Molecular characterisation of South African isolates of grapevine fanleaf virus and a new, associated satellite RNA

by
Renate Luise Lamprecht

Dissertation presented for the degree of Doctor of Philosophy in the Faculty of Natural Sciences at Stellenbosch University

Supervisor: Prof J. T. Burger
Co-Supervisor: Dr D. Stephan

Faculty of Natural Sciences
Department of Genetics
December 2013

Declaration

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

December 2013

Declaration by the candidate:

With regards to Chapter 3 of this dissertation, which also appeared as a published article: "RL Lamprecht, HJ Maree, D Stephan, JT Burger (2012). Complete nucleotide sequence of a South African isolate of Grapevine fanleaf virus. Virus Genes, 45, 406-410." the nature and scope of my contribution were as follows:

Nature of contribution	**Extent of contribution(%)**
Design of sequencing strategy Complete nucleotide sequence determination of GFLV-SAPCS3 RNA1 and RNA2 Phylogenetic analysis Recombination analysis Write up of article	85%

The following co-authors have contributed to Lamprecht et al., 2012.

Name	**Email address**	**Nature of Contribution**	**Extent of contribution (%)**
Dr HJ Maree	hjmaree@sun.ac.za	Design of the 5'and 3' UTR sequence determination Preparation and critical reading of the manuscript	5%
Dr D Stephan	dstephan@sun.ac.za	Conceptual design of study Research funding Preparation and critical reading of the manuscript	10%
Prof JT Burger	jtb@sun.ac.za		

With regards to Chapter 4 of this dissertation, which also appeared as a published article: "RL Lamprecht, M Spaltman, D Stephan, T Wetzel, JT Burger (2013). Complete nucleotide sequence of a South African isolate of Grapevine fanleaf virus and its associated satellite RNA. Viruses, 5 (7), 1815-1823." the nature and scope of my contribution were as follows:

Nature of contribution	**Extent of contribution(%)**
Screening for and discovery of the two satRNAs Full-length sequence determination of the GFLV-SACH44 satRNA Completion of the GFLV-SACH44 RNA1 and RNA2 sequences Phylogenetic analysis Construction of the GFLV satRNA infectious clone Infectivity testing of the GFLV satRNA infectious clone Write up of article	75%

The following co-authors have contributed to Lamprecht et al., 2013.

Name	**Email address**	**Nature of Contribution**	**Extent of contribution (%)**
Ms M Spaltman	15660060@sun.ac.za	Produced the initial GFLV-SACH44 RNA1 and RNA2 sequencing data	5%
Dr D Stephan	dstephan@sun.ac.za	Conceptual design of study Preparationa and critical reading of the manuscript	5%

Dr D Wetzel	Thierry.wetzel@agroscience.rlp.de	Facilities at AlPlanta Research funding Infectivity testing of the satRNA infectious clones Northern blot analysis Critical reading of the prepared manuscript	10%
Prof JT Burger	jtb@sun.ac.za	Conceptual design of study Research funding Preparation and critical reading of the manuscript	5%

Signature of candidate:

Date: 15/10/2013

Declaration by co-authors:
The undersigned hereby confirm that
1. the declaration above accurately reflects the nature and extent of the contributions of the candidate and the co-authors to Chapter 6 (p 101-),
2. no other authors contributed to Chapters 6 (p 101-) besides those specified above, and
3. potential conflicts of interest have been revealed to all interested parties and that the necessary arrangements have been made to use the material in Chapter 6 (p 101-) of this dissertation.

Signature	**Institutional affiliation**	**Date**
	Department of Genetics, Stellenbosch University, South Africa	
	Department of Genetics, Stellenbosch University, South Africa	
	Department of Genetics, Stellenbosch University, South Africa	
	Department of Genetics, Stellenbosch University, South Africa	
	RLP Agroscience, AlPlanta – Institute for Plant Research, Breitenweg 71, 67435, Neustadt an der Weinstrasse, Germany	

