Identification and characterisation of Grapevine leafroll-associated virus 3 genomic and subgenomic RNAs

by

Hans Jacob Maree

Dissertation presented for the degree of Doctor
of Philosophy in Genetics, Stellenbosch University

Promoter: Prof. JT Burger
Co-promoter: Dr. M-J Freeborough
Faculty of Science
Department of Genetics
December 2010

Declaration

By submitting this dissertation electronically, I declare that the entirety of the work contained therein is my own, original work, and that I have not previously in its entirety or in part submitted it for obtaining any qualification.

December 2010

Abstract

Grapevine leafroll-associated virus 3 (GLRaV-3) is the type strain for the genus *Ampelovirus*, family *Closteroviridae*. There has been only one report that claimed the complete nucleotide sequence of GLRaV-3 (isolate NY-1, AF037268). Here we report the complete sequence of the South African GLRaV-3, isolate GP18 (EU259806) and show a significantly extended 5' end. We used RLM-RACE to determine the 5' end of GP18 and found the 5' UTR to be 737 nt compared to 158 nt in the NY-1 sequence. This extended UTR was found in all other South African isolates of GLRaV-3 that were tested. In two collaborative studies the existence of the extended 5' UTR was confirmed and further investigated. In the first study (Coetzee *et al.*, 2010), metagenomic data generated by next generation sequencing (Illumina Genome Analyzer II) was analysed for GLRaV-3 specific sequences. Sequences similar to the GP18 isolate confirmed the sequence of the extended 5' UTR. In the second study (Jooste *et al.*, 2010), three genetic variants were identified and their respective 5' UTRs studied. Great diversity was observed between the 5' UTRs of the different genetic variants, however within a variant the 5' UTR was found to be highly conserved. Grapevine leafroll-associated virus 3 is a positive sense, single stranded RNA virus that has been shown, like other closteroviruses, to produce subgenomic (sg) RNAs during replication. These sgRNAs are deployed for the expression of the ORFs on the 3' half of the genome. In this study a dsRNA blot confirmed the presence of three, 3' co-terminal sgRNAs species [sgRNA(ORF3/4), sgRNA(ORF5) and sgRNA(ORF6)] in GLRaV-3-infected plant material when using a probe directed at the coat protein gene. The specific 5' terminal nucleotides for these sgRNAs as well as four additional sgRNAs [sgRNA(ORF7), sgRNA(ORF8), sgRNA(ORF9) and sgRNA(ORF10-12)] were determined by RLM-RACE for GLRaV-3 isolate GP18. The construction of a GLRaV-3 mini-replicon, analogous to RNA1 of Lettuce infectious yellows virus, for the evaluation of putative sg-promoters is also described.

Opsomming

Grapevine leafroll-associated virus 3 (GLRaV-3) is 'n lid van die *Closteroviridae* familie en die hooflid vir die genus *Ampelovirus*. Tot dusver was daar net een studie wat die volledige nukleïensuurvolgorde van GLRaV-3 gerapporteer het (isolaat NY-1, AF037268). In hierdie studie rapporteer ons die volledige volgorde van 'n Suid-Afrikaanse GLRaV-3, isolaat nl. GP18 (EU259806) wat noemenswaardig langer is aan die 5' kant. RLM-RACE is gebruik om die 5' eindpunt van GP18 te bepaal en daar is gevind dat die 5' ongetransleerde streek (UTR) 737 nt lank is in vergelyking met die 158 nt van die NY-1 volgorde. Die verlengde 5' UTR is gevind in alle Suid-Afrikaanse monsters wat getoets is. Die verlengde 5' UTR is bevestig en verder bestudeer tydens twee samewerkingsprojekte. In die eerste studie (Coetzee *et al.*, 2010), is metagenomiese data gegenereer deur volgende-generasie volgordebepaling (Illumina Genome Analyzer II) en geanaliseer vir GLRaV-3 spesifieke volgordes. Volgordes soortgelyk aan die GP18 isolaat het die verlengde 5' UTR volgorde bevestig. In die tweede studie (Jooste *et al.*, 2010), is drie genetiese variante van GLRaV-3 geidentifiseer en hulle onderskeie 5' UTR volgordes bepaal en bestudeer. Daar is groot diversiteit tussen die 5' UTRs van die verskillende genetiese variante gevind, maar tussen isolate van dieselfde variant is die volgordes gekonserveerd. Grapevine leafroll-associated virus 3 is 'n positiewe-sin, enkelstring RNA virus wat al voorheen bewys is om, soos ander closterovirusse, subgenomiese (sg) RNAs te produseer tydens replisering. Hierdie sgRNAs word ingespan vir die uitdrukking van die ORFs op die 3' helfte van die virusgenoom. In hierdie studie is 'n dsRNA klad gebruik om die voorkoms van 3' ko-terminale sgRNAs [sgRNA(ORF3/4), sgRNA(ORF5) and sgRNA(ORF6)] te bevestig in GLRaV-3 geinfekteerde plantmateriaal deur gebruik te maak van 'n peiler teen die kapsiedproteïengeen. Die spesifieke 5' terminale nukleotides vir hierdie sgRNAs sowel as vier additionele sgRNAs [sgRNA(ORF7), sgRNA(ORF8), sgRNA(ORF9) and sgRNA(ORF10-12)] is bepaal deur gebruik te maak van RLM-RACE op die GLRaV-3 isolaat GP18. Die konstruksie van 'n GLRaV-3 mini-repliserings konstruk, analoog aan die RNA1 van Lettuce infectious yellows virus, vir die evaluasie van moontlike sg-promotors word ook beskryf.

List of abbreviations

+ss	positive-sense single stranded
A	Adenine
aa	amino acid
AMV	Avian Myeloblastosis Virus
BLAST	Basic Local Alignment Search Tool
BMV	Brome mosaic virus
BMYV	Beet mild yellowing virus
bp	base pair
BYDV	Barley yellow dwarf virus
BYV	Beet yellows virus
C	Cytosine
CaMV	Cauliflower mosaic virus
cDNA	copy DNA
kb	kilobase
CP	coat protein
CTAB	cetyltrimethylammonium bromide
CTV	Citrus tristeza virus
DNA	deoxyribonucleic acid
DRNA	defective RNA
dsRNA	double stranded RNA
EDTA	ethylene-diamine-tetra-acetic acid
et al.	et alibi
g	gram
G	Guanine
GAMaV	Grapevine asteroid mosaic-associated virus
GFkV	Grapevine fleck virus
GFLV	Grapevine fanleaf virus
GLRaV-2	Grapevine leafroll-associated virus 2
GLRaV-3	Grapevine leafroll-associated virus 3
GRGV	Grapevine redglobe virus
gRNA	genomic RNA
GRP	Gross regional product
GRSPaV	Grapevine rupestris stempitting associated virus
GRVFV	Grapevine rupestris vein feathering virus
GUS	beta-glucuronidase
GUSi	beta-glucuronidase with intron
GVA	Grapevine virus A
GVB	Grapevine virus B
HEL	helicase

Hsp70h	heatshock protein 70 homologue
ICTV	International Committee on Taxonomy of Viruses
ICVG	The International Council for the Study of Viruses and Virus-like Diseases of Grapevine
II	Internal initiation
kcal	kilo calorie
kV	kilovolt
LChV-2	Little cherry virus 2
LIYV	Lettuce infectious yellows virus
L-Pro	leader protease
LRD	Grapevine Leafroll Disease
M	molar
MAQ	Mapping and Assembly with Quality assembler
MET	methyltransferase
ng	nanogram
NGO	Non-governmental organization
nt	nucleotide
ORF	open reading frame
PCR	polymerase chain reaction
PT	premature termination
R&D	Research and development
RdRp	RNA-dependent RNA polymerase
RLM-RACE	RNA ligase mediated rapid amplification of cDNA ends
RNA	Ribonucleic acid
RT-PCR	Reverse transcription PCR
SAWIS	South African Wine Industry Statistics
sgRNA	subgenomic RNA
sp	species
spp	species plural
SSCP	single stranded conformation polymorphism
T	Thymine
Ta	annealing temperature
ug	micro gram
UTR	untranslated region
VIGS	virus induced gene silencing
Winetech	Wine Industry Network of Expertise and Technology
X-Gluc	5-bromo-4-chloro-3-indolyl-β-D-glucuronic acid, cyclohexylammonium salt
μF	microfarad
Ω	ohm

Acknowledgements

I wish to extend my gratitude towards the following people for their various contributions and assistance during this study:

Prof. Johan Burger, for guidance and the opportunity to study in the Vitislab.

Dr. Michael-John Freeborough, for guidance and endless pep talks.

Ms. Elize Jooste and Ms. Beatrix Coetzee for inviting me to be part of their projects.

Dr. Jean-Marc Chelton, Prof. D. Jasper G. Rees, Dr. Dariusz E. Goszczynski, Prof Gerhard Pietersen, Mrs. Helena F.J. Gardner and Prof. Dirk U. Bellstedt for being collaborators and co-authors.

Prof. Gerhard Pietersen, Prof. Pere Mastre, Prof. Edgar Maiss, Mr. Kassie Kasdorf and Mr. Tobie Oosthuizen for providing various materials used in this study.

Mr. Willem Botes (Dept Genetics, SU), Mr. Ferdi van Zyl (SAPO) and Mr. Paul Stemmet (SAPO) for providing glasshouse space to store infected vines.

The owners of the various wine farms for allowing us to collect samples (Vergelegen, Kanonkop, KWV).

Prof. Dirk Bellstedt and Mr Chris Visser for assistance in phylogenetic analyses.

Ms N. Mgocheki for the identification and photographs of parasitoids and mealybugs.

And all the members of the Vitis lab for a comfortable working environment.

Finally, to my wife and family, for moral support and encouragement, I thank you all.

Contents

Chapter 1: Introduction

1.1 General introduction

The international wine industry is a major contributor to the global economy. According to the 33rd report of South African Wine Industry Statistics (SAWIS) published in 2009, the South African wine industry has been stable for the last three years (2006-2008). South Africa has approximately 124 000 hectares under vines, producing more than a billion litres of wine every year, contributing R3.5 bn (5.6%) to the State Revenue. In 2006 South Africa was ranked the 7th largest wine producing country, contributing 3.6% to the global production. In the Western Cape Province, the industry employs more than 250 000 people and greatly contributes to the provinces' GRP (www.SAWIS.co.za).

This global industry is threatened by many pests and pathogens such as insects, fungi, bacteria, nematodes, phytoplasmas and viruses. All these pathogens, except for viruses and phytoplasmas can be controlled by agrochemicals. The most destructive grapevine viruses are those involved in grapevine leafroll disease (LRD), Rugose wood disease and Fanleaf degeneration (Martelli and Boudon-Padieu, 2006). Vines infected with viruses cannot be cured and viticulturists have to resort to actively managing these diseases in their vineyards. Control methods consist of planting uninfected vines, clean pruning techniques and controlling the virus vectors with pesticides. All these methods help only in containing the virus spread with limited success.

To help prevent the devastating affects of grapevine viruses an intimate knowledge of the virus infection cycle, mode of spread (vector), geographical distribution and level of infection is required. The International Council for the Study of Viruses and Virus-like Diseases of Grapevine (ICVG) has long recognised Grapevine leafroll-associated virus 3 (GLRaV-3), the main causative agent in leafroll disease, as one of the most economically important viruses, and since 2003 it has been regarded as the foremost virus problem facing the grapevine industry. The NGO, Wine Industry Network of Expertise and Technology (Winetech) that co-ordinates and facilitates R&D for the South African wine industry has recognised leafroll disease as the most threatening of the grapevine virus diseases in their Vision 20/20 initiative.

Research on GLRaV-3 lags behind that of other economically important grapevine viruses like Grapevine fanleaf virus (GFLV) and has largely focused on epidemiology and the development of detection techniques. However, in 2004 Ling *et al.* published the genome sequence of GLRaV-3 isolate NY-1. Using this information, researchers were able to investigate the genome as a whole in an attempt to understand the role of this virus in LRD at the molecular level.

1.2 Aims and objectives

This study attempted to elucidate some of the fundamental questions pertaining to the virus replication strategy and mechanics of replication. The results obtained give us a better understanding of GLRaV-3 replication that will hopefully assist in the creation of new approaches to combat LRD. The main aim of this project was to address the lack of knowledge that existed for the genomic and sub-genomic (sg)RNA of GLRaV-3 and how these are utilised in virus replication. To achieve this goal the following objectives were set out:

° To identify grapevine plants that were singly infected with GLRaV-3, to be used as starting material.

° To determine the complete genome sequence of a South African isolate of GLRaV-3, to be able to compare to previous genome sequences and map 5' ends accurately.

° To determine the 5' ends of the sgRNAs associated with GLRaV-3 replication using RNA ligase mediated rapid amplification of cDNA ends (RLM-RACE) and map on the genomic sequence generated.

° To construct a GLRaV-3 mini-replicon for the evaluation of putative sg-promoters involved in the production of sgRNAs in a transient expression system in *Nicotiana benthamiana*.

1.3 Breakdown of thesis chapters

The thesis is divided into six chapters; a general introduction and literature overview followed by three research chapters and a general conclusion. Each chapter is introduced and referenced separately.

Chapter 1: Introduction

General introduction, aims and objectives of the study with a breakdown of the thesis chapters. The scientific outputs generated during the study and the contribution by Mr. Maree is stated.

Chapter 2: Literature overview

An overview of the literature relating to leafroll disease, GLRaV-3, positive-sense, single stranded RNA virus replication (focussing on the role of sgRNA) and closterovirus infectious clones are given.

Chapter 3: Sequencing and analysis of the complete genome of a South African Grapevine leafroll-associated virus 3 isolate, GP18.

In this chapter the sequencing of the South African isolate GP18 is described. The discovery that the 5' end extended 579 nt further than previously reported was further investigated. Relevant results from two collaborative studies are also included in this chapter, with additional analysis not included in the original publications.

Chapter 4: Mapping of the 5' terminal nucleotides of Grapevine leafroll-associated virus 3 sgRNAs.

In this chapter the use of RLM-RACE to map the 5' terminal nucleotides of seven positive-sense sgRNAs, for the expression of ORFs 3-12 of GLRaV-3 is described.

Chapter 5: Construction of a Grapevine leafroll-associated virus 3 mini-replicon.

In this chapter the construction of a GLRaV-3 mini-replicon is described. This is the first report of the construction of a mini-replicon for a member of the genus *Ampelovirus*. An attempt to utilize the GLRaV-3 mini-replicon to evaluate the activity of the putative sg-promoter of sgRNA(ORF6) using a GUS gene expression assay is also reported.

Chapter 6: Conclusion

General concluding remarks and future prospects.

1.4 Research output and author contributions:

The following papers, conference proceedings and conference posters were generated during the study.

1.4.1 Publications

1. **Maree H.J.**, Freeborough, M-J., Burger, J.T., 2008. Complete nucleotide sequence of a South African isolate of grapevine leafroll-associated virus 3 reveals a 5' UTR of 737 nucleotides. Archives of Virology 153:755-757.

   This paper forms the basis of Chapter 3 and is in its entirety the work of Mr Maree.

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

   This paper was partially included in Chapter 3 as supporting data for the sequence of GP18 and GP18-like viruses, and confirmation of the extended 5' UTR. The data from this study was also used for further analyses, not included in the publication, but included in Chapter 3. Mr Maree was involved in the experimental design and formed an integral part of the analysis team interpreting the data generated.

3. Jooste, A.E.C., **Maree, H.J.**, Bellstedt, D.U., Goszczynski, D.E., Pietersen, G., Burger, J.T., 2010. Genetic variation of Grapevine leafroll-associated virus 3 (GLRaV-3) in leafroll infected vineyards of South Africa. DOI 10.1007/s00705-010-0793-y.

   This paper was partially included in Chapter 3 as supporting data for the existence of the extended 5' UTR and variation observed in this area between molecular variants. Mr Maree was involved in the experimental design and supplied technical support in the sequencing of the 3rd molecular variant, PL20. He also assisted in the determination of the 5' ends of all three molecular variants described as well as the analysis of sequence data.

4. **Maree, H.J.**, Gardner, H.F.J., Freeborough, M-J., Burger, J.T., 2010. Mapping of the 5' terminal nucleotides of Grapevine leafroll-associated virus 3 sgRNAs. Virus Research 151, 252-255.

   This paper forms the basis of Chapter 4 and is completely the work of Mr Maree. The dsRNA blot image used in the publication was generated by H. Gardner.

1.4.2 Conference proceedings

1. **Maree, H.J.**, Jooste, A.E.C., Stephan, D., Freeborough, M-J., Burger, J.T. Characterisation of the genomic and subgenomic RNA of Grapevine leafroll-associated virus 3 (GLRaV-3). 16th meeting of the International Council for the Study of Virus and Virus-like Diseases of the Grapevine (ICVG), 31 August - 4 September 2009, Dijon, France. p222. ISSN 0369-8173.

   This proceeding includes work described in Chapters 3, 4 and 5 and is the in its entirety the work of Mr Maree.

2. Jooste, A.E.C., **Maree, H.J.**, Pietersen, G., Goszczynski, D.E., Burger, J.T. Identification and distribution of three divergent molecular variants of Grapevine leafroll-associated virus 3 (GLRaV-3) in South African vineyards. 16th meeting of the International Council for the Study of Virus and Virus-like Diseases of the Grapevine (ICVG), 31 August - 4 September 2009, Dijon, France. p273. ISSN 0369-8173.

   Mr Maree was involved in the experimental design and supplied technical support in the sequencing of the 3rd molecular variant, PL20. He also assisted in the determination of the 5' ends of all three molecular variants described, as well as the analysis of sequence data.

3. Coetzee, B., Freeborough, M-J., **Maree, H.J.**, Celton, J-M., Rees, D.J.G., Burger, J.T. Virome of a vineyard: ultra deep sequence analysis of diseased grapevines. 16th meeting of the International Council for the Study of Virus and Virus-like Diseases of the Grapevine (ICVG), 31 August - 4 September 2009, Dijon, France. p216. ISSN 0369-8173.

   Mr Maree was involved in the design of the experiments and the analysis and interpretation of the data generated.

4. **Maree, H.J.**, Freeborough, M-J., Burger, J.T., Characterisation of the Grapevine leafroll-associated virus 3 sgRNAs. 46th Congress of the South African Society for Plant Pathology and the 6th Congress of the African Mycological Association, 25-28 January 2009, Villa Via Hotel, Gordon's Bay, South Africa. P.70. ISBN: 13 978-1-86849-376-0.

   This proceeding includes work described in Chapters 3 and 4 and is the in its entirety the work of Mr Maree.

1.4.3 Posters

1. **Maree H.J.**, Freeborough, M-J., Burger, J.T. Characterisation of the Grapevine leafroll-associated virus 3 sgRNAs. Cape Biotechnology Forum, Somerset West, South Africa, 30 November – 2 December 2008. Poster: PP06

   This conference proceeding includes work described in Chapters 3 and 4 and is the in its entirety the work of Mr Maree.

2. **Maree, H.J.**, Freeborough, M-J., Burger, J.T., Characterisation of the Grapevine leafroll-associated virus 3 replication mechanism. Agricultural Biotechnology International Conference (ABIC), Cork, Ireland, 24-27 August 2008. Poster: 2.07 p.20

   This conference proceeding includes work described in Chapters 3 and 4 and is the in its entirety the work of Mr Maree.

1.5 References

Ling, K.S., Zhu, H.Y., Gonsalves, D., 2004. Complete nucleotide sequence and genome organization of grapevine leafroll-associated virus 3, type member of the genus ampelovirus. J. Gen. Virol. 85, 2099-2102.

Martelli, G.P., Boudon-Padieu, E., 2006. Directory of infectious diseases of grapevines and viroses and virus-like diseases of grapevine: Bibliographic report 1998-2004. Opinions Mediterraneennes Serie B: Studies and Research.

Chapter 2: Literature overview

2.1 Introduction

The grapevine has been an important supplementary source of food and drink for millennia. Grapes are mainly used for the production of wine but also consumed fresh, dried or pressed into juice. Archaeological evidence of early viticulture and viniculture can be traced back as far as the Chalcolithic and mid-Bronze Age (Figueiral *et al.*, 2010). Vinification residues found in clay jars from 7000 BCE and archeobiological remains of pressed grapes from the 5th millennium BCE are evidence of early winemaking (McGovern, 2003; Valamoti *et al.*, 2007). Unfortunately, grapevine is also the crop plant most susceptible to intracellular pathogens, of which many cause disorders that reduce plant vigour and longevity as well as yield and quality of the harvest. Infectious intracellular agents like viruses, viroids, and phloem- or xylem-limited prokaryotes are some of the most important pathogens affecting grapevine. The International Council for the Study of Virus and Virus-like Diseases of the Grapevine (ICVG) was established in 1964 to serve as a platform to discuss research methodologies and results related to grapevine viral diseases. Since then the ICVG has had 15 meetings the latest (16th) in Dijon, France in 2009. The ICVG now recognises more than 70 infectious agents of grapevine which include more than 60 viruses from more than 8 families (Martelli and Boudon-Padieu, 2006). The diseases caused by these viruses can be divided into five main categories: Infectious degeneration (GFLV, European and Mediterranean nepoviruses, American nepoviruses), Leafroll (closteroviruses), Rugose wood complex (vitiviruses and foveaviruses), Graft incompatibility (Grapevine leafroll-associated virus 2 (GLRaV-2) and Grapevine virus B (GVB)) and Fleck complex (Grapevine fleck virus (GFkV), Grapevine redglobe virus (GRGV), Grapevine asteroid mosaic-associated virus (GAMaV) and Grapevine rupestris vein feathering virus (GRVFV)). Some of these diseases are caused by viruses that individually cause no symptoms in the plant but in combination with other viruses, i.e. in a virus complex, cause specific diseases (Martelli and Boudon-Padieu, 2006). Worldwide and also in South Africa the most important of these viral diseases is Grapevine Leafroll Disease (LRD) (Pietersen, 2004). Recently, metagenomic and next generation sequencing have demonstrated that the complexity of grapevine viral diseases might extend even further than originally though as new viruses as well as viruses

previously not known to infect grapevine are being identified (Al Rwahnih *et al.*, 2009; Coetzee *et al.*, 2010; Prosser *et al.* 2007).

This literature review is divided into four major parts to give an overview of the information available and to give credit to the research that formed the basis from were this research was undertaken. The first two parts broadly deals with LRD and Grapevine leafroll-associated virus 3 (GLRaV-3) specifically, for a comprehensive review refer to the report by Charles *et al.* compiled in 2006 for the New Zealand winegrowers. The third part focuses on the role of sgRNAs in viral replication, highlighted with examples from the family *Closteroviridae*. The fourth part of the literature overview is on the use of infectious clones as molecular tools to study members of the family *Closteroviridae* and the use of derived deletion mutants (mini-replicons).

2.2 Grapevine Leafroll Disease

2.2.1 History

Reports of aberrant grapevine leaf morphology date back to the mid 1800s and were initially thought to be physiological in origin (Martelli and Boudon-Padieu, 2006). It was not until 1935 when Sheu demonstrated that the disorder "Rollkrankheit" was graft transmissible, that a pathogen was suspected (referenced by Charles *et al.*, 2006). Several studies (geographically separated) over the years have investigated diseases now all believed to be LRD and are regarded as synonyms: White Emperor disease (English), "Rollkrankheit" and "Blattrollkrankheit" (German), "enroulement" (French), "accartocciamento" and "accartocciamento fogliare" (Italian), "enrollamiento de la hoja" and "enrollado" (Spanish) and "Enrolamento de la folha" (Portuguese) (Martelli and Boudon-Padieu, 2006). Since then LRD has been detected around the world and is currently the most widespread of the grapevine viral diseases.

2.2.2 Symptoms

The phenotypic symptoms of LRD in red *Vitis vinifera* cultivars are diagnostic and easily identified. The older leaves on the plant turn red prematurely, progressing to a dark purple while the primary and secondary veins remain green, this symptom will spread through the plant as the season progresses. Later in the season the leaves become brittle and roll downwards. In white cultivars the symptoms are less noticeable. Leaves also roll

downward but do not turn red but rather yellow or become chlorotic. Rootstock cultivars seem to be mainly symptomless (Martelli and Boudon-Padieu, 2006). The severity of the phenotypic symptoms also seems to be linked to cultivar and the combination of scion and rootstock used (Golino, 1993). See Figure 2.1 for typical LRD symptoms in red and white cultivars.

Figure 2.1: Typical grapevine leafroll diseased vines from a red cultivar on the left and a white cultivar on the right.

2.2.3 Detection

The detection and identification of viruses form a critical part of the defence against these debilitating diseases. The earliest method used to identify virus diseases was through indexing. This method is based on the assumption that a virus will be transmissible through the graft union and induce symptoms on the indicator shoot. Indexing has been successfully used for LRD, typically using *V. vinifera* cv Cabernet Franc as the indicator. The drawbacks to this method are that it takes 1-3 years before a result is obtained and it does not provide any additional information on the viruses infecting the plant being tested.

The need to be able to identify not only the disease but also the different viruses involved was greatly satisfied by the introduction of serological techniques. The most well known of these techniques are the enzyme-linked immunosorbent assay (ELISA). It has been successfully used to detect GLRaV-3 infection in field collected samples (Ling *et al.*, 2000; Zee *et al.*, 1987). The development of virus and even strain-specific antibodies makes ELISA a very useful detection system that is currently the method of choice in industry for routine screening. The advantages of ELISA are that it is fast, reliable, relatively inexpensive and up-scalable to process large numbers.

RT-PCR based techniques are generally regarded as the most sensitive detection methods currently available (Dovas and Katis, 2003; Gambino and Gribaudo, 2006; La Notte *et al.*, 1997; Ling *et al.*, 2001; Osman and Rowhani, 2006; Osman *et al.*, 2007; Osman *et al.*, 2008). It is sensitive enough to be unaffected by seasonal fluctuations in virus titre that might affect the result in other tests. The strength of these techniques is also their drawback. It is highly specific and it is possible to determine genetic variants but it is also possible to generate false negatives as one unfortunate mismatch in the wrong position of a primer might end in a false negative result. RT-PCR is reliant on good sequence information and primer design with the use of multiple sequence alignments and possibly degenerate primers.

2.2.4 Grapevine leafroll-associated viruses

The initial research performed to determine the virus particles associated with LRD, identified potyvirus-like, closterovirus-like and isometric virus-like particles (Castellano and Martelli, 1984; Namba *et al.*, 1979; Tanne *et al.*, 1977). The transmission of the closterovirus-like particles was the final proof that the virus associated with LRD is a closterovirus or closteroviruses (Rosciglione and Gugerli, 1989; Tanne, 1988). Since the first purification of LRD closterovirus particles, many additional viruses have been identified, as determined through serology or more recently through nucleotide similarities. To date there are possibly 11 viruses associated with LRD (Table 2.1). In South Africa GLRaV-3 is the most prevalent virus associated with LRD (Pietersen, 2004).

Table 2.1: List of Grapevine leafroll-associated viruses

Name	Reference *
Grapevine leafroll-associated virus 1	Gugerli *et al.*, 1984
Grapevine leafroll-associated virus 2	Gugerli *et al.*, 1984
Grapevine leafroll-associated virus 3	Rosciglione and Gugerli, 1986
Grapevine leafroll-associated virus 4	Hu *et al.*, 1990
Grapevine leafroll-associated virus 5	Walter and Zimmerman, 1991
Grapevine leafroll-associated virus 6	Gugerli *et al.*, 1997
Grapevine leafroll-associated virus 7	Choueiri *et al.*, 1996
Grapevine leafroll-associated virus 8	Does not exist
Grapevine leafroll-associated virus 9	Alkowni, *et al.*, 2002
Grapevine leafroll-associated virus Pr	Maliagka *et al.*, 2008
Grapevine leafroll-associated virus De	Maliagka *et al.*, 2008
Grapevine leafroll-associated virus ?	Abou Ghanem-Sabanadzovic at al., 2004

* As referenced by Martelli and Boudon-Padieu, 2006.

