between 72.5 and 84.1% of merged sequences were retained. The positive results may be attributed to the large overlapping region (~257nt) between the forward and reverse reads, decreasing the overall error rate of the sequences.

Figure 4.2D shows the filtered sequences to be of high quality, with the mean quality above a Phred score of Q34 across the length of the sequences. Table 4.1 summarises the output statistics for the initial quality control steps.

Table 4.1: The number of raw read pairs and the output statistics for adapter trimming, paired-end read-merging and quality filtering, per sample. The number and percentage of read pairs/sequences remaining after each step are indicated.

Sample	Read pairs	Adapter trimming	Merging			Filtering
		Read pairs (%)	Sequences (%)	Mean length (nt)		Sequences (%)
				Merged	Alignment	
KP1.27O	822 991	821 077 (99.77)	809 247 (98.56)	344.05	257.95	648 062 (80.08)
KP3.29O	605 323	596 757 (98.58)	585 438 (98.10)	344.07	257.93	452 976 (77.37)
KP4.25O	418 677	415 142 (99.16)	412 938 (99.47)	343.06	258.93	346 995 (84.03)
KP4.30O	360 436	345 194 (95.77)	337 655 (97.82)	343.88	258.11	245 002 (72.56)
KP1.29Y	749 507	747 607 (99.75)	733 354 (98.09)	348.87	253.12	570 834 (77.84)
KP3.20Y	574 723	570 700 (99.30)	562 322 (98.53)	344.70	257.29	442 030 (78.61)
KP5.37Y	600 627	593 080 (98.74)	582 899 (98.28)	343.46	258.53	454 289 (77.94)
KP7.30Y	468 081	461 992 (98.70)	455 867 (98.67)	345.20	256.80	372 393 (81.69)

A total of 3 532 581 pooled filtered sequences were reduced to 279 558 unique sequences, of which nearly 76% were singletons. High-throughput datasets typically contain a large number of singletons (Lindahl *et al*., 2013). As derivatives of randomly occurring PCR and sequencing errors are unlikely to be reproduced by chance, sequences present exactly once are mostly artefacts (Edgar, 2013; Tedersoo *et al*., 2010). Singletons increase with sequencing depth and are a primary source of spurious OTUs (Edgar, 2013). As a result, estimates of diversity by rarefaction analysis may be questioned. It is therefore imperative that such sequences be discarded before clustering (Edgar, 2013). Filtering singletons has implications for diversity metrics that incorporate low-abundance sequence counts; however, such metrics are unlikely to be accurate if the sequences are erroneous (Bokulich *et al*., 2013; Dickie, 2010). In this study, a pre-clustering filtering step was applied, and a total of 67 603 high-quality unique sequences were retained.

The ITSx software identified 67 537 sequences representing the fungal ITS2 region, of which 99.7% was successfully extracted. The remaining 225 sequences were regarded

potentially problematic for lacking a 5.8S gene, and excluded by default. While the 'partial' parameter may recover incomplete ITS2 sequences, the removal of such sequences serves as an additional quality control measure (Bengtsson-Palme *et al*., 2013). The length of the extracted ITS2 sequences ranged between 108 and 347nt, with a mean and standard deviation of 181 ± 43nt, and a median length of 178nt. A study by Nilsson *et al*. (2008) reported a median length of 173nt for fungal ITS2 sequences in the international nucleotide sequence databases, further supporting the results of this study.

Blaalid *et al*. (2013) evaluated the performance of ITS1 and ITS2 as fungal barcodes and determined that a 97% identity threshold is appropriate for estimating the number of species in a mock dataset of known taxonomic composition, for both regions. Clustering at 97% identity generated 674 OTUs. Of the 67 312 non-redundant ITS2 sequences, 8.1% were detected as chimeric by the *de novo* approach implemented in UPARSE-OTU. Reference-based filtering with UCHIME discarded 119 plausible chimeric OTUs. A number of chimeric sequences may have also been filtered during previous steps, either as erroneous sequences (E_{max} > 1.0) or singletons. The filtered sequences were then mapped to the non-chimeric OTUs. The bulk of the data was recovered during this step, with 97.5% of the sequences assigned to the 555 OTUs, indicating that most erroneous sequences had not deviated substantially from the original template.

The chimera content was moderate, despite the high concentration of template DNA amplified, and number of PCR cycles; both of these factors increase chimera formation rates (Haas *et al*., 2011; Lahr & Katz, 2009). It is plausible that a large portion of the chimeras reported are false positives. This was observed in a recent study that evaluated the performance of chimera filtering methods on mock ITS2 data, where the verified chimera content represented only 0.2% of the data (Bjørnsgaard-Aas *et al*., 2016). In this study, a more conservative detection approach was used to improve the integrity of the data; if undetected, these artefacts are misinterpreted as novel taxa (Edgar *et al*., 2011). Additional filtering against a reference dataset, as performed in this study, is also recommended by other fungal ITS metabarcoding pipelines (Bálint *et al*., 2014; Gweon *et al*., 2015; Větrovský & Baldrian, 2013).

Differentiating between PCR chimeras and true biological sequences remains a significant challenge in amplicon data analysis. Putative chimeric sequences may be manually scrutinised by conventional BLAST analysis (https://blast.ncbi.nlm.nih.gov/), as described by Nilsson *et al*. (2012). However, considering the general limitations of

bioinformatic detection methods, future studies should aim to reduce chimera formation rates experimentally.

4.3.3.2 Data analysis using QIIME

The OTUs were taxonomically classified with BLAST against the dynamic set of the UNITE database, consisting of sequence clusters, termed species hypotheses, delimited at thresholds that had been pre-defined by taxonomic experts (Kõljalg *et al*., 2013). Filtering at an abundance threshold of 0.005% removed 346 low-confidence OTUs. These clusters are probably the result of carryover contamination between sequencing runs. According to Illumina reports, up to 0.1% of reads typically contaminate subsequent MiSeq runs (Nelson *et al*., 2014). The remaining 209 OTUs represented a total of 3 428 170 high-quality sequences across the eight samples. The sequence count per sample ranged between 234 369 and 635 137, with a mean and standard deviation of 428 521±117 652. The number of OTUs observed for the old- and young-vine sample groups ranged from 93 to 116 and 115 to 133, respectively.

A total of 128 OTUs were shared among the sample groups, of which 48 were observed in all eight samples. Only 81 OTUs were specific to either group (Figure 4.3). Two OTUs were observed in all four young-, and none of the old-vine samples; however, in both cases the clusters accounted for only one or two sequences in at least two of the four samples. Although these OTUs were not discarded during any of the previous filtering steps, the low sequence counts decrease the confidence in true positive detection. In future studies, a more conservative abundance threshold should be applied. Due to the small sample size, no further statistical tests were performed to identify OTUs that are associated with either sample group, as the results may be considered uninformative.

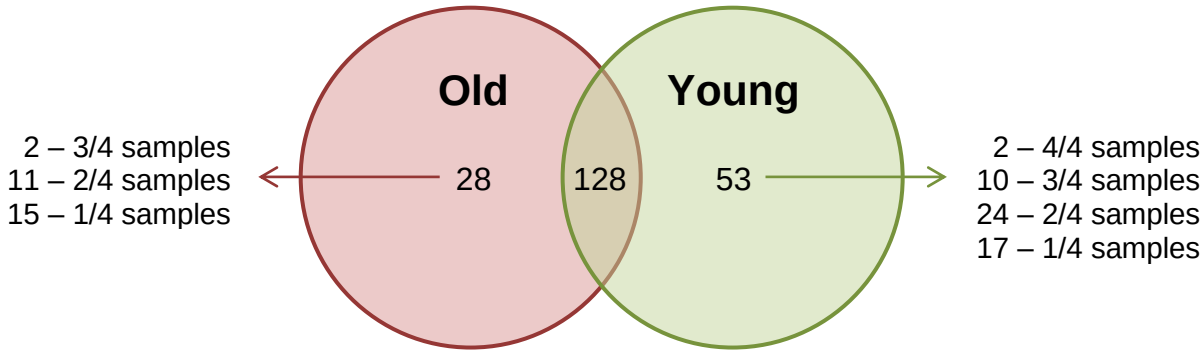

Figure 4.3: Venn diagram displaying the number of OTUs shared among, and specific to the old- and young-vine sample groups. The majority of the unique OTUs were observed in less than three out of the four grouped samples.