Abstract

Grapevine fanleaf virus (GFLV) is one of the oldest, most widespread and devastating viruses infecting grapevine, and occurs globally where *Vitis vinifera* is grown. In South Africa (SA) GFLV is predominant in the Breede River Valley, one of the highest wine producing regions in SA. To date, only three GFLV isolates have been completely sequenced internationally, and limited sequence information is available for SA GFLV isolates. In this study, the first full-length GFLV genome sequence from a South African isolate, GFLV-SAPCS3, was determined. Full-length sequences were used for phylogenetic analysis and revealed that the SA isolates are separate from other sequenced GFLV isolates. Full-length sequences were also used to investigate putative intra- and interspecies recombination events involving GFLV-SAPCS3 RNA1 and RNA2 between GFLV and Arabis mosaic virus (ArMV) isolates. Using two different recombination analysis software packages, the most notable of the putative recombination events involving GFLV-SAPCS3 indicated that the GFLV-SAPCS3 RNA2 5' UTR might have evolved from an interspecies recombination event between GFLV-F13-type and ArMV Ta-type isolates. The presence of satellite RNAs (satRNA) associated with South African GFLV isolates was also investigated. In a collaborative study (see Chapter 4 for details), more than a 100 GFLV- infected grapevine plants were screened for satRNAs. SatRNAs were present in only two plants, containing isolates GFLV-SACH44 and GFLV-SACH47. The full-length nucleotide sequences of the GFLV-SACH44 genomic RNAs 1 and 2, and the associated satRNA were determined. No significant sequence variation could be detected between the GFLV isolates that had the presence of a satRNA and those that had not. The GFLV-SACH44 RNA2 5' UTR also had the same conserved sequence that was found in GFLV-SAPCS3, which suggests that GFLV-SACH44, like GFLV-SAPCS3, may have arisen from a common ancestor, which may have originated from an interspecies recombination event. The GFLV-SACH44 satRNA was found to be more closely related to the ArMV large satRNA than to the satRNA associated with GFLV-F13. A full-length cDNA clone of GFLV-SACH44 satRNA was constructed and its replication and systemic

spread in herbaceous hosts, when mechanically co-inoculated with two GFLV isolates as helper viruses, was demonstrated. Replication of the GFLV-SACH44 satRNA cDNA clone was however abolished when co-inoculated with an ArMV helper virus, even though it is phylogenetically more closely related to ArMV satRNAs. The full-length satRNA clones were modified to be used as vectors for expression and/or silencing of foreign genes, by inserting the green fluorescence protein (GFP) full-length or partial sequences downstream of the open reading frame of the satRNA. These constructs were cloned into a binary vector to allow for agro-infiltration into plants. Full-length cDNA clones of GFLV-SAPCS3 RNA1 and RNA2 were constructed to be used in conjunction with modified GFLV-SACH44 satRNA full-length clones. The full length GFLV-SAPCS3 RNA1 and RNA2 clones were however not infectious in *Nicotiana benthamiana* after agro-infiltration and therefore the evaluation of the modified satRNA expression and silencing constructs had to be aborted. Attempts to understand this failure revealed that, among other point mutations, four frameshifts had occurred in the RNA1 full-length clone, rendering the transcripts untranslatable, and hence non-infectious. Strategies to correct the mutations are discussed. Once these mutations have been corrected this study can continue in evaluating the use of the satRNA component for expression and silencing analysis.

Opsomming

Grapevine fanleaf virus (GFLV) is een van die oudste, mees wydverspreide en mees verwoestende virusse wat wingerd affekteer en word wêreldwyd waar *Vitis vinifera* verbou word, gevind. In Suid Afrika (SA) kom GFLV veral in die Breederivier vallei, een van die mees produktiewe wyn-produserende areas in SA, voor. Tot dusver is daar net drie GFLV isolate waarvan die volledige nukleïensuurvolgorde internasionaal bepaal is. Die nukleïensuurvolgorde informasie vir SA GFLV isolate is redelik beperk. In hierdie studie was die eerste volledige nukleïensuurvolgorde van ‘n SA GFLV isolaat, GFLV-SAPCS3, bepaal. Die volledige nukleïensuurvolgordes was vir filogenetiese analise gebruik, en vermeende intra- en interspesie rekombinasie gebeurtenisse, wat GFLV-SAPCS3 RNA1 en RNA2 betrek, tussen GFLV en Arabis mosaic virus (ArMV) isolate was ondersoek. Twee verskillende rekombinasie-analise sagteware programme was gebruik en die noemenswaardigste van die vermeende rekombinasie gebeurtenisse, met betrekking tot GFLV-SAPCS3, het aangedui dat die GFLV-SAPCS3 RNA2 5’ ontransleerde streek (UTR) waarskynlik van ‘n interspesie rekombinasie gebeurtenis tussen ‘n GFLV-F13-tipe en ‘n ArMV-Ta-tipe isolaat ontwikkel het. Die teenwoordigheid van satelliet RNAs (satRNAs), wat met SA GFLV isolate geassosieer is, was ook ondersoek. Meer as ‘n 100 GFLV ge-infekteerde wingerd plante was in ‘n samewerkingsprojek (sien Hoofstuk 4 vir besonderhede) getoets vir die teenwoordigheid van satRNAs. SatRNAs was net in twee plante teenwoordig, in isolate GFLV-SACH44 en GFLV-SACH47. Die vollengte nukleïensuurvolgordes van GFLV-SACH44 RNA1, RNA2 en geassosieerde satRNA was bepaal. Geen beduidende volgorde variasie tussen die GFLV isolate wat satRNAs bevat het, en die GFLV isolate sonder satRNA was waargeneem nie. Die GFLV-SACH44 RNA2 5’ UTR het ook die gekonserveerde volgorde, wat in GFLV-SAPCS3 teenwoordig was, gehad en dit dui daarop dat GFLV-SACH44, soos GFLV-SAPCS3, van dieselfde stamvader, wat tydens ‘n vorige rekombinasie gebeurtenis ontstaan het, mag ontwikkel het. Die GFLV-SACH44 satRNA was meer naverwant aan die ArMV satRNAs as aan die satRNA, wat met GFLV-F13. ‘n Vollengte cDNA kloon van die GFLV-SACH44 satRNA was ontwikkel