2.2.5 Effects of LRD on grapevines

Numerous studies have been published that reported on the effect of LRD on the yield and quality of grapes on infected vines. These studies largely used four parameters to evaluate the grapes: yield, sugar level, titratable acid and anthocyanin accumulation. The results from these studies varied greatly, emphasising that there are several contributing factors that should also be taken into consideration when comparing these studies. Factors like *Vitis* spp. and cultivar, rootstock and scion combination, vine age, virus infection status (not only grapevine leafroll-associated viruses but also additional viruses) and climate (Golino, 1993; Lee and Martin, 2009). In general these studies indicate that LRD infected vines have a reduction in yield of up to 80% (Credi and Babini, 1997; Komar *et al.*, 2007; Over de Linden and Chamberlain, 1970) with some notable exceptions where no effect was observed (Mannini *et al.*, 1998). They also seem to confirm a reduction in sugar level with an increase in titratable acid (Cabaleiro *et al.*, 1999; Komar *et al.*, 2007; Over de Linden and Chamberlain, 1970). A reduction in the concentration of anthocyanin in the berry skins, that reduce the quality of the wine made from these grapes have been reported by several research groups (Guidoni *et al.*, 1997; Lee and Martin, 2009; Mannini *et al.*, 2000; Over de Linden and Chamberlain, 1970).

The overall health of grapevines is negatively affected by the virus infection. Diseased vines have been shown to have significant physiological symptoms like reduced photosynthetic ability and vigour that could have a negative effect on the grape yield and quality (Bertamini *et al.*, 2004; Cabaleiro *et al.*, 1999; Guidoni *et al.*, 1997; Mannini *et al.*, 1996; Mannini *et al.*, 2000; Sampol *et al.*, 2003).

2.2.6 Transmission and spread of LRD

Leafroll disease is graft transmissible and mainly spreads through the propagation of infected material. Grapevine leafroll-associated viruses are not mechanically transmissible with the exception of GLRaV-2 that has been shown to be transmissible to an herbaceous host (*Nicotiana benthamiana*) (Castellano *et al.*, 1995). The natural vectors for these viruses have been demonstrated to be mainly mealybugs, but possibly also soft scale insects (Figure 2.2). The vector for each of the grapevine leafroll-associated viruses is not specific and most likely linked to geographical distribution of the insects (Martelli and Boudon-Padieu, 2006). In Figure 2.3 typical spreading patterns can be seen.

Figure 2.2: Close-up photograph of a healthy female *Planococcus ficus* adult with a first instar nymph indicated by the arrow.

Figure 2.3: Panoramic photo of a grapevine leafroll diseased vineyard in the Stellenbosch area showing typical spread patterns of virus infection. The arrows indicate the gradient of spread from adjacent blocks (short distance spread) and the circles, foci of infection (long distance spread) that is spreading outwards.

2.2.7 LRD management

No natural immunity to LRD has been found in *Vitis* spp and it currently seems unlikely that such resistance exist. Some species or cultivars might be more tolerant to virus infection (esp. rootstocks) than others with only mild symptom expression but it has been shown that even in these vines the grapes are negatively affected. Currently, the best approach to maintain general vineyard health is to manage the disease and to plant sanitised material. Managing the disease could involve strategies to eliminate the vector through rigorous spray regimes or biological control with parasitoids like *Anagyrus pseudococci* (Figure 2.4), roguing of infected vines and ensuring that there are no additional stresses on the vine (Charles *et al.*, 2006). The planting of virus-free material have been shown to be the most effective method available (Martelli and Boudon-Padieu, 2006). Sanitation of propagation material is performed by heat therapy but chemotherapy has also been shown to be a possibility (Panattoni *et al.*, 2007).

In the future, molecular approaches such as genetic engineering of disease resistance into grapevine might play a prominent role in disease management. Several researchers are investigating the potential of transgenic grapevines resistant to grapevine viruses like GFLV (Krastanova *et al.*, 1995; Maghuly *et al.*, 2006) and GLRaV-2 and -3 (Freeborough, 2003; Orecchia *et al.*, 2008; Xue *et al.*, 1999). The results from these studies are promising but unfortunately years away from being commercially available, environmental and ethical concerns as well as consumer preferences not withstanding.
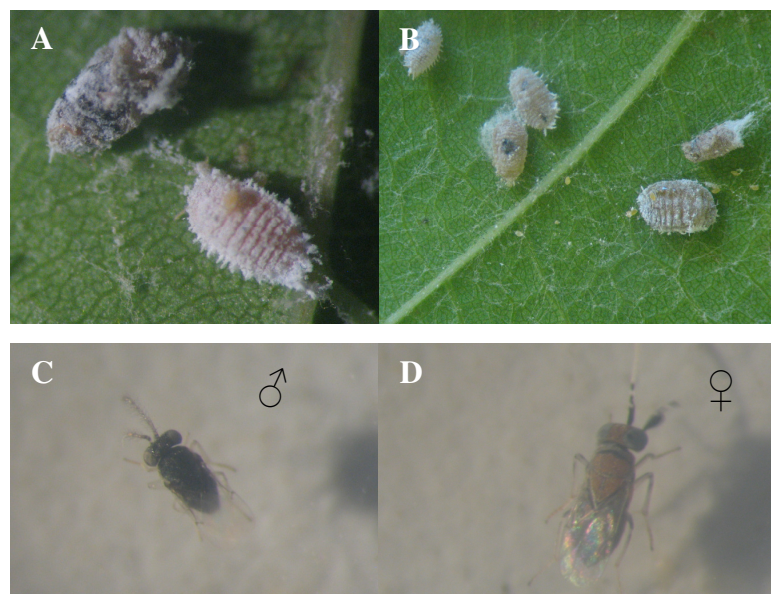
Figure 2.4: A) Photograph of a healthy female *Planococcus ficus* on the right and a mummy of a parasitoid on the left. B) Female *P. ficus* that have already been parasitised, the black spots indicate where the egg of the parasitoid has been laid. C and D) Photographs of the male and female parasitoids *Anagyrus pseudococci* hatched from the parasitised mealybugs.

2.3 *Grapevine leafroll-associated virus* 3

2.3.1 General

Grapevine leafroll-associated virus 3 is generally regarded as the single most economically important grapevine virus in South Africa and possibly world wide (Freeborough and Burger, 2008). The first GLRaV-3 particles were purified from isolate NY-1 by Zee *et al.* in 1987. These purified particles were used to produce antiserum that was developed into an ELISA for field testing. The serological typing of different virus isolates established that there were five distinct recognised serotypes designated GLRaV I, GLRaV II, GLRaV III, GLRaV IV and GLRaV V in 1995 (Boscia *et al.*, 1995). At the International Committee on Taxonomy of Viruses (ICTV) meeting in 1995 it was decided to change the Roman numerals in virus acronyms to Arabic numerals separated by a hyphen from the letters. This changed the acronyms of grapevine leafroll-associated viruses to GLRaV-1 to -5, as it is currently used.

2.3.2 Properties of GLRaV-3

2.3.2.1 Morphology

Grapevine leafroll-associated virus 3 is a phloem limited, flexuous filamentous virus with a non-enveloped virion approximately 1800-2000nm in length (Figure 2.5) (Hoefert and Gifford, 1967; Karasev, 2000; Tidona and Darai, 2001).
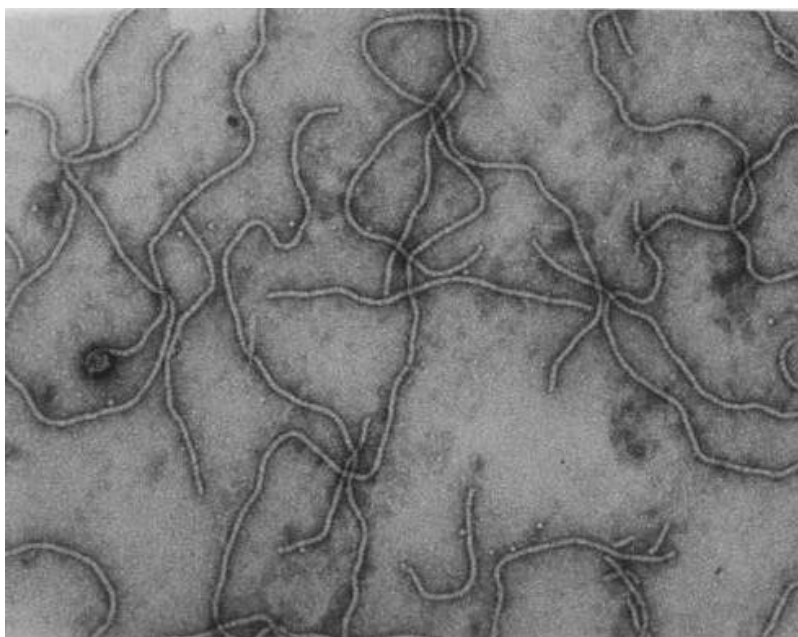
Figure 2.5: Transmission electron micrograph of negatively stained, purified GLRaV-3 particles, using 1% (w/v) aqueous uranyl acetate staining. Picture taken by G.G.F Kasdorf.

2.3.2.2 Genome

In 2004, Ling *et al.* published the first complete sequence of GLRaV-3 isolate NY-1(Ling *et al.*, 2004). Grapevine leafroll-associated virus 3 has a positive sense single stranded (+ss) RNA genome and the genome length was determined to be 17919nt divided into 13 ORFs with 5' and 3' UTRs of 158 and 277 nucleotides, respectively. The genome organisation confirmed that GLRaV-3 is a closterovirus as it was similar to the closterovirus convention established by Dolja *et al.* (1994) (Ling *et al.*, 1998). Since then more isolates of GLRaV-3 had been sequenced: isolate GP18 from South Africa (Maree *et al.*, 2008), isolate CL-766 from Chile (Engel *et al.*, 2008) and isolates 621, 623, and PL20 from South Africa (Jooste *et al.*, 2010). The sequencing of GLRaV-3 isolate GP18 as well as additional proof that the original genome sequence (NY-1) might not be complete is discussed in Chapter 3 of this thesis.

The 13 ORFs of GLRaV-3 were designated ORF1a and 1b and ORFs 2-12 according to the convention set out by Agranovsky *et al.* (1994) (Ling *et al.*, 1998). The putative functions of the different ORFs were determined by sequence comparisons to known proteins (Ling *et al.*, 1998). ORF1a encodes a large polypeptide (Superfamily 1) with four distinct protein domains: leader papain-like protease (L-Pro)(Ling *et al.*, 2004), methyltransferase (Ling *et al.*, 1998), AlkB (Maree *et al.*, 2008) and a Helicase (Ling *et al.*, 1998). OFR1b encodes an RNA-dependent RNA polymerase (RdRp) belonging to the Supergroup 3 RdRps. There is no counterpart for the small peptide putatively encoded by ORF2 in the other closteroviruses. At this position in Citrus tristeza virus (CTV) and Lettuce infectious yellows virus (LIYV) a much larger ORF is found with no sequence similarity (Karasev *et al.*, 1995; Klaassen *et al.*, 1995). ORF3 potentially encodes a small transmembrane hydrophobic protein similar to other closteroviruses. An Hsp70-homologue protein is encoded by ORF4 and was identified by amino acid sequence similarity. Eight conserved domains (A-H) were identified of which three (A-C) are believed to contain an ATPase domain typical of closteroviral Hsp70 chaperone-like proteins. The function of ORF5 could not be confirmed as a potential Hsp90-homologue as expected in other closteroviruses (Ling *et al.*, 1998). These two proteins, along with the movement coat protein (analogous to the GLRaV-3 divergent coat protein), have been shown to form part of the virion tail assembly in BYV, which is responsible for the systemic spread of the virus (Dolja *et al.*, 2006). ORF6 encodes the coat protein (CP) and contains the four amino acids (N, R, G, and D) conserved in all closterovirus coat proteins (Ling *et al.*, 1997). ORF7 potentially encodes a divergent coat protein (dCP) and is identified by four

conserved closteroviral coat protein amino acids (N, R, G and D) on the C-terminus (Ling *et al.*, 1998). It is interesting to note that the order of the CP and the dCP is the same as for the bi-partite closteroviruses (e.g. LIYV, criniviruses) and reversed compared to other mono-partite closteroviruses (e.g. BYV and CTV, closterviruses) (Karasev, 2000). The function of the remaining ORFs 8 to 12 was not determined by Ling *et al.* (1998). Similarity of ORFs 8, 9 and 10 to analogous ORFs of BYV and CTV makes it likely that these ORFs encode for viral silencing suppressors and systemic movement proteins, but remains to be proven experimentally. The small ORFs 11 and 12 are unique to GLRaV-3 and not found in other closteroviruses. The intergenic region found in GLRaV-3 is also atypical of closteroviruses and due to its high GC content is expected to have extensive RNA secondary structure (Karasev, 2000).

2.3.2.3 Genome variation

Several studies have been conducted to determine genetic variability in the GLRaV-3 genome (Fuchs *et al.*, 2009; Turturo *et al.*, 2005). In the study by Fuchs *et al.* (2009) they were able to identify five genetic variant groups. Four of the groups displayed low genetic variation with nucleotide sequence similarity of >90% to isolate NY-1, while the fifth group , represented by a single isolate from New Zealand (NZ-1), displayed high sequence variation with only 74.1% similarity to NY-1. The variant groups were designated NY-1, C5-1, GP18, MT48-2 and NZ-1. Differential symptom expression has been observed that might be linked to genetic variation but it remains to be associated with a specific viral genotype (Habili *et al.*, 2009).

In South Africa two genetic variant groups have previously been identified; represented by isolates 621 and 623 (Jooste and Goszczynski, 2005). Analysis of isolates 621 and 623 found them to group with isolate NY-1 and isolate GP18, respectively. A third South African genetic variant had recently been found and sequenced by Jooste *et al.* (2010). In that study, three genetic variants of GLRaV-3 were fully sequenced and phylogenetically analysed. In Chapter 3 of this thesis the variation in the 5' UTRs of these sequence variants are discussed.

2.3.3 Taxonomy

In 2002, the family *Closteroviridae* was revised by the ICTV study group on closteroviruses and allied viruses. The taxonomic composition of the family *Closteroviridae* was restructured to incorporate biological data as suggested by Karasev

(2000). The grouping of viruses according to their mono- vs. bi-partite genomes was abandoned and replaced by a system where viruses are grouped according to their transmission vector; aphid, whitefly or mealybug. This necessitated the establishment of a new grouping for closteroviruses that are transmitted by mealybugs. The new genus suggested by Martelli *et al.* is *Ampelovirus* (from ampelos, Greek for grapevine), with GLRaV-3 as the type species and was approved by the ICTV in 2002 (Martelli *et al.*, 2002; Mayo, 2002).

2.3.4 Vector transmission of GLRaV-3

Grapevine leafroll-associated virus 3 can only infect *Vitis* spp., and in South African vineyards is predominantly spread semi-persistently by the mealybug, *Planococcus ficus* (Bar-Joseph *et al.*, 1997; Karasev, 2000). More recently it has been shown that GLRaV-3 was present in the salivary glands of *Planococcus citri,* challenging the notion of semi-persistence and putting forward the hypothesis that GLRaV-3 is transmitted by the insect *via* a circulative mechanism (Cid *et al.*, 2007). Although GLRaV-3 cannot be transmitted by mechanical means, it can by transmitted by grafting and is rapidly spread by the planting of infected propagation material (Pietersen, 2004).

Several studies have been conducted on the spread of GLRaV-3 by insect vectors. Combined, these studies determined that GLRaV-3 could be transmitted by: *Planococcus ficus* (Engelbrecht and Kasdorf, 1990), *Planococcus Citri* (Cabaleiro *et al.*, 2008), *Pseudococcus longispinus* and *Pseudococcus calceolariae* (Petersen and Charles, 1997; Tanne, 1988). *Planococcus ficus* and *Pseudococcus longispinus* have also been shown to be such effective vectors that a single nymph is capable of transmitting the virus under experimental conditions (Douglas and Krüger, 2008).

2.3.5 Replication of GLRaV-3

The replication mechanism of GLRaV-3, as with most woody plant phloem-limited viruses, has not been studied in detail (Zee *et al.*, 1987). It is assumed that GLRaV-3 will follow a similar replication strategy to other closteroviruses like CTV and BYV, which have been studied comprehensively. The next section (2.4) will describe the replication strategies of these viruses.

Some studies did investigate GLRaV-3 replication by determining the distribution of GLRaV-3 in the plant throughout the growing season. Monis *et al.* (1996) demonstrated that the highest concentration of virus accumulated in the petioles of older leaves and in

cane material (Monis and Bestwick, 1996). Ling *et al.* (2001) confirmed this and added that the most reliable source of GLRaV-3 infected tissue is bark scrapings of mature canes. Citrus tristeza virus (Hilf *et al.*, 1995) and BYV (He *et al.*, 1997) express the ORFs located on the 3' half of their genomes *via* sgRNA, and similarly it is hypothesised that ORFs 3-12 of GLRaV-3 are also expressed *via* sgRNAs. Subgenomic RNAs are RNA molecules of viral origin, shorter than the genome, analogous to mRNA utilised for the expression of the proximal ORF on the RNA molecule. The presence of GLRaV-3 specific sgRNAs in leafroll-diseased vines have been observed by several research groups but have not been characterised further (Habili and Razaian, 1995; Hu *et al.*, 1990; Ling *et al.*, 1997; Monette and James, 1990; Mossop *et al.*, 1985; Prosser *et al.*, 2007; Rezaian *et al.*, 1991; Saldarelli *et al.*, 1994). The only study that added some information on the composition and characteristics of the sgRNAs associated with GLRaV-3 infection was done by Saldarelli *et al.* (1994). In their study they made a cDNA library from dsRNA isolated from an Italian isolate of GLRaV-3. The cDNA clones were then used to make RNA probes for northern blots. A probe transcribed from pGEM23ds showed high specificity to GLRaV-3 dsRNA and was able to detect several bands expected to be the genomic RNA and sgRNA; thereby proving that the sgRNAs are derived from the genomic RNA (Saldarelli *et al.*, 1994). This probe (23ds) was sequenced by Habili *et al.* (1995) and used to study GLRaV-3 spread in Australia. Analysis of the probe sequence by Ling *et al.* (1998) showed that the sequence comprised of a 5' portion of ORF5 and a portion of the 3' UTR. This led them to believe that this sequence is proof that GLRaV-3, like CTV also produces defective RNA (DRNA). Chapter 4 of this thesis adds to the limited knowledge by describing the identification of the 5' ends of seven positive sense GLRaV-3 sgRNAs.

2.4 Role of sgRNAs in the infection cycle of closteroviruses

2.4.1 Replication and expression of viral ORFs.

To understand the role of sgRNAs in the infection cycle of closteroviruses it is important to have a working knowledge of what is currently understood as a general outline of the infection cycle of these viruses.

The majority of plant viruses identified to date have +ssRNA genomes with great organisational variety of the ORFs and terminal structures. The 5' and 3' terminal structures have been shown to play a vital role in replication and includes structures like a cap or genome linked protein (VPg) at the 5' end and a poly(A)-tail or tRNA-like structure

at the 3' end (Goldbach *et al.*, 1991). Through multiple amino acid sequence alignments of the RdRps of plant viruses, Koonin (1991) was able to identify three viral supergroups. The closteroviruses sorted into supergroup III with other Tymo-, Rubi-, and Tobamo-like viruses (Bustamante and Hull, 1998). The organisation of the ORFs and the relation between the non-structural proteins would indicate a common evolutionary origin and a similar replication strategy (Bar-Joseph *et al.*, 1997; Bustamante and Hull, 1998; Koonin, 1991). Closteroviruses, like most other members of supergroup III are capped, uses sgRNA during replication, and do not have a poly(A)-tail.

The infection cycle of closteroviruses comprise of the following stages (Bustamante and Hull, 1998; Dolja *et al.*, 2006): i) entry into the host cell and un-coating of the virion, ii) translation of the viral replicase polyprotein by the host cell, iii) transcription of genomic and sgRNA by the viral RdRp, iv) translation of viral ORFs *via* sgRNA and v) encapsidation of viral genomic RNA and spread of virions.

2.4.1.1 Entry into the host cell and un-coating of virion

Closteroviruses are naturally spread mainly by insects from the order Homoptera in a semi-persistent manner (Bar-Joseph *et al.*, 1997; Karasev, 2000). Upon entry into the cell the replication cycle is initiated by the un-coating of the virion and exposure of the genomic +ssRNA to the cellular replication proteins.

2.4.1.2 Translation of viral replicase polyprotein by host cell

The viral RNA replicase is directly translated from the viral genomic RNA molecule by the host cell ribosomes (Karasev *et al.*, 1989). It is also possible that the virion tail proteins associated with the 5' end of the genome could play a role by securing the translational initiation proteins from the host (Dolja *et al.*, 2006). The translated viral replicase contains three domains: a methyltransferase (MET), a helicase (HEL) and an RdRp, which is translated *via* a +1 frameshift in very low quantities (Koonin and Dolja, 1993). The replicase polyprotein is then processed into the L-Pro and the replicase components that are predominantly MET-HEL with low quantities of MET-HEL-RdRp (Agranovsky *et al.*, 1994). The replicase proteins restructure the endoplasmic reticulum to form vesicles for the transcription of the viral genomic and sgRNA (Dolja *et al.*, 2006).

The L-Pro is not only involved in its own autocatalytic processing but plays a vital role in the accumulation of viral RNA in the host (Peng and Dolja, 2000). The mechanism by which the viral accumulation is enhanced is still being elucidated. More recently the

activity of the tandem leader protease of GLRaV-2 has been demonstrated to be host-specific (Liu *et al.*, 2009).

The AlkB domain is not ubiquitously present in plant viruses and found in only in a few members (mostly woody plant infecting viruses) from different genera. This domain has been identified in GLRaV-3 as well as several other ampeloviruses (Dolja, 2009). The AlkB proteins from bacteria and mammals are iron(II)- and 2-oxoglutarate-dependent dioxygenases that are able to reverse methylation damage in DNA and RNA (Aas *et al.*, 2003). Van den Born *et al.* (2008) demonstrated, for the first time, that the AlkB domains of plant viruses play a vital role in the repair of methylation damage of ssRNA and dsRNA.

2.4.1.3 Transcription of genomic and sgRNA by viral RdRp

The RdRp transcribe full-length negative sense (complementary) genomic RNA molecules from the genome that forms the templates for the transcription of new progeny genomic RNA molecules (Bustamante and Hull, 1998). It also transcribes sgRNAs utilised in the expression of various ORFs. The production of sgRNA is regulated and timed for optimal infection efficiency (Dolja *et al.*, 2006). The method by which the RdRp transcribes the sgRNAs will be discussed separately (2.4.2).

2.4.1.4 Translation of viral ORFs via sgRNA

The timing of viral ORF expression *via* sgRNA is essential to the success of the infection (Hagiwara *et al.*, 1999). Generally, the ORFs involved in suppression of host silencing mechanisms are expressed first, to be followed by the structural proteins, short and long distance movement proteins and other proteins (Dolja *et al.*, 2006).

2.4.1.5 Encapsidation of viral genomic RNA and virus spread

The mass accumulation of structural proteins initiates the encapsidation of the genomic RNA. These newly formed virions accumulate to high concentrations in the host cell to be potentially acquired by an insect vector for spread to another host plant. Some of the virions spread through the plasmodesmata to adjacent cells (cell-to-cell movement) until they reach the phloem tissue from where it is possible to travel throughout the plant (long distance systemic movement) (Dolja *et al.*, 2006).

2.4.2 Expression of 3' half of genome via sgRNA

Approximately half the genome of closteroviruses is dedicated to ORFs coding for proteins involved in the replication the genome. In GLRaV-3, ORF1a and 1b, span from nucleotide 738 to nucleotide 9066 out of 18498 nucleotides (Maree *et al.*, 2008) making the production of sgRNAs essential for the expression of the ORFs located on the 3' half of the genome. The mechanism by which sgRNAs generally are produced is currently not known and two mechanisms have been proposed: internal initiation (II) and premature termination (PT) (Bertamini *et al.*, 2004; Bustamante and Hull, 1998; Miller and Koev, 2000). The II method relies on the RdRp to initiate the production of sgRNA from the negative-sense genomic RNA molecule (Jaspars, 1998; Miller *et al.*, 1985), while the PT method proposes that there is a premature termination of the negative-sense genomic RNA during replication. These shortened negative-sense RNA molecules would then serve as templates for sgRNA production (Palukaites *et al.*, 1983; Sit *et al.*, 1998). In Figure 2.6 a diagrammatic representation of these two methods can be seen.

The level and timing of the sgRNAs' transcription is controlled by sgRNA-promoters (sg-promoters) or sgRNA control elements. Viruses using the II method of sgRNA production have been shown the utilise sg-promoters and have been the focus of many studies. It has been shown to have limited nucleotide conservation between viruses or within a specific virus. In Barley yellow dwarf virus (BYDV) all three sg-promoters have different primary and secondary structures and positions relative to the start site (Miller and Koev, 2000). However, the sg-promoters of Brome mosaic virus (BMV) have been demonstrated to be conserved on nucleotide level (Siegel *et al.*, 1997) and the secondary structure of the viral sg-promoters is critical for promoter activity (Haasnoot *et al.*, 2000).

The method by which closteroviruses transcribe their ORFs have not been clarified and warrants more research on a wider range of viruses. Investigations into the infection cycle of CTV and BYV revealed that they produce positive and negative sense sgRNAs, making it more likely that they, and potentially all closteroviruses, are utilising the PT method to produce sgRNA (Dolja *et al.*, 1990; Hilf *et al.*, 1995). Gowda *et al.* (2001) studied the activity of the sgRNA controller elements in CTV but were unable to establish if these elements are promoters or terminators of replication (Gowda *et al.*, 2001). It is hypothesised that CTV has a highly complex infection cycle that could potentially produce
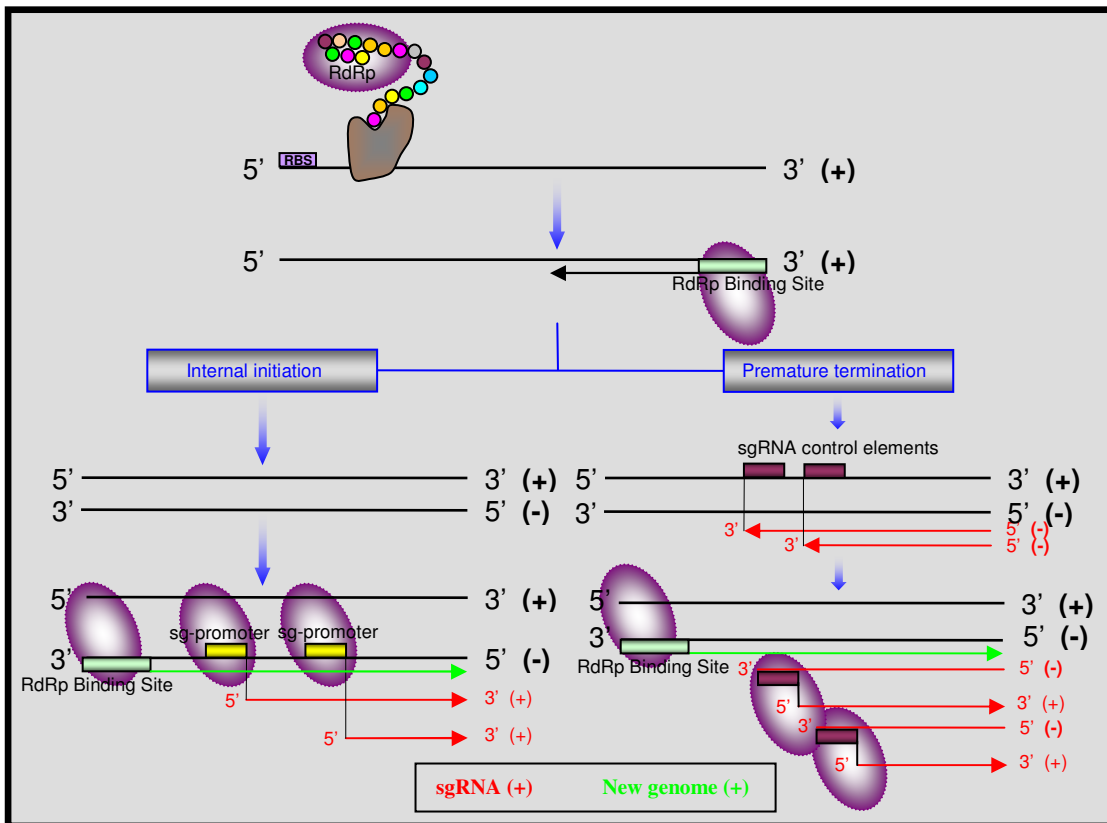
Figure 2.6: Diagrammatic representation of the two proposed methods for the production 3' co-terminal sgRNAs. In virus replication the 5' genes (ORF1a and 1b) is translated directly from the positive sense gRNA, producing MET-HEL and MET-HEL-RdRp proteins. The RdRp is able to bind to the 3' end of the gRNA and produces negative sense RNA molecules. There are two proposed methods for the production of the sgRNA. **Internal initiation:** RdRp produces full-length negative sense RNA that serves as template for the production of sgRNA. The sg-promoters serve as binding sites for the RdRp from where transcription can be initiated. **Premature termination:** RdRp produces not only full-length negative sense RNA molecules, but also negative sgRNAs. The point of termination is sometimes referred to as the sgRNA control elements. The negative sgRNAs can then serve as templates for the production of positive sgRNA used for translation.

as many as 33 different RNA species, these include: 3' and 5' co-terminal sgRNA of positive and negative sense as well as DRNA (Bar-Joseph *et al.*, 1997; Gowda *et al.*, 2001; He *et al.*, 1997; Hilf *et al.*, 1995; Mawassi *et al.*, 1995a; Mawassi *et al.*, 1995b). The complexity of the CTV infection cycle hampers progress in elucidating the mechanism by which sgRNA controller elements regulate sgRNA transcription. These elements do not seem to be conserved at the nucleotide sequence level within individual viruses or between related viruses but possibly have conserved features in their secondary structure. The variation observed in these structures might contribute to the regulation and timing of gene expression (Dolja *et al.* 2006). Conversely, for BYV two conserved sg-promoter motifs were identified in a study by Vitushkina *et al.* (2007). These motifs were found to be conserved in three orthologous genes of CTV and two orthologous genes of Beet yellow stunt virus.