The fungal community in the old and young vines was further assessed by measuring the alpha (within-sample) diversity. First, the OTU table was rarefied to an even sample depth; 234 369 sequences per sample, to account for differences in sequencing depth. This was followed by multiple rarefaction analysis, a process whereby the OTU table is subjected to random subsampling at different sequencing depth intervals. To determine whether the sampling effort was adequate to characterise the complete mycobiome, rarefaction curves (Figure 4.4) were generated using the Chao1 richness metric as a measure of alpha diversity.

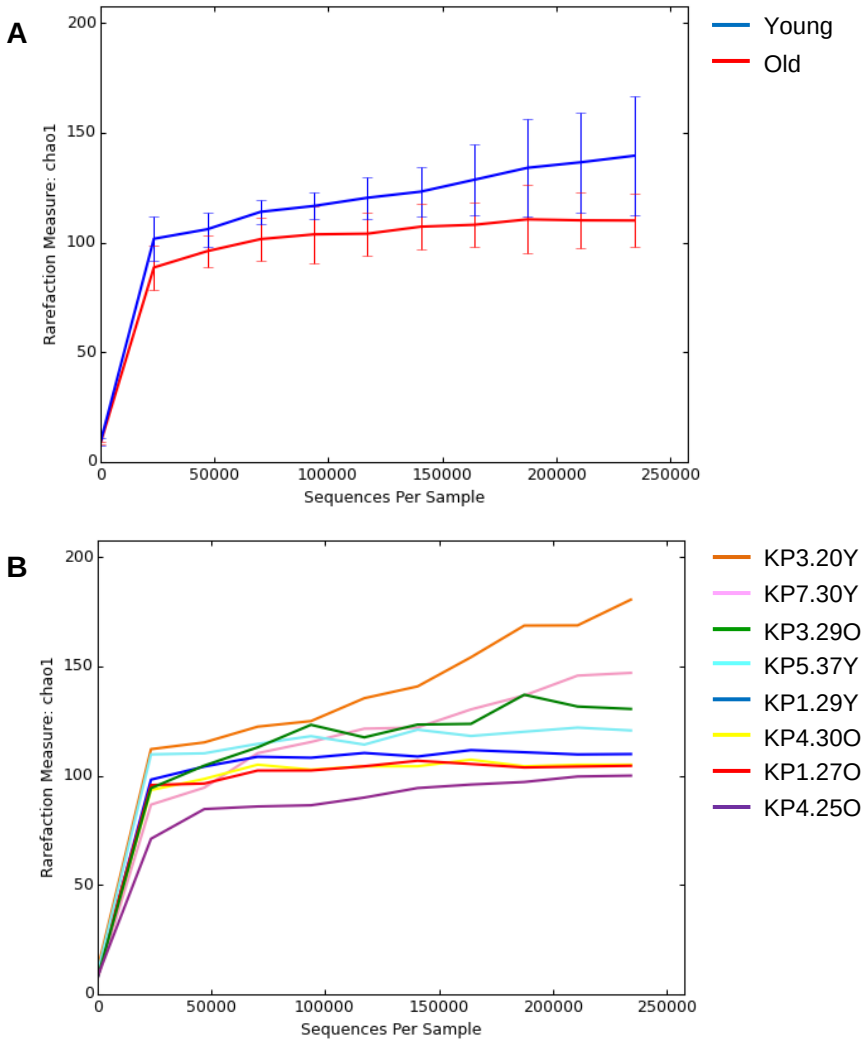

Figure 4.4: Rarefaction curves of the estimated OTU richness for the A) old- and young-vine sample groups and B) individual samples, as calculated by the Chao1 alpha diversity metric. The sequencing depth is shown on the x-axis. The sequencing depth intervals and step size was calculated by the script based on the specified maximum rarefaction depth.

The rarefaction graphs illustrate that 234 369 sequences per sample achieved adequate coverage of the complete fungal community in the old and young vines, with the exception of sample KP3.20Y. For this sample, the number of OTUs continued to

increase with further subsampling, indicating that there still exists a portion of the grapevine microbiota that has yet to be uncovered. The Chao1 rarefaction curves followed similar trends for the eight samples, but the estimated richness was greater for the young-vine sample group. Two additional alpha diversity metrics were calculated; the observed number of OTUs, as another measure of richness (based on OTU counts), and the Shannon diversity index. Diversity indices provide more information about the microbial community, accounting for both abundance and evenness of the taxa present.

All three measures revealed a higher fungal diversity in the young vines. However, comparison of the mean alpha diversity by non-parametric two-sample t-tests showed no significant differences between the sample groups, indicated by the p-values (>0.05) in Table 4.2. To date, little research has been done to characterise the microbiomes of old and young vines. Interestingly, a previous study that compared the bacterial diversity in three- and 15-year-old grapevine plants by culture-dependent techniques reported greater richness at the genus level for the younger vines (Andreolli *et al*., 2016). These results give more insight into the microbial diversity in vines of different ages, and are comparable with the findings of the present study, in terms of microbial richness. Bruez *et al*. (2016) characterised the endophytic fungal communities colonising the woody tissues of 42- and 58-year-old vines and reported higher species richness for the older vines, although not statistically significant.

Table 4.2: Comparison of the alpha diversity measures for the old- and young-vine sample groups by means of non-parametric two-sample t-tests.

Alpha diversity metric	Old		Young		p-value[a]
	Mean	Standard deviation	Mean	Standard deviation	
Chao1 richness	110.00	11.99	139.51	27.25	0.126
Observed OTUs	101.75	8.53	115.50	7.76	0.085
Shannon diversity index	3.37	0.13	3.46	0.38	0.758

[a]p-values were calculated through Monte Carlo permutations.

As no significant differences were observed between the alpha diversity of the sample groups, and considering the small sample size, reducing the statistical power, the focus of this study was to present a taxonomic description of the mycobiomes of old and young vines.

Of the 209 OTUs that passed the abundance filtering step, 195 were assigned to the fungal kingdom, and 14 could not be classified through BLAST analysis against the

UNITE database. Consistent with previous grapevine microbiome studies, Ascomycota was the most dominant phylum, accounting for 98.6% of the sequences (Morgan *et al*., 2017; Pinto *et al*., 2014; Setati *et al*., 2015; Taylor *et al*., 2014; Wang *et al*., 2015). Taxa of the Basidiomycota phylum comprised less than 0.6% of the data. With regards to the number of observations, the respective phyla represented approximately 79 and 12% of the OTUs. Only five OTUs could not be classified at the phylum level, while 79% of the OTUs were classified at the genus level. The greatest constraint of taxonomic inference is the limited reference data, increasing the rates of false positive errors, and restricting low-level classification of the OTUs.

The OTUs were grouped into phylotypes based on their taxonomic identities. Figure 4.5 illustrates the relative sequence abundance of the taxa present in the samples, ranging from the phylum to species level. Taxa that had a relative abundance greater than or equal to 1% in at least one of the eight samples are discussed in greater depth. The data revealed a complex grapevine mycobiome, with no significant differences observed between the old and young vines, in terms of community composition. Several fungal genera associated grapevines were detected, most notably *Acremonium*, *Alternaria*, *Aureobasidium*, *Cladosporium*, *Cytospora*, *Diplodia*, *Epicoccum*, *Neofusicoccum* and *Stemphylium*. The most abundant species detected across the eight samples were *Alternaria* spp., *Aureobasidium pullulans*, *Cladosporium exasperatum*, *Epicoccum nigrum* and *Stemphylium herbarum* (Figure 4.5). The basidiomycetous yeasts, *Cryptococcus*, *Meira*, *Rhodotorula* and *Sporobolomyces* were also observed; however, at low frequencies.

Filamentous fungi of the genera *Alternaria* and *Cladosporium* are pathogens of postharvest diseases, but were also previously identified as quiescent fungi in the berries, and endophytes in the leaves, shoots and wood (Casieri *et al*., 2009; Dugan *et al*., 2002; González & Tello, 2011; Pancher *et al*., 2012; Steel *et al*., 2013). Considering that this study characterised the fungal communities colonising the phloem and xylem tissues of the canes, these species are most likely endophytic. As discussed in section 2.5, the pleosporaceous species, *A*. *alternata* and *E*. *nigrum* are potential biocontrol agents of downy mildew (Kortekamp, 1997; Musetti *et al*., 2006).