en die replisering en sistemiese verspreiding in sagte plante, nadat dit met twee GFLV isolate as helper virusse saam ge-inokuleer was, was gedemonstreer. Replisering van die GFLV-SACH44 satRNA cDNA kloon was egter ontwrig toe dit saam met ‘n ArMV helper virus saam ge-inokuleer was, al is dit filogeneties meer verwant aan ArMV satRNAs. Die vol-lengte satRNA klone was gemodifiseer om as vektore vir uitdrukking en/of uitdowing van transgene te dien, deur om vol-lengte of gedeeltelike groen fluoressensie proteïen (GFP) nukleïensuurvolgordes aan die einde van die satRNA leesraam te koppel. Hierdie konstrukte was in ‘n binêre vektor gekloon om agro-infiltrasie in plante toe te laat. Vol-lengte cDNA klone van GFLV-SAPCS3 RNA1 en RNA2 was ontwikkel om in samewerking met die gemodifiseerde GFLV-SACH44 satRNA konstrukte gebruik te word. Die vol-lengte GFLV-SAPCS3 RNA1 en RNA2 klone het egter nie in *Nicotiana benthamiana* gerepliseer na agro-infiltrasie nie, daarom was die evaluasie van die gemodifiseerde satRNA konstrukte gestaak. Pogings om die mislukking te verstaan, het daarop gewys dat, behalwe punt mutasies, vier leesraam versteurings in die RNA1 vollengte kloon voorgekom het, wat ontransleerbare transkripte, en dus nie-repliserende konstrukte tot gevolg gehad het. Strategieë om die mutasies te korrigeer is bespreek. Sodra die mutasies gekorrigeer is, kan die studie voortgaan om te evalueer of die satRNA komponent vir uitdrukking en uitdowing analise gebruik kan word.

List of Abbreviations

A	Adenine
aa	amino acid
Ala	Alanine
Amp	Ampicillin
AMV	Avian Myeloblastosis Virus
Arg	Arginine
ArMV	Arabis mosaic virus
BaMV	Bamboo mosaic virus
BLAST	Basic Local Alignment Search Tool
BMV	Brome Mosaic Virus
bp	basepair
C	Cytosine
CaMV	Cauliflower mosaic virus
cDNA	copy DNA
CMV	Cucumber mosaic virus
CP	coat protein
CPMV	Cowpea mosaic virus
cv	cultivar
Cys	Cysteine
DAS-ELISA	double-antibody sandwich enzyme linked immunosorbent assay
DNA	deoxyribonucleotide acid
dNTP	deoxyribonucleotide triphosphate
dpi	days post inoculation
dsRNA	double stranded RNA
ER	endoplasmic reticulum
G	Guanine
GBLV	Grapevine Bulgarian latent virus
GCMV	Grapevine chrome mosaic virus
GDef	Grapevine deformation virus
GDP	gross domestic product
GFLV	Grapevine fanleaf virus
GFP	green fluorescent protein
GLRaV-2	Grapevine leafroll-associated virus 2
GLRaV-3	Grapevine leafroll-associated virus 3
Glu	Glutamine
Gly	Glycine
GRSPaV	Grapevine rupestris stempitting associated virus
GTRSV	Grapevine Tunisian ringspot virus
GVA	Grapevine virus A
HEL	Helicase
HP	homing protein
ICTV	International Committee on Taxonomy of Viruses
ICVG	The International Council for the Study of Viruses and Virus-like Diseases of Grapevine
Kan	Kanamycin

kb	kilobase
LB	Luria Broth
MP	movement protein
ng	nanogram
nm	nanometer
nt	nucleotide
OD	optical density
ORF	open reading frame
P	polyprotein
PCA	polymerase cyclic assembly
PCR	polymerase chain reaction
PDS	Phytoene denaturase
Pro	proteinase
PTGS	post-transcriptional gene silencing
PVX	Potato virus X
RACE	rapid amplification of cDNA ends
RCA	rolling circle amplification
RdRp	RNA-dependent RNA polymerase
RF	replicative-form
RISC	RNA-induced silencing complex
RNA	Ribonucleic acid
rpm	revolutions per minute
RpRSV	Raspberry ringspot virus
RT	reverse transcriptase
RT-PCR	reverse transcription PCR
SA	South Africa
SAWIS	SA Wine Industry Information & Systems
siRNA	short interfering RNA
SMART	switching mechanism at the 5’ end of the RNA transcript
ss	single stranded
SVISS	satellite virus-induced silencing system
T	Thymine
Ta	annealing temperature
TBRSV	Tomato bushy stunt virus
T-DNA	transfer DNA
Tet	Tetracyclin
TMV	Tobacco mosaic virus
ToBRSV	Tomato blackring virus
ToRSV	Tomato ringspot virus
ToRSV	Tomato ringspot virus
TRSV	Tobacco ringspot virus
U	unit
UTR	untranslated region
VIGS	virus-induced gene silencing
VPg	virus genome-linked protein

Acknowledgements

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

Prof. Johan Burger, for his guidance and the opportunity to study in his prestigous group.