It is clear that the expression of the 3' half of closteroviral genomes are not completely understood yet. Despite this gap in the current knowledge, it has however been demonstrated that the production of sgRNAs are central to the expression strategy.

2.5 Infectious clones of plant RNA viruses

Infectious clones of plant RNA viruses provide a valuable platform to study viral functional genomics as well as gain insight into the replication and expression of viral ORFs through mutagenesis (Boyer and Haenni, 1994; Nagyová and Šubr, 2007). Plant viruses have relatively small genomes that make them particularly suitable to be assembled into cDNA clones, thereby making them easier to manipulate. Infectious clones are usually assembled in a bacterial plasmid, from cDNA fragments generated by RT-PCR. These clones could then be manipulated to suit a particular investigation. Even though many viruses have been converted into infectious clones it remains a difficult task with many obstacles (Boyer and Haenni, 1994). The design and the assembly strategy of the intended infectious clone needs to be carefully planned to ensure that the cDNA clone represents the wild type sequence. It is also necessary to exclude non-viral nucleotides between the promoter elements and the cDNA clone especially at the 5' end (Boyer and Haenni, 1994). Constructed cDNA clones are prone to mutations that are introduced during reverse transcription, assembly or when transformed into *Escherichia coli*. Clones are often unstable or toxic in *E. coli* which may lead to random rearrangements and point mutations that could render the clone non-infectious. Several strategies have been employed to overcome the mutation and instability problems associated with these clones. These include the use of high fidelity long template PCR, the incorporation of eukaryotic introns at critical genome regions, the use of a population cloning strategy and the inclusion of frameshifts (López-Moya and García, 2000; Satyanarayana *et al.*, 2003; Yamshchikov *et al.*, 2001; Yu and Wong, 1998).

2.5.1 Types of infectious clones

The infectious clones of RNA viruses can be divided into two types depending on the site of transcription: infectious RNA (*in vitro*) and infectious cDNA (*in vivo*). Both contain the viral genome but differ in the regulatory sequences flanking the viral genome. Infectious clones producing infectious RNA *in vitro* contain the viral genome under the control of a bacterial phage promoter typically T7, but also λpm, SP6 and T3 (Nagyová and Šubr,

2007). The phage promoter is then used to transcribe large quantities of viral RNA from the cDNA clone. To ensure optimal infectivity, the transcribed RNA must be identical to the viral wild type RNA. This implies that no additional nucleotides should be inserted between the viral cDNA 5' end and the promoter's transcription initiation site, since that would add nucleotides to the 5' end of the transcribed RNA (Boyer and Haenni, 1994). Some drawbacks to the infectious RNA approach are the sensitivity of the transcribed RNA to degradation and the difficulties associated with mechanical inoculation. However, the advantage of this technique is that the transcribed RNA functions as mRNA that can be directly translated in the cytoplasm and does not require delivery to the nucleus (Nagyová and Šubr, 2007). The *in vivo* transcription of infectious RNA from the viral cDNA clones is achieved using the CaMV35S promoter. *In vivo* transcription has several advantages: it is simpler to perform and is less expensive and less sensitive to degradation compared to *in vitro* transcription. The main disadvantage is that the cDNA needs to be delivered to the nucleus for transcription to occur (Boyer and Haenni, 1994).

2.5.2 Transfection of infectious clones

There are several methods available to transfect plants and plant tissues with infectious cDNA clones or RNA: Agroinfection, biolistics, electroporation, liposome-mediated transfection, microinjection and mechanical inoculation (Nagyová and Šubr, 2007).

Agroinfection is based on the ability of *Agrobacterium* species (mainly *A. tumefaciens*) to infect plant cells and transfer its T-DNA to the nucleus (Leiser *et al.*, 1992). By incorporating the infectious clone into the T-DNA ensures that it will be delivered to the nucleus where it will be transcribed. *Agrobacterium* can be infiltrated into plant tissues by syringe, vacuum infiltration or agrodrenching (Brigneti *et al.*, 2004; Ekengren *et al.*, 2003; Liu and Lomonossoff, 2002; Vaghchhipawala and Mysore, 2008).

Biolistics is mainly used to transfect tissues of plants that are not a natural host of *Agrobacterium* (Turnage *et al.*, 2002). Nucleic acids are precipitated onto gold or tungsten particles and shot with compressed helium, under vacuum, into plant tissues.

Electroporation, liposome-mediated transfection and microinjection are methods available for the transfection of protoplasts. During electroporation a high voltage pulse is applied to a solution of recombinant nucleic acids and protoplasts. Liposome-mediated transfection introduces nucleic acids to the protoplasts non-invasively while stably packaging these nucleic acids in liposomes (Lurquin and Rollo, 1993). Microinjection is another protoplast transfection method that can transfer nucleic acids to individual cells. This method is time-

consuming and requires expensive devices that make this a specialised approach (Kost *et al.*, 1995; Reich *et al.*, 1986).

Mechanical inoculation of nucleic acids to leaf surfaces is usually used for the transfection of *in vitro* RNA transcripts. In this method the leaf surface is abraded and the nucleic acids applied to enter the cells through the damaged cell walls in a way similar to how viruses are mechanically transmitted (Hull, 2002). This method might be less effective than the other methods but it is inexpensive and fast.

2.5.3 Infectious clones of closteroviruses and their deletion mutants

Infectious clones of several grapevine-infecting viruses have been constructed that have aided research on virus replication and plant-pathogen interactions. They include: the nepovirus GFLV (Viry *et al.*, 1993), the vitiviruses GVA (Galiakparov *et al.*, 1999; Saldarelli *et al.*, 2000) and GVB (Moskovitz *et al.*, 2007; Saldarelli *et al.*, 2000) and the closterovirus GLRaV-2 (Liu *et al.*, 2009). These infectious clones hold great potential to be converted into expression or Virus induced gene silencing (VIGS) vectors. The viti- and nepovirus infectious clones are more sensitive to foreign nucleic acids inserted into their genomes than closteroviruses, which have greater genetic capacity to accommodate these nucleic acids (Dolja *et al.*, 2006; Folimonov *et al.*, 2007). GLRaV-2 is the only grapevine-infecting closterovirus that is able to infect an herbaceous host (*N. benthamiana*), making it an ideal model virus to study plant-virus interaction. Currently there are no ampelovirus infectious clones and in Chapter 5 the first steps towards such a clone is described.

Several closteroviruses have been converted into infectious clones: LIYV (Klaassen *et al.*, 1995), BYV (Peremyslov *et al.*, 1998), CTV (Satyanarayana *et al.*, 1999) and GLRaV-2 (Liu *et al.*, 2009).

The first closterovirus to be converted into an infectious clone was LIYV, which has a bipartite genome (Klaasen *et al.*, 1996). These researchers were able to construct full-length cDNA constructs for RNA 1 and RNA 2, under the control of the T3 promoter. RNA 1 *in vitro* transcripts were proven to be replication competent in protoplasts (*N. benthamiana*), independent from RNA 2, while RNA 2 was shown to be dependent on RNA 1 co-inoculation.

Peremyslov *et al.* (1998) constructed a full-length infectious clone for BYV and demonstrated that the RNA transcripts are infectious in protoplasts (*Nicotiana tabacum* cv. Xanthi). The BYV infectious clone was used to confirm the function of several ORFs: ORF 1a and 1b are the replicase proteins and ORF8 (p21) has suppressor activity.

Additionally, they found that the six ORFs located on the 3' terminus of the genome are dispensable for replication. The gene expression profile of BYV infection was determined by investigating the transcriptional regulation of BYV ORFs using infectious constructs containing GUS at different positions in the genome (Hagiwara *et al.*, 1999). Hagiwara *et al.* (1999) confirmed that ORFs 2 to 8 are not required for replication. The construction of a BYV mini-replicon, analogous to LIYV RNA 1, proved to be a valuable molecular tool in closterovirus research (Figure 2.7). Through mutational analysis, the role of the L-Pro in BYV replication was determined (Peng and Dolja, 2000). The activity of the BYV p21 and five heterologous suppressors were evaluated using a BYV mini-replicon assay (Chiba *et al.*, 2006). Through the use of the BYV infectious clones and mini-replicons, the function and regulation of several BYV ORFs and domains were determined.
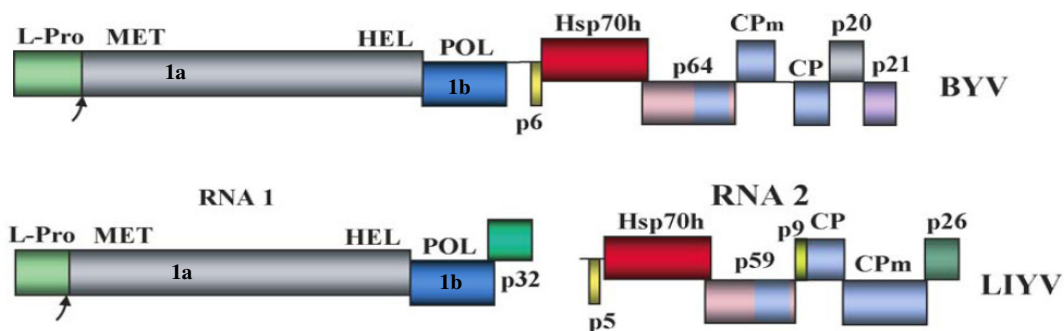
Figure 2.7: Genome maps of BYV (monopartite) and LIYV (bipartite), both members of the *Closteroviridae* family. Conserved proteins are colour coded. These maps clearly display the conserved proteins between LIYV RNA1 and ORFs 1a and 1b of BYV that were demonstrated to be essential for replication (Hagiwara *et al.*, 1999; Klaasen *et al.*, 1996). This figure is adapted from Figure 2 in Dolja *et al.* (2006).

The first woody plant closterovirus to be converted to an infectious clone was CTV (Satyanarayana *et al.*, 1999). A full-length CTV infectious cDNA clone was constructed from which replication-competent RNA transcripts could be transcribed. The large genome of CTV made the clone difficult to manipulate and prompted the construction of a mini-replicon similar to BYV. Chimaeric CTV mini-replicons, containing terminal sequences from different genotypes, were then used to investigate its effect on replication (Satyanarayana *et al.*, 1999).

The latest closterovirus to be converted to an infectious clone was GLRaV-2 (Liu *et al.*, 2009). Lui *et al.* (2009) generated full-length, reporter tagged (GFP/GUS) clones to demonstrate that they were able to infect *N. benthamiana* plants systemically. Through the use of these infectious clones and their mini-replicon derivatives they investigated the

function of the tandem papain-like leader proteases (L1 and L2) in the GLRaV-2 infection cycle. They demonstrated that L1 is vital for infection in *N. benthamiana*, the experimental host, and that L2 only plays a minor role. However, in *Vitis vinifera*, the natural host, both L1 and L2 was found to be essential, indicating host-specific requirements.

2.6 Conclusion

Leafroll disease is arguably the most important grapevine viral disease, but the inherent difficulties associated with studying viruses of woody plants that are phloem-limited, have hampered advances in research. The most important virus associated with LRD is GLRaV-3, which unfortunately only infects *Vitis* spp. Research on GLRaV-3 has mainly focussed on epidemiology and the development of detection techniques, while studies on the genome and viral replication lagged far behind compared to other grapevine viruses. The genome of GLRaV-3 was only sequenced in 2004 (Ling *et al.*, 2004) and even though several research groups detected sgRNAs, these were not further pursued. The closterovirus CTV has had similar difficulties, but these were overcome largely due to its great economic impact on the citrus industry and its use in cross protection. Research on CTV has made great progress in elucidating its infection cycle and the role of sgRNAs in replication. The use of an infectious clone and its deletion mutants have accelerated the understanding of the viral replication mechanisms. Although, GLRaV-3 is an ampelovirus, its close evolutionary relation to the closteroviruses (CTV, BYV and GLRaV-2) allows us to draw from this research and apply it to GLRaV-3. In the research chapters that will follow the lack of knowledge on the genomic and subgenomic RNA of GLRaV-3 will be addressed.

2.7 References

Aas, P.A., Otterlei, M., Falnes, P.O., Vagbo, C.B., Skorpen, F., Akbari, M., et al, 2003. Human and bacterial oxidative demethylases repair alkylation damage in both RNA and DNA. Nature 421, 859–863.

Agranovsky, A.A., Koonin, E.V., Boyko, V.P., Maiss, E., Frötschl, R., Lunina, N.A., et al, 1994. Beet yellows closterovirus: Complete genome structure and identification of a leader papain-like thiol protease. Virology 198, 311-324.

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, 395-401.

Bar-Joseph, M., Yang, G., Gafny, R., Mawassi, M., 1997. Subgenomic RNAs: The possible building blocks for modular recombination of *Closteroviridae* genomes. Semin.Virol. 8, 113-119.

Bertamini, M., Muthuchelian, K., Nedunchezhian, N., 2004. Effect of grapevine leafroll on the photosynthesis of field grown grapevine plants (Vitis vinifera L. cv. lagrein). J. Phytopathol. 152, 145-152.

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

Boyer, J., Haenni, A., 1994. Infectious transcripts and cDNA clones of RNA viruses. Virology 198, 415-426.

Brigneti, G., Martin-Hernandez, A., Jin, H., Chen, J., Baulcombe, D., Baker, B., et al, 2004. Virus-induced gene silencing in solanum species. The Plant Journal 39, 264–272.

Bustamante, P.I., Hull, R., 1998. Plant virus gene expression strategies. Electronic Journal of Biotechnology 1, ISSN: 0717-3458.

Cabaleiro, C., Couceiro, C., Pereira, S., Cid, M., Barrasa, M., Segura, A., 2008. Spatial analysis of epidemics of Grapevine leafroll-associated virus 3. Eur. J. Plant Pathol. 121, 121–130.

Cabaleiro, C., Segura, A., Garcia-Berrios, J.J., 1999. Effects of Grapevine leafroll-associated virus 3 on the physiology and must of Vitis vinifera L. cv. albarifio following contamination in the field. Am. J. Enol. Vitic. 50, 40-44.

Castellano, M.A., Abou-Ghanem, N., Martelli, G.P., Boscia, D., Savino, V., 1995. Cytopathology of two filamentous grapevine viruses and their intracellular identification by gold immunolabelling. Journal of Plant Diseases and Protection 102, 23-33.

Castellano, M.A., Martelli, G.P., 1984. Ultrastructure and nature of vesiculated bodies associated with isometric virus-like particles in diseased grapevines. J. Ultrastruct. Res. 89, 56-64.

Charles, J.G., Cohen, D., Walker, J.T.S., Forgie, S.A., Bell, V.A., Breen, K.C., 2006. A review of Grapevine leafroll-associated virus type 3 (GLRaV-3) for the New Zealand wine industry. Report to New Zealand Winegrowers. HortResearch Client Report 18447.

Chiba, M., Reed, J.C., Prokhnevsky, A.I., Chapman, E.J., Mawassi, M., Koonin, E.V., et al, 2006. Diverse suppressors of RNA silencing enhance agroinfection by a viral replicon. Virology 346, 7-14.

Cid, M., Pereira, S., Cabaleiro, C., Faoro, F., Segura, A., 2007. Presence of Grapevine leafroll-associated virus 3 in primary salivary glands of the mealybug vector Planococcus citri suggests a circulative transmission mechanism. Eur. J. Plant Pathol. 118, 23–30.

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

Credi, R., Babini, A.R., 1997. Effects of virus and virus-like infections on growth, yield, and fruit quality of albana and trebbiano romagnolo grapevines. Am. J. Enol. Vitic. 48, 7-12.

Dolja, V.V., 2009. New developments in understanding gene function and evolution of the grapevine closteroviruses. Extended abstracts 16th ICVG conference, Dijon, France., 29-31.

Dolja, V.V., Karasev, A.V., Agranovsky, A.A., 1990. Organization of the Beet yellows closterovirus genome. in: Brinton, M.A., Heinz, F.X. (eds.), new aspects of positive-strand RNA viruses. ASM press, Washington, DC, 31–35.

Dolja, V.V., Karasev, A.V., Koonin, E.V., 1994. Molecular biology and evolution of closteroviruses: Sophisticated build-up of large RNA genomes. Annu. Rev. Phytopathol. 32, 261-285.

Dolja, V.V., Kreuze, J.F., Valkonen, J.P.T., 2006. Comparative and functional genomics of closteroviruses. Virus Res. 117, 38-51.

Douglas, N., Krüger, K., 2008. Transmission efficiency of Grapevine leafroll-associated virus 3 (GLRaV-3) by the mealybugs Planococcus ficus and Pseudococcus longispinus (hemiptera: Pseudococcidae). Eur. J. Plant Pathol. 122, 207–212.

Dovas, C.I., Katis, N.I., 2003. A spot multiplex nested RT-PCR for the simultaneous and generic detection of viruses involved in the aetiology of grapevine leafroll and rugose wood of grapevine. J. Virol. Methods 109, 217-226.

Ekengren, S.K., Liu, Y., Schiff, M., Dinesh-Kumar, S.P., Martin, G.B., 2003. Two MAPK cascades, NPR1, and TGA transcription factors play a role in pto-mediated disease resistance in tomato. Plant J. 36, 905–917.

Engel, E.A., Girardi, C., Escobar, P.F., Arredondo, V., Domínguez, C., Pérez-Acle, T., et al, 2008. Genome analysis and detection of a Chilean isolate of Grapevine leafroll-associated virus 3. Virus Genes 37, 110-118.

Engelbrecht, D.J., Kasdorf, G.G.F., 1990. Transmission of grapevine leafroll disease and associated closteroviruses by the vine mealybug, Planococcus ficus. Phytophylactica 22, 341-346.

Figueiral, I., Bouby, L., Buffat, L., Petitot, H., Terral, J-F., 2010. Archaeobotany, vine growing and wine producing in Roman Southern France: The site of Gasquinoy (Beziers, Hérault). Journal of Archaeological Science 37, 139-149.

Folimonov, A.S., Folimonova, S.Y., Bar-Joseph, M., Dawson, W.O., 2007. A stable RNA virus-based vector for citrus trees. Virology 368, 205-216.

Freeborough, M.J., 2003. A pathogen-derived resistance strategy for the broad-spectrum control of Grapevine leafroll-associated virus infection. University of Stellenbosch, PhD Thesis.

Freeborough, M.J., Burger, J.T., 2008. Leafroll: Economic implications. Wynboer, http://www.wynboer.co.za/recentarticles/200812-leafroll.php3.

Fuchs, M., Martinson, T.E., Loeb, G.M., Hoch, H.C., 2009. Survey for the three major leafroll disease-associated viruses in Finger Lakes vineyards in New York. Plant Dis. 93, 395-401.

Galiakparov, N., Tanne, E., Sela, I., Gafny, R., 1999. Infectious RNA transcripts from Grapevine virus A cDNA clone. Virus Genes 19, 235-242.

Gambino, G., Gribaudo, A., 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, 1223-1229.

Goldbach, R., Le Gall, O., Wellink, J., 1991. Alpha-like viruses in plants. Seminars in Virology 2, 19-25.

Golino, D.A., 1993. Potential interactions between rootstocks and grapevine latent viruses. Am. J. Enol. Vitic. 44, 148-152.

Gowda, S., Satyanarayana, T., Ayllon, M.A., Albiach-Marti, M.R., Mawassi, M., Rabindran, S., et al, 2001. Characterization of the cis-acting elements controlling subgenomic mRNAs of Citrus tristeza virus: Production of positive- and negative-stranded 3'-terminal and positive-stranded 5'-terminal RNAs. Virology 286, 134–151.

Guidoni, S., Mannini, F., Ferrandion, A., Argamante, N., Di Stefano, R., 1997. The effect of grapevine leafroll and rugose wood sanitation on agronomic performance and berry and leaf phenolic content of a nebbiolo clone (Vitis vinifera L.). Am. J. Enol. Vitic. 48, 438-442.

Haasnoot, P.C.J., Brederode, F.T., Olsthoorn, R.C.L., Bol, J.F., 2000. A conserved hairpin structure in alfamovirus and bromovirus subgenomic promoters is required for efficient RNA synthesis in vitro. RNA 6, 708-716.

Habili, N., Cameron, I., Randles, J., 2009. A mild strain of Grapevine leafroll-associated virus 3 is present in desirable clones of crimson seedless table grapes in Western Australia. Extended abstracts 16th ICVG conference, Dijon, France., 237-238.

Habili, N., Razaian, M.A., 1995. Cloning and molecular analysis of double-stranded RNA associated with grapevine leafroll disease. Ann. appl. biol. 127, 95-103.

Hagiwara, Y., Peremyslov, V.V., Dolja, V.V., 1999. Regulation of closterovirus gene expression examined by insertion of a self-processing reporter and by northern hybridization. J. Virol. 73, 7988-7993.

He, X., Rao, A.L.N., Creamer, R., 1997. Characterization of beet yellows closterovirus-specific RNAs in infected plants and protoplasts. Phytopathology 87, 347-352.

Hilf, M.E., Karasev, A.V., Pappu, H.R., Gumpf, D.J., Niblett, C.L., Garnsey, S.M., 1995. Characterization of Citrus tristeza virus subgenomic RNAs in infected tissue. Virology 208, 576-582.

Hoefert, L.L., Gifford, E.M., 1967. Trabeculae in the grapevine infected with leafroll virus. Amer. J. Bot. 54, 257-261.

Hu, J.S., Gonsalves, D., Teliz, D., 1990. Characterization of closterovirus-like particles associated with grapevine leafroll disease. J. Phytopathology 128, 1-14.

Hull, R., 2002. Matthews' plant virology, 4th edition, Academic press.

Jaspars, E.M., 1998. A core promoter hairpin is essential for subgenomic RNA synthesis in alfalfa mosaic alfamovirus and is conserved in other bromoviridae. Virus Genes 17, 233-242.

Jooste, A.E.C., Goszczynski, D.E., 2005. Single-strand conformation polymorphism (SSCP), cloning and sequencing reveals two major groups of divergent molecular variants of Grapevine leafroll-associated virus 3 (GLRaV-3). Vitis 44, 39-43.

Jooste, A.E.C., Maree, H.J., Bellstedt, D.U., Goszczynski, D.E., Pietersen, G., Burger, J.T., 2010. Three Grapevine leafroll-associated virus 3 (GLRaV-3) variants identified from South African vineyards show high variability in their 5'UTR. Archives of Virology, DOI 10.1007/s00705-010-0793-y

Karasev, A.V., 2000. Genetic diversity and evolution of closteroviruses. Annu. Rev. Phytopathol. 38, 293-324.

Karasev, A.V., Agranovsky, A.A., Rogov, V.V., Miroshnichenko, N.A., Dolja, V.V., Atabekov, J.G., 1989. Virion RNA of Beet yellows closterovirus: Cell-free translation and some properties. J. Gen. Virol. 70, 241–245.

Karasev, A.V., Boyko, V.P., Gowda, S., Nikolaeva, O.V., Hilf, M.E., Koonin, E.V., et al, 1995. Complete sequence of the Citrus tristeza virus RNA genome. Virology 208, 511-520.

Klaassen, V.A., Boeshore, M.L., Koonin, E.V., Tian, T., Falk, B.W., 1995. Genome structure and phylogenetic analysis of Lettuce infectious yellows virus, a whitefly-transmitted, bipartite closterovirus. Virology 208, 99-110.

Komar, V., Vigne, E., Demangeat, G., Fuchs, M., 2007. Beneficial effect of selective virus elimination on the performance of Vitis vinifera cv. chardonnay. Am. J. Enol. Vitic. 58, 202-210.

Koonin, E.V., Dolja, V.V., 1993. Evolution and taxonomy of positive strand RNA viruses: Implications of comparative analysis of amino acid sequences. Crit. Rev. Biochem. Mol. Biol. 28, 375-430.

Koonin, E.V., 1991. The phylogeny of RNA-dependent RNA polymerases of positive-strand RNA viruses. J. Gen. Virol. 72, 2197–2206.

Kost, B., Galli, A., Potrykus, I., Neuhaus, G., 1995. High-efficiency transient and stable transformation by optimized DNA microinjection into Nicotiana tabacum protoplasts. J. Exp. Botany 46, 1157-1167.

Krastanova, S., Perrin, M., Barbier, P., Demangeat, G., Cornuet, P., Bardonnet, N., et al, 1995. Transformation of grapevine rootstocks with the coat protein gene of Grapevine fanleaf nepovirus. Plant Cell Reports 14, 550-554.

La Notte, P., Minafra, A., Saldarelli, P., 1997. A spot-PCR technique for the detection of phloem-limited grapevine viruses. J. Virol. Methods 66, 103-108.

Lee, J., Martin, R.R., 2009. Influence of grapevine leafroll associated viruses (GLRaV-2 and -3) on the fruit composition of Oregon Vitis vinifera L. cv. pinot noir: Phenolics. Food Chemistry 112, 889–896.

Leiser, R.M., Ziegler-Graff, V., Reutenauer, A., Herrbach, E., Lemaire, O., Guilley, H., et al, 1992. Agroinfection as an alternative to insects for infecting plants with Beet western yellows luteovirus. Proc. Natl. Acad. Sci. USA 89, 9136-9140.

Ling, K.S., Zhu, H.Y., Alvizo, H., Hu, J.S., Drong, R.F., Slightom, J.L., et al, 1997. The coat protein gene of Grapevine leafroll-associated closterovirus 3: Cloning, nucleotide sequencing and expression in transgenic plants. Arch. Virol. 142, 1101-1116.

Ling, K.S., Zhu, H.Y., Drong, R.F., Slightom, J.L., McFerson, J.R., Gonsalves, D., 1998. Nucleotide sequence of the 3'-terminal two-thirds of the Grapevine leafroll-associated virus-3 genome reveals a typical monopartite closterovirus. J. Gen. Virol. 79, 1299-1307.

Ling, K.S., Zhu, H.Y., Gonsalves, D., 2004. Complete nucleotide sequence and genome organization of Grapevine leafroll-associated virus 3, type member of the genus ampelovirus. J. Gen. Virol. 85, 2099-2102.

Ling, K.S., Zhu, H.Y., Jiang, Z.Y., Gonsalves, D., 2000. Effective application of DAS-ELISA for detection of Grapevine leafroll-associated closterovirus 3 using a polyclonal antiserum developed from recombinant coat protein. Eur. J. Plant Pathol. 106, 301-309.