E. *nigrum* and *A*. *pullulans* are common inhabitants of grapevine (González & Tello, 2011; Martini *et al*., 2009; Pancher *et al*., 2012). The species are particularly dominant in vineyards cultivated by organic practices, due to their tolerance to copper and sulphur, both of which are regularly applied as fungicides in organic viticulture (Grube *et*

al., 2011). *A*. *pullulans* has been isolated from endophytic and epiphytic communities, and is an inhibitor of postharvest fungal pathogens (Schena *et al*., 1999). In this study, the yeast-like fungus had a relative abundance ranging from 3.14 to 37.40% across the samples (Figure 2.6).

The family Botryosphaeriaceae was well represented in the old and young vines. *Diplodia pseudoseriata* was detected in seven samples, and was the dominant species in KP1.27O and KP4.30O, with a relative abundance of 30.30 and 22.72%, respectively (Figure 4.6). *Neofusicoccum australe* was present at low frequencies (<2%) in one old and three young vines. Botryosphaeriaceous species are mostly wood pathogens, but have also been described as endophytes (Slippers & Wingfield, 2007). More than 20 species in this family, including *N*. *australe*, are associated with grapevine trunk diseases of the Botryosphaeria dieback complex (Úrbez-Torres, 2011). The presence of *D*. *pseudoseriata* in grapevine has not been reported. This species was first isolated from the healthy tissues of various Myrtaceae trees native to Uruguay, but was also associated with a stem canker (Pérez *et al*., 2010).

Other species observed in at least four samples included *Cytospora* spp., *Lophiostoma winteri*, *Paraconiothyrium africanum* and four partially classified fungal taxa, namely Amphisphaeriaceae sp., Dothideomycetes sp., Ascomycota sp. and Fungi sp. (Figure 2.6). The genus *Cytospora* was somewhat better represented in the old vines, in terms of the number of species and observations. *Cytospora* species were previously isolated from wood cankers of declining grapevines, and as endophytes from asymptomatic vines (Fotouhifar *et al*., 2010; González & Tello, 2011; Lawrence *et al*., 2017). Similarly, *Lophiostoma* and *Paraconiothyrium* species were detected in the woody tissues of asymptomatic and esca-symptomatic vines (Bruez *et al*., 2016; Halleen *et al*., 2007; Hofstetter *et al*., 2012). The dothideomycetous species, observed in three old- and all four young-vines samples, was particularly abundant in KP4.25O, KP3.20Y and KP5.37Y, with a relative abundance of 10.41, 12.74 and 15.84%, respectively. The sequences classified as Dothideomycetes sp. were subjected to conventional BLAST analysis; however, the identity of the species could not be further elucidated.

As seen in Figure 2.5, the order Xylariales was better represented in the young vines. An unidentified Xylariales species accounted for more than 50% of the fungal community in KP1.29Y, while *Discostroma* sp. was the dominant species detected in KP5.37Y, with a relative abundance of 23.99%. Conventional BLAST analysis of the sequence classified as Xylariales sp. revealed 100% identity to strains of *Circinotrichum*

maculiforme (family Xylariaceae). *Circinotrichum* species are commonly associated with decomposing leaf litter (Shanthi & Vittal, 2010). Considering the abundance of the species in KP1.29Y, it is plausible that the collected sample material was contaminated with plant debris and not adequately surface sterilised. The presence of *Discostroma* in grapevine is not well documented. The genus *Pestalotiopsis* comprises pathogenic fungi that have been associated with grapevine trunk diseases (Jayawardena *et al*., 2015; Maharachchikumbura *et al*., 2016). Interestingly, species of *Discostroma* and *Pestalotiopsis* have been isolated from fynbos plants (Lee *et al*., 2006).

Other grapevine pathogens observed at low abundances (<1%), included *Botryotinia fuckeliana* (*Botrytis cinerea*) (KP4.25O, 0.37%), *Diaporthe ampelina* (*Phomopsis viticola*) (KP7.30Y, 0.67%), *Neofusicoccum parvum* (KP1.29Y, 0.12%), *Phaeomoniella chlamydospora* (KP5.37Y, 0.07%) and *Cryptovalsa ampelina* (KP4.30O, 0.22%). *B. fuckeliana*, or anamorph *B. cinerea*, although previously isolated from endophytic communities, is considered an important grapevine pathogen (Casieri *et al*., 2009; González & Tello, 2011; Pancher *et al*., 2012). This species typically colonises the grape endosphere as quiescent infections. At *véraison*, the fungus may resume growth and induce bunch rot (Dugan *et al*., 2002). The other four species are associated with grapevine trunk diseases (Bertsch *et al*., 2013).

The majority of the fungal species detected in the present study have been associated with either grapevine trunk diseases or postharvest rot. Whether any of the species are indeed pathogenic, quiescent or endophytic in the old and young vines was not determined. The fungal disease status of the vines at the time of sampling had not been established. In this study, samples were collected at a commercial vineyard that produces ‘old-vine’ wines. Grapevines tend to accumulate a number of diseases over the course of their lifespan. As the majority of the vines are older than 40 years, the presence of several grapevine trunk pathogens in the vineyard was not unexpected. Conversely, a number of endophytic species that act as antagonists of fungal diseases were observed at high abundances. Future studies investigating the balance between endophytic and pathogenic communities, and how these microbial interactions shape the mycobiome would further support the development of commercial vine probiotics. This fell outside of the scope of the present study.