Dr. Dirk Stephan, for his technical expertise, guidance and words of encouragement.

Dr. Thierry Wetzel (collaborator, AlPlanta Insitute for Plant Research, Neustadt an der Weinstrasse, Germany), for his guidance and support during the research visit to AlPlanta, Germany, and for critical reading of this thesis.

Dr. Hano Maree, for always providing friendly advice and intellectual conversations.

Members and friends of the Vitis lab, for creating a stimulating and friendly atmosphere to work in, and all the social events.

The owners of De Wetshof, for allowing us to collect samples.

Frank Poole (Plant Health Diagnostic Services, Directorate Plant Health, National Department of Agriculture, Stellenbosch, South Africa) for providing *Nicotiana benthamiana* and *Chenopodium quinoa* seeds.

The personal financial assistance of the National Research Foundation (NRF) and Stellenbosch University is acknowledged. Opinions expressed and conclusions arrived at, are those of the authors and are not necessarily to be attributed to the NRF. The financial assistance of the Federal Ministry of Education and Research (BMBF) (Germany) towards this research is hereby also acknowledged.

And last, but not least, to my parents, family and friends for their moral support during all my years of study. I thank you all dearly.

Dedicated to Dr. Dirk Stephan

23 December 1972 – 28 January 2013

Contents

Chapter 1

Introduction

1.1 General introduction

Grapevine is cultivated worldwide primarily for wine production, which is a major contributor to the global economy. Grapevine cultivars occupy ~10 million hectares worldwide, growing in several types of soils and in different climates, and as a result this precious crop is subjected to many pests and pathogens. Viruses are some of the most destructive pathogens, as they are difficult to control. Grapevine fanleaf virus (GFLV) is one such virus and causes fanleaf degeneration disease which is one of the world's most severe grapevine diseases. However, in South Africa (SA), GFLV is not an economically important threat to the SA wine industry at present, and as a result, research on SA-isolates of GFLV has been lacking and still is limited. The natural nematode vector, *Xiphinema index,* is however present, suggesting that the potential threat cannot be ignored and that research on SA isolates of GFLV should be conducted. On an international level, molecular characterisation of these viruses are important to understand the variability among GFLV isolates and closely related viruses. Studying variability among viruses is key to understanding the origins of these viruses and how they keep evolving. A complete understanding of the molecular structure and variability among GFLV isolates is also imperative in developing control strategies against fanleaf degeneration disease, locally and internationally.

Studying sequence information alone is not adequate to understand the mode of action of a virus. Infectious clones based on wild-type viruses have been constructed to gain insight in host-pathogen relationships. To date, only a few grapevine viruses have been converted into infectious clones that can be used to study the virus in herbaceous hosts,

and only two of these grapevine virus infectious clones are able to infect grapevine. As a result, research on grapevine virus diseases is lagging behind those of other economically important fruit crops, such as citrus and apples. The availability of more infectious clones based on grapevine viruses would advance our knowledge of grapevine viruses and to combat their diseases.

The aim of this study was to determine the full-length sequences of South African GFLV isolates as there is a lack of GFLV full-length sequences in the public domain. We also investigated whether GFLV-associated satellite RNAs are present in South Africa, as very little is known about these molecular parasites. To further characterise these GFLV isolates, this study attempted to construct infectious clones based on the SA GFLV isolates. The study therefore has two main phases; sequencing of the complete genomes of two SA GFLV variants and the construction of infectious full-length cDNA clones. To achieve these goals the following objectives were set out:

- Determine the complete genome sequences of two SA isolates of GFLV.
- Determine the phylogenetic relationships between SA GFLV isolates and other described GFLV isolates.
- Develop a satellite RNA detection technique to survey grapevines for the natural occurrence of GFLV-associated satellite RNAs.
- If satellite RNAs are found, determine the full-length sequences and compare them to other characterised satellite RNAs.
- Construct full-length infectious cDNA clones based on the SA GFLV genomic and satellite RNA sequences.

1.2 Summary of thesis chapters

This thesis is divided into six chapters, including a general introduction and a literature review, followed by three research chapters and a general conclusion. Each chapter has its own introduction and conclusion, and is referenced individually.

Chapter 1: Introduction

In this chapter a general introduction to the thesis is given, which also includes the aims and objectives of the study, as well as a general breakdown of the chapters for the remainder of the thesis.