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

Liu, L., Lomonossoff, G., 2002. Agroinfection as a rapid method for propagating Cowpea mosaic virus-based constructs. J. Virol. Methods 105, 343-348.

Liu, Y., Peremyslov, V.V., Medina, V., Dolja, V.V., 2009. Tandem leader proteases of Grapevine leafroll-associated virus-2: Host-specific functions in the infection cycle. Virology 383, 291-299.

López-Moya, J.J., García, J.A., 2000. Construction of a stable and highly infectious intron-containing cDNA clone of Plum pox potyvirus and its use to infect plants by particle bombardment. Virus Res. 68, 99-107.

Lurquin, P.F., Rollo, F., 1993. Liposome-mediated delivery of nucleic acids into plant protoplast. Methods Enzymol. 221, 409-415.

Maghuly, F., Leopold, S., Machado, A.D., Fernandez, E.B., Khan, M.A., Gambino, G., et al, 2006. Molecular characterization of grapevine plants transformed with GFLV resistance genes: II. Plant Cell Rep 25, 546–553.

Mannini, F., Argamante, N., Credi, R., 1996. Improvements in the quality of grapevine nebbiolo clones obtained by sanitation. Acta Horticulturae 427, 319-324.

Mannini, F., Argamante, N., Credi, R., 2000. Leaf morphological modifications induced by different viruses in clones of Vitis vinifera cultivars. Acta Horticulturae 528, 765-768.

Mannini, F., Gerbi, V., Credi, R., 1998. Heat-treated virus-infected grapevine clones: Agronomical and enological modifications. Acta Horticulturae 473, 155-163.

Maree, H.J., Freeborough, M.J., Burger, J.T., 2008. Complete nucleotide sequence of a south african isolate of Grapevine leafroll-associated virus 3 reveals a 5′ UTR of 737 nucleotides. Arch. Virol. 153, 755-757.

Martelli, G.P., Agranovsky, A.A., Bar-Joseph, M., Boscia, D., Candresse, T., Coutts, R.H.A., et al, 2002. The family *Closteroviridae* revised. Arch. Virol. 147, 2039-2044.

Martelli, G.P., Boudon-Padieu, E., 2006. Directory of infectious diseases of grapevines and viroses and virus-like diseases of grapevine: Bibliographic report 1998-2004. Opinions Mediterraneennes Serie B: Studies and Research.

Mawassi, M., Karasev, A.V., Mietkiewska, E., Gafny, R., Lee, R.F., Dawson, W.O., et al, 1995a. Defective RNA molecules associated with Citrus tristeza virus. Virology 208, 383-387.

Mawassi, M., Mietkiewska, E., Hilf, M.E., Ashoulin, L., Karasev, A.V., Gafny, R., et al, 1995b. Multiple species of defective RNAs in plants infected with Citrus tristeza virus. Virology 214, 264-268.

Mayo, M.A., 2002. A summary of taxonomic changes recently approved by ICTV. Arch. Virol. 147, 1655-6.

McGovern, P.E., 2003. Ancient wine: The search for the origins of viniculture. Princeton University Press.

Miller, W.A., Dreher, T.W., Hall, T.C., 1985. Synthesis of Brome mosaic virus subgenomic RNA in vitro by internal initiation on (+) sense genomic RNA. Nature 313, 68-70.

Miller, W.A., Koev, G., 2000. Synthesis of subgenomic RNAs by positive-strand RNA viruses. Virology 273, 1-8.

Monette, P.L., James, D., 1990. The use of Nicotiana benthamiana as an herbaceous receptor host for closteroviruses from leafroll affected grapevines. Am. J. Enol. Vitic. 41, 201-203.

Monis, J., Bestwick, R.K., 1996. Detection and localization of grapevine leafroll-associated closteroviruses in greenhouse and tissue culture grown plants. Am. J. Enol. Vitic. 47, 199-205.

Moskovitz, Y., Goszczynski, D.E., Bir, L., Fenigstein, A., Czosnek, H., Mawassi, M., 2007. Sequencing and assembly of a full-length infectious clone of Grapevine virus B and its infectivity on herbaceous plants. Arch. Virol. 153, 323-328.

Mossop, D.W., Elliott, D.R., Richards, K.D., 1985. Association of clostervirus-like particles and high molecular weight double-stranded RNA with grapevines affected by leafroll disease. N.Z. J. Agric. Res. 28, 419-425.

Nagyová, A., Šubr, Z., 2007. Infectious full-length clones of plant viruses and their use for construction of viral vectors. Acta Virol. 51, 223–237.

Namba, S., Yamashita, S., Doi, Y., Yora, K., Terai, Y., Yano, R., 1979. Grapevine leafroll virus, a possible member of closteroviruses. Ann. Phytopath. Soc. Japan 45, 497-502.

Orecchia, M., Nolke, G., Saldarelli, P., Dell'Orco, M., Uhde-Holzem, K., Sack, M., et al, 2008. Generation and characterization of a recombinant antibody fragment that binds to the coat protein of Grapevine leafroll-associated virus 3. Arch. Virol. 153, 1075–1084.

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

Osman, F., Leutenegger, C., Golino, D., Rowhani, A., 2008. Comparison of low-density arrays, RT-PCR and real-time TaqMan® RT-PCR in detection of grapevine viruses. J. Virol. Methods 149, 292-299.

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

Over de Linden, A.J., Chamberlain, E.E., 1970. Effect of grapevine leafroll virus on vine growth and fruit yield and quality. New Zealand Journal Of Agricultural Research 13, 689-698.

Palukaites, P., Garcia-Arenal, F., Sulzinski, M.A., Zaitlin, M., 1983. Replication of Tobacco mosaic virus VII. further characterization of single and double stranded virus related RNAs from TMV-infected plants. Virology 131, 533-545.

Panattoni, A., D'Anna, F., Triolo, E., 2007. Antiviral activity of tiazofurin and mycophenolic acid against Grapevine leafroll-associated virus 3 in Vitis vinifera explants. Antiviral Res. 73, 206-211.

Peng, C.W., Dolja, V.V., 2000. Leader proteinase of the Beet yellows closterovirus: Mutation analysis of the function in genome amplification. J. Virol. 74, 9766–9770.

Peremyslov, V.V., Hagiwara, Y., Dolja, V.V., 1998. Genes required for replication of the 15.5-kilobase RNA genome of a plant closterovirus. J. Virol. 72, 5870-5876.

Petersen, C.L., Charles, J.G., 1997. Transmission of grapevine leafroll-associated closteroviruses by Pseudococcus longispinus and P. calceolariae. Plant pathology 46, 509-515.

Pietersen, G., 2004. Spread of grapevine leafroll disease in South Africa - a difficult, but not insurmountable problem. Wynboer, http://www.wynboer.co.za/recentarticles/articles.php3.

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

Reich, T.J., Iyer, V.N., Scobie, B., Miki, B.L., 1986. A detailed procedure for the intranuclear microinjection of plant protoplasts. Can. J. Bot. 64, 1255-1258.

Rezaian, M.A., Krake, L.R., Cunying, Q., Hazzalin, C.A., 1991. Detection of virus-associated dsRNA from leafroll infected grapevines. J. Virol. Methods 31, 325-334.

Rosciglione, B., Gugerli, P., 1989. Transmission of grapevine leafroll disease and an associated closterovirus to health grapevine by mealybug Planococcus ficus Signoret. Proceedings of the 9th meeting of the ICVG, Kiryat Anavim 1987, 67-69.

Saldarelli, P., Dell'Orco, M., Minafra, A., 2000. Infectious cDNA clones of two grapevine viruses. Arch. Virol. 145, 397–405.

Saldarelli, P., Minafra, A., Martelli, G.P., Walter, B., 1994. Detection of grapevine leafroll-associated closterovirus III by molecular hybridization. Plant pathology 43, 91-96.

Sampol, B., Bota, J., Riera, D., Medrano, H., Flexas, J., 2003. Analysis of the virus-induced inhibition of photosynthesis in Malmsey grapevines. New Phytologist 160, 403-412.

Satyanarayana, T., Gowda, S., Ayllón, M.A., Dawson, W.O., 2003. Frameshift mutations in infectious cDNA clones of Citrus tristeza virus: A strategy to minimize the toxicity of viral sequences to Escherichia coli . Virology 313, 481-491.

Satyanarayana, T., Gowda, S., Boyko, V.P., Albiach-Marti, M.R., Mawassi, M., Navas-Castillo, J., et al, 1999. An engineered closterovirus RNA replicon and analysis of heterologous terminal sequences for replication. Proc. Natl. Acad. Sci. USA 96, 7433-7438.

Siegel, R.W., Adkins, S., Kao, C.C., 1997. Sequence-specific recognition of a subgenomic RNA promoter by a viral RNA polymerase. Proc. Natl. Acad. Sci. USA 94, 11238–11243.

Sit, T.L., Vaewhongs, A.A., Lommel, S.A., 1998. RNA-mediated transactivation of transcription from a viral RNA. Science 281, 829-832.

Tanne, E., 1988. Evidence for the transmission by mealybugs to healthy grapevines of a closter-like particle associated with grapevine leafroll disease. Phytoparasitica 16, 288.

Tanne, E., Sela, I., Klein, M., Harpaz, I., 1977. Purification and characterization of a virus associated with the grapevine leafroll disease. Phytopathology 67, 442-447.

Tidona, G.A., Darai, G., 2001. The Springer index of viruses. Springer. 235-248

Turnage, M.S., Muangsan, N., Peele, C.G., Robertson, D., 2002. Geminivirus-based vectors for gene silencing in Arabidopsis. Plant J. 30, 107–117.

Turturo, C., Saldarelli, P., Yafeng, D., Digiaro, M., Minafra, A., Savino, V., et al, 2005. Genetic variability and population structure of Grapevine leafroll-associated virus 3 isolates. J. Gen. Virol. 86, 217-224.

Vaghchhipawala, Z., Mysore, K.S., 2008. Agroinoculation: A simple procedure for systemic infection of plants with viruses. in: Foster G, editor. Plant Virology Protocols: From viral sequence to protein function. Humana press, Totowa, NJ 451, 555–562.

Valamoti, S.M., Koukouli-Chrysanthaki, C., Malamidou, D., Mangafa, M., 2007. Grape-pressings from northern Greece: The earliest wine in the Aegean?. Antiquity 81, 54.

Van den Born, E., Omelchenko, M.V., Bekkelund, A., Leihne, V., Koonin, E.V., Dolja, V.V., et al, 2008. Viral AlkB proteins repair RNA damage by oxidative demethylation. Nucleic Acids Res. 36, 5451.

Viry, M., Serghini, M.A., Hans, F., Ritzenthaler, C., Pinck, M., Pinck, L., 1993. Biologically active transcripts from cloned cDNA of genomic grapevine fanleaf nepovirus RNAs. J. Gen. Virol. 74, 169-174.

Vitushkina, M.V., Rogozin, I.B., Jelkmann, W., Koonin, E.V., Agranovsky, A.A., 2007. Completion of the mapping of transcription start sites for the five-gene block subgenomic RNAs of Beet yellows closterovirus and identification of putative subgenomic promoters. Virus Res. 128, 153-158.

Xue, B., Ling, K., Reid, C.L., Krastanova, S., Sekiya, M., Momol, E.A., et al, 1999. Transformation of five grapevine rootstocks with plant virus genes and a virE2 gene from Agrobacterium tumefaciens. In Vitro Cell. Dev. Biol.- Plant 35, 226-231.

Yamshchikov, V., Mishin, V., Cominelli, F., 2001. A new strategy in design of (+) RNA virus infectious clones enabling their stable propagation in E. coli. Virology 281, 272-280.

Yu, H.H., Wong, S.M., 1998. Synthesis of biologically active cDNA clones of cymbidium mosaic potexvirus using a population cloning strategy. Arch. Virol. 143, 1617–1620.

Zee, F., Gonsalves, D., Goheen, A., Kim, K.S., Pool, R., Lee, R.F., 1987. Cytopathology of leafroll-diseased grapevines and the purification and serology of associated closteroviruslike particles. Phytopathology 77, 1427-1434.

Chapter 3: Sequencing and analysis of the complete genome of a South African Grapevine leafroll-associated virus 3 isolate, GP18.

3.1 Introduction

Grapevine leafroll-associated virus 3 (GLRaV-3) is the type strain for the genus *Ampelovirus*, family *Closteroviridae*. (Martelli *et al.*, 2002). It is an economically important virus that is known to only infect *Vitis spp.* and that has a negative impact on the wine and table grape industries world wide. In South Africa it is the main causative agent of Grapevine Leafroll Disease (LRD) (Pietersen, 2004).

To date, there have been only two reports that claim the complete nucleotide sequence of GLRaV-3. Ling *et al.* reported the complete sequence of isolate NY-1 (AF037268) in 2004 and Engel *et al.* the complete sequence of isolate CL-766 (EU344893) in 2008. The single stranded 17919 nt RNA genome was reported to be organised into 13 ORFs (Ling *et al.*, 2004; Ling *et al.*, 1998). Comparison of the genome organisation to other closteroviruses led to the establishment of a new taxonomic group transmitted by mealybugs namely ampelovirus (Ling *et al.*, 1998). The extreme 5' and 3' ends of NY-1 was determined by Poly(A)-tailing whereas for CL-766, primers were designed to the NY-1 sequence at the extremities. The 5' untranslated region (5' UTR) for both isolates was reported to be 158 nt in length (Engel *et al.*, 2008; Ling *et al.*, 2004).

Several studies have been conducted to determine genetic variability in the GLRaV-3 genome. These studies focussed on selected genome sections typically in the RNA dependent RNA polymerase (RdRp), Heat shock protein 70 homologue (Hsp70h) and the coat protein (CP) and determined the variability by single stranded conformation polymorphism (SSCP) analysis and sequencing. Turturo *et al.*, (2005) calculated a relatively low variability in their sample set (45) collected from 14 different countries. Their results indicated that there is a single dominant variant and that mixed infections are common (Turturo *et al.*, 2005). Recently, an extensive survey was conducted by Fuchs *et al.*, (2009) to determine the genetic variation between GLRaV-3 isolates within a specific geographical area. The study determined that the genetic variation in the Hsp70h gene was relatively low with nucleotide sequence homology ranging from 92.5-98.3% between isolates from the same region; but when 25 samples of this region were compared to available sequence data from other geographical regions the variation increased, with nucleotide sequence homology of 74.1-100% (Fuchs *et al.*, 2009). A phylogenetic analysis

of these isolates was able to identify five genetic variant groups designated NY-1, C5-1, GP18, MT48-2 and NZ-1.

In South Africa two genetic variant groups were previously identified; represented by isolates 621 (GQ352631) and 623 (GQ352632) (Jooste and Goszczynski, 2005). Isolate 621 and 623 were found to be similar to NY-1 and GP18, respectively. The diversity observed in South African GLRaV-3 isolates and the lack of a complete sequence for the second genetic variant group prompted the sequencing of a full-length South African isolate.

In this chapter the sequencing of the South African GLRaV-3 isolate GP18 is described. The finding that the 5' end extended 579 nt further than previously reported was further investigated. Relevant results from two collaborative studies are also included in this chapter with additional analysis not included in the original publications (Coetzee *et al.*, 2010; Jooste *et al.*, 2010).

3.2 Material and methods

3.2.1 Source material

Grapevine material (*Vitis vinifera* cv Cabernet Sauvignon) was harvested in the Somerset West and Paarl wine-producing regions of South Africa from monitored vineyards. Vines displaying LRD symptoms for the first time were selected and tested for virus infection (Pietersen, pers comm). These vines were newly infected in a healthy vineyard, most likely through long distance transport by the natural insect vector (*Planococcus ficus*). All 18 vines (GP1-18) were tested for virus infection by ELISA and RT-PCR. ELISA was performed to test for GLRaV-1, -2 and -3 and RT-PCR for GLRaV-1, -2, -3, -5 and -9, Grapevine virus A (GVA), Grapevine virus B (GVB), Grapevine rupestris stempitting associated virus (GRSPaV) and Grapevine fanleaf virus (GFLV). Vines that tested positive for GLRaV-3 only were rooted and maintained in an insect free greenhouse facility.

3.2.2 Sequencing of isolate GP18

Double stranded RNA (dsRNA) was extracted with a cellulose (Whatman, CF-11) extraction protocol as described previously by Hu *et al.* (1990). Phloem scrapings of

wooded GLRaV-3-infected grapevine canes, isolate GP18, was used for the extraction. The quality and quantity of the dsRNA was evaluated by electrophoresis.

Primers were designed with the Oligo Explorer software v 1.2 to cover the GLRaV-3 genome (nt 1 835 – 17 905) in ten overlapping amplicons using the NY-1 sequence as reference (Table 3.1).

Amplicons were generated by RT-PCR using AMV reverse transcriptase and a high fidelity DNA polymerase. Standard molecular techniques were used to clone amplicons into pDrive cloning vector. Clones were then sequenced using the SP6 and T7 primer sites located on the vector. For the large amplicons additional primers were designed on the isolate-specific sequence and used to complete the sequence of each of the amplicons.

To determine the 5' and 3' ends of the genome, Poly(A)-tailing was performed on the dsRNA as described by Meng *et al.*, (2005) using a genome specific primer and a modified oligo(dT) primer (Table 3.1).

Additional to the Poly(A)-tailing, RLM-RACE (FirstChoice® RLM-RACE kit, Ambion, USA) was also used to determine the 5' end of the isolate GP18 genome. Nested reverse primers were designed to the 5' end of the consensus sequence determined for GP18 (Table 3.1). Total RNA was extracted from mature canes containing isolate GP18, as described by White *et al.* (2008). The quality and quantity of the extracted total RNA was determined by spectrophotometry and electrophoresis and 12μg was used per RLM-RACE reaction. Reactions were performed as suggested by the manufacturer. The amplicon was cloned and four clones sequenced. The experiment was repeated and 5 additional clones were sequenced.

3.2.3 Analysis of GP18 genome

Sequence data was analysed with BioEdit software and a consensus sequence compiled (Hall, 1999). ORF borders were determined using the NCBI ORF finder function. Proteins and protein-domains were predicted with the Pfam 22.0 domain search software (Finn *et al.*, 2006). A partial sequence of the Hsp70h for GP18 as well as sequence data available on GenBank from geographically diverse areas were phylogenetically analysed using PAUP 4.0b10 (Swofford and Sullivan, 2003). Hsp70h sequence data of isolates 621, 623 and PL-20 from the collaborative study with Jooste *et al.* (2010) was also included.

Table 3.1: Primer sets used to sequence the GLRaV-3 isolate GP18

Amplicon	Primer name	Position*	Primer sequence
	Sequencing		
1a2	ORF1a2 (1835) For	2414	GTCCAGTGACCTTCTAACTG
	ORF1a2 (4175) Rev	4754	CTCTTCCTCGACATAAGTGT
1a34	ORF1a3 (3956) For	4535	GTATTACCGGGACTTTGAC
	ORF1a4 (7172) Rev	7751	TCAAACGTGACGAGGTTAC
1b	LR3 ORF1b Rev	9173	GATGGCACGCCTAAGAGAAG
	LR3 ORF1b For	7156	GGTCGTTGAAGGGAAAAGG
2+3	LR3 ORF2+3 Rev	10717	GGGGAAAAGCAGATTGTGC
	LR3 ORF2+3 For	9081	TTCTGTGCCTCGGTTCTTC
4	LR3 ORF4 Rev	12551	GAGAGCGTGGTAGTCAATCC
	LR3 ORF4_2 For	10458	GCTGAGCGAAGGTGATATCG
5	LR3 ORF5 Rev	13924	GCATCCCCCACTCTAACTC
	LR3 ORF5 For	12039	GGAACGGTGTCTGTTATCG
6	LR3 ORF6 Rev	15420	CCAGGGTAGAGGTTCTGATC
	LR3 ORF6 For	13615	GGGACGTTAGCGTATGACAC
7	LR3 ORF7 Rev	16427	GTCTCGAAACGACTTTACCG
	LR3 ORF7 For	14656	GTCCGACGTACGATCTGTTC
8+9	LR3 ORF8+9 Rev	17501	AGCGCGTCGTATCATCAAC
	LR3 ORF8+9 For	16018	CACTGTGCGATCCTTCATG
10-12	LR3 ORF10-12 Rev	18484	TCGATAAGTTAGCCTCGTAA
	LR3 ORF10-12 For	17097	CGATAGAAACAGCCAGAAGT
	Poly(A) tailing		
PolyA 5'	ORF1a1 (2020) Rev	2599	GGGAACATAAGAGCTATGTC
PolyA 3'	LR3 17743 For	17097	CGAGGTAAGATGACTAAACT
	Spanning RT-PCR		
Spanning RT-PCR	LR3 xtra 5' END For	159	GCTGTTGTTAGTAGTTTCTGTTGT
	LR3 ORF1a 365 Rev	944	CGTCCGCTTCACCCCTTTGG
	5' RLM-RACE		
5' RLM-RACE	LR3 ORF1a 450 Rev	1029	TCTTACCATCCCCTCTCAAT
	LR3 ORF1a 365 Rev	944	CGTCCGCTTCACCCCTTTGG

*5' Binding position relative to the GP18 sequence

3.2.4 Metagenomic sequencing

In a collaborative project, dsRNA from a diseased vineyard was subjected to metagenomic sequencing. In brief, 50ng of dsRNA were isolated from 44 randomly selected vines in a virus-infected vineyard and subjected to metagenomic sequencing using the Illumina Genome Analyzer II. Paired-end sequence data was assembled into scaffolds using the short read assembler Velvet 0.7.31 (Zerbino and Birney, 2008). Scaffolds were identified

and assigned using BLAST analysis. Re-assemblies were also performed with Mapping and Assembly with Quality (MAQ) assembler v 0.7.1 (Li *et al.*, 2008) using the Easyrun command. For a detailed description of the methods refer to Coetzee *et al.* (2010).

Additional *de novo* assemblies were performed experimenting with different parameters in order to improve the length of the GLRaV-3-specific scaffolds. See Table 3.2 for the parameter settings for the various Velvet assembly runs. The scaffolds generated from the different runs were assembled into contigs with CAP3 contig assembly software using the default parameters (Huang and Madan, 1999). The NCBI BLAST analysis to the non-redundant nucleotide database was used to identify contigs that showed high sequence similarity to the GP18 isolate.

Table 3.2: Velvet run parameters

Run #	k-mer	Paired end	sd	cov_cutoff	exp_cov	Number of scaffolds	Largest scaffold
35762	19	N	-	10	100	6849	3135(NODE_3466)
35139	19	N	-	100	500	1162	3095(NODE_153)
35188	19	Y	20	100	500	967	3095(NODE_147)
35764	21	N	-	10	100	5683	1534(NODE_3703)
35146	21	N	-	100	1000	806	7495(NODE_101)
35331	21	Y	20	100	1000	663	8593(NODE_325)
35766	23	N	-	10	100	4751	1808(NODE_3290)
35150	23	N	-	100	3000	579	8624(NODE_100)
35336	23	Y	20	100	3000	434	8624(NODE_91)
35775	25	N	-	10	100	4083	1808(NODE_2698)
35172	25	N	-	100	4000	466	7124(NODE_92)
35353	25	Y	20	100	4000	339	11987(NODE_796)
35777	27	N	-	10	100	3372	1808(NODE_2113)
35174	27	N	-	100	400	408	7500(NODE_82)
35356	27	Y	20	100	4000	301	8620(NODE_211)
35779	29	N	-	10	100	2729	1540(NODE_472)
35176	29	N	-	100	6000	331	7147(NODE_81)
35358	29	Y	20	100	6000	249	18635(NODE_40)
35782	31	N	-	10	100	2484	2756(NODE_399)
35178	31	N	-	100	7000	246	10293(NODE_31)
35360	31	N	20	100	7000	179	18645(NODE_247)
35494	33	N	20	-	50	3207	1500(NODE_19016)
35784	33	N	-	10	100	2484	2756(NODE_399)
35183	33	N	-	100	7000	246	10293(NODE_31)
35364	33	N	20	100	7000	179	18645(NODE_247)
35366	45	N	20	100	7000	179	18645(NODE_247)
						43376	

3.2.5 Further investigation into the occurrence of the extended 5' UTR

RLM-RACE was also used to determine the 5' ends of GLRaV-3, Grapevine leafroll-associated virus 2 (GLRaV-2) and Grapevine rupestris stempitting-associated virus (GRSPaV) within the same reaction using sample material from a vine infected with all three viruses. The primers were designed as specified by the manufacturer and reactions performed as prescribed.

Additionally, a "spanning RT-PCR" was developed that spanned the 5' end of the NY-1 sequence (nts 159 - 944 of GP18, Table 3.1) to prove that the extended sequence observed is truly of GLRaV-3 genomic origin. Four samples (K1, K2, GP16 and KK1) from different origins were screened and the 786 nt amplicon generated was cloned and sequenced.

In another collaborative project the 5' ends of three genetic variants of GLRaV-3 were determined and analysed (Jooste *et al.*, 2010). The 5' ends of the isolates 621, 623 and PL20 (GQ352333) were determined using RLM-RACE. For a detailed description of the methods used, refer to Jooste *et al.* (2010). A multiple sequence alignment of the 5' ends of these three groups was submitted to the RNAalifold server for secondary structure prediction with default parameters. RNAalifold is based on the algorithm developed by Zuker and Stiegler (1981).

MAQ reassemblies were performed using the 5' UTRs determined for the three genetic variants as reference sequences.

3.3 Results

3.3.1 Source material

The ELISA was able to detect GLRaV-3 in all the samples. Samples GP(1, 2, 3, 4, 8, 10, 12, 13, 15, 17 and 18) had relatively high titres compared to samples GP(6, 7, 9, 11, 14,16). RT-PCR was able to detect GLRaV-3 in GP(3, 5, 6, 8, 9, 10, 12, 14, 16, 18) and GVA in GP3. All the samples tested negative for GLRaV-1, -2, -5 and -9, GVB, GRSPaV and GFL. Sample GP18 was selected as source material for sequencing and designated isolate GP18.

3.3.2 Sequencing of isolate GP18

In figure 3.1A a schematic representation of the complete genome sequence of isolate GP18 is shown. In figure 3.1B the amplicons used to compile this sequence is depicted.

The 10 overlapping amplicons spanning most of the genome were cloned and sequenced and a consensus sequence constructed that excluded the primer sequences.

Poly(A)-tailing was used to determine the 5' and 3' ends of the genome of isolate GP18 and the 3' end was found to be similar to that of NY-1. However, the 5' end was consistently found to be 49 nt short of the 5' end found for NY-1. By adjusting the PCR conditions to be sub-optimal (Ta = 48ºC), a range of amplicons were generated. Sequencing of these amplicons not only showed sequence that extended beyond the 49 nt short sequence found with the optimised reaction, but also novel sequence that extended beyond the 5' end of the NY-1 sequence. The poly(A) polymerase negative control also generated a similar range of amplicons. Sequencing of these amplicons confirmed that they are similar to the sub-optimal Poly(A)-tailing amplicons. A multiple sequence alignment of the amplicon sequences also indicate that just upstream of these amplicons the GLRaV-3 sequence contains a high concentration of thymines (Figure 3.2).

Hundred and seventy micrograms of total RNA was extracted from 2g of phloem scrapings and had an $A_{260/280}$ ratio of 2.14. Amplicons generated from the RLM-RACE reactions were sequenced and demonstrated that the 5' UTR of GP18 extended 579 nt further than the 5' end of NY-1.