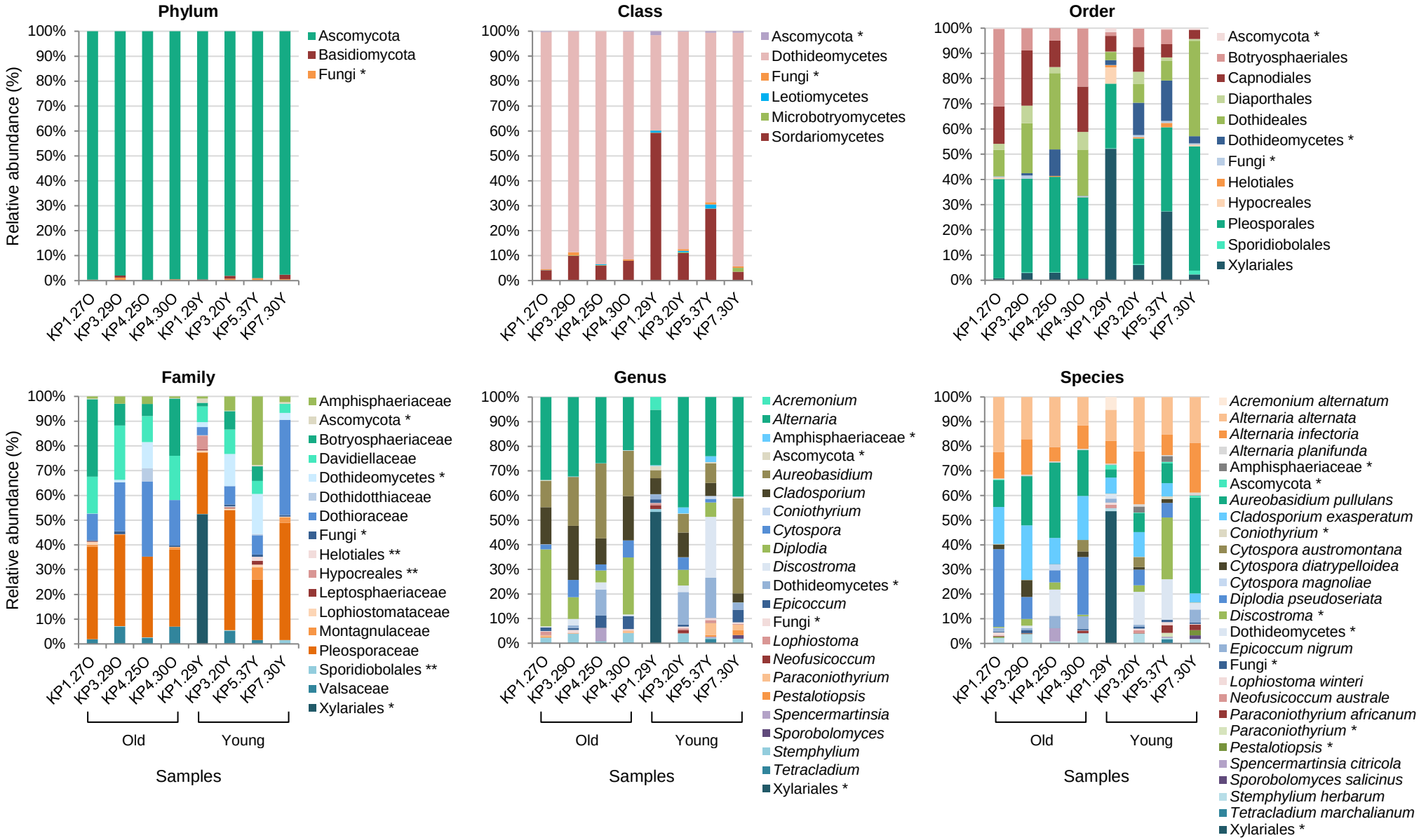

Figure 4.5: Relative sequence abundance of the fungal taxa detected in the samples, ranging from the phylum to species level. Taxa that had a relative abundance of ≥1% in at least one of the eight samples are indicated in the legend. *Unidentified within taxonomic group, ** *Incertae sedis*, uncertain placement within taxonomic group.

KP1.27O	KP3.29O	KP4.25O	KP4.30O	KP1.29Y	KP3.20Y	KP5.37Y	KP7.30Y	Species	Genus	Family	Order	Class	Phylum
30.30	8.72	4.77	22.72	0.00	6.11	5.63	0.01	*Diplodia pseudoseriata*	*Diplodia*	Botryosphaeriaceae	Botryosphaeriales	Dothideomycetes	Ascomycota
0.11	0.00	0.00	0.00	1.30	1.17	0.09	0.00	*Neofusicoccum australe*	*Neofusicoccum*				
14.50	21.55	10.41	17.33	6.13	9.55	5.17	3.51	*Cladosporium exasperatum*	*Cladosporium*	Davidiellaceae	Capnodiales		
10.41	19.52	29.93	18.08	3.14	7.32	7.68	37.40	*Aureobasidium pullulans*	*Aureobasidium*	Dothioraceae	Dothideales		
0.00	0.00	5.17	0.00	0.00	0.00	0.41	0.00	*Spencermartinsia citricola*	*Spencermartinsia*	Dothidotthiaceae	Pleosporales		
0.16	0.00	0.00	0.00	0.00	0.10	1.10	0.00	*Coniothyrium* sp.	*Coniothyrium*	Leptosphaeriaceae			
1.27	0.00	0.00	0.10	0.74	0.58	1.08	0.27	*Lophiostoma winteri*	*Lophiostoma*	Lophiostomataceae			
0.26	0.23	0.00	0.94	0.00	0.00	2.98	2.17	*Paraconiothyrium africanum*	*Paraconiothyrium*	Montagnulaceae			
0.05	0.00	0.00	0.04	0.00	0.00	1.52	0.00	*Paraconiothyrium* sp.					
21.60	16.76	19.91	11.24	11.78	20.96	14.70	17.83	*Alternaria alternata*	*Alternaria*	Pleosporaceae			
10.20	14.04	5.66	9.20	8.73	20.43	7.88	19.50	*Alternaria infectoria*					
0.39	0.48	0.33	0.38	0.33	0.90	0.35	1.02	*Alternaria planifunda*					
1.41	0.82	4.94	5.11	1.45	0.76	0.04	4.97	*Epicoccum nigrum*	*Epicoccum*				
2.18	3.76	0.79	4.03	1.06	3.60	0.48	1.76	*Stemphylium herbarum*	*Stemphylium*				
0.09	0.93	10.41	0.00	1.90	12.74	15.84	2.75	Dothideomycetes sp.	unidentified	unidentified	unidentified		
0.00	0.00	0.00	0.00	0.07	0.00	1.61	0.00	*Tetracladium marchalianum*	*Tetracladium*	*Incertae sedis*	Helotiales	Leotiomycetes	
0.02	0.08	0.00	4.52	0.00	3.85	0.00	0.00	*Cytospora austromontana*	*Cytospora*	Valsaceae	Diaporthales	Sordariomycetes	
1.87	6.73	0.01	2.21	0.00	0.96	1.45	0.00	*Cytospora diatrypelloidea*					
0.00	0.00	2.35	0.00	0.00	0.00	0.00	0.00	*Cytospora magnoliae*					
0.00	0.00	0.00	0.00	4.85	0.00	0.00	0.00	*Acremonium alternatum*	*Acremonium*	*Incertae sedis*	Hypocreales		
0.35	2.63	2.85	0.62	0.02	2.51	23.99	0.06	*Discostroma* sp.	*Discostroma*	Amphisphaeriaceae	Xylariales		
0.45	0.11	0.00	0.00	0.00	0.04	0.03	2.02	*Pestalotiopsis* sp.	*Pestalotiopsis*				
0.00	0.16	0.10	0.08	0.24	2.36	2.30	0.20	Amphisphaeriaceae sp.	unidentified				
0.00	0.00	0.00	0.00	50.37	0.28	0.06	0.01	Xylariales sp.	unidentified	unidentified			
0.32	0.00	0.00	0.14	1.58	0.16	0.57	0.64	Ascomycota sp.	unidentified	unidentified	unidentified	unidentified	
0.00	0.01	0.10	0.02	0.04	0.02	0.03	1.40	*Sporobolomyces salicinus*	*Sporobolomyces*	*Incertae sedis*	Sporidiobolales	Microbotryomycetes	Basidiomycota
0.31	1.20	0.00	0.53	0.08	0.75	0.85	0.61	Fungi sp.	unidentified	unidentified	unidentified	unidentified	unidentified

Figure 4.6: Heatmap of the fungal species that had a relative abundance of ≥1% in at least one of the eight samples. The taxonomic classifications, as determined by the UNITE database, are indicated. The variation in colour intensity represents the range in the percentage relative sequence abundance [scale: Min. (0.00) - Max. (50.37)].

4.4 Conclusion

There is anecdotal evidence from sensory panels suggesting that old vines produce wines of greater depth and complexity compared to young vines. The aim of the study was to determine whether there is a microbial contribution to this difference in wine quality. The mycobiomes of four old and four young Pinotage vines were characterised by deep amplicon sequencing of the fungal ITS2 region. The data revealed the presence of several taxa of filamentous and yeast-like fungi commonly associated with grapevine and other woody crops, including *Alternaria*, *Aureobasidium*, *Cladosporium*, *Epicoccum* and the Botryosphaeriaceae family. *Alternaria* spp. and *A*. *pullulans* were the most abundant species identified per sample. Rarefaction analysis indicated that the sequencing depth was adequate to capture the total fungal community for all samples, except KP3.20Y. The Chao1 richness and Shannon index revealed greater diversity in the young-vine sample group, although not statistically significant. It may be speculated that old vines have a more established, less diverse mycobiome, having selected for specific fungal species in response to environmental stresses. No differences were observed between the old and young vines, with regards to the community composition. Due to the small sample size, no further statistical tests were performed to identify taxa associated with the old vines; that may indirectly contribute to the unique old-vine character of the final product.

Several pathogens of grapevine trunk- and postharvest diseases, and endophytic species with biocontrol properties were detected across the eight samples. These results would suggest that the vines, although probably afflicted by fungal diseases, have remained viable for the production of high-quality wines; however, this was not established in the present study. To date, little research has been performed to characterise the fungal communities colonising the phloem and xylem tissues of the canes, as most grapevine microbiome studies have focused on the microbial consortia of the grapes and various stages of wine fermentation. The generated data has contributed to the study of the grapevine endosphere; the grape microbiome is not dissociated from the rest of the grapevine niche, and the fungal communities in the vascular tissues may be actively involved in shaping the microbial composition of the grapes, thereby indirectly influencing the character of the wine. Future research should focus on the movement of microbes through the vascular tissues, and to what extent the microbial interactions in the endosphere contribute to the favourable characteristics of

old-vine wines. The results of this study indicate that the mycobiome does not impact wine quality; however, future studies should use a larger sample size.