Chapter 2: Literature review

An overview of the literature relating to the history and importance of winemaking in South Africa, aspects of Grapevine fanleaf virus including the general classification of the virus, replication and cell-to-cell movement, transmission by vectors as well as the variability and recombination events between viral isolates. The remainder of the chapter is an overview on the development of plant virus infectious clones and their uses as plant based viral vectors for the heterologous expression of foreign genes, or as virus-induced gene silencing vectors for silencing of endogenous genes.

Chapter 3: Complete nucleotide sequence of a South African isolate of Grapevine fanleaf virus

In this chapter the sequencing of the South African GFLV isolate GFLV-SAPCS3 is described. Symptomatology of GFLV-SAPCS3 on the herbaceous hosts *Chenopodium quinoa* and *Nicotiana benthamiana* is described. Phylogenetic analysis and the putative recombination events involving GFLV-SAPCS3 are also described.

Chapter 4: A new satellite RNA associated with a Grapevine fanleaf virus isolate found in South Africa

In this chapter the discovery of GFLV-associated satellite RNAs is described. Satellite RNA detection techniques were developed that resulted in the detection of two new naturally GFLV-associated satRNAs in a South African vineyard. The satellite RNA of one of these isolates, GFLV-SACH44, was completely sequenced. A full-length infectious cDNA clone was constructed based on this satRNA, and infectivity in herbaceous hosts, following mechanical inoculation together with GFLV helperviruses, was demonstrated.

Chapter 5: Construction of a full-length cDNA clone of a South African isolate of Grapevine fanleaf virus and a satellite RNA

In this chapter the construction of full-length cDNA clones of GFLV-SAPCS3 RNA1 and RNA2, and modifications to the GFLV-SACH44 satRNA full-length infectious cDNA clone, are discussed. These clones were constructed to be agro-infiltrated into herbaceous hosts and to be used in future as expression and silencing vectors for *Vitis vinifera*.

Chapter 6: Concluding remarks and future prospects

This chapter includes concluding remarks and future prospects of the molecular tools that were developed in this study.

1.3 Research outputs and author contributions

1.3.1 Peer-reviewed publications

RL Lamprecht, HJ Maree, D Stephan, JT Burger (2012). Complete nucleotide sequence of a South African isolate of Grapevine fanleaf virus. Virus Genes, 45, 406-410.

RL Lamprecht, M Spaltman, D Stephan, T Wetzel, JT Burger (2013). Complete nucleotide sequence of a South African isolate of Grapevine fanleaf virus and its associated satellite RNA. Viruses, 5 (7), 1815-1823.

1.3.2 Conference proceedings

RL Lamprecht, D Stephan, HJ Maree, T Wetzel, JT Burger. Molecular characterization of South African isolates of Grapevine fanleaf virus and an associated new satellite RNA. Presented at the 17th Congress of the International Council for the Study of Virus and Virus-like Diseases of the Grapevine (ICVG), held at the University of California Davis Campus, California, USA, 7-14 October 2012.

RL Lamprecht, D Stephan, HJ Maree, T Wetzel, JT Burger. Molecular characterization of South African isolates of Grapevine fanleaf virus and an associated new satellite RNA. Presented at the 34th South African Society for Enology and Viticulture (SASEV) conference, held at Allee Bleue, Simondium, Cape Town, South Africa, 14-16 November 2012.

1.3.3 Posters

RL Lamprecht, D Stephan, J Du Preez, HJ Maree and JT Burger. Molecular characterisation of South African isolates of Grapevine fanleaf virus. Presented at The 32nd South African Society for Enology and Viticulture Conference, held at the Lord Charles Hotel, Somerset West, South Africa. 18-19 November 2010.

Chapter 2

Literature review

2.1. Introduction

The grape is a very important food crop and grapes are internationally traded as table grapes (fresh, or dried as raisins), for the production of grape juice, but mostly for the production of wine. The use of grapes for wine making, or viniculture, is a long process using several techniques, and humans as early as in the Neolithic period (8500-4000 BC) already had the knowledge and tools for viniculture (McGovern, 2004). *Vitis vinifera* is the most common species used for table and wine grape production, while other *Vitis* species, including *V. labrusca*, *V. riparia*, *V. rotundifolia*, *V. aestivalis* and *V. mustagenesis* are also used for table and/or wine grape production.

Vitis vinifera is a very important agricultural commodity in South Africa (SA), and internationally. The first vineyards were planted in SA by the Dutch settler, Jan van Riebeeck, in 1655, ordered by the Dutch East Indian Company. The vineyards were originally planted to help the sailors on the spice route ward off scurvy (www.sahistory.org.za). In recent years, SA wines have become highly competitive on the international markets. Wine is produced in four out of the nine provinces in SA, of which the Western Cape produces the most wine. There are six main grapevine growing and wine producing regions in the of the Western Cape Province (Figure 2.1); the Breede River Valley, Cape South Coast, Coastal Region, Klein Karoo, Olifants River and Boberg. The Breede River Valley contributes approximately 30% of the total

production of the local wine industry, and includes Worcester, which produces close to 27% of South Africa's total volume of wine and spirits (www.wosa.co.za/sa).