3.3.3 Analysis of GP18 genome

Analysis of the GP18 genome showed the same ORF borders as was found for NY-1 (Ling *et al.*, 1998). The only exception being the greater overlap of ORF1a and ORF1b. The assignment of proteins and protein domains were also the same as for NY-1 (Figure 3.1A and Table 3.3). The L-Pro domain was identified manually by multiple sequence alignment as Pfam was unable to do so (Figure 3.3). Phylogenetic analysis of the Hsp70h showed three genetic variant groups with bootstap values higher than 80, and with the GP18 isolate in a separate group to NY-1 (Figure 3.4).
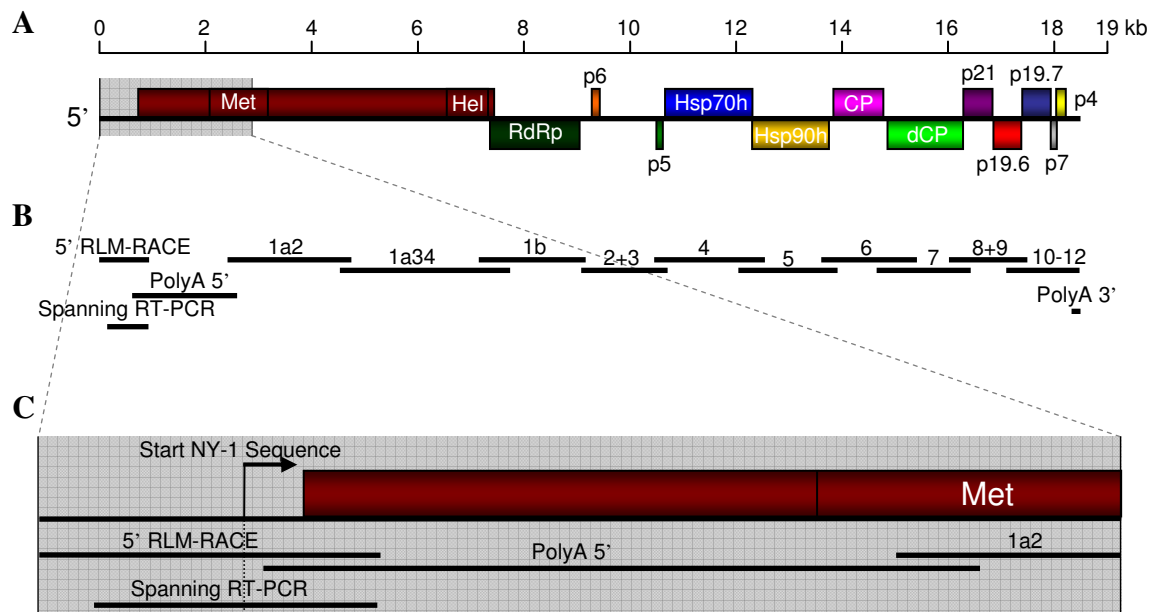
Figure 3.1: A) Schematic representation of the genome organisation of GLRaV-3 GP18 (drawn to scale). B) Lines indicate the regions cloned and sequenced by different techniques. Lines 1a2, 1a34, 1b, 2+3, 4, 5, 6, 7, 8+9, 10-12 are representative of the clones generated by RT-PCR to sequence the majority of the genome. "PolyA 5'" and "PolyA 3'" represent the clones generated using poly A tailing and "5' RLM-RACE" show the area amplified using RLM-RACE. "Spanning RT-PCR" represents the area generated by RT-PCR to indicate that other isolates also has the extended 5' UTR. C) Enlargement of the 5' area also shows the start of the NY-1 sequence compared to the GP18 sequence.

Table 3.3. Position of untranslated regions (UTRs) and open reading frames (ORFs) on the GLRaV-3, isolate GP18 sequence and the percentage nucleotide and amino acid sequence identity to isolate NY-1 (AF037268).

UTR or ORF	Position on the genome	Size (nt)	% Sequence identity to NY-1 isolate	
			nt	Aa
5' UTR	1 - 737	737	80	-
1a (Met/Hel)	738 - 7451	6714	94	95
1b (RdRp)	7369 - 9066	1698	95	97
2 (p6)	9287 - 9442	156	90	80
3 (p5)	10509 - 10646	138	93	98
4 (Hsp70)	10665 - 12314	1650	94	97
5 (Hsp90)	12307 - 13758	1452	92	93
6 (CP)	13848 - 14789	942	92	95
7 (dCP)	14852 - 16285	1434	92	90
8 (p21)	16296 - 16853	558	93	97
9 (p19.6)	16850 - 17383	534	91	89
10 (p19.7)	17390 - 17929	540	90	86
11 (p4)	17932 - 18042	111	90	89
12 (p7)	18039 - 18221	183	97	97
3' UTR	18222 - 18498	277	97	-
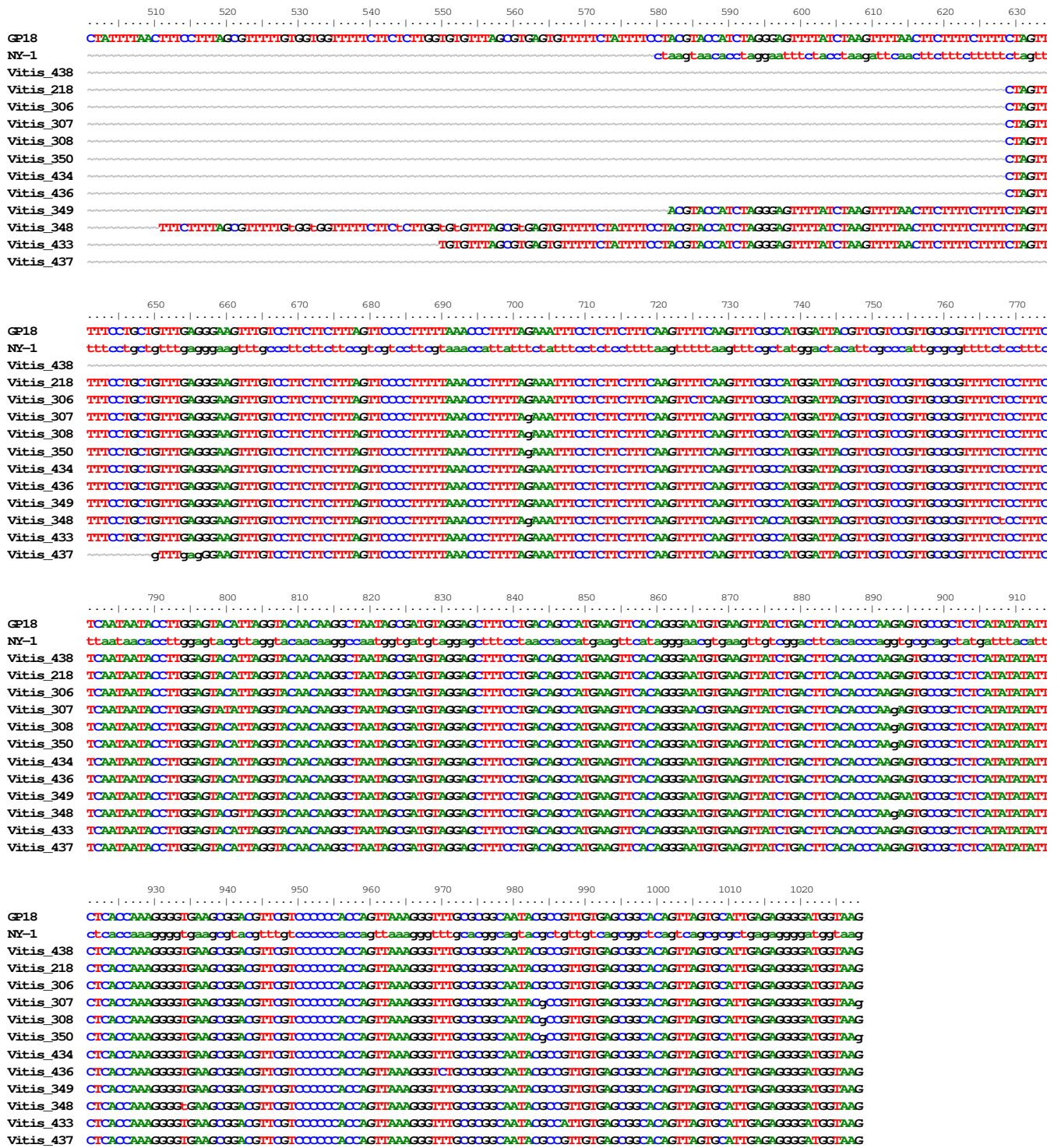
Figure 3.2: Multiple alignment of sequenced amplicons generated in Poly(A) tailing experiments using the oligo(dT) primer and LR3 ORF1a 365 Rev. Sequences of GP18 and NY-1 is depicted in top two lines. Sequences 218, 306-308 were generated from cloned amplicons from optimised Poly(A)-tailing RT-PCR reactions. Sequences 348-350 are sequence data from amplicons sequenced from suboptimal RT-PCR conditions. Sequences 433(Neg), 434(Neg), 436(Pos), 437(Pos) and 438(Pos) were generated in the comparative RT-PCR on Poly(A)-tailing positive and negative reactions.
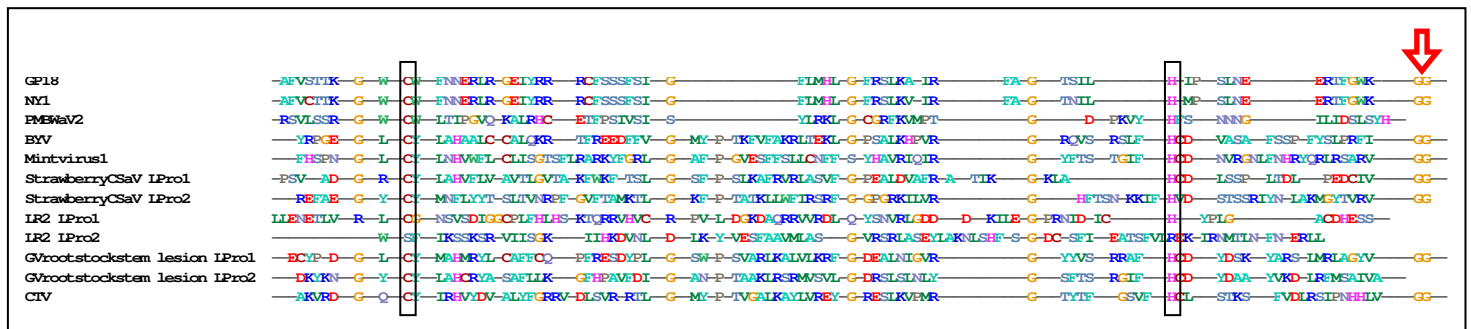
Figure 3.3: Multiple amino acid sequence alignment of Leader Protease domains of different closteroviruses. The predicted catalytic amino acids of the papain-like protease are blocked and the predicted cleavage site is indicated by the arrow.

3.3.4 Metagenomic sequencing

The metagenomic sequencing yielded 837MB of data from $>1.9 \times 10^7$ pair-end sequence reads. *De novo* assembly of the short reads yielded 449 scaffolds using the following parameters: hash length of 23, coverage cut-off of 50, expected coverage of 1,000 and a minimum scaffold length of 100. BLAST analysis of these scaffolds showed that 124 scaffolds were associated with GLRaV-3. Sixty-six of these scaffolds aligned preferentially across the genome of isolate GP18. The largest scaffold found to be similar to GP18, was node 611 (1765 nt). MAQ re-assembly with the GP18 sequence as reference determined that 4 242 321 reads can be assembled to an average depth of 10 008 nt with a 100% genome coverage. Plotting the number of sequence reads assembled by MAQ on the GP18 genome demonstrates that the depth of coverage is variable across the genome with a maximum depth of >70 000 reads (Figure 3.5A).

The additional *de novo* assemblies that were performed, yielded a combined total of 43 376 scaffolds. A large portion of these scaffolds were redundant and reduced by creating consensus contigs with the CAP3 contig assembly software. The scaffolds were reduced to 4 085 contigs, ranging from 45 to 18 653 nt in length. The largest contig that aligned with high homology to the GP18 sequence was Contig-3843 (18 635 nt). This contig spans the GP18 sequence from nt position 21 to nt position 18 472 (99% coverage), with an identity homology of 97.1%. The contig contains 72 nt insertions, 4 nt deletions and an incomplete duplication of 114 nt spanning from position 14 523 - 14 636 in the contig. Compared to the GP18 sequence, the duplication was homologous to nucleotides 14 501 – 14 614, inserted between nucleotides 14613 and 14614. MAQ reassembly with Contig-3843 as reference sequence was able to assemble 3 131 958 reads to an average depth of 6 842 nt with a 99% reference sequence coverage (Figure 3.5B). MAQ reassembly was also
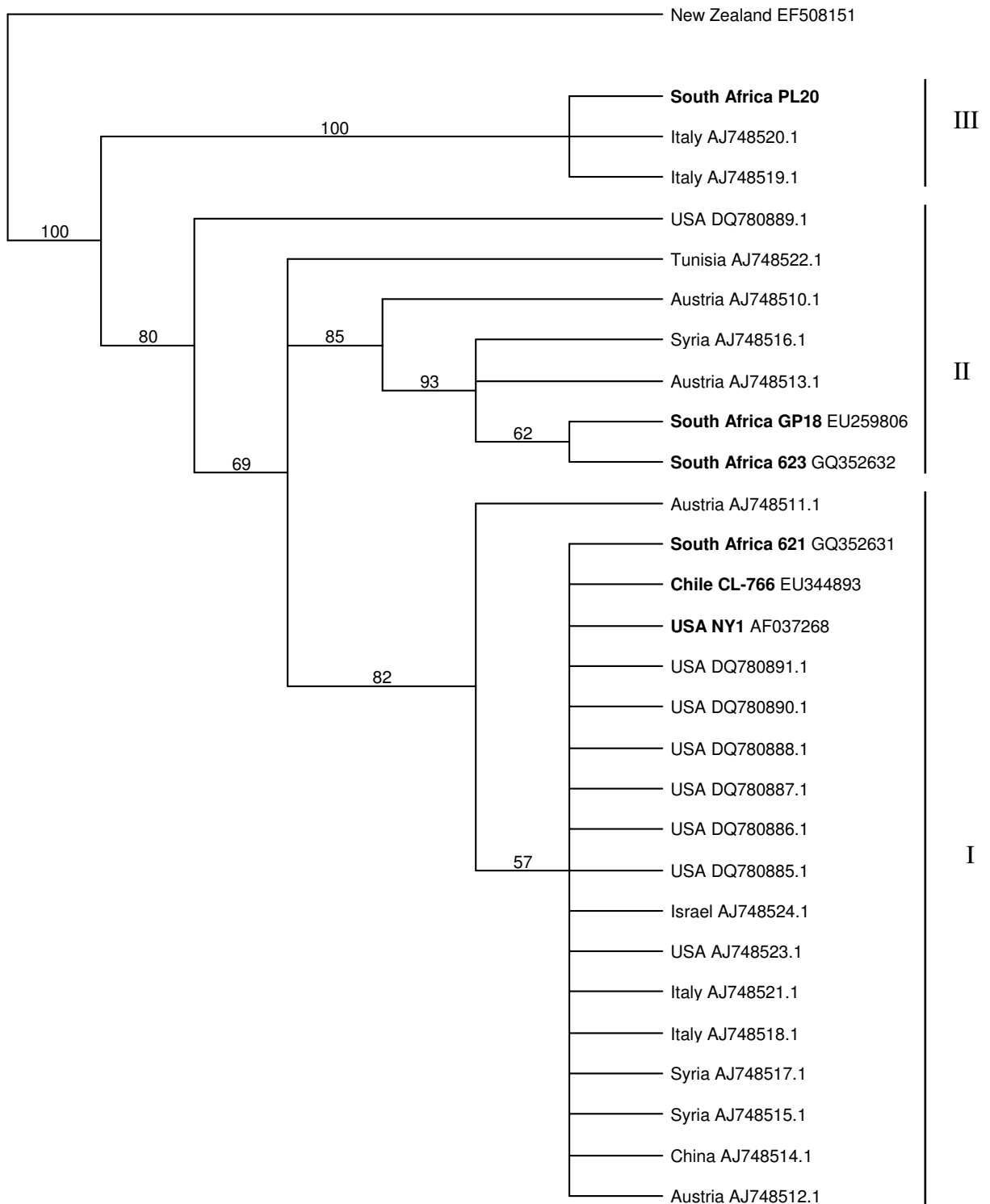
Figure 3.4: Phylogenetic tree constructed using partial sequence data of the Hsp70h gene (ORF4). The three genetic variant groups identified in the Jooste *et al.* (2010) are indicated with roman numerals. Isolates discussed in this chapter are highlighted in bold. Phylogenetic analysis of the aligned partial Hsp70h sequence matrix was performed using PAUP 4.0b10. A heuristic search (1,000 replicates) using TBR branch swapping with all characters weighted equally was performed to search for the shortest possible trees from both data matrices. The CI and RI were 0.906 and 0.786 respectively. A bootstrap analysis (1,000 replicates) using TBR branch swapping was performed to establish clade support. Branches with bootstrap values C75% were considered well supported, whilst values between 75% and 50% were considered moderately supported. Values below 50% were considered weakly supported and, in line with other phylogenetic analyses, were not indicated on phylograms.
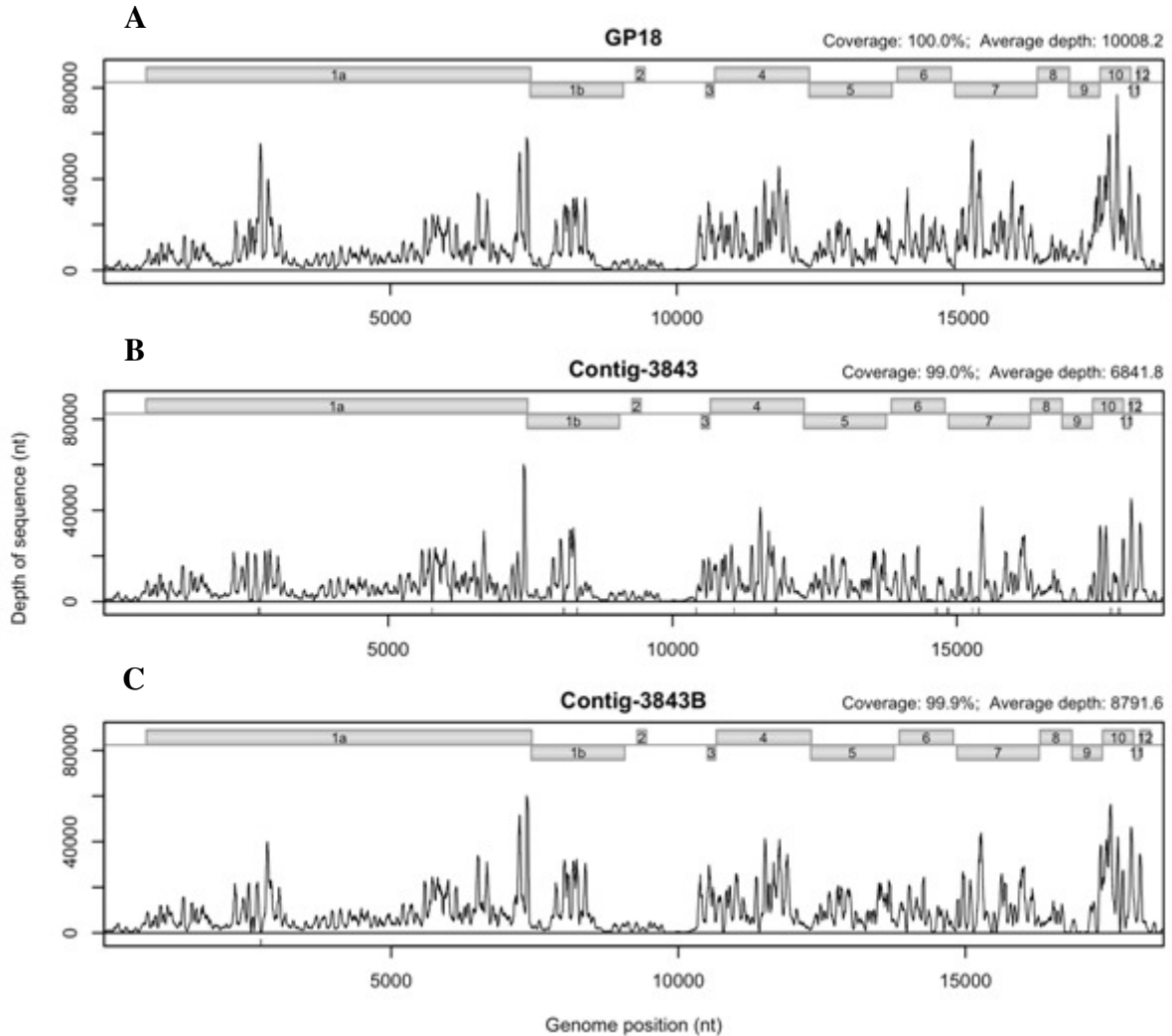
Figure 3.5: MAQ re-assembly reads as plotted different reference sequences. A) Complete GP18 genome B) Contig3843 generated from CAP3 contig assembly and C) Contig3843B a "repaired" Contig3843.

performed on a "repaired" Contig-3843 (Contig-3843B), where all the insertions and the duplication were removed, as reference sequence. In this reassembly, 3 731 803 reads were assembled to an average depth of 8 792 nt covering 99.9% of the reference sequence (Figure 3.5C).

3.3.5 Further investigation into the occurrence of the extended 5' UTR

RLM-RACE data from the vine infected by multiple viruses was able to determine the 5' ends of GLRaV-3, GLRaV-2 and GRSPV in the same reaction. The sequence data from the amplicons showed identical 5' ends for GLRaV-2 (AY881628) and GRSPV (AY881627) as were previously published. However, for GLRaV-3 the sequence confirmed the extended 5' UTR identified for isolate GP18.

Sequencing results from the 786 nt amplicon generated from the "spanning RT-PCR" using four additional samples (K1, K2, GP16 and KK1) clearly showed the extended 5' UTR that continues beyond the 5' end reported for NY-1 (Figure 3.1B).

The 5' ends of three additional South African genetic variants 621, 623 and PL20 were successfully determined using RLM-RACE. The sequence data demonstrated that the isolates contain extended 5' UTRs similar to isolate GP18. The 5' UTRs demonstrated great diversity among the variant groups. Sequencing of additional samples from each of the genetic variants groups indicated that the 5' UTRs are strictly conserved within a particular variant group (Figure 3.6). A complex consensus secondary structure for these 5' UTRs was predicted by RNAalifold with a minimum free energy for the structure of -163.54 kcal/mol (Figure 3.7).
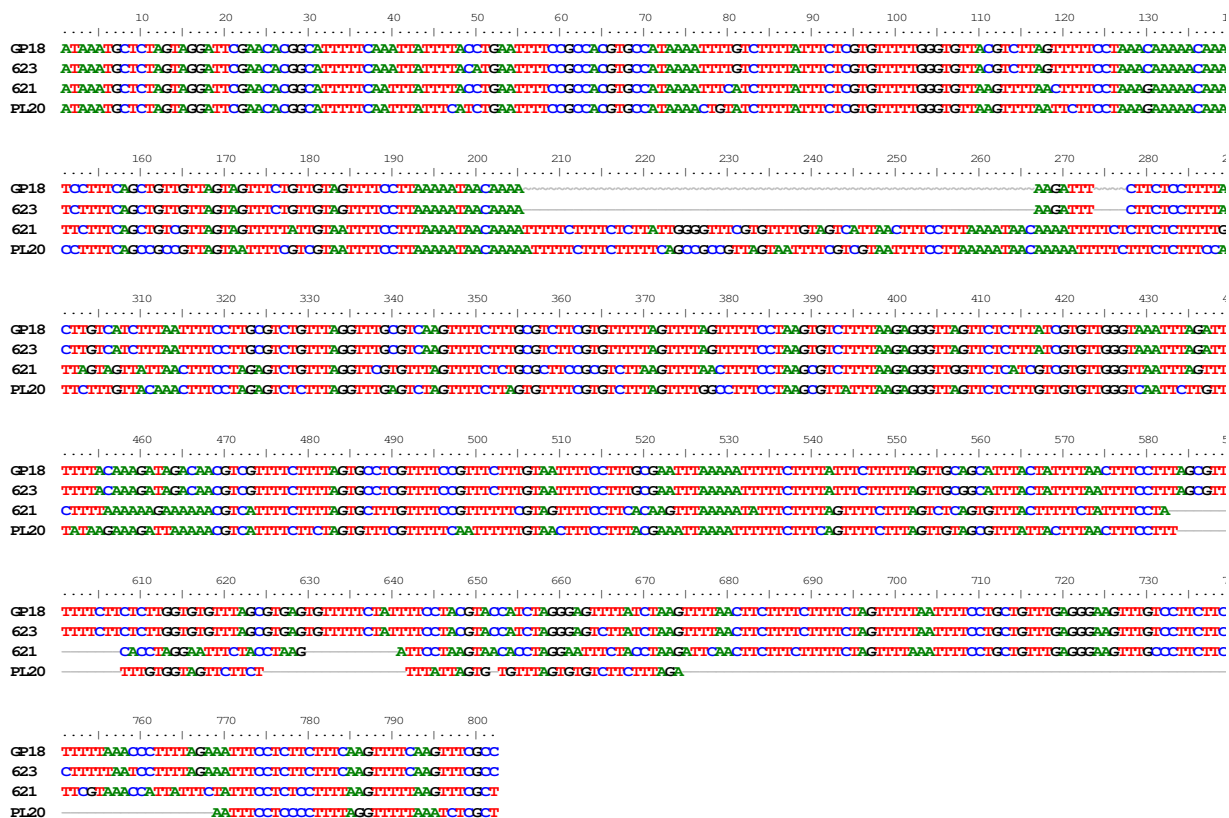
Figure 3.6: Multiple alignment of the 5' UTR sequences of the 4 complete sequences GP18, 623 621 and PL20. The genetic variability between the groups are clear with large deletions/insertions when compared.
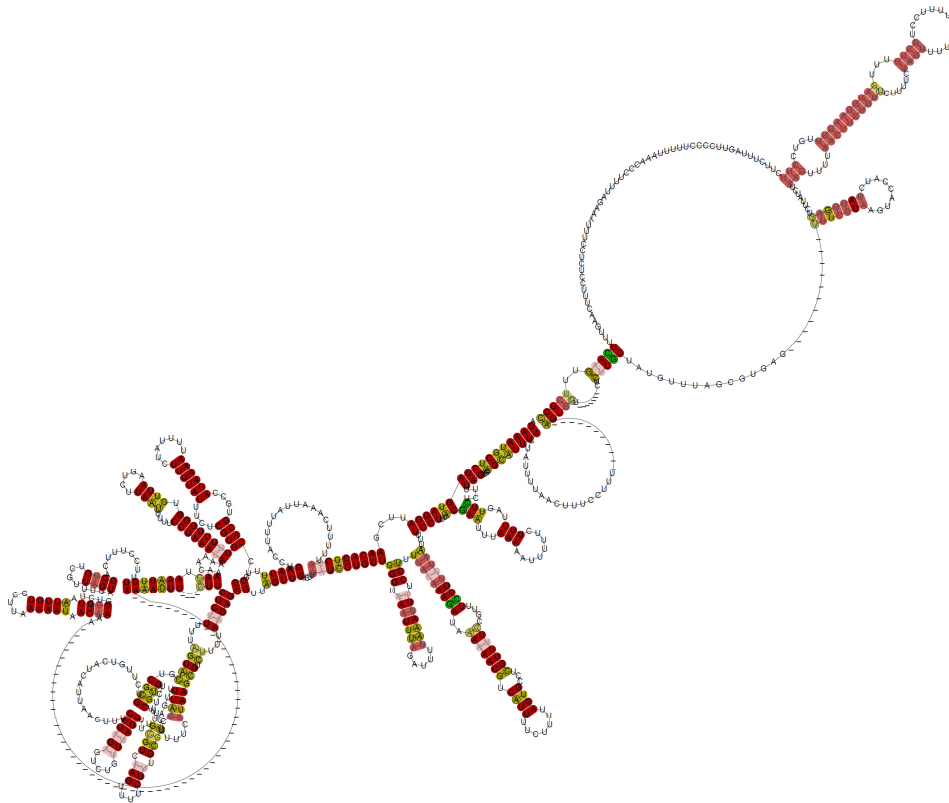
Figure 3.7: RNA secondary structure of the 5' UTR of GLRaV-3 as predicted by RNAalifold Vienna using a multiple alignment of the 5' UTRs of isolates GP18, 621 and PL-20. The structure was predicted to have a free energy of -163.54 kcal/mol.