4.5 References

Altschul, S.F., Gish, W., Miller, W., Myers, E.W., Lipman, D.J., 1990. Basic Local Alignment Search Tool. J. Mol. Biol. 215(3), 403-410.

Andreolli, M., Lampis, S., Zapparoli, G., Angelini, E., Vallini, G., 2016. Diversity of bacterial endophytes in 3 and 15 year-old grapevines of *Vitis vinifera* cv. Corvina and their potential for plant growth promotion and phytopathogen control. Microbiol. Res. 183, 42-52.

Bálint, M., Schmidt, P., Sharma, R., Thines, M., Schmitt, I., 2014. An Illumina metabarcoding pipeline for fungi. Ecol. Evol. 4(13), 2642-2653.

Barata, A., Malfeito-Ferreira, M., Loureiro, V., 2012. The microbial ecology of wine grape berries. Int. J. Food Microbiol. 153(3), 243-259.

Bengtsson-Palme, J., Ryberg, M., Hartmann, M., Branco, S., Wang, Z., Godhe, A., De Wit, P., Sánchez-García, M., Ebersberger, I., De Sousa, F., Amend, A., Jumpponen, A., Unterseher, M., Kristiansson, E., Abarenkov, K., Bertrand, Y.J.K., Sanli, K., Eriksson, K.M., Vik, U., Veldre, V., Nilsson, R.H., 2013. Improved software detection and extraction of ITS1 and ITS2 from ribosomal ITS sequences of fungi and other eukaryotes for analysis of environmental sequencing data. Methods Ecol. Evol. 4(10), 914-919.

Bertsch, C., Ramírez-Suero, M., Magnin-Robert, M., Larignon, P., Chong, J., Abou-Mansour, E., Spagnolo, A., Clément, C., Fontaine, F., 2013. Grapevine trunk diseases: Complex and still poorly understood. Plant Pathol. 62(2), 243-265.

Bjørnsgaard-Aas, A., Davey, M.L., Kauserud, H., 2016. ITS all right mama: Investigating the formation of chimeric sequences in the ITS2 region by DNA metabarcoding analyses of fungal mock communities of different complexities. Mol. Ecol. Resour. 17(4), 730-741.

Blaalid, R., Kumar, S., Nilsson, R.H., Abarenkov, K., Kirk, P.M., Kauserud, H., 2013. ITS1 versus ITS2 as DNA metabarcodes for fungi. Mol. Ecol. Resour. 13(2), 218-224.

Bokulich, N.A., Subramanian, S., Faith, J.J., Gevers, D., Gordon, J.I., Knight, R., Mills, D.A., Caporaso, J.G., 2013. Quality-filtering vastly improves diversity estimates from Illumina amplicon sequencing. Nat. Methods. 10(1), 57-59.

Bokulich, N.A., Thorngate, J.H., Richardson, P.M., Mills, D.A., 2014. Microbial biogeography of wine grapes is conditioned by cultivar, vintage, and climate. Proc. Natl. Acad. Sci. U.S.A. 111(1), E139-E148.

Bokulich, N.A., Collins, T.S., Masarweh, C., Allen, G., Heymann, H., Ebeler, S.E., Mills, D.A., 2016. Associations among wine grape microbiome, metabolome, and fermentation behavior suggest microbial contribution to regional wine characteristics. MBio. 7(3), e00631-16. DOI: 10.1128/mBio.00631-16.

Bolger, A.M., Lohse, M., Usadel, B., 2014. Trimmomatic: A flexible trimmer for Illumina sequence data. Bioinformatics. 30(15), 2114-2120.

Bruez, E., Baumgartner, K., Bastien, S., Travadon, R., Guérin-Dubrana, L., Rey, P., 2016. Various fungal communities colonise the functional wood tissues of old grapevines externally free from grapevine trunk disease symptoms. Aust. J. Grape Wine Res. 22(2), 288-295.

Campisano, A., Antonielli, L., Pancher, M., Yousaf, S., Pindo, M., Pertot, I., 2014. Bacterial endophytic communities in the grapevine depend on pest management. PLoS ONE. 9(11), e112763. DOI: 10.1371/journal.pone.0112763.

Caporaso, J.G., Kuczynski, J., Stombaugh, J., Bittinger, K., Bushman, F.D., Costello, E.K., Fierer, N., Peña, A.G., Goodrich, J.K., Gordon, J.I., Huttley, G.A., Kelley, S.T., Knights, D., Koenig, J.E., Ley, R.E., Lozupone, C.A., McDonald, D., Muegge, B.D., Pirrung, M., Reeder, J., Sevinsky, J.R.,

Turnbaugh, P.J., Walter, W.A., Widmann, J., Yatsunenko, T., Zaneveld, J., Knight, R., 2010. QIIME allows analysis of high-throughput community sequencing data. Nat. Methods. 7(5), 335-336.

Casieri, L., Hofstetter, V., Viret, O., Gindro, K., 2009. Fungal communities living in the wood of different cultivars of young *Vitis vinifera* plants. Phytopathol. Mediterr. 48(1), 73-83.

Chao, A., 1984. Nonparametric estimation of the number of classes in a population. Scand. J. Statist. 11(4), 265-270.

Dickie, I.A., 2010. Insidious effects of sequencing errors on perceived diversity in molecular surveys. New Phytol. 188(4), 916-918.

Dugan, F.M., Lupien, S.L., Grove, G.G., 2002. Incidence, aggressiveness and *in planta* interactions of *Botrytis cinerea* and other filamentous fungi quiescent in grape berries and dormant buds in Central Washington State. J. Phytopathol. 150(7), 375-381.

Eddy, S.R., 2011. Accelerated profile HMM searchers. PLoS Comput. Biol. 7(10), e1002195. DOI: 10.1371/journal.pcbi.1002195.

Edgar, R.C., 2010. Search and clustering orders of magnitude faster than BLAST. Bioinformatics. 26(19), 2460-2461.

Edgar, R.C., Haas, B.J., Clemente, J.C., Quince, C., Knight, R., 2011. UCHIME improves sensitivity and speed of chimera detection. Bioinformatics. 27(16), 2194-2200.

Edgar, R.C., 2013. UPARSE: Highly accurate OTU sequences from microbial amplicon reads. Nat. Methods. 10(10), 996-998.

Edgar, R.C., Flyvbjerg, H., 2015. Error filtering, pair assembly and error correction for next-generation sequencing reads. Bioinformatics. 31(21), 3476-3482.

Fonesca, V.G., Nichols, B., Lallias, D., Quince, C., Carvalho, G.R., Power, D.M., Creer, S., 2012. Sample richness and genetic diversity as drivers of chimera formation in nSSU metagenetic analyses. Nucleic Acids Res. 40(9), e66. DOI: 10.1093/nar/gks002.

Fotouhifar, K.B., Hedjaroude, G.A., Leuchtmann, A., 2010. ITS rDNA phylogeny or Iranian strains of *Cytospora* and associated teleomorphs. Mycologia. 102(6), 1369-1382.

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

Grube, M., Schmid, F., Berg, G., 2011. Black fungi and associated bacterial communities in the phyllosphere of grapevine. Fungal Biol. 115(10), 978-986.

Gweon, H.S., Oliver, A., Taylor, J., Booth, T., Gibbs, M., Read, D.S., Griffiths, R.I., Schonrogge, K., 2015. PIPITS: An automated pipeline for analyses of fungal internal transcribed spacer sequences from the Illumina sequencing platform. Methods Ecol. Evol. 6(8), 973-980.

Haas, B.J., Gevers, D., Earl, A.M., Feldgarden, M., Ward, D.V., Giannoukos, G., Ciulla, D., Tabbaa, D., Highlander, S.K., Sodergren, E., Methé, B., DeSantis, T.Z., The Human Microbiome Consortium, Petrosino, J.F., Knight, R., Birren, B.W., 2011. Chimeric 16S rRNA sequence formation and detection in Sanger and 454-pyrosequenced PCR amplicons. Genome Res. 21(3), 494-504.