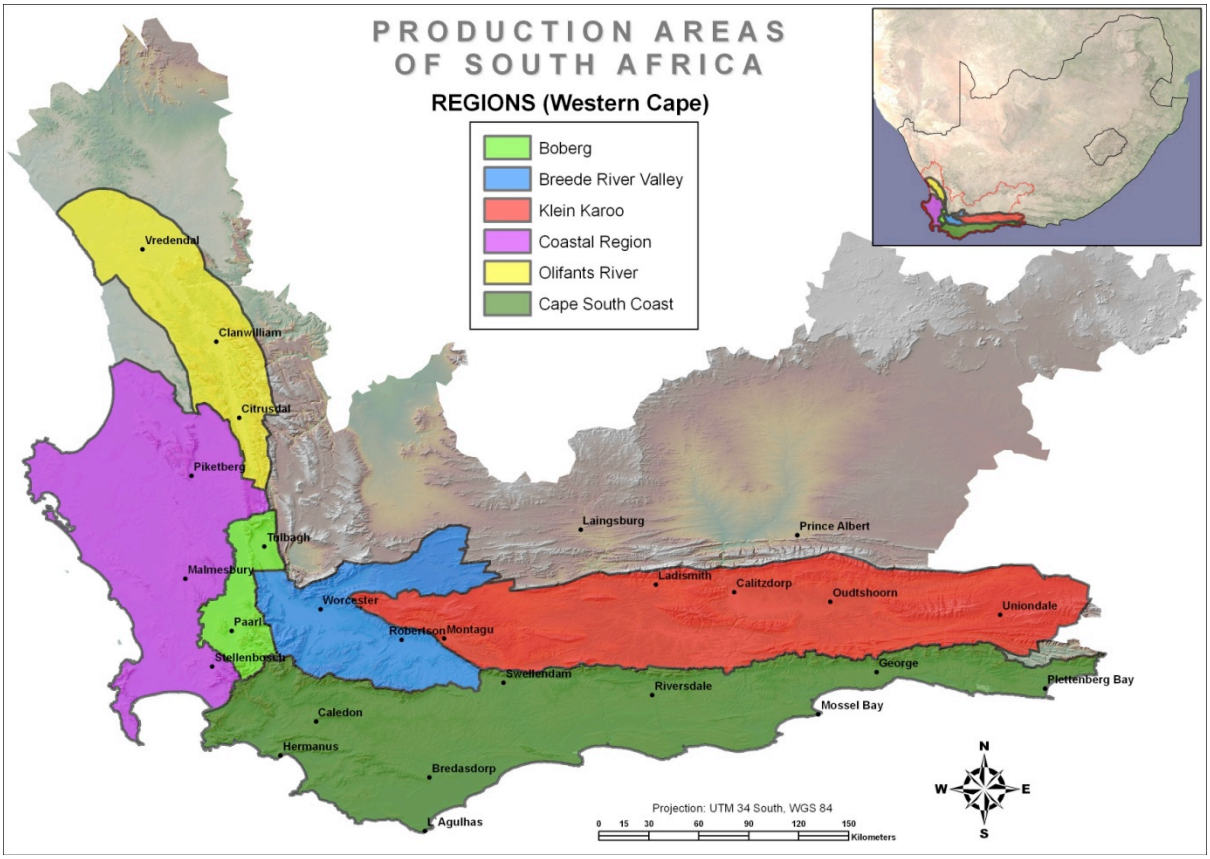


Figure 2.1: Map of South Africa indicating the six major wine growing regions in the Western Cape Province. (http://www.sawis.co.za/cert/productionareas.php)

The SA wine industry is showing a sustainable and increasingly positive trend over recent years. According to a macro-economic impact study, commissioned by SA Wine Industry Information & Systems (SAWIS) and published in December 2009 (http://www.wosa.co.za/sa/stats_worldwide.php), the wine industry contributed R26.2 billion to the regional economy (R4.3 billion was generated indirectly through wine-tourism activities centered in the Cape winelands), which amounts to 2.2% of the total gross domestic product (GDP) of SA in 2008. Of the R26.2 billion GDP created in SA by the wine industry, about R14.2 billion remained in the Western Cape Province (approximately 54%) to benefit its residents. In 2008, the wine industry supported employment opportunities of 275 606 people, of which 58% were unskilled, 29% semi-

skilled and 13% skilled. In terms of production, SA ranks as number eight in overall volume production, and produces 3.8% of the world's wine (2011). South African wine exports have grown by 219% between 1998 and 2010, and produced 409.0 million litres of wine for export in 2012. This illustrates the exceptional ability of the wine industry as a creator of economic growth for SA (SAWIS figures, 2008-2013).