3.4 Discussion

The complete genome of GLRaV-3 isolate GP18 was sequenced and found to be longer at the 5' end, extending the 5' UTR 579 nt beyond that reported for the NY-1 isolate. The extended 5' UTR was also demonstrated to be present in all other isolates of GLRaV-3 tested.

Isolate GP18 was sequenced by employing three RT-PCR based techniques. RT-PCR for the bulk (87%) of the genome and two techniques for the 5' and 3' ends of the genome, Poly(A)-tailing and RLM-RACE. A consensus sequencing spanning nucleotides 1 835 – 17 905 of the NY-1 sequence was constructed from the sequence data of 10 overlapping cloned amplicons generated with RT-PCR.

Poly(A)-tailing on dsRNA was successful in determining the 3' end of the GP18 genome and was found to be similar to NY-1. This technique was however not able to determine the 5' end of the genome. Using optimised PCR conditions, the 5' end was consistently

determined to be at position 50 of the NY-1 sequence. By reducing the annealing temperature by 7ºC it was possible to generate a range of amplicons that indicated the 5' end extended beyond the 5' end of the NY-1 sequence. The poly(A)-polymerase negative control under these sub-optimal conditions also yielded the same amplicons indicating that these fragments can be generated independently of poly-adenylation by yeast poly(A)-polymerase. A multiple sequence alignment of these fragments revealed that the fragments are preceded by a high number of thymine/uracil nucleotides that seem to serve as binding sites for the oligo(dT) primer on the complementary negative strand during RT-PCR. This result highlights the potential hazard of using a single, sequence-dependant technique.

The 5' end of the GP18 genome was determined using RLM-RACE on total RNA. The amplicon generated from the RLM-RACE reaction was significantly larger than expected but it was cloned and four clones sequenced. The reaction was repeated and an additional five clones sequenced. Sequence data from all nine clones confirmed the same sequence that contains the first 365 nucleotides of the NY-1 sequence and an additional 579 nt upstream of the 5' end reported for the NY-1 sequence.

The complete genome sequence of isolate GP18 was found to be 18498 nucleotides long with a 93% sequence identity with the NY-1 sequence over nucleotides 580-18498. The putative function for some of the ORFs was predicted using the Pfam 22.0, domain search software. In table 3.3 the nucleotide and amino acid sequence identities for each of the ORFs between GP18 and NY-1 are listed. The predicted function and domains of GP18 are indicated and found to be similar to NY-1 (Ling *et al.*, 2004).

The GP18 sequence has an extended 5' UTR compared to NY-1, which is 737 nt long with an unusual sequence composition. It has an adenine/uracil content of 68.4% with a high uracil content of 48.5%. The only other member of the genus *Ampelovirus* that has been completely sequenced is Little cherry virus 2 (LChV-2, AF531505), which was found to contain a region of 539 nt 5' of the ORF1a. The 5' region of LChV-2 and the 5' UTR of GLRaV-3 GP18 are much larger than the 5' UTRs of other members of the *Closteroviridae* family, GLRaV-2, closterovirus (AY881628) 105 nt, BYV, closterovirus (AF190581) 107 nt, CTV, closterovirus (DQ272579) 107 nt and LIYV, crinivirus (NC_003617) 97 nt. The 5' UTR of GLRaV-3 GP18 contains two small ambisense ORFs with no similarity to the LChV-2 ORF0 and it is likely that these ORFs are not expressed (Rott and Jelkmann, 2005). This leaves one to speculate about the function of such a large 5' UTR, and which warrants further investigation.

The domains predicted by the Pfam software for the replicase (ORF1a) were similar to those previously described for the NY-1 isolate; with two differences. The software was able to detect the presence of an AlkB (2OG-Fe(II) oxygenase) domain (aa 1 938 – 2 199) in ORF 1a not previously described, but was unable to detect a p-protease (L-Pro) domain. This analysis was also performed on NY-1 with the same result. The presence of the AlkB domain is not unique to GP18, but present in all GLRaV-3 sequences tested. The presence of the L-Pro domain could not be determined by this software, possibly due to a lack of sequence information in the database. The position of the L-Pro domain and possible critical amino acids were determined by multiple sequence alignment and found to be similar to NY-1 (Figure 3.3). A further significant difference between the sequence of GP18 and NY-1 is the 82 nt overlap between ORF1a and ORF1b. In the GP18 sequence ORF1b can still be expressed as a +1 frameshift. Furthermore, analysis of the expression of ORF1b needs to be evaluated to determine the mechanism by which this protein is transcribed.

Similar to the phylogenetic analysis by Fuchs *et al.* (2009) and Jooste *et al.* (2010) using a partial sequence of the Hsp70h gene, three main genetic variant groups were found. GP18 was grouped with genetic variant 623 which was found by Jooste *et al.*(2010) to be the most abundant genetic variant group in South Africa. This group is separate to the group that contains isolates NY-1 and 621 and the group containing isolate PL20.

Analysis of short read sequence data generated by metagenomic sequencing of a diseased vineyard yielded 449 scaffolds with *de novo* assembly. GLRaV-3 was identified as the most abundant virus in the sample as 124 scaffolds aligned with GLRaV-3 isolates, accounting for 58.5% of the short sequence read data. Isolate GP18 was found to be the most homologous to 66 of these scaffolds accounting for 56% of the short sequence read data associated with GLRaV-3, identifying a GP18-like GLRaV-3 as the most abundant virus in this environmental sample. MAQ reassemblies using available full-length GLRaV-3 sequences as reference sequences confirmed the dominance of a GP18-like GLRaV-3 in the sample. MAQ reassembler was able to align 4 242 321 reads on the reference sequence with 100% genome coverage to an average depth of 10 008 nt. The MAQ reassembly was able to generate a full-length GP18-like consensus sequence from the environmental sample that confirmed the presence of the 5' UTR observed for GP18. It was observed, by plotting the number of reads on the reference sequence, that there is great variability in the

depth of coverage across the reference sequence. This could be due to the incomplete dissociation of the dsRNA used as starting material or because the variant dominant in the sample varies at the positions with low coverage in the GP18 reference. See Coetzee *et al.* (2010) for a more complete discussion of this data set and other viruses that were identified.

From figure 3.5, where the number of reads are plotted on the reference sequence, it is clear that although the average depth is very high, there is great variability in the depth that might explain the lack of a full-length scaffold for GLRaV-3. In an attempt to generate a full-length GLRaV-3 scaffold from the original data set of short sequence reads, additional *de novo* assemblies using different parameters were performed. Some of these assemblies were able to generate much larger scaffolds than those using the original parameter settings, but, interestingly, it was found that there is also a loss of data for some of the other virus sequences by using different parameter settings. The parameter settings in the original analysis were a good compromise and yielded the most usable data from a single analysis. None of the *de novo* assemblies were able to generate a full-length scaffold for GLRaV-3. Combined, all the *de novo* assemblies yielded 43376 scaffolds with a high degree of redundancy. By compiling all the *de novo* scaffolds into a contig assembly, it was possible to reduce the redundancy and construct larger contigs. Compiling *de novo* assembled scaffolds into contigs results in a loss of all quantitative and qualitative information linked to each of the individual scaffolds. Some indication of the quantitative and qualitative information associated with each of the contigs can be determined by using MAQ reassembly with the contigs that was generated as reference sequence. Contig-3843 was found to be a nearly complete GLRaV-3 contig with 97% homology to isolate GP18. This contig however, contained many assembly errors such as insertions, deletions and an 114 nt duplication when aligned to the GP18 sequence. When MAQ reassembly was performed on Contig-3843, the number of reads, coverage and depth of coverage was significantly lower than when isolate GP18 was used as reference sequence. Contig-3843 was also "repaired" by removing all the insertions and the duplication to create Contig-3843B and was subjected to MAQ reassembly. The number of reads, coverage and depth of coverage did increase, but was still lower than when GP18 was used. It is expected that a contig assembled from the sample sequence data would have a 100% coverage and a higher number of reads aligning to it when compared to a sequence from a different isolate. This discrepancy might be due to the composition of the sample that most likely contains

several closely related GLRaV-3 variants that would make *de novo* assembly difficult. It is important to note that the sequence of these contigs do not represent an isolate in the environmental sample but are rather a representation of the most dominant sequences found. The *de novo* assemblies from this data most likely construct chimeric contigs representing a range of isolates that are closely related and possibly quasispecies. Even though this chimeric contig cannot be directly compared to the GP18 isolate it still provides us with a GP18-like sequence that is more closely related to GP18 than NY-1 is to GP18 and more importantly it corroborates the existence of the extended 5' UTR.

The unexpected result of the extended 5' UTR warranted further investigation and confirmation. To exclude the possibility that the extended 5' UTR might be due to an experimental artefact, RLM-RACE was performed on a vine that was infected by three different viruses. Sequence data from the amplicons generated confirmed the 5' ends of GLRaV-2 and GRSPaV as published, as well as the 5' end determined for GP18 and reported here. This result excluded the possibility that the extended 5' UTR is due to an RLM-RACE experimental artefact.

In an attempt to determine if all isolates of GLRaV-3 contain this extended 5' UTR an RT-PCR was designed that amplified a portion of the extended 5' UTR found for GP18 and the first 365 nt of the known sequence of NY-1. Sequencing results from the amplicons generated from the "spanning RT-PCR" using four additional samples clearly showed the extended 5' UTR that continues beyond the 5' end reported for NY-1. Using this primer set, additional samples have been tested for the presence of this extended 5' UTR and all were found to contain the extension. Some of these samples were also sequenced and found to be highly conserved (data not shown).

The presence of such an unusual 5' UTR was further supported by the collaborative project with Jooste *et al.* (2010) in which the 5' ends of three genetic variant isolates were determined by RLM-RACE. All three variants were found to contain 5' UTRs similar in length to GP18, but with great sequence diversity.

Using these 5' UTR sequences as reference sequences, it was determined that the environmental sample contained the GP18 and possibly the PL20 variants, with GP18

being the dominant variant with 100% coverage and the highest number of reads aligning to it.

In Figure 3.6 the multiple alignment of the 5' UTRs of the three genetic groups clearly demonstrate the areas where some variants have insertions or deletions compared to the other variants, flanked by regions that are highly conserved. Additional sequence data from samples in each of the genetic variant groups show that the 5' UTRs are highly conserved within a genetic group. A multiple alignment of the three genetic variant 5' UTRs were used to predict a possible secondary structure for the 5' UTR of GLRaV-3. The complex secondary structure predicted had multiple conserved stems and loops and a structure free energy of -163.54 kCal/mol (Figure 3.7). Interpretation of this highly complex structure is not possible without an analogous example or biological data. Currently it is only possible to speculate on the function of such an unusually large 5' UTR for a plant virus, however it's possible role in replication cannot be discounted. Even though GLRaV-3 is a capped virus, the possibility exist that this 5' UTR plays a role in sequestering ribosomes, independently from the cap, similar to the IRES sequence found in the 5' UTRs of picorna viruses (insect virus). Alternatively, it might play a role in insect transmission. The 5' UTR might also be involved in the binding of the tail assembly proteins. In the related closterovirus, BYV and GLRaV-2 the tail assembly was found to be influenced by the sequence of the 5' terminal sequence of the replicase (Alzhanova *et al.*, 2007; Liu *et al.*, 2009). Until experimental data is available all these possibilities are only conjecture.

This chapter describes the sequencing of the first truly complete GLRaV-3 genome. The South African isolate GP18 was sequenced and found to contain an unusually large 5' UTR. Further investigation and two collaborate projects unequivocally proved that the 5' UTR is not only present in the GP18 isolate, but possibly in all GLRaV-3 isolates. The function of this large 5' UTR is however unknown and needs to be investigated further.

3.5 Acknowledgements

3.6 References

Alzhanova, D.V., Prokhnevsky, A.I., Peremyslov, V.V., Dolja, V.V., 2007. Virion tails of Beet yellows virus: Coordinated assembly by three structural proteins. Virology 359, 220-226.

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

Finn, R.D., Mistry, J., Schuster-Böckler, B., Griffiths-Jones, S., Hollich, V., Lassmann, T., et al, 2006. Pfam: Clans, web tools and services. Nucleic Acid Res. Database Issue 34, D247-D51.

Fuchs, M., Martinson, T.E., Loeb, G.M., Hoch, H.C., 2009. Survey for the three major leafroll disease-associated viruses in Finger Lakes vineyards in New York. Plant Dis. 93, 395-401.

Hall, T., 1999. BioEdit: A user-friendly biological sequence alignment editor and analysis program for windows 95/98/NT. Nucl Acids Symp Ser 41, 95-98.

Hu, J.S., Gonsalves, D., Teliz, D., 1990. Characterisation of closterovirus-like particles associated with grapevine leafroll disease. J. Phytopathology 128, 1-14.

Huang, X., Madan, A., 1999. CAP3: A DNA sequence assembly program. Genome Res. 9, 868.

Jooste, A.E.C., Maree, H.J., Bellstedt, D.U., Goszczynski, D.E., Pietersen, G., Burger, J.T., 2010. Three Grapevine leafroll-associated virus 3 (GLRaV-3) variants identified from South African vineyards show high variability in their 5'UTR. DOI 10.1007/s00705-010-0793-y.

Jooste, A., Goszczynski, D., 2005. Single-strand conformation polymorphism (SSCP), cloning and sequencing reveals two major groups of divergent molecular variants of Grapevine leafroll-associated virus 3 (GLRaV-3). Vitis 44, 39-43.

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

Ling, K.S., Zhu, H.Y., Gonsalves, D., 2004. Complete nucleotide sequence and genome organisation of Grapevine leafroll-associated virus 3, type member of the genus ampelovirus. J.Gen.Virol. 85, 2099-2102.

Ling, K., Zhu, H., Drong, R., Slightom, J., McFerson, J., Gonsalves, D., 1998. Nucleotide sequence of the 3'-terminal two-thirds of the Grapevine leafroll-associated virus 3 genome reveals a typical monopartite closterovirus. J.Gen.Virol. 79, 1299-1307.

Liu, Y., Peremyslov, V.V., Medina, V., Dolja, V.V., 2009. Tandem leader proteases of Grapevine leafroll-associated virus 2: Host-specific functions in the infection cycle. Virology 383, 291-299.

Martelli, G., Agranovsky, A., Bar-Joseph, M., Boscia, D., Candresse, T., Coutts, R., et al, 2002. The family *Closteroviridae* revised. Arch.Virol. 147, 2039-2044.

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

Pietersen, G., 2004. Spread of grapevine leafroll disease in South Africa - a difficult, but not insurmountable problem. Wynboer, http://www.wynboer.co.za/recentarticles/articles.php3.

Rott, M., Jelkmann, W., 2005. Little cherry virus-2: Sequence and genomic organisation of an unusual member of the *Closteroviridae*. Arch.Virol. 150, 107-123.

Swofford, D.L., Sullivan, J., 2003. Phylogeny inference based on parsimony and other methods using PAUP*. The phylogenetic handbook: A practical approach to DNA and protein phylogeny, 160–206.

Turturo, C., Saldarelli, P., Yafeng, D., Digiaro, M., Minafra, A., Savino, V., et al, 2005. Genetic variability and population structure of Grapevine leafroll-associated virus 3 isolates. J. Gen. Virol. 86, 217-224.

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, 1424-1428.

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

Zuker, M. and Stiegler, P., 1981. Optimal computer folding of large RNA sequences using thermodynamics and auxiliary information. Nucleic Acid Res. 9, 133-148.

3.7 Recommended internet resources

Vienna RNA webserver: http://rna.tbi.univie.ac.at/

http://www.genelink.com/tools/gl-oe.asp

http://pfam.sanger.ac.uk/

Chapter 4: Mapping of the 5' terminal nucleotides of Grapevine leafroll-associated virus 3 sgRNAs.

4.1 Introduction

Grapevine leafroll-associated virus 3 (GLRaV-3) is a phloem limited, positive sense, single stranded (+ss) RNA virus. It is the type strain for the genus *Ampelovirus,* family *Closteroviridae* and is known to infect only *Vitis spp.* (Martelli *et al.*, 2002). It is of economical importance to the wine and table grape industries as virus infected grapevines have reduced grape quality and yield (Freeborough and Burger, 2008).

The replication of large +ssRNA viruses is complex and these viruses can employ various replication strategies including the use of polyproteins, +1 frameshifts, subgenomic RNA (sgRNA) and defective RNA (DRNA) to translate their proteins. To express genes located at the 3' terminal end of the genome, many +ssRNA viruses make use of sgRNAs. The mechanism how these sgRNAs are produced in the case of GLRaV-3 however still needs to be elucidated. There are two proposed models for the production of sgRNA: internal initiation and premature termination. Irrespective of which mechanism a virus utilises, the production of sgRNA is essential for the expression of the 3' ORFs. For a review on positive sense RNA virus replication refer to Miller and Koev (Miller and Koev, 2000).

Closteroviruses have complex genomes which range in size from 15.5 to 19.5 kb with between 10 and 14 ORFs (Martelli *et al.*, 2002). In a review by Dolja *et al.* "a working model of the infection cycle for a 'generic' closterovirus" is proposed (Dolja *et al.*, 2006). In this model, it is proposed that 3' co-terminal sgRNAs are produced to facilitate the translation of the internal genes on the 3' half of the genome and are approximately the same number as the number of ORFs (Dolja *et al.*, 2006).

Molecular research on GLRaV-3 has made slow progress compared to other closteroviruses such as Beet yellows virus (BYV) (Dolja, 2003) or other grapevine-infecting viruses like Grapevine virus A (GVA) (Mawassi, 2007), probably due to the lack of an alternative herbaceous host (Monette and James, 1990). In recent years, research on GLRaV-3 focused on genetic variation for applications such as phylogeny, epidemiology and detection (Fuchs *et al.*, 2009; Jooste and Goszczynski, 2005; Osman *et al.*, 2007; Turturo *et al.*, 2005).

To date, no studies have investigated the composition and characteristics of the sgRNAs associated with GLRaV-3 infection. Although the presence of GLRaV-3-specific sgRNAs

in leafroll-diseased vines have been observed, they have not been further characterised. (Hu *et al.*, 1990; K. S. Ling *et al.*, 1997; Rezaian *et al.*, 1991; Saldarelli *et al.*, 1994). The replication mechanism of GLRaV-3 can be inferred from research conducted on the related closteroviruses like Citrus tristeza virus (CTV) (Hilf *et al.*, 1995) and BYV (He *et al.*, 1997) and it is hypothesised that ORFs 3-12 are expressed via sgRNAs produced by an unknown mechanism.

In this study the use of RLM-RACE to map the 5' terminal nucleotides of GLRaV-3 positive-sense sgRNAs, representing ORFs 3-12 is demonstrated. These are the first steps towards characterising the sgRNAs involved in replication of this ampelovirus.

4.2 Material and methods

4.2.1 Double stranded RNA extraction and DIG double stranded RNA blot

Double stranded RNA (dsRNA) was extracted using a cellulose (Whatman, CF-11) extraction protocol as described previously (Hu *et al.*, 1990). Phloem scrapings of lignified grapevine canes (*Vitis vinifera* cv Merlot) from the Stellenbosch area (Western Cape Province, South Africa) infected with the same genetic variant of GLRaV-3 as GP18 were used for the extraction. The quality and quantity of the dsRNA was evaluated by electrophoresis. The dsRNA purified from 20g of phloem scrapings was separated in a single lane on a 1% agarose-TAE gel, overnight at 40V.

All protocols used for the dsRNA blot were performed according to the Southern blotting protocol described in the DIG manual supplied by the manufacturer (Roche) except where stated differently. The probe was PCR-labelled using digoxigenin (DIG)-dUTP. Primers to the CP gene (ORF6) of GLRaV-3 (CP01For-GCGATGGCATTTGAACTGAA and CP01Rev-ATCGATCGTAGCTACTTCTTTTGC) were used for probe synthesis. The dsRNA was blotted after separation on a 0.8% TAE agarose gel by alkaline vertical downward transfer (Brown, 1999). CDP-Star detection substrate (Roche) was used and the membrane was exposed to ECL Hyper film for 30' (Amersham). DIG molecular weight marker VII (Roche) was used to determine the approximate sizes of the sgRNA bands observed in the blot. The migration distances of the different fragments were compared to a standard curve drawn using the molecular weight marker and the sizes calculated (Table 1).

4.2.2 Total RNA extraction

Total RNA was extracted from the phloem scrapings of lignified canes infected with the GP18 isolate (Maree et al 2008) (Somerset West region, South Africa) using a modified CTAB method (White *et al.*, 2008). Isolate GP18 was selected as it was the only South African full-length sequence available for this study. The purity and concentration of the purified total RNA was determined spectrophotometrically. The integrity of the total RNA was analysed by electrophoresis on a non-denaturing 1% Agarose-TAE gel.

4.2.3 RLM-RACE

The 5' RNA Ligase Mediated Rapid Amplification of cDNA Ends (RLM-RACE) reactions were performed with the FirstChoice® RLM-RACE kit (Ambion, USA) as instructed by the manufacturer. Nested reverse primers were designed on the GP18 sequence (EU259806) using Oligo Explorer (version 1.1.0, http://www.genelink.com/tools/gl-oe.asp) on the 5' ends of ORF2-10 and ORF12. See table 2 for primer sequence and position of primers on the GLRaV-3 genome, as well as the amplicon associated with every primer combination. Twelve micrograms of total RNA was subjected to RLM-RACE and the amplicons generated were cloned using TA-cloning into the pDrive vector (Qiagen). Plasmid DNA was purified using the GeneJet miniprep kit (Fermentas). For each amplicon at least 5 clones were sequenced using the SP6 or T7 sequencing primer sites on the pDrive vector. A summary of the number of clones sequenced and the 5' ends predicted by the sequencing results can be seen in table 3. Sequences were aligned and analysed using BioEdit software with isolate GP18 as the reference sequence (Hall, 1999).

4.3 Results

4.3.1 DIG dsRNA blot

The dsRNA yield from the cellulose extraction protocol was of sufficient quantity and quality for blotting and displayed an intact, large molecular weight band at the expected size (~18.5kb) for GLRaV-3 genomic RNA (gRNA) after separation in an agarose gel. The dsRNA blot showed four distinct bands, one band at the expected size for the gRNA and 3 smaller bands expected to be sgRNA (Figure 1B). The DIG DNA molecular weight marker was used to draw a standard curve that was used to approximate the sizes of the sgRNA bands. For a summary of the predicted sizes see table 1. In the table the predicted sizes from the dsRNA blot are compared to the sizes predicted by RLM-RACE.

4.3.2 RLM-RACE

The CTAB extraction method yielded 170 μg of total RNA from 2g of phloem scrapings with an A260/A280 ratio of 2.14. Electrophoresis of total RNA also indicated that the RNA was intact and of sufficient quality to be used in the RLM-RACE reactions. Primer combinations used to determine the 5' ends of the different sgRNAs and the corresponding amplicons can be seen in table 2. All amplicons were cloned and sequenced. The primer combination for amplicon-1 produced a fragment. The same fragment was also produced in the RLM-RACE negative control. The primer combination for amplicon-2 was unable to produce an amplicon. Primer combinations for amplicon-3 to -11 were able to amplify specific fragments and clones of these amplicons were sequenced and used to predict the 5' ends of the 7 different sgRNAs. All the clones sequenced, predicted the same nucleotide as the 5' end of the corresponding sgRNA, with one exception. Three from fifteen clones from amplicon-4 did not contain the adapter sequence as was found for amplicon-1.

The 5' terminal nucleotides of the 7 sgRNAs of GLRaV-3 were mapped on the GP18 sequence and compared to the NY-1 sequence (AF037268). These 5' nucleotides were all found to be purines and conserved between GP18 and NY-1. The predicted sgRNAs are plotted in figure 1C.

Table 1: Predicted sizes of GLRaV-3 sgRNA using RLM-RACE or dsRNA blot results.

sgRNA	RLM-RACE	dsRNA Blot	%Diff
sgRNA(ORF3/4)	8021	7974	0.59
sgRNA(ORF5)	6313	6000	4.96
sgRNA(ORF6)	4698	4615	1.77

Table 2: Primer combinations used in RLM-RACE to generate amplicons for sgRNA 5' end prediction.

Amplicon number	GLRaV-3 ORF	Inner primer [a]			Outer primer [b]		
		Position[c]	Name	Sequence	Position[c]	Name	Sequence
1	2	9341	ORF2 (8762) Rev	CGACAAGAGTAGGAAGGGT	9447	ORF2 (8868) Rev	CCATCCTACCTTGACACAG
2	3	10401	ORF3 (9822) Rev	TATCCCAATCTAGCACGAA	10516	ORF3 (9937) Rev	TCATCCATAAATAACTACGC
3	3	10516	ORF3 (9937) Rev	TCATCCATAAATAACTACGC	10863	ORF4 (10284) Rev	TAACATACAACCTTCCCTCT
4	4	10828	ORF4 (10249) Rev	GCTTTCCCCGCTGCTTTACC	10863	ORF4 (10284) Rev	TAACATACAACCTTCCCTCT
5	5	12495	ORF5 (11916) Rev	AGATTGCGAAGTTGATGTAA	12543	ORF5 (11964) Rev	GGTAGTCAATCCGCCTTTCA
6	6	13955	ORF6 (13376) Rev	GAAACTCGCCTTACTAAACT	14000	ORF6 (13421) Rev	GATTCCTGTTAACTCCGCCT
7	7	14991	ORF7 (14412) Rev	TAAGCCCTCGTATAACTATC	15079	ORF7 (14500) Rev	CATAAGACCTGAAGCCAACT
8	8	16461	ORF8 (15882) Rev	AATACTCTTTCATAACGGTG	16537	ORF8 (15958) Rev	CTTATCATCTCACCTTCCTT
9	9	17039	ORF9 (16460) Rev	AACTACACCTGGCTATGAGA	17105	ORF9 (16526) Rev	TTTCTATCGTCGCCTTACAG
10	10	17572	ORF10 (16993) Rev	AGTCGTCGCTGTAGTAGTTA	17622	ORF10 (17043) Rev	TAAGCGAAGGCGGAGTCTAT
11	12	18200	ORF12 (17621) Rev	ACCTTCTCCTTTTCTACAT	18227	ORF12 (17648) Rev	GGATTTTTACCCATTACTG

ND - Not determined, a- in combination with the RLM-RACE inner primer, b- in combination with the RLM-RACE outer primer, c- 5' nt of primer compared to GP18 sequence (EU259806).

Table 3: Characteristics of RLM-RACE predicted sgRNAs for GLRaV-3 ORFs 3-12.