Halleen, F., Mostert, L., Crous, P.W., 2007. Pathogenicity testing of lesser-known vascular fungi of grapevine. Australas. Plant Pathol. 36(3), 277-285.

Hofstetter, V., Buyck, B., Croll, D., Viret, O., Coloux, A., Gindro, K., 2012. What if esca disease of grapevine were not a fungal disease? Fungal Divers. 54(1), 51-67.

Huse, S.M., Welch, D.M., Morrison, H.G., Sogin, M.L., 2010. Ironing out the wrinkles in the rare biosphere through improved OTU clustering. Environ. Microbiol. 12(7), 1889-1898.

Jayawardena, R.S., Zhang, W., Liu, M., Maharachchikumbura, S.S.N., Zhou, Y., Huang, J., Nilthong, S., Wang, Z., Li, X., Yan, J., Hyde, K.D., 2015. Identification and characterization of *Pestalotiopsis*-like fungi related to grapevine diseases in China. Fungal Biol. 119(5), 348-361.

Kõljalg, U., Nilsson, R.H., Abarenkov, K., Tedersoo, L., Taylor, A.F.S., Bahram, M., Bates, S.T., Bruns, T.D., Bengtsson-Palme, J., Callaghan, T.M., Douglas, B., Drenkhan, T., Eberhardt, U., Dueñas, M., Grebenc, T., Griffith, G.W., Hartmann, M., Kirk, P.M., Kohout, P., Larsson, E., Lindahl, B.D., Lücking, R., Martín, M.P., Matheny, P.B., Nguyen, N.H., Niskanen, T., Oja, J., Peay, K.G., Peintner, U., Peterson, M., Põldmaa, K., Saag, L., Saar, I., Schüßler, A., Scott, J.A., Senés, C., Smith, M.E., Suija, A., Taylor, D.L., Telleria, M.T., Weiss, M., Larsson, K., 2013. Towards a unified paradigm for sequence-based identification of fungi. Mol. Ecol. 22(1), 5271-5277.

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.

Kuczynski, J., Stombaugh, J., Walters, W.A., González, A., Caporaso, J.G., Knight, R., 2011. Using QIIME to analyze 16S rRNA gene sequences from microbial communities. Curr. Protoc. Bioinformatics. 36, 10.7.1-10.7.2.

Kunin, V., Engelbrekston, A., Ochman, H., Hugenholtz, P., 2010. Wrinkles in the rare biosphere: Pyrosequencing errors can lead to artificial inflation of diversity estimates. Environ. Microbial. 12(1), 118-123.

Lahr, D.J., Katz, L.A., 2009. Reducing the impact of PCR-mediated recombination in molecular evolution and environmental studies using a new-generation high-fidelity DNA polymerase. BioTechniques. 47(4), 857-866.

Lawrence, D.P., Travadon, R., Pouzoulet, J., Rolshausen, P.E., Wilcox, W.F., Baumgartner, K., 2017. Characterization of *Cytospora* isolated from wood cankers of declining grapevines in North America, with the descriptions of two new *Cytospora* species. Plant Pathol. 66(5), 713-725.

Lee, S., Crous, P.W., Wingfield, M.J., 2006. Pestalotioid fungi from Restonaceae in the Cape Floral Kingdom. Stud. Mycol. 55, 175-187.

Lindahl, B.D., Nilsson, R.H., Tedersoo, L., Abarenkov, K., Carlsen, T., Kjøller, R., Kõljalg, U., Pennanen, T., Rosendahl, S., Stenlid, J., Kauserud, H., 2013. Fungal community analysis by high-throughput sequencing of amplified makers - A user's guide. New Phytol. 199(1), 288-299.

Lodhi, M.A., Ye, G., Weeden, N.F., Reisch, B.I., 1994. A simple and efficient method for DNA extraction from grapevine cultivars and *Vitis* species. Plant Mol. Biol. Rep. 12(1), 6-13.

Magurran, A.E., 2004. Measuring biological diversity. Blackwell Publishing. Oxford, England, UK.

Maharachchikumbura, S.S.N., Larignon, P., Hyde, K.D., Al-Sadi, A.M., Liu, Z., 2016. Characterization of *Neopestalotiopsis*, *Pestalotiopsis* and *Truncatella* species associated with grapevine trunk diseases in France. Phytopathol. Mediterr. 55(3), 380-390.

Martini, M., Musetti, R., Grisan, S., Polizzotto, R., Borselli, S., Pavan, F., Osler, R., 2009. DNA-dependent detection of the grapevine fungal endophytes *Aureobasidium pullulans* and *Epicoccum nigrum*. Plant Dis. 93(10), 993-998.

Marzano, M., Fosso, B., Manzari, C., Grieco, F., Intranuovo, M., Cozzi, G., Mulè, G., Scioscia, G., Valiente, G., Tullo, A., Sbisà, E., Pesole, G., Santamaria, M., 2016. Complexity and dynamics of the winemaking bacterial communities in berries, musts, and wines from Apulian grape cultivars through time and space. PLoS ONE. 11(6), e0157383. DOI: 10.1371/journal.pone.0157383.

Morgan, H.H., Du Toit, M., Setati, M.E., 2017. The grapevine and wine microbiome: Insights from high-throughput amplicon sequencing. Front. Microbiol. 8, 820. DOI: 10.3389/fmicb.2017.00820.

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

Nelson, M.C., Morrison, H.G., Benjamino, J., Grim, S.L., Graf, J., Analysis, optimization and verification of Illumina-generated 16S rRNA gene amplicon surveys. PLoS ONE. 9(4), e94249. DOI: 10.1371/journal.pone.0094249.

Nilsson, R.H., Kristiansson, E., Ryberg, M., Hallenberg, N., Larsson, K., 2008. Intraspecific ITS variability in the kingdom fungi as expressed in the international sequence databases and its implications for molecular species identification. Evol. Bioinform. Online. 4, 193-201.

Nilsson, R.H., Ryberg, M., Abarenkov, K., Sjökvist, E., Kristiansson, E., 2009. The ITS region as a target for characterization of fungal communities using emerging sequencing technologies. FEMS Microbiol. Lett. 296(1), 97-101.

Nilsson, R.H., Tedersoo, L., Abarenkov, K., Ryberg, M., Kristiansson, E., Hartmann, M., Schoch, C.L., Nylander, J.A.A., Bergsten, J., Porter, T.M., Jumpponen, A., Vaishampayan, P., Ovaskainen, O., Hallenberg, N., Bengtsson-Palme, J., Eriksson, K.M., Larsson, K., Larsson, E., Kõljalg, U., 2012. Five simple guidelines for establishing basic authenticity and reliability of newly generated fungal ITS sequences. MycoKeys. 4, 37-63.

Nilsson, R.H., Tedersoo, L., Ryberg, M., Kristiansson, E., Hartmann, M., Unterseher, M., Porter, T.M., Bengtsson-Palme, J., Walker, D.M., De Sousa, F., Gamper, H.A., Larsson, E., Larsson, K., Kõljalg, U., Edgar, R.C., Abarenkov, K., 2015. A comprehensive, automatically updated fungal ITS sequence dataset for reference-based chimera control in environmental sequencing efforts. Microbes Environ. 30(2), 145-150.

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.

Perazzolli, M., Antonielli, L., Storari, M., Puopolo, G., Pancher, M., Giovannini, O., Pindo, M., Pertot, I., 2014. Resilience of the natural phyllosphere microbiota of the grapevine to chemical and biological pesticides. Appl. Environ. Microbiol. 80(12), 3585-3596.

Pérez, C.A., Wingfield, M.J., Slippers, B., Altier, N.A., Blanchette, R.A., 2010. Endophytic and canker-associated Botryosphaeriaceae occurring on non-native *Eucalyptus* and native Myrtaceae trees in Uruguay. Fungal Divers. 41(1), 53-69.

Pinto, C., Pinho, D., Sousa, S., Pinheiro, M., Egas, C., Gomes, A.C., 2014. Unravelling the diversity of grapevine microbiome. PLoS ONE. 9(1), e85622. DOI: 10.1371/journal.pone.0085622.