Vitis vinifera is globally dispersed and is unfortunately susceptible to a wide variety of plant pathogenic fungi, bacteria and viruses. These pathogens affect the plant's health and also affect the yield and quality of the fruit. It was recently reported that 63 different grapevine-infecting viruses have been described. None of the other woody crops are affected by such a high number of viruses (Martelli, 2012). These viruses either cause disease individually, or, more often, as a complex of viruses. One such virus, Grapevine fanleaf virus (GFLV), is the most globally dispersed of all grapevine-infecting viruses (Andret-Link et al., 2004). Grapevine fanleaf virus is the causative agent of grapevine degeneration disease, but it was also recently linked to Chardonnay translucent vein clearing disease, a devastating disease, which is a mixed infection of GFLV, Grapevine rupestris stem pitting-associated virus (GRSPav) and Tomato ringspot virus (ToRSV) (Qiu et al., 2004). In SA, GFLV is prevalent in the Breede River valley and is at present not a serious threat to the SA wine industry. As a result, research on SA GFLV isolates is limited compared to the other more economically important grapevine viruses, such as Grapevine Leafroll-associated virus 3 (GLRaV-3) and Grapevine virus A (GVA). Molecular characterisation of globally dispersed GFLV isolates is important to determine the variability among different isolates. Studying the sequence variability among GFLV isolates is important for understanding how these viruses have evolved and continue to evolve (Arenal et al., 2001), and may aid in the development of better control strategies against GFLV.

Knowledge of genome sequences alone is not adequate to study viruses. A recent shift in approach has been the construction of full-length infectious clones of plant RNA viruses. With these tools researchers are able to modify genomes in order to gain insight into the molecular structure, replication strategies and gene function of these

viruses. Several grapevine virus infectious clones that have been shown to be infectious in herbaceous hosts are available. Only two of these infectious clones, based on GVA (Muruganantham et al., 2009) and Grapevine Leafroll-associated virus 2 (GLRaV-2) (Kurth et al., 2012) were however infectious in *Vitis vinifera*. Both these clones were successfully converted to virus-induced gene silencing (VIGS) vectors for functional genomics in grapevine. These molecular tools are invaluable in studying grapevine viruses and are also used to decipher the wealth of genomic data that are currently generated through new sequencing technologies.

This literature review is divided in two main sections. In the first section GFLV is discussed with regards to its genome organisation, replication and transmission by vectors. The genetic variability and recombination events between different isolates are also discussed. The second section gives an overview of the construction and importance of infectious clones, and how infectious clones are converted to expression and/or VIGS vectors.

2.2. Grapevine fanleaf virus

Grapevine fanleaf disease was first reported in *V. vinifera* in the 1880s (Rathay 1882, 1883) and is one of the oldest and most devastating viral diseases of grapevines (Hewit, 1954; Andret-Link et al., 2004). Grapevine fanleaf virus causes fanleaf disease that includes a wide range of symptoms in grapevine. Fanleaf disease, or also known as grapevine degeneration disease, is characterised by malformations of the leaves, such as open marginal and petiolar sinuses, prominent marginal teeth (hence the name fanleaf), asymmetrical blades and irregular veins. It also causes malformations in the canes, such as uneven internode spacing, double nodes, zigzag growth and abnormal branching. The disease also causes dramatic chlorotic discolouration on the leaves, such as mottling, yellowing and line patterns. The disease causes the berries to be reduced in number and size, to ripen irregularly and the fruit quality is also affected by a decrease in sugar content (Hewit, 1954; Martelli and Savino, 1988). The disease occurs worldwide where *V. vinifera* is cultivated and is responsible for significant economic

losses by reducing crop yields by as much as 80%. The prevalence of the disease is dependent on the virulence of the isolate, the susceptibility of the grapevine variety, as well as several environmental factors (Andret-Link et al., 2004).

In South Africa, GFLV is mostly restricted to the Breede River valley in the Western Cape due to the prevalence of its nematode vector, *Xiphinema index,* in this region. The disease is spread in this area through nematode-infested irrigation water (Malan and Hugo, 2003). Three distinct types of grapevine fanleaf symptoms (Figure 2.2) were observed on grapevines grown in the Breede River valley and Stellenbosch areas: yellow mosaic, fanleaf and vein banding (Liebenberg et al., 2009).

Figure 2.2: Leaf symptoms on GFLV-infected grapevine plants observed in the Western Cape in SA; A) Yellow mosaic, B) Fanleaf and C) Vein banding (Liebenberg et al., 2009).