GLRaV-3 ORF[a]	Predicted sgRNA	Amplicon number	Size of the sgRNA 5' UTR	5' nt in isolate		sgRNA size	Clones sequenced	dsRNA Blot
				GP18	NY-1[c]			
2 (9287)	b	1	-	-	-	-	10/10	
3 (10509)	-	2	No amplification	-	-	-	-	
3 (10509)	sgRNA(ORF3/4)	3	32	G-10477	G-9898	8021	6/6	Yes
4 (10665)	sgRNA(ORF3/4)	4	188	G-10477	G-9898	8021	12/15	Yes
5 (12307)	sgRNA(ORF5)	5	122	G-12185	G-11606	6313	9/9	Yes
6 (13848	sgRNA(ORF6)	6	48	A-13800	A-13221	4698	5/5	Yes
7 (14852)	sgRNA(ORF7)	7	37	G-14815	G-14236	3683	9/9	ND
8 (16296)	sgRNA(ORF8)	8	23	A-16273	A-15694	2225	5/5	ND
9 (16850)	sgRNA(ORF9)	9	92	G-16754	G-16175	1740	6/6	ND
10 (17390)	sgRNA(ORF10-12)	10	125	A-17265	A-16686	1233	7/7	ND
11 (17932)	sgRNA(ORF10-12)	-	667	-	-	1233	-	ND
12 (18039)	sgRNA(ORF10-12)	11	774	A-17265	A-16686	1233	5/5	ND

a- Position of ATG in brackets and nucleotide positions relative to GP18 sequence (EU259806), b- Amplification products were shown to be as a results false positive, c- position of sgRNA 5' nt relative to NY-1 sequence (AF037268) as determined through sequence similarity, ND - Not determined.
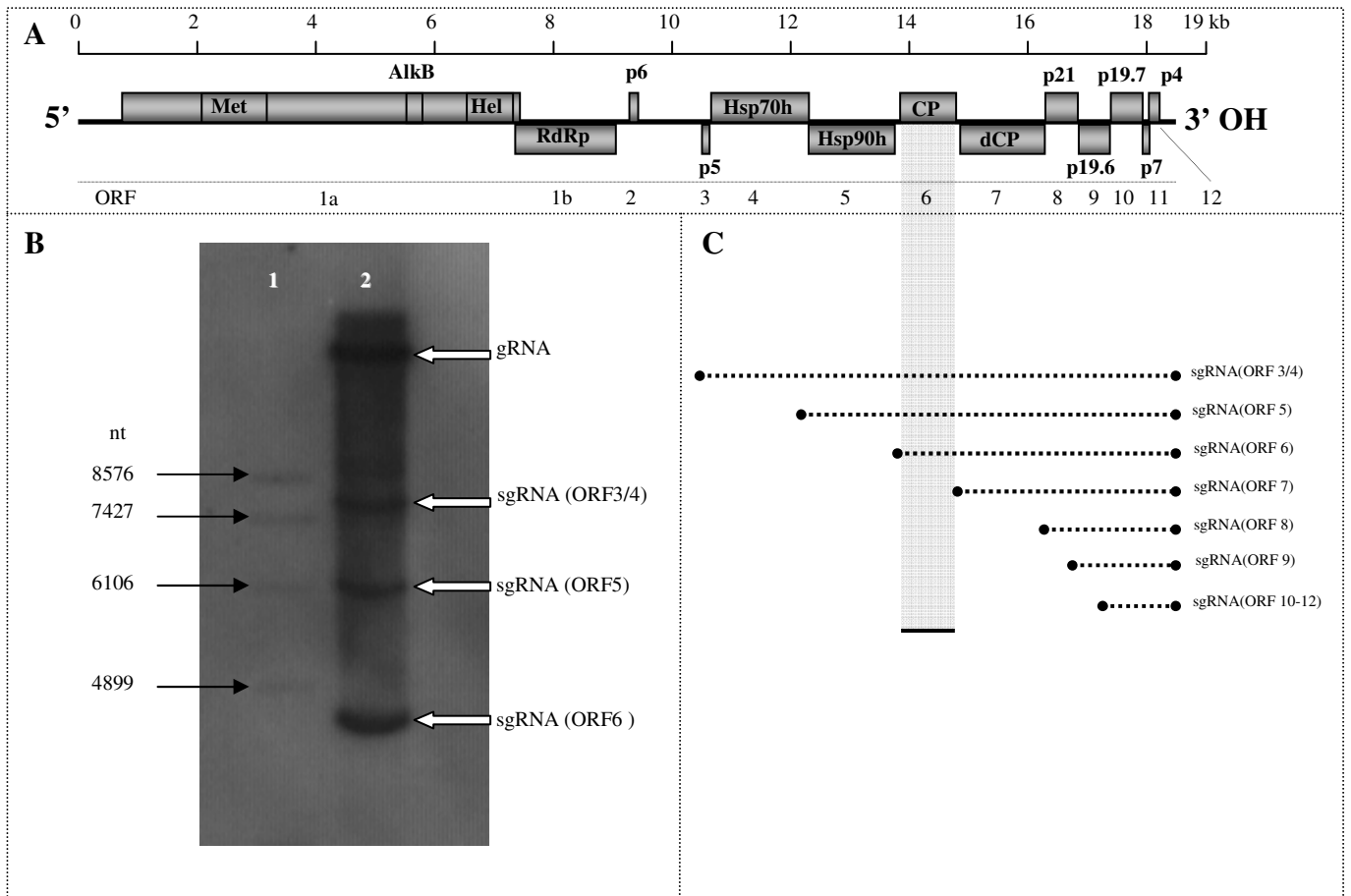
Figure 1: A) Diagrammatic representation of the GLRaV-3 genome. B) Double stranded RNA blot. In lane 1 DIG molecular weight marker VII with nt sizes as indicated. Lane 2 dsRNA sample with hybridised bands as indicated and annotated according to RLM-RACE predictions. C) Diagrammatic representation of the RLM-RACE predicted sgRNAs. The position of probe used in dsRNA is marked with a grey box.

4.4 Discussion

In this study we identified three GLRaV-3 sgRNAs by dsRNA blotting with a probe directed at the coat protein gene. The 5' nts of these three sgRNAs associated with ORFs 3-6 as well as four additional putative sgRNAs associated with ORFs 7-12 were determined with RLM-RACE.

The yield of dsRNA extracted from virus infected grapevine, as with other woody plants, was low but sufficient for blotting. The DIG-labelled probe directed towards the CP gene of GLRaV-3 detected four distinct bands. The high molecular weight band corresponds to the viral gRNA (~18500nt) and the three smaller bands to the sgRNAs most likely associated with ORFs 3-6. These bands were labelled sgRNA(ORF3/4) (~7974nt), sgRNA(ORF5) (~6000nt) and sgRNA(ORF6) (~4615nt) according to their respective sizes

and sequencing results from the amplicons generated with RLM-RACE. The sizes of these fragments were approximated with a standard curve drawn using the DIG molecular marker and their migration distances. The size of the different fragments approximated by the dsRNA blot and the sizes predicted by the RLM-RACE experiments can be seen in table 1 and was found to vary by less than 5%. The dsRNA blot confirmed that the sgRNAs of ORFs 3-6 are 3' co-terminal, which correlates with other closteroviruses like BYV and CTV (He *et al.*, 1997; Hilf *et al.*, 1995).

Nested reverse primers were designed for the RLM-RACE to the 5' ends of ORFs 2-10 and ORF 12. Amplicons were generated with all of the primer sets except for amplicon-2 (Table 2). The amplicons (1, 3-11) were sequenced and analysed, and used to determine the 5' nt of the sgRNA corresponding to a particular ORF (Table 3). The 5' proximal gene was assigned as the ORF associated with each sgRNA. In the case of ORFs 3 and 4 and ORFs 10-12 it was co-assigned as data indicated that these ORFs may be expressed from the same sgRNAs. All the clones sequenced for each amplicon predicted the same nucleotide to be the 5' transcription initiation site for the sgRNA with 2 exceptions.

Amplicon-1, designed to determine the 5' nt of a possible sgRNA(ORF2) was consistently generated, however the RLM-RACE negative control amplified the same fragment. Sequencing of all amplicon-1 fragments clearly demonstrated that all the amplicons lacked the RNA adapter sequence ligated to the RNA 5' ends during the RLM-RACE reaction. Three sequenced clones for amplicon-4 also lacked the RNA adapter sequence. Sequence homology between the RLM-RACE Inner primer and the sequence upstream of G-9001 for amplicon-1 and A-10516 for amplicon-4 indicates that amplification is possible in the absence of the RNA adapter. It was thus concluded that the amplification of amplicon-1 and the three clones sequenced from amplicon-4 resulted from mispriming of the RLM-RACE Inner primer and should be regarded as false positive clones. The results obtained do not exclude the possibility of ORF 2 to be translated via a sgRNA(ORF2) as the 5' end of the sgRNA might be beyond the priming site for the RLM-RACE Inner primer. However, the existence of a sgRNA(ORF2) is not supported by the dsRNA blot. Primer sets to generate amplicon-2, -3 and -4 were designed to determine the 5' ends of the sgRNAs produced in the expression of ORFs 3 and 4. Sequence analysis of amplicons-3 and -4 revealed that they predicted the same 5' nt to be the 5' end of a sgRNA. In BYV the analogous proteins p6 (ORF 2) and Hsp70h (ORF 3) was determined to be expressed from two different sgRNAs by RLM-RACE (Peremyslov and Dolja, 2002). Sequence analysis

of 15 clones derived from amplicon-3 and -4 indicate that the 5' nt for the proposed sgRNA(ORF3/4) is G-9898. The dsRNA blot was unable to resolve more than one band and an in depth investigation is required to determine if ORFs 3 and 4 are expressed from different sgRNAs like BYV or from the same sgRNA molecule. Sequence analysis of amplicons -5 and -6 determined the exact 5' nt for sgRNA(ORF5) and sgRNA(ORF6) to be at G-12185 and A-13800, respectively. All the clones sequenced for each of these amplicons predicted the same 5' nt to be the 5' end of the respective sgRNAs.

The dsRNA blot indicated the existence of three 3' co-terminal sgRNAs for the expression of ORF 3, 4, 5 and 6. It is believed that ORFs 7-12 are also translated via sgRNA similarly to other closteroviruses like BYV and CTV (He *et al.*, 1997; Hilf *et al.*, 1995). To determine the 5' ends of these putative sgRNAs possibly utilised in the expression of ORFs 7-12, amplicons-7 to -11 were generated. Amplicons-7 to -9, consistently predicted the 5' ends of 3 sgRNAs potentially involved in the expression of ORFs 7, 8 and 9 each from their own putative sgRNAs (designated sgRNA(ORF7), sgRNA(ORF8) and sgRNA(ORF9)) with 5' ends mapped at G-14815, A-16273, and G-16754 respectively. Sequence data from amplicons-10 and -11 predicted that ORFs 10, 11 and 12 are expressed from a single sgRNA. Even though no amplicon was designed for ORF 11 because of its small size, the result indicates that ORF 10-12 are most likely translated from the same sgRNA. Eleven clones in all mapped the 5' end of putative sgRNA(10-12) at A-17265. Amplicons-7 to -11 thus predicts the existence of 4 additional positive sense sgRNAs responsible for the expression of ORFs 7-12.

Bioinformatic analysis of the sequences upstream of the mapped 5' ends of the sgRNAs was unable to reveal any conserved sequence or secondary structure within the GLRaV-3 genome. Comparisons with known viral sg-promoters and transcription control elements were unable to determine the position and critical bases of any possible sg-promoters. As bioinformatic tools are refined and more sequence data become available it may be possible in the future to identify such elements.

Collectively the RLM-RACE data predict the existence of 7, most likely 3' co-terminal positive sense sgRNAs for the expression of the 3' ORFs 3-12 of GLRaV-3. The exact 5' nts were mapped on the GP18 genome (Table 3) and found to all be purines. Compared to the NY-1 isolate these bases were found to be conserved. Similar to CTV, all the

amplicons sequenced indicated that the sgRNAs were continuous with the genome with no common 5' leader sequence as observed in the evolutionary related Nidoviruses (Karasev *et al.*, 1997). Data also indicated the interesting possibility that ORF 3 and 4, and ORFs 10-12 are expressed from polycistronic sgRNAs.

4.5 Acknowledgements

4.6 References

Brown, T., 1999. Southern blotting. Curr. Protoc. Mol. Biol. 68, 2.9.1-2.9.20.

Dolja, V.V., 2003. Beet yellows virus: The importance of being different. Mol. Plant Pathol. 4, 91-98.

Dolja, V.V., Kreuze, J.F., Valkonen, J.P.T., 2006. Comparative and functional genomics of closteroviruses. Virus Res. 117, 38-51.

Freeborough M-J, Burger J.T., 2008. Leafroll: Economic implications. Wynboer. http://www.wynboer.co.za/recentarticles/200812-leafroll.php3

Fuchs, M., Martinson, T.E., Loeb, G.M., Hoch, H.C., 2009. Survey for the three major leafroll disease-associated viruses in Finger Lakes vineyards in New York. Plant Dis. 93, 395-401.

Hall, T., 1999. BioEdit: A user-friendly biological sequence alignment editor and analysis program for windows 95/98/NT. Nucl. Acids Symp. Ser. 41, 95-98.

He, X., Rao, A.L.N., Creamer, R., 1997. Characterization of Beet yellows closterovirus-specific RNAs in infected plants and protoplasts. Phytopathology 87, 347-352.

Hilf, M.E., Karasev, A.V., Pappu, H.R., Gumpf, D.J., Niblett, C.L., Garnsey, S.M., 1995. Characterization of Citrus tristeza virus subgenomic RNAs in infected tissue. Virology 208, 576-582.

Hu, J.S., Gonsalves, D., Teliz, D., 1990. Characterisation of closterovirus-like particles associated with grapevine leafroll disease. J. Phytopathology 128, 1-14.

Jooste, A., Goszczynski, D., 2005. Single-strand conformation polymorphism (SSCP), cloning and sequencing reveals two major groups of divergent molecular variants of Grapevine leafroll-associated virus 3 (GLRaV-3). Vitis 44, 39-43.

Karasev, A., Hilf, M., Garnsey, S., Dawson, W., 1997. Transcriptional strategy of closteroviruses: Mapping the 5'termini of the Citrus tristeza virus subgenomic RNAs. J. Virol. 71, 6233-6236.

Ling, K., Zhu, H., Drong, R., Slightom, J., McFerson, J., Gonsalves, D., 1998. Nucleotide sequence of the 3'-terminal two-thirds of the Grapevine leafroll-associated virus-3 genome reveals a typical monopartite closterovirus. J. Gen. Virol. 79, 1299-1307.

Ling, K.S., Zhu, H.Y., Alvizo, H., Hu, J., Drong, R., Slightom, J., et al, 1997. The coat protein gene of grapevine leafroll associated closterovirus-3: Cloning, nucleotide sequencing and expression in transgenic plants. Arch. Virol. 142, 1101-1116.

Maree, H.J., Freeborough, M.J., Burger, J.T., 2008. Complete nucleotide sequence of a South African isolate of Grapevine leafroll-associated virus 3 reveals a 5′ UTR of 737 nucleotides. Arch. Virol. 153, 755-757.

Martelli, G., Agranovsky, A., Bar-Joseph, M., Boscia, D., Candresse, T., Coutts, R., et al, 2002. The family *Closteroviridae* revised. Arch. Virol. 147, 2039-2044.

Mawassi, M., 2007. The vitivirus Grapevine virus A: A "small" but surprising virus. Phytoparasitica 35, (5) 425-428.

Miller, W.A., Koev, G., 2000. Synthesis of subgenomic RNAs by positive-strand RNA viruses. Virology 273, 1-8.

Monette, P.L., James, D., 1990. The use of Nicotiana benthamiana as an herbaceous receptor host for closteroviruses from leafroll-affected grapevines. Am. J. Enol. Vitic. 41, 201-203.

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

Peremyslov, V.V., Dolja, V.V., 2002. Identification of the subgenomic mRNAs that encode 6-kDa movement protein and Hsp70 homolog of beet yellows virus. Virology 295, 299-306.

Rezaian, M.A., Krake, L.R., Cunying, Q., Hazzalin, C.A., 1991. Detection of virus-associated dsRNA from leafroll infected grapevines. J. Virol. Methods 31, 325-334.

Saldarelli, P., Minafra, A., Martelli, G.P., Walter, B., 1994. Detection of grapevine leafroll-associated closterovirus III by molecular hybridization. Plant Pathology 43, 91-96.

Turturo, C., Saldarelli, P., Yafeng, D., Digiaro, M., Minafra, A., Savino, V., et al, 2005. Genetic variability and population structure of Grapevine leafroll-associated virus 3 isolates. J. Gen. Virol. 86, 217-224.

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, 1424-1428.

4.7 Recommended internet resources

https://www.roche-applied-science.com/PROD_INF/index.jsp?id=techMan&pageid=/PROD_INF/MANUALS/DIG_MAN/dig_toc.htm

http://www.genelink.com/tools/gl-oe.asp

Chapter 5: Construction of a Grapevine leafroll-associated virus 3 mini-replicon.

5.1 Introduction

The inherent difficulties associated with studying woody plant viruses in their natural hosts necessitate the development of infectious clones that are able to replicate in herbaceous hosts. An infectious clone of an RNA virus is a full-length cDNA resembling the viral gRNA that can be used to induce an infection in a susceptible plant. The viral cDNA is normally cloned downstream of a CaMV 35S promoter, transferred to a binary vector, transformed into *Agrobacterium* and infiltrated into a plant host. After T-DNA transfer, the viral genome will be transcribed *in vivo,* the virus will replicate and move systemically throughout the plant. Such cDNA clones can be utilised to improve our understanding of the functional role of viral ORFs as well as the biological characteristics of the represented viruses.

One of the first plant viruses to be converted into an infections clone for the study of viral replication was Brome mosaic virus (BMV) (Ahlquist *et al.*, 1984). Since then many more infectious clones representing plant viral genomes have been constructed. Infectious clones of the following closteroviruses have been constructed: Citrus tristeza virus (CTV), Beet yellows virus (BYV), Lettuce infectious yellows virus (LIYV), and Grapevine leafroll-associated virus 2 (GLRaV-2) (Klaassen et al., 1995; Liu et al., 2009; Peremyslov et al., 1998; Satyanarayana et al., 1999).

Infectious clones are useful molecular tools that can be utilised in expression systems to produce target proteins at high concentrations or in virus induced gene silencing (VIGS) systems for functional genomic studies of the host plant. Constructed clones are frequently found not to be infectious due to spontaneous rearrangements and mutations induced by the bacterial host during propagation (Boyer and Haenni, 1994). It has been reported that even a single point mutation within a viral cDNA clone can lead to reduced infectivity or complete abolishment of infectivity (Boyer and Haenni, 1994). Additional errors that influence the infectivity of a cDNA clone can also be introduced during the reverse transcription or PCR-based assembly steps. Infectious clones assembled from the genomes of larger RNA viruses, are more cumbersome to manipulate. Such clones are also more susceptible to errors being incorporated, due to their large size. To avoid this problem, replication competent mini-replicons were constructed by various research groups for

different viruses including the closteroviruses CTV, GLRaV-2 and BYV (Liu *et al.*, 2009; Peremyslov *et al.*, 1998; Satyanarayana *et al.*, 1999). These mini-replicons contained only the replicase proteins and the 5' and 3' regulatory elements, analogous to the RNA1 of the bipartite closterovirus LIYV (Klaassen *et al.*, 1995). They are useful for studies in protoplasts or agroinfiltrated plant material and are unable to move systemically (Liu *et al.*, 2009; Peremyslov *et al.*, 1998; Satyanarayana *et al.*, 1999).

The construction of CTV and GLRaV-2 infectious clones and their deletion mutants enabled researchers to accelerate the progression of functional knowledge and increase the understanding of the replication cycle of these woody plant-infecting viruses. Such studies would have been impossible to perform without these respective clones (Liu *et al.*, 2009; Satyanarayana *et al.*, 1999).

The construction of an infectious clone or mini-replicon of GLRaV-3 is necessary to further investigate the results generated in Chapter 3 and 4 of this thesis. In Chapter 3 the sequence of isolate GP18 was determined and found to contain a 5' UTR of 737 nt (Coetzee *et al.*, 2010; Jooste *et al.*, 2010; Maree *et al.*, 2008). Great diversity in this large 5' UTR was also observed among different genetic variants (Jooste *et al.*, 2010). Such a large 5' UTR is unusual for closteroviruses and a molecular tool, like an infectious clone will greatly assist in determining its function and to establish if the genetic variation observed has any functional significance. An infectious clone is also required to investigate the production of the seven sgRNAs identified in Chapter 4 (Maree *et al.*, 2010). It still remains to be proven if the sgRNAs of GLRaV-3 are transcribed by sg-promoters or controlled by sgRNA controller elements (Dolja *et al.*, 2006). In this study, the assumption is made that, the sgRNAs of GLRaV-3 are transcribed by sg-promoters situated on the negative genomic RNA strand and will be referred to as such.

Here we report the first attempt to construct a GLRaV-3 mini-replicon, a member of the *Ampelovirus* genus.

5.2 Material and methods

5.2.1 Assembly of the GLRaV-3 mini-replicon

Total RNA was extracted from the phloem tissue of mature canes infected with GLRaV-3 isolate GP18 using a CTAB protocol (Maree *et al.*, 2008; White *et al.*, 2008). Primers were designed with Oligo Explorer (version 1.1.0, developed by Teemu Kuulasma) using the GP18 sequence (EU259806) as reference.

The GLRaV-3 mini-replicon was designed to include the 5' UTR, ORF1a, ORF1b and 3' UTR of the GLRaV-3 genome. The mini-replicon was assembled from four amplicons using the sub-cloning vector pL140 (Supplied by Edgar Maiss) as backbone. The pL140 vector contains a duplicated CaMV 35S promoter and a hammerhead ribozyme (Shintaku *et al.*, 1996) (Figure 5.1). The first three amplicons were designed to be assembled from the 5' half of the genome (nucleotides 1-9100) designated LR3-Rep-1, -2 and -3. The fourth amplicon representing the 3' UTR (nucleotides 18222-18498) was designated LR3-Rep3UTR and was designed to be separated from the replication ORFs by a small spacer and two unique restriction sites SgfI and AflII (Figure 5.1B and 5.1C).

All the amplicons were generated with RT-PCR from the extracted total RNA using AMV and a high fidelity DNA polymerase and was initially cloned into a standard TA-cloning vector (pDrive, Invitrogen). For the final design of the GLRaV-3 mini-replicon see figure 5.1. See table 5.1 for the characteristics of primers used in the assembly.

The two amplicons, LR3-Rep-1 and LR3-Rep-2, were first assembled in an intermediate cloning vector, pLitmus38 (supplied by E. Maiss). LR3-Rep-1 was transferred from the pDrive cloning vector into pLitmus38 using the restriction enzymes BamHI and HindIII (sites incorporated on primer flaps) to create pLitmus+I. LR3-Rep-2 was then transferred from pDrive into pLitmus+I using two endogenous GLRaV-3 restriction sites XbaI and NgoMIV to create pLitmus+I+II.

The other two amplicons LR3-Rep-3 and LR3-Rep3UTR were first assembled in pL140. LR3-Rep3UTR was transferred from pDrive into pL140 using XhoI and Bsp120I (sites incorporated on primer flaps) to generate pL140+3UTR. LR3-Rep-3 was then transferred from pDrive into pL140+3UTR using AscI and SgfI (sites incorporated on primer flaps) to generate pL140+III+3UTR.

To complete the GLRaV-3 mini-replicon, the first two assembled amplicons (LR3-Rep-1 and -2) were cloned from pLitmus+I+II with restriction enzymes AscI and NgoMIV into pL140+III+3UTR to result in pL140_LR3_Minireplicon.

The construct pL140_LR3_Minireplicon containes two restriction sites, SgfI and AflII separated by a GGG spacer that were used as an entry site for test constructs. The GLRaV-3 mini-replicon was transferred into a pBIN_SN binary vector (derived from pBIN19, supplied by Edgar Maiss) using SmaI and NotI. The final assembled pBIN_LR3_Minireplicon was confirmed by sequencing.

5.2.2 Assembly of sg-promoter test constructs

Test constructs were assembled to evaluate the activity of the sgRNA(ORF6) sg-promoter using β-glucuronidase (GUS) assays (Figure 1D). The sg-promoter of sgRNA(ORF6) responsible for the expression of the CP was selected as it is generally regarded as one of the most active sg-promoters. As positive test construct, pBIN_LR3_Minireplicon_ ORF6 sg-promoter/GUSi containing the putative sg-promoter of sgRNA(ORF6) and the GUSi gene (GUS gene with an intron) was assembled. As negative test construct, pBIN_LR3_Minireplicon_ ΔORF6 sg-promoter/GUSi was generated that was identical to the positive test construct but lacked the sg-promoter. The inserts used to generate these test constructs were first assembled and cloned into pDrive before subsequent transfer into the pL140_LR3_Minireplicon using SgfI and AflII. These assembled constructs were finally transferred into pBIN_SN using SmaI and NotI.

The "ORF6 sg-promoter/GUSi" insert (positive test construct) was assembled by overlap primer extension of two amplicons. The amplicons were assembled using primers with a 23nt overlap (Table 5.1) (Higuchi *et al.*, 1988). The sg-promoter amplicon, spanning from nucleotides 13651 to 13859 in the GLRaV-3 genome, was generated from total RNA by RT-PCR. The sg-promoter amplicon included 150 nts upstream of the predicted 5' transcription initiation site for sgRNA(ORF6) (Chapter 4 this thesis), the sgRNA(ORF6) 5' UTR for ORF6 as well as the start codon and first three codons of the ORF6 (CP). The GUSi amplicon was generated by use of a high fidelity DNA polymerase (Pyrobest, Takara) from the plasmid p35S:GUSi (supplied by Pere Mastre). Amplicons were gel-purified and 25ng of each were used in overlap extension PCR using high fidelity DNA polymerase (ExTaq, Takara). The resulting amplicon was gel-extracted and cloned into pDrive.

The negative test construct insert "ΔORF6 sg-promoter/GUSi" contains only the GUSi without the sg-promoter for ORF6. The GUSi amplicon was amplified from p35S:GUSi adding restriction sites SgfI and AflII and cloned into pDrive.

Both test construct inserts were confirmed by sequencing before subsequent cloning into pL140-LR3-Minireplicon and then pBIN_LR3_Minireplicon. The resulting test constructs were designated: pBIN_LR3_Minireplicon_ORF6 sg-promoter/GUSi (positive test construct) and pBIN_LR3_Minireplicon_ΔORF6 sg-promoter/GUSi (negative test construct).

5.2.3 Evaluation of GLRaV-3 sgRNA(ORF6) sg-promoter in Nicotiana benthamiana

The positive and negative test constructs were transformed, using electroporation, into electro-competent *Agrobacterium tumefaciens* cells C58C1(pCH32) using the following settings: capacitance 25 µF, resistance 200 Ω and voltage 1,5 kV (Annamalai and Rao, 2006). Transient expression assays in *N. benthamiana* were performed *via* agroinfiltrations as described by Voinnet *et al.* in 1998. Test constructs were co-infiltrated with a clone of the silencing suppressor of Beet mild yellowing virus (BMYV), P0 under control of the CaMV 35S promoter (supplied by Edgar Maiss). Plants were maintained and sample leaves taken at 2, 4 and 6 days post infection (dpi). The sample leaves were then infiltrated with GUS-staining solution (100mM NaH_2PO_4, 0.5mM K-Ferrocyanid, 0.5mM K-Ferricyanid, 100mM EDTA, 0.1% Triton X100, pH7 with added 0.5mg/mL X-Gluc) and incubated O/N at 37ºC with mild agitation and decoloured with 100% EtOH until completely white.

5.3 Results

5.3.1 Assembly of the GLRaV-3 mini-replicon

A GLRaV-3 mini-replicon spanning nucleotides 1 to 9100 and 18222 to 18498 was successfully assembled and engineered to be under control of a duplicated CaMV 35S promoter and a ribozyme. The complete clone pBIN_LR3_Minireplicon was sequenced and found to contain a single nucleotide insertion in the 5' UTR and 27 nucleotide changes when compared to the GP18 sequence (Table 5.2). Sequencing results confirmed that the ORFs and other regulatory elements were intact and contained no significant mutations such as premature stop codons or frameshifts. By comparing the GLRaV-3 mini-replicon sequence to the GP18 and other GLRaV-3 isolates (623 - GQ352632, NY-1 - AF037268, CL-766 - EU344893 and 621 - GQ352631), 13 nucleotide changes were found to be unique to the GLRaV-3 mini-replicon. The 27 nucleotide changes, when compared to the GP18 sequence, resulted in 13 amino acid changes, eight in the replicase and five in the RdRp. Seven of these aa changes were found to be unique to the GLRaV-3 mini-replicon. The grey boxes in table 5.2 highlights the nucleotide and amino acid changes that are unique to the GLRaV-3 mini-replicon.

Figure 5.1: Diagrammatic representation of the construction of the GLRaV-3 mini-replicon (not to scale) in pL140. These assemblies were cloned into pBIN_SN using the SmaI and NotI sites before they were transformed into Agrobacterium for transient expression studies. A) Diagrammatic representation of the genome of GLRaV-3 with "☆" indicating the positions of the sgRNA 5' ends and the grey arrow the 5' end of the NY-1 sequence compared the GP18 sequence. B) Figure shows the four amplicons used to construct the mini-replicon and their relative positions on the GLRaV-3 genome as well as the restriction enzymes used in assembly. C) The completed GLRaV-3 mini-replicon under the control of a duplicated CaMV 35S promoter and a ribozyme in the pL140 backbone is shown. D) The three test inserts and their construct names when cloned into pL140_LR3_Minireplicon are shown.