Pinto, C., Pinho, D., Cardoso, R., Custódio, V., Fernandes, J., Sousa, S., Pinheiro, M., Egas, C., Gomes, A.C., 2015. Wine fermentation microbiome: A landscape from different Portuguese wine appellations. Front. Microbiol. 6, 905. DOI: 10.3389/fmicb.2015.00905.

Schena, L., Ippolito, A., Zahavi, T., Cohen, L., Nigro, F., Droby, S., 1999. Genetic diversity and biocontrol activity of *Aureobasidium pullulans* isolates against postharvest rots. Postharvest Biol. Technol. 17(3), 189-199.

Schoch, C.L., Seifert, K.A., Huhndorf, S., Robert, V., Spouge, J.L., Levesque, C.A., Chen, W., Fungal Barcoding Consortium, 2012. Nuclear ribosomal internal transcribed spacer (ITS) region as a universal DNA barcode marker for fungi. Proc. Natl. Acad. Sci. U.S.A. 109(16), 6241-6246.

Setati, M.E., Jacobson, D., Bauer, F.F., 2015. Sequence-based analysis of the *Vitis vinifera* L. cv. Cabernet Sauvignon grape must mycobiome in three South African vineyards employing distinct agronomic systems. Front. Microbiol. 6, 1358. DOI: 10.3389/fmicb.2015.01358.

Shannon, C.E., 1948. A mathematical theory of communication. Bell Syst. Tech. J. 27(3), 379-423, 623-656.

Shanthi, S., Vittal, B.P.R., 2010. Fungi associated with decomposing leaf litter of cashew (*Anacardium occidentale*). Mycology. 1(2), 121-129.

Slippers, B., Wingfield, M.J., 2007. Botryosphaeriaceae as endophytes and latent pathogens of woody plants: Diversity, ecology and impact. Fungal Biol. Rev. 21(2-3), 90-106.

Steel, C.C., Blackman, J.W., Schmidtke, L.M., 2013. Grapevine bunch rots: Impact on wine composition, quality, and potential procedures for the removal of wine faults. J. Agric. Food Chem. 61(22), 5188-5206.

Taylor, M.W., Tsai, P., Anfang, N., Ross, H.A., Goddard, M.R., 2014. Pyrosequencing reveals regional differences in fruit-associated fungal communities. Environ. Microbiol. 16(9), 2848-2858.

Úrbez-Torres, J.R., 2011. The status of Botryosphaeriaceae species infecting grapevines. Phytopathol. Mediterr. 50(Suppl.), S5-S45.

Vega-Avila, A.D., Gumiere, T., Andrade, P.A.M., Lima-Perim, J.E., Durrer, A., Baigori, M., Vazquez, F., Andreote, F.D., 2015. Bacterial communities in the rhizosphere of *Vitis vinifera* L. cultivated under distinct agricultural practices in Argentina. Antonie Van Leeuwenhoek. 107(2), 575-588.

Větrovský, T., Baldrian, P., 2013. Analysis of soil fungal communities by amplicon pyrosequencing: Current approaches to data analysis and the introduction of the pipeline SEED. Biol. Fertil. Soils. 49(8), 1027-1037.

Wang, C., García-Fernández, D., Mas, A., Esteve-Zarzoso, B., 2015. Fungal diversity in grape must and wine fermentation assessed by massive sequencing, quantitative PCR and DGGE. Front. Microbiol. 6, 1156. DOI: 10.3389/fmicb.2015.01156.

Westcott, S.L., Schloss, P.D., 2015. *De novo* clustering methods outperform reference-based methods for assigning 16S rRNA gene sequence to operational taxonomic units. Peer J. 3, e1487. DOI: 10.7717/peerj.1487.

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, 315-322. Academic Press, San Diego, CA, USA.

Zarraonaindia, I., Owens, S.M., Weisenhorn, P., West, K., Hampton-Marcell, J., Lax, S., Bokulich, N.A., Mills, D.A., Martin, G., Taghavi, S., Van der Lelie, D., Gilbert, J.A., 2015. The soil microbiome influences grapevine-associated microbiota. MBio. 6(2), e02527-14. DOI: 10.1128/mBio.02527-14.

Internet sources

https://support.illumina.com/downloads/16s_metagenomic_sequencing_library_preparation.html

http://www.bioinformatics.babraham.ac.uk/projects/fastqc/

http://drive5.com/useach/manual/uparse_pipeline.html

https://unite.ut.ee/repository.php

http://qiime.org/scripts/

https://blast.ncbi.nlm.nih.gov/

Chapter 5: Conclusion

A number of vineyards in South Africa have remained economically productive beyond their life expectancy. These vineyards continue to produce superior wines despite the prolonged exposure to environmental stresses. These Recent years have seen growing interest in the propagation of old vines. Reports from sensory panels indicate that these old vines produce wines of greater depth and complexity compared to young vines. To date, limited research has been performed to establish which factors may be responsible for this difference in wine quality. Potential contributing factors include variations at the genome or transcriptome level, or differences in the viral and microbial component of the vines. The focus of the present study was to characterise the viral and microbial profiles of old and young vines in the Stellenbosch region, using next-generation sequencing (NGS) in a metagenomics approach.

Viruses of the families *Closteroviridae*, *Betaflexiviridae* and *Tymoviridae*, and viroids of the family *Pospiviroidae* were detected. As expected, the old vines had a more diverse virus community, having been exposed to viral pathogens for a longer period. However, it is surprising that the young vines have accumulated multiple viruses in a relative short time, considering that the vines were subjected to heat therapy prior to propagation and planting. Grapevine leafroll-associated virus 3 (GLRaV-3) was detected in every sample with the highest abundance, and may be ascribed to the high incidence of grapevine leafroll disease in this vineyard. Grapevine Syrah virus 1 (GSyV-1), and possibly grapevine rupestris vein feathering virus (GRVFV), was identified for the first time in South African grapevines.

Several genera of filamentous and yeast-like fungi commonly associated with grapevine and other woody crops were identified, including *Alternaria*, *Aureobasidium*, *Cladosporium* and *Epicoccum*. The most dominant species detected across the eight samples were *Alternaria* spp. and *Aureobasidium pullulans*. The basidiomycetous yeasts, *Cryptococcus*, *Rhodotorula* and *Sporobolomyces* were also observed; however, at low abundances. A number of pathogens associated with grapevine trunk diseases and postharvest rots, and endophytes that act as antagonists of fungal pathogens were present in the samples. The young-vine samples showed a tendency of greater fungal diversity, although not statistically significant. It can be hypothesised that the environment has played an important role is shaping the fungal community of the old vines, resulting in a more established, less diverse mycobiome than that of young vines.

No differences in community composition were observed between the old and young vines. Future research should focus on identifying microbes that are unique to old or young vines, and microbes that are shared among vines, irrespective of age.

This study had various limitations, most notably the quality and amount of data generated for the virome part of the study, and the small sample size, which had implications for diversity and statistical analyses. An initial objective of this study was to determine the bacterial diversity in the old and young vines by sequencing two variable regions of the 16S rRNA gene. However, due to host chloroplast and mitochondrion contamination, as a result of primer cross-reactivity, this objective could not be accomplished and is therefore reserved for future research.

The virome data indicated the presence of a possible divergent variant of GLRaV-3 that requires further investigation. The detection of GRVFV for the first time in South African grapevines also needs to be validated experimentally. Considering the low sequence identity of numerous GRVFV-like contigs with the available GRVFV genome sequences, it is plausible that a divergent variant of GRVFV or possibly a novel virus within *Tymoviridae* had been identified. It would therefore be of importance to characterise the complete genome of the GRVFV-like species, and to study the diversity of the virus in greater depth by phylogenetic comparison with other tymoviruses. Research on the incidence and biological impact of the identified tymoviruses would contribute to the study of grapevine disease complexes.

Future investigations that could build on this study include the characterisation of endophytic fungal species by means of culture-dependent techniques. To improve the significance of the data, a larger sample size should be used, and samples of other grapevine tissues and cultivars included in the study. Furthermore, a biogeographical study of old South African vines may contribute to research on wine *terroir*. This would require the collection of samples from vineyards geographically distributed across various wine-producing regions of the Western Cape. Future studies should focus on identifying core microbes that are associated with old vines; that may contribute to the unique character of old-vine wines. This could ultimately lead to the development of commercial vine probiotics, designed to improve the wine quality for younger vines.