Table 5.1 Details of amplicons used to assemble the GLRaV-3 mini-replicon

Amplicon	Size	Name	RE sites added	Sequence*	Position GP18
LR3-Rep-1	2550	Rep LR3-1 For	BamHI, AscI	AT GGATCC GGCGCGCC ATAAATGCTCTAGTAGGATTCG	1
		Rep LR3-1.2 Rev	HindIII	AG AAGCTT TATCGTACACATCCACCATA	2526
LR3-Rep-2	3397	Rep LR3-2 For	none	TTCTAGAGCCGATGCGGGT	2486
		Rep LR3-2 Rev	none	TTGGTGACCTCTCATCCGA	5883
LR3-Rep-3	3297	Rep LR3-3 For	AscI	AA GGCGCGCC GGCGTAGGCAGGATGTTGC	5823
		Rep LR3-3 Rev	SgfI	AA GCGATCGC CGAAGAACCGAGGCACAGA	9100
LR3-Rep3UTR	311	Rep LR3-3' UTR For	XhoI, SgfI,AflII	AG CTCGAG GCGATCGC GGG CTTAAG AAATCCTTCAATAAATTTAAAATA	18222
		Rep LR3-3' UTR Rev	Bsp120I	AA GGGCCC GACCTAACTTATTGTCGATAAG	18498
ORF6-sgPromGUS	227	LR3 sgProm6 For	SgfI	T GCGATCGC TCTAAGGTGAGAAGTAT	13651
		sgPr6 GUS Junct Rev		**caggacgtaa TTCAAACGCCATC**GCGTCCA	13859
	2021	sgPr6 GUS Junct For		**GATGGCGTTTGAA ttacgtcctg**tagaaac	
		GUS Rev	AflII	t cttaag tcattgtttgcctccctgct	
GUS	2017	GUS For	SgfI	t gcgatcgc atgttacgtcctgtagaaac	
		GUS Rev	AflII	t cttaag tcattgtttgcctccctgct	
ORF6wt	1183	LR3 sgProm6 For	SgfI	T GCGATCGC TCTAAGGTGAGAAGTAT	13651
		LR3 ORF6(CP) Rev	AflII	TT CTTAAG TTCACCGATTTATGGACAT	14817

* Restriction sites are underlined. Overlapping primer sequences are shown in bold. The GUSi sequence is shown in lowercase.

5.3.2 Assembly of sg-promoter test constructs

Two test constructs to evaluate the activity of the putative sg-promoter of sgRNA(ORF6) were successfully assembled and cloned into pBIN_LR3_Minireplicon.

5.3.3 Evaluation of GLRaV-3 sgRNA(ORF6) sg-promoter in Nicotiana benthamiana

Nicotiana benthamiana leaves co-infiltrated with a test construct and the P0 suppressor were harvested and assayed for GUS expression. Leaves taken 2 dpi clearly showed small foci of GUS expression in the plant cells for both test constructs (Figure 5.2A and 5.2C). At 4 dpi the negative test construct showed lower expression of GUS compared to the positive test construct (Figure 5.2B and 5.2D). At 6 dpi no foci of GUS expression were observed with any of the test constructs. The infiltrations were repeated several times, but were not reliably reproducible. Additionally, observed GUS expression levels were inconsistent and attempts to improve on this were unsuccessful. The results depicted in figure 5.2 show the greatest variation observed in GUS expression between the two constructs.
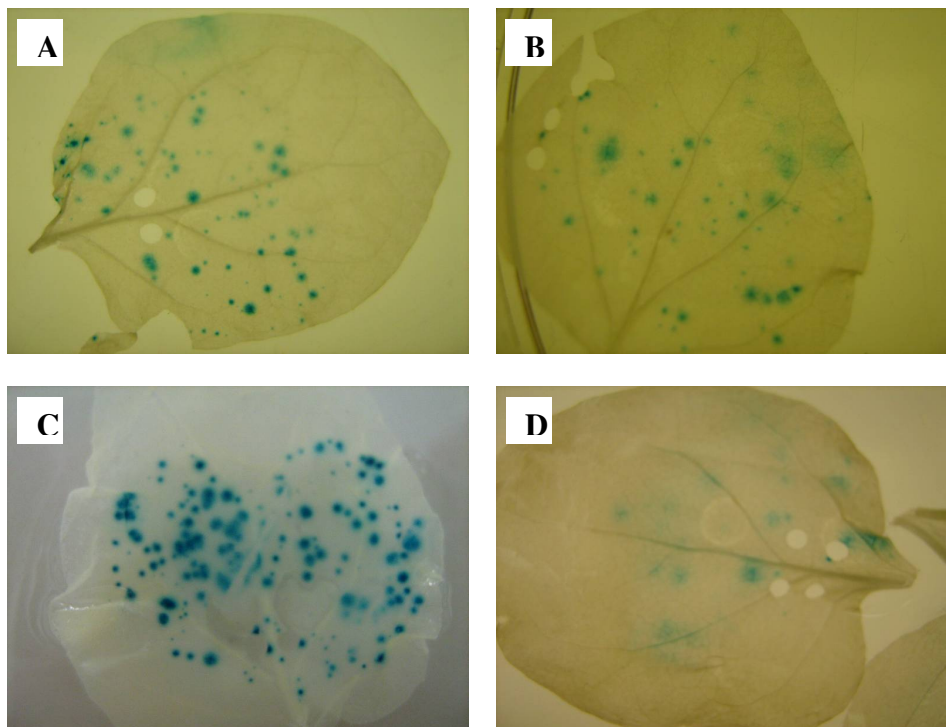
Figure 5.2: *N. benthamiana* leaves infiltrated with GLRaV-3 mini-replicon constructs to evaluate the activity of the putative sg-promoter of sgRNA(ORF6). In panes A and C are representative photos of leaves 2 dpi and 4 dpi respectively when infiltrated with pL140_LR3_Minireplicon_ ORF6 sg-promoter/GUSi. In panes B and D are representative photos of leaves 2 dpi and 4 dpi respectively when infiltrated with pL140_LR3_Minireplicon_ ΔORF6 sg-promoter/GUSi.

Table 5.2: Nucleotide and amino acid differences of the GLRaV-3 mini-replicon and selected GLRaV-3 isolates.

Nucleotide differences

Position (GP18)	Mini-replicon	GP18	632	621	NY-1	CL-766
151	T	INS	INS	INS	INS	INS
156	C	T	T	T	-	-
359	C	T	T	T	-	-
497	C	T	T	T	-	-
989	A	T	T	T	T	T
1039	C	T	T	T	T	T
2235	G	A	A	A	A	A
2714	A	T	A	A	A	A
3109	G	A	A	A	A	A
3414	A	G	A	A	A	A
3503	T	C	T	T	T	T
3534	G	A	G	G	G	G
3685	C	T	T	T	T	T
3771	A	G	A	A	A	A
3908	G	A	G	A	A	G
4010	T	C	T	T	T	T
4786	A	G	A	A	A	A
4929	G	A	G	G	G	G
6253	G	A	A	A	A	A
7780	C	T	T	T	T	T
7785	C	T	T	C	C	C
7993	G	A	A	A	A	A
8037	G	A	G	G	G	G
8270	A	G	A	A	A	A
8501	T	C	T	T	T	T
8532	C	T	T	T	T	T
8549	T	C	T	T	T	T
9033	C	T	T	T	T	T

Amino acid differences — Replicase

Position (GP18)	Mini-replicon	GP18	632	621	NY-1	CL-766
101	T	M	M	M	M	M
659	E	D	D	D	D	D
791	G	D	D	D	D	D
893	R	G	R	R	R	R
933	A	T	A	A	A	A
983	S	L	L	L	L	L
1012	T	A	T	T	T	T
1839	C	Y	Y	Y	Y	Y

Amino acid differences — RdRp

Position (GP18)	Mini-replicon	GP18	632	621	NY-1	CL-766
114	R	C	C	C	C	C
185	A	T	T	T	T	T
277	K	R	K	K	K	K
354	L	P	L	L	L	L
370	L	P	L	L	L	L

5.4 Discussion

In this chapter the construction of the first GLRaV-3 mini-replicon is described. This is the first report of the construction of a mini-replicon for a member of the genus *Ampelovirus*. An attempt to utilise the GLRaV-3 mini-replicon to evaluate the activity of the putative sg-promoter of sgRNA(ORF6) using a GUS gene expression assay is also reported.

It has been demonstrated that deletion mutants (including mini-replicons) of other woody plant viruses like CTV and GLRaV-2 were able to replicate in the absence of the genes located on the 3' halves of their genomes (Liu *et al.*, 2009; Satyanarayana *et al.*, 1999). The choice to forego the construction of a complete infectious clone and assemble a mini-replicon for GLRaV-3 first, was recognised as a high risk approach, but is justified in light of the published successes for other closteroviruses like BYV, CTV and GLRaV-2 (Liu *et al.*, 2009; Peremyslov *et al.*, 1998; Satyanarayana *et al.*, 1999).

A GLRaV-3 mini-replicon was assembled to contain the 5' UTR, ORF1a, ORF1b and the 3' UTR. These elements were placed under the control of a duplicated CaMV 35S promoter and a ribozyme with two unique restriction sites (SgfI and AflII) engineered between ORF1b and the 3' UTR.

Sequencing of the GLRaV-3 mini-replicon revealed several nucleotide and amino acid differences when compared to the GP18 sequence. By comparing the sequence of the GLRaV-3 mini-replicon to other GLRaV-3 isolate sequences (623, NY-1, CL-766 and 621) it is clear that some of these differences were shared with other GLRaV-3 isolate sequences. There are several possible explanations for these changes. Some of the viral sequences might be toxic to the bacterial intermediate host that could have resulted in spontaneous rearrangements, which might affect the ability of the GLRaV-3 mini-replicon to replicate in plants. The error prone nature of the RdRp can also be a contributing factor as it leads to a population of gRNA templates in the plant cell of which some are lethal mutants. All these templates have an equal chance to be amplified during RT-PCR and incorporated in the final assembly. Additionally, the plant material used as source material were propagated from the grapevine plant from where isolate GP18 was originally sequenced. Since the original RNA extraction and sequencing of GP18, the plants have been maintained for more the three years in a greenhouse facility with no seasonal fluctuations. The virus infection has thus been persistently replicating for the equivalent of six growing seasons. The variation observed in the GLRaV-3 mini-replicon is thus not

necessarily errors but potentially genetic drift accumulated over this time. Another contributing factor is the lack of proof-reading activity in the reverse transcriptase that might contribute to the differences observed in the GLRaV-3 mini-replicon by adding errors through slippage or by transcribing incorrect bases. Together, these two enzymes, reverse transcriptase and RdRp, can compound differences in the final assembly, compared to the original GP18 sequence, of which some might cause the construct to be non-replicating.

The evaluation of the putative sg-promoter for sgRNA(ORF6) was conducted with two GUSi containing GLRaV-3 mini-replicon constructs. The constructs used were the same except for the sg-promoter deletion mutant that was used as the negative control. It was expected that the sg-promoter present in the positive test construct will drive the expression of GUS in the infiltrated leaves. *Nicotiana bethamiana* leaves were co-infiltrated with a GUS construct together with a suppressor of silencing (P0). Leaves harvested 2 and 4 dpi for both constructs showed small foci of GUS expression with very little difference between the sg-promoter positive and negative constructs. Leaves taken 6dpi did not show clear GUS expression. This was unexpected as the negative control does not contain the sg-promoter sequence that is believed to be essential for GUS expression in this context. The GUS expression observed in the sg-promoter negative control cannot be explained and might indicate the presence of some other regulatory element remnant still present in the GLRaV-3 mini-replicon, which would make the GLRaV-3 mini-replicon unsuitable for use as a molecular tool to evaluate the activity of sg-promoters, in its current state. The inconsistency observed in the GUS expression as well as the GUS expression levels using these vectors were not optimal and needs further optimisation.

It is clear that the GLRaV-3 mini-replicon still requires several optimisation steps before it can be regarded as a useful molecular tool to study GLRaV-3 or possibly even functional genomics of the host, *V. vinifera*. Several aspects of the GLRaV-3 mini-replicon can be further investigated to try and improve replication of the construct and expression of a reporter gene. Bioinformatic analysis did not reveal why basal levels of GUS expression were observed with the sg-promoter negative control constructs, but ideally, if any such elements could be identified and removed it might be possible to have a complete negative control construct. It was clear from the experiment described in this chapter that we were unsuccessful in detecting sg-promoter activity for the region tested (nucleotides 13651 to

13800). This might be due to the inability of the GLRaV-3 mini-replicon to replicate or that the size of the sg-promoter region being evaluated was too small. The possibility that there are other factors involved for the sg-promoter to be functional, like tissue specific requirements or additional proteins of host or viral origin, cannot be discounted. The first steps would however be to extend the test area upstream of the 5' transcription initiation site for the sg-promoter for sgRNA(ORF6) and to also include other putative sg-promoters in the evaluation. Once activity has been detected deletion studies could be performed to narrow in on the sg-promoter sites. The use of silencing suppressors has been shown to be essential for the active replication of closteroviral mini-replicon constructs (Liu *et al.*, 2009; Peremyslov *et al.*, 1998; Satyanarayana *et al.*, 1999). The P0 silencing suppressor from BMYV was used as it has been demonstrated to be very effective as a heterologous viral suppressor and enhancing the replication of GVA infectious clones in *N. bethamiana* (Du Preez, 2010). In a study by Chiba *et al.* (2006) the effect of five heterologous viral suppressors were evaluated for a BYV mini-replicon. It was demonstrated that heterologous viral suppressors can potentially be more effective than native viral suppressors. The efficacy of heterologous viral suppressors (like GLRaV-2, p24) to enhance replication of the GLRaV-3 mini-replicon needs to be evaluated. The native suppressors for GLRaV-3 have not been determined yet and candidates are currently being tested. The role of the natural host, *V. vinifera* in the replication cycle of the virus and its infectious clone derivatives should not be underestimated. The use of a model plant like *N. bethamiana* can potentially lead to an incomplete view of the mechanisms involved in viral replication. It was found by Liu *et al.* (2009) while investigating the tandem papain-like leader proteases using GLRaV-2 mini-replicon constructs that the activity of the L1 and L2 proteases are affected by the host plant.

If the GLRaV-3 mini-replicon is still not able to replicate after all these measures have been taken a *de novo* assembly strategy using population cloning should be considered (Yu and Wong, 1998). The rationale, in first constructing a mini-replicon before a full-length cDNA clone should also be re-evaluated.

In this chapter we report the first steps towards the construction of a functional GLRaV-3 mini-replicon for use as a molecular tool in the investigation of the replication mechanisms of GLRaV-3.

5.5 Acknowledgements

5.6 References

Ahlquist, P., French, R., Janda, M., Loesch-Fries, L.S., 1984. Multicomponent RNA plant virus infection derived from cloned viral cDNA. Proc. Natl. Acad. Sci. USA. 81, 7066-7070.

Annamalai, P., Rao, A.L.N., 2006. Delivery and expression of functional viral RNA genomes in planta by agroinfiltration. Current Protocols In Microbiology 1 16B, 2.1-2.15.

Boyer, J., Haenni, A., 1994. Infectious transcripts and cDNA clones of RNA viruses. Virology 198, 415-426.

Chiba, M., Reed, J.C., Prokhnevsky, A.I., Chapman, E.J., Mawassi, M., Koonin, E.V., et al, 2006. Diverse suppressors of RNA silencing enhance agroinfection by a viral replicon. Virology 346, 7-14.

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

Dolja, V.V., Kreuze, J.F., Valkonen, J.P.T., 2006. Comparative and functional genomics of closteroviruses. Virus Res. 117, 38-51.

Du Preez, J., 2010. The development and characterisation of grapevine virus-based expression vectors. University of Stellenbosch, PhD Thesis.

Higuchi, R., Krummel, B., Saiki, R., 1988. A general method of in vitro preparation and specific mutagenesis of DNA fragments: Study of protein and DNA interactions. Nucleic Acids Res 16, 7351-7367.

Jooste, A.E.C., Maree, H.J., Bellstedt, D.U., Goszczynski, D.E., Pietersen, G., Burger, J.T., 2010. Three Grapevine leafroll-associated virus 3 (GLRaV-3) variants identified from south african vineyards show high variability in their 5'UTR. Arch. Virol., DOI 10.1007/s00705-010-0793-y.

Klaassen, V.A., Boeshore, M.L., Koonin, E.V., Tian, T., Falk, B.W., 1995. Genome structure and phylogenetic analysis of Lettuce infectious yellows virus, a whitefly-transmitted, bipartite closterovirus. Virology 208, 99-110.

Liu, Y., Peremyslov, V.V., Medina, V., Dolja, V.V., 2009. Tandem leader proteases of Grapevine leafroll-associated virus-2: Host-specific functions in the infection cycle. Virology 383, 291-299.

Maree, H.J., Freeborough, M.J., Burger, J.T., 2008. Complete nucleotide sequence of a south african isolate of Grapevine leafroll-associated virus 3 reveals a 5′ UTR of 737 nucleotides. Arch. Virol. 153, 755-757.

Maree, H.J., Gardner, H.F.J., Freeborough, M., Burger, J.T., 2010. Mapping of the 5' terminal nucleotides of Grapevine leafroll-associated virus 3 sgRNAs. Virus Res., Manuscript Submitted.

Peremyslov, V.V., Hagiwara, Y., Dolja, V.V., 1998. Genes required for replication of the 15.5-kilobase RNA genome of a plant closterovirus. J. Virol. 72, 5870-5876.

Satyanarayana, T., Gowda, S., Boyko, V.P., Albiach-Marti, M.R., Mawassi, M., Navas-Castillo, J., et al, 1999. An engineered closterovirus RNA replicon and analysis of heterologous terminal sequences for replication. Proc. Natl. Acad. Sci. USA 96, 7433-7438.

Shintaku, M.H., Carter, S.A., Bao, Y., Nelson, R.S., 1996. Mapping nucleotides in the 126-kDa protein gene that control the differential symptoms induced by two strains of Tobacco mosaic virus. Virology 221, 218–225.

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, 1424-1428.

Yu, H.H., Wong, S.M., 1998. Synthesis of biologically active cDNA clones of cymbidium mosaic potexvirus using a population cloning strategy. Arch. Virol. 143, 1617–1620.

5.7 Recommended internet resources

http://www.genelink.com/tools/gl-oe.asp

Chapter 6: Conclusion

Grapevine leafroll disease (LRD) is the most widely spread of all the grapevine viral diseases and probably also the most economically important. Advances in research have been hampered by the inherent difficulties associated with studying woody plant viruses that are phloem limited. Grapevine leafroll-associated virus 3 (GLRaV-3), the main causative agent of LRD in South Africa, has no known herbaceous host. Such an experimental host would greatly assist studies on the replication mechanisms of GLRaV-3 and its interactions with the natural host. Current research on GLRaV-3 is mainly focussed on epidemiology and the development of detection techniques. The genome sequence of GLRaV-3 was only first reported in 2004 (Ling *et al.*, 2004) and even though sgRNAs were detected by several research groups, it was not further characterised. The study presented in this thesis focussed on the further characterisation of the gRNA and sgRNA of GLRaV-3. The sequencing of a South African isolate of GLRaV-3, GP18, as well as mapping the 5' nt of seven sgRNAs is presented. The construction of a GLRaV-3 mini-replicon to study viral replication is also described.

The complete genome of GLRaV-3 isolate GP18 was sequenced and found to have a 5' UTR of 737 nt, extending 579 nt beyond the sequence reported for the NY-1 isolate. The sequence was compiled from combined data generated by three RT-PCR based techniques. Unlike, Ling *et al.* (2004) that used Poly(A)-tailing on dsRNA, the 5' end of the GP18 genome was determined using RLM-RACE on total RNA. The complete genome sequence of isolate GP18 was found to be 18498 nucleotides long with a 93% sequence identity with the NY-1 sequence over nucleotides 580-18498. The GP18 5' UTR has an unusual sequence composition that suggested that it is highly structured. To ascribe function to this large 5' UTR would be speculative, and warrants further investigation. Phylogenetic analysis, using a partial sequence of the Hsp70h gene, grouped isolate GP18 with genetic variant 623, representing the most abundant genetic variant group in South Africa. This group is separate from the groups that contain isolates NY-1 and 621, and isolate PL20.

Analysis of short read sequence data generated by metagenomic sequencing of a diseased vineyard identified GLRaV-3 as the most abundant virus in the environmental sample, with a GP18-like genetic variant of GLRaV-3 being the most dominant. Additional *de novo* assemblies, using a range of parameter settings, and contig assemblies were

performed to generate larger contigs representing near complete GLRaV-3 genomes. Contig-3843 was found to represent a nearly complete GLRaV-3 genome with 97% homology to isolate GP18. The *de novo* assemblies from this data most likely represent chimeric contigs from a range of templates that are closely related. Even though this chimeric contig cannot be directly compared to the GP18 isolate, it still provides us with a sequence that is closely related to GP18 that corroborates the existence of the extended 5' UTR.

The unexpected result of the extended 5' UTR required further investigation and confirmation. RLM-RACE was performed on a vine that was infected by three different viruses to prove that the extended 5' end is not due to an experimental artefact. Sequence data from the amplicons generated confirmed the 5' ends of GLRaV-2 and GRSPaV as published, as well as the 5' end determined for GP18. To determine if all isolates of GLRaV-3 contain this extended 5' UTR a "spanning RT-PCR" was designed that amplified a portion of the extended 5' UTR found for GP18 and the first 365 nt of the known sequence of NY-1. Sequencing results confirmed the extended 5' UTR in all the samples tested.

The presence of such an unusual 5' UTR was further supported by the collaborative project with Jooste *et al.* (2010) in which the 5' ends of three genetic variant isolates were determined by RLM-RACE. All three variants were found to contain 5' UTRs similar in length to GP18, but with great sequence diversity between variants and high conservation within a genetic variant group. These different 5' UTRs were analysed and predicted to have a complex secondary structure with multiple conserved stems and loops. Interpretation of this highly complex structure is not possible without an analogous example or biological data.

The sgRNAs produced by GLRaV-3 during its replication cycle were further characterised by mapping the 5' nt of seven sgRNAs using RLM-RACE. The 5' nts of three sgRNAs, identified by dsRNA blotting with a probe directed at the coat protein gene, associated with ORFs 3-6 as well as four additional putative sgRNAs associated with ORFs 7-12 were determined.

Amplicons generated in the RLM-RACE reactions were sequenced and analysed, and the 5' nt of each of the sgRNAs determined. Primer sets were unable to determine if ORFs 3 and 4 are expressed form the same or from different sgRNAs. The 5' nts for sgRNA(ORF3/4), sgRNA(ORF5) and sgRNA(ORF6) were determined to be G-9898, G-

12185 and A-13800, respectively. It is believed that ORFs 7-12 are also translated via sgRNA similarly to other closteroviruses like BYV and CTV (He *et al.*, 1997; Hilf *et al.*, 1995). Analysis of sequenced data from amplicons designed to determine the 5' nts of these putative sgRNAs associated with ORFs 7-12 predicted the 5' ends of four sgRNAs. ORFs 7, 8 and 9 are believed to be translated from their own putative sgRNAs (designated sgRNA(ORF7), sgRNA(ORF8) and sgRNA(ORF9)) with 5' ends mapped at G-14815, A-16273, and G-16754, while ORFs 10-12 are likely translated from the same sgRNA with the 5' end of putative sgRNA(10-12) mapped at A-17265.

The seven positive sense sgRNAs predicted by RLM-RACE are most likely 3' co-terminal with purines at the 5' ends. Similar to CTV, all the amplicons sequenced indicated that the sgRNAs were continuous with the genome with no common 5' leader sequence as observed in the evolutionary related Nidoviruses (Karasev *et al.*, 1997). In this study we were unable to determine if ORFs 3 and 4 and ORFs 10-12 are expressed from separated sgRNAs or from the same polysistronic sgRNAs. This is an interesting finding and warrants an in depth investigation.

Bioinformatic analyses were unable to identify any conserved sequences or secondary structures upstream of the mapped 5' ends, which might provide some information on the position, function or structure of any possible sg-promoters or transcription control elements. As more sequence data becomes available and bioinformatic databases are expanded and tools refined, re-evaluation of these sequences might be able to identify such elements.

The close evolutionary relation of ampeloviruses and closteroviruses, allows us to draw from the research performed on CTV and apply it to GLRaV-3. Research on CTV has made great progress in elucidating its infection cycle and the role of sgRNAs in replication, despite having similar difficulties than with GLRaV-3. The construction of an infectious clone and its deletion mutants for CTV proved to be a valuable molecular tool to elucidate the viral replication mechanisms. The construction of a GLRaV-3 infections mini-replicon was initiated in an attempt to investigate the replication mechanisms of GLRaV-3 by evaluating the activity of the putative sg-promoters.

The mini-replicon was constructed to contain the 5' UTR, ORF1a, ORF1b and the 3' UTR of GLRaV-3, isolate GP18, under the control of a duplicated CaMV 35S promoter and a hammerhead ribozyme. Sequencing of the GLRaV-3 mini-replicon revealed several nucleotide and amino acid differences when compared to the GP18 sequence, none

affecting the ORFs. Some of these differences were found to be conserved when compared to other GLRaV-3 sequences and in Chapter 5 several possible explanations for these changes are explored, but ultimately the effect on infectivity has to be determined experimentally.

To evaluate the activity of the putative sg-promoter for sgRNA(ORF6) a GUS reporter GLRaV-3 mini-replicon construct under the control of the putative sg-promoter was made. *Nicotiana bethamiana* leaves were co-agroinfiltrated with a GUS construct and suppressor of silencing. It was expected that the sg-promoter present in the positive test construct will drive the expression of GUS in the infiltrated leaves, unfortunately the sg-promoter deletion mutant, used as the negative control, was also able to express GUS. This makes the GLRaV-3 mini-replicon unsuitable for use as a molecular tool to evaluate the activity of sg-promoters in its current state and requires several optimisation steps. Once such a tool has been optimised it will proof useful to study GLRaV-3 or possibly even the host, *V. vinifera*.

This study expanded the knowledge base of GLRaV-3 through investigations into the viral genomic and subgenomic RNAs. Our understanding of the genome of GLRaV-3 was altered by the sequencing of a South African isolate that indicated, for the first time, that the 5' UTR extended further than originally indicated. Apart from several references to the presence of sgRNAs, little research had been conducted on the sgRNA of GLRaV-3. Here we describe our efforts to extend our knowledge of these sgRNAs by mapping the 5' nt of seven positive sense sgRNAs using RLM-RACE.

In future studies, it would be interesting to investigate the role in viral replication of the large 5' and intergenic UTRs found in GLRaV-3. Repair of the GLRaV-3 mini-replicon to a functional molecular tool and optimisation of transfections protocols would greatly assist these investigations. An in-depth study into the sgRNAs involved in the translation of ORFs 3 and 4, and ORFs 10 to 12 would be of interest. Further investigations on the possible silencing suppressor activity of ORFs 8, 9 or 10 are ongoing and if found to be more effective than GLRaV-2, p24 will be used in co-infiltration experiments. Advances in *V. vinifera* transfection will allow future experiments to be conducted not only in an experimental host but also in the natural host.

References

He, X., Rao, A.L.N., Creamer, R., 1997. Characterization of Beet yellows closterovirus-specific RNAs in infected plants and protoplasts. Phytopathology 87, 347-352.

Hilf, M.E., Karasev, A.V., Pappu, H.R., Gumpf, D.J., Niblett, C.L., Garnsey, S.M., 1995. Characterization of Citrus tristeza virus subgenomic RNAs in infected tissue. Virology 208, 576-582.

Jooste, A.E.C., Maree, H.J., Bellstedt, D.U., Goszczynski, D.E., Pietersen, G., Burger, J.T., 2010. Three Grapevine leafroll-associated virus 3 (GLRaV-3) variants identified from South African vineyards show high variability in their 5'UTR. Archives of Virology, DOI 10.1007/s00705-010-0793-y.

Karasev, A., Hilf, M., Garnsey, S., Dawson, W., 1997. Transcriptional strategy of closteroviruses: Mapping the 5'termini of the Citrus tristeza virus subgenomic RNAs. J. Virol. 71, 6233-6236.

Ling, K.S., Zhu, H.Y., Gonsalves, D., 2004. Complete nucleotide sequence and genome organization of Grapevine leafroll-associated virus 3, type member of the genus ampelovirus. J. Gen. Virol. 85, 2099-2102.