Supplementary data

Table S1: Grapevine virus and viroid species included in the read-mapping analysis. The corresponding length (bp), GenBank accession number and taxonomic family of each reference sequence are indicated. This list was adapted from the directory of virus and virus-like diseases of the grapevine and their agents (Martelli, 2014).

Species (virus or viroid)	Length (bp)	Accession[c]	Family
Alfalfa mosaic virus	3 644 2 593 2 037	NC_001495 NC_002024 NC_002025	*Bromoviridae*
Arabis mosaic virus	7 334 3 820	NC_006057 NC_006056	*Secoviridae*
Artichoke Italian latent virus[a]	1 828	X87254	*Secoviridae*
Australian grapevine viroid	369	NC_003553	*Pospiviroidae*
Beet cryptic virus 3[a]	1 607	S63913	*Partitiviridae*
Blackberry virus S[a]	6 463	FJ915122	*Tymoviridae*
Blueberry leaf mottle virus[a]	1 908 3 082	U20622 U20621	*Secoviridae*
Broad bean wilt virus 1	5 817 3 446	NC_005289 NC_005290	*Secoviridae*
Broad bean wilt virus 2	5 951 3 607	NC_003003 NC_003004	*Secoviridae*
Carnation mottle virus	4 003	NC_001265	*Tombusviridae*
Cherry leafroll virus	7 918 6 360	NC_015414 NC_015415	*Secoviridae*
Cucumber mosaic virus	3 357 3 050 2 216	NC_002034 NC_002035 NC_001440	*Bromoviridae*
Grapevine Algerian latent virus	4 731	NC_011535	*Tombusviridae*
Grapevine Anatolian ringspot virus	7 288 4 607	NC_018383 NC_018384	*Secoviridae*
Grapevine angular mosaic virus[a]	381	AY590305	*Bromoviridae*
Grapevine-associated chrysovirus[a]	2 879 945 1 371 451	GU108588 GU108589 GU108596 GU108597	*Chrysoviridae*
Grapevine-associated mycovirus[a]	1 225 1 016 495	GU108587 GU108600 GU108601	Unclassified
Grapevine-associated narnavirus 1[a]	622	GU108590	*Narnaviridae*
Grapevine asteroid mosaic-associated virus[a]	1 852	AJ249357	*Tymoviridae*
Grapevine berry inner necrosis virus	7 241	NC_015220	*Betaflexiviridae*
Grapevine Bulgarian latent virus	7 452 5 821	NC_015492 NC_015493	*Secoviridae*
Grapevine chrome mosaic virus	7 212 4 441	NC_003622 NC_003621	*Secoviridae*

Species (virus or viroid)	Length (bp)	Accession[c]	Family
Grapevine deformation virus	7 386 3 753	NC_017939 NC_017938	*Secoviridae*
Grapevine endophyte endornavirus	12 154	NC_019493	*Endornaviridae*
Grapevine fanleaf virus[b]	7 341 3 816	KC900162 KC900163	*Secoviridae*
Grapevine fleck virus	7 564	NC_003347	*Tymoviridae*
Grapevine leafroll-associated virus 1	18 659	NC_016509	*Closteroviridae*
Grapevine leafroll-associated virus 2	16 494	NC_007448	*Closteroviridae*
Grapevine leafroll-associated virus 3[b]	18 498	EU259806	*Closteroviridae*
Grapevine leafroll-associated virus 4	13 830	NC_016416	*Closteroviridae*
Grapevine leafroll-associated virus 7	16 404	NC_016436	*Closteroviridae*
Grapevine Pinot gris virus	7 275	NC_015782	*Betaflexiviridae*
Grapevine red blotch-associated virus	3 206	NC_022002	*Geminiviridae*
Grapevine Red Globe virus[a]	2 006	AF521977	*Tymoviridae*
Grapevine rupestris stem pitting-associated virus	8 744	NC_001948	*Betaflexiviridae*
Grapevine rupestris vein feathering virus[a]	6 617	AY706994	*Tymoviridae*
Grapevine Syrah virus 1	6 506	NC_012484	*Tymoviridae*
Grapevine vein clearing virus	7 753	NC_015784	*Caulimoviridae*
Grapevine virus A[b]	7 360	DQ787959	*Betaflexiviridae*
Grapevine virus B[b]	7 599	EF583906	*Betaflexiviridae*
Grapevine virus D[a]	963	Y07764	*Betaflexiviridae*
Grapevine virus E[b]	7 568	GU903012	*Betaflexiviridae*
Grapevine virus F[b]	7 547	KP114220	*Betaflexiviridae*
Grapevine yellow speckle viroid 1	366	NC_001920	*Pospiviroidae*
Grapevine yellow speckle viroid 2	363	NC_003612	*Pospiviroidae*
Hop stunt viroid	302	NC_001351	*Pospiviroidae*
Peach rosette mosaic virus[a]	7 977	AF016626	*Secoviridae*
Petunia asteroid mosaic virus[a]	1 238	AY500881	*Tombusviridae*
Potato virus X	6 435	NC_011620	*Alphaflexiviridae*
Raphanus sativus cryptic virus 3	1 609 1 581	NC_011705 NC_011706	*Partitiviridae*
Raspberry ringspot virus	7 935 3 914	NC_005266 NC_005267	*Secoviridae*
Southern tomato virus	3 437	NC_011591	*Amalgaviridae*
Strawberry latent ringspot virus	7 496 3 842	NC_006964 NC_006965	*Secoviridae*
Tobacco mosaic virus	6 395	NC_001367	*Virgaviridae*
Tobacco necrosis virus D	3 762	NC_003487	*Tombusviridae*
Tobacco ringspot virus	7 514 3 929	NC_005097 NC_005096	*Secoviridae*

Species (virus or viroid)	Length (bp)	Accession[c]	Family
Tomato black ring virus	7 358	NC_004439	*Secoviridae*
	4 633	NC_004440	
Tomato mosaic virus	6 383	NC_002692	*Virgaviridae*
Tomato ringspot virus	8 214	NC_003840	*Secoviridae*
	7 271	NC_003839	
Tomato spotted wilt virus	8 897	NC_002052	*Bunyaviridae*
	4 821	NC_002050	
	2 916	NC_002051	

[a]No reference sequence (RefSeq) or complete genome sequence was available on GenBank at the time of this study; the largest sequence was included. [b]The genome sequence of a South African isolate was used. [c]The sequences of all available genomic segments were included for multipartite viruses.

Protocol S1: Cetyltrimethylammonium bromide DNA extraction method

CTAB buffer: 2% [w/v] CTAB, 100mM Tris-HCl pH 8, 1.4M NaCl, 20mM EDTA pH 8

Centrifugation steps are performed at 4°C

1. Pulverise 100mg plant material in liquid nitrogen and transfer to 2ml Eppendorf tube
2. Add 25mg (2.5% [w/v]) PVP-10
3. Add 1ml pre-heated CTAB buffer containing 30µl [3% [v/v]) β-mercaptoethanol
4. Incubate in waterbath for 15 minutes at 60°C
5. Centrifuge at 13200rpm for 10 minutes
6. Pipette aqueous phase to new 2ml Eppendorf tube
7. Add 0.2µl RNase A (stock 10mg/µl)
8. Incubate in waterbath for 15 minutes at 37°C
9. Add equal volume cold chloroform-isoamyl alcohol (24:1)
10. Centrifuge at 13200rpm for 10 minutes
11. Pipette aqueous phase to new 2ml Eppendorf tube
12. Repeat x2 steps 9-11
13. Add 0.8 volume cold isopropanol
14. Centrifuge at 13200rpm for 5 minutes
15. Carefully decant or pipette supernatant
16. Add 500µl cold 70% [v/v] ethanol
17. Centrifuge at 13200rpm for 5 minutes
18. Carefully decant or pipette ethanol
19. Spin down to pipette remaining ethanol
20. Air dry pellet for 10 minutes at room temperature
21. Resuspend in 40µl sterile water treated by reverse osmosis
    (volume dependent on pellet size)