2.2.1. Taxonomy and morphology

Grapevine fanleaf virus belongs to the genus *Nepovirus* in the family *Secoviridae*. *Secoviridae* is a newly assigned family of plant viruses in the order *Picornaviridae* that includes the genera *Comovirus*, *Fabavirus*, *Nepovirus*, *Cheravirus*, *Sadwavirus*, *Sequivirus* and *Waikavirus* (Sanfacon et al., 2009). Nepoviruses, or nematode transmitted polyhedral viruses, have non-enveloped icosahedral virions that are 28 nm in diameter (Figure 2.3). Incomplete virus particles, or empty capsids, are often present (indicated in as the darker particles in Figure 2.3). More than 46 viruses are classified in this genus, with Tobacco ringspot virus (TRSV) as the type member. Nepoviruses are

divided into three subgroups based on the sizes of their RNA2. Species that infect grapevine in subgroup A are GFLV, Arabis mosaic virus (ArMV), TRSV, Grapevine deformation virus (GDefV) and Raspberry ringspot virus; in subgroup B are Artichoke Italian latent virus, Grapevine chrome mosaic virus, Grapevine Anatolian ringspot virus and Tomato black ring virus; and in subgroup C are Blueberry leaf mottle virus, Cherry leaf roll virus, Grapevine Bulgarian latent virus, Grapevine Tunisian ringspot virus, Peach rosette mosaic virus and ToRSV (Digiaro et al., 2007).

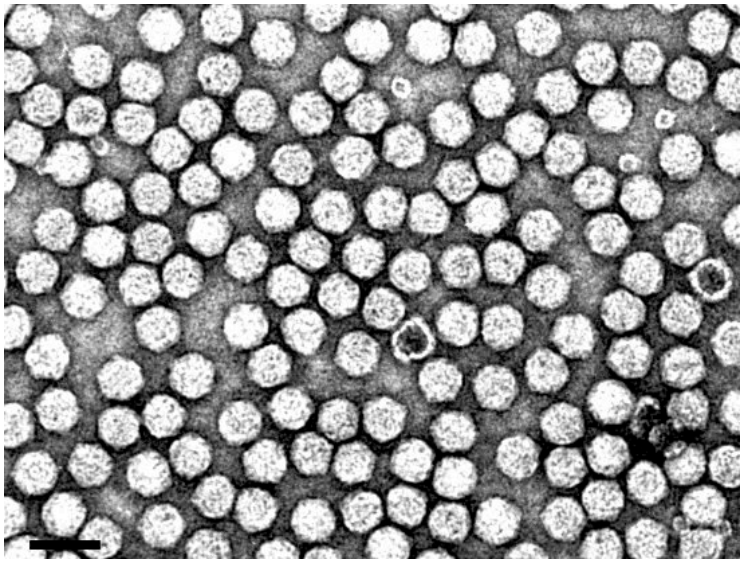


Figure 2.3: Electron micrograph showing the isometric particles of Grapevine fanleaf virus. Bar = 50nm (http://www.dpvweb.net/dpv/showdpv.php?dpvno=385)

2.2.2. Genome structure

The GFLV genome is composed of two single-stranded positive-sense linear RNA fragments, RNA1 and RNA2 (Quacquarelli et al., 1976). RNA1 encodes proteins necessary for replication, while RNA2 codes for products involved in cell-to-cell movement and coating of the viral RNAs (Ritzenthaler et al., 2002). Each RNA has a single open reading frame (ORF) that encodes its own polyprotein, which is in turn proteolytically cleaved into functional proteins by the RNA1 encoded protease (Margis et al., 1993). Both RNAs are covalently attached to a small protein (VPg) at its 5' end

(Pinck et al., 1988) and polyadenylated at its 3' end (Serghini et al., 1990; Ritzenthaler et al., 1991).

RNA1 is approximately 7.3 kb in length and contains one ORF of 6855 nucleotides (nt) (Figure 2.4). RNA1 includes genes for the maturation products of the proteinase co-factor (also known as protein 1A), helicase ($1B^{Hel}$), the viral genome-linked protein ($1C^{VPg}$), proteinase ($1D^{Pro}$) and the RNA-dependendent RNA polymerase ($1E^{Pol}$), which are all generated by *cis* processing at Cys/Ala, Cys/Ser, Gly/Glu and Arg/Gly cleavage sites, respectively (Pinck et al., 1991; Ritzenthaler et al., 1991; Margis et al., 1994). These proteins are the only ones required for RNA1 replication (Viry et al., 1993), and function in *trans* for RNA2 replication.

RNA2 is approximately 3.8 kb in length and contains one ORF of 3330 to 3333 nt (Figure 2.4). The maturation products of RNA2 are the homing protein ($2A^{HP}$), the movement protein ($2B^{MP}$) and the coat protein ($2C^{CP}$), which are generated by *trans* processing at Cys/Ala and Arg/Gly cleavage sites respectively (Quacquarelli et al. 1976; Pinck et al., 1988; Serghini et al., 1990; Ritzenthaler et al., 1991; Margis et al., 1993). Protein $2A^{HP}$ is required for RNA2 replication together with the RNA1-encoded proteins, and acts as a homing protein by leading the RNA2 to the virus replication sites (Ritzenthaler et al., 2002). Protein $2B^{MP}$ is important for the formation of tubular structures through which viral particles are delivered to adjacent uninfected cells (Ritzenthaler et al., 1995), and $2C^{CP}$ assembles into the virus capsid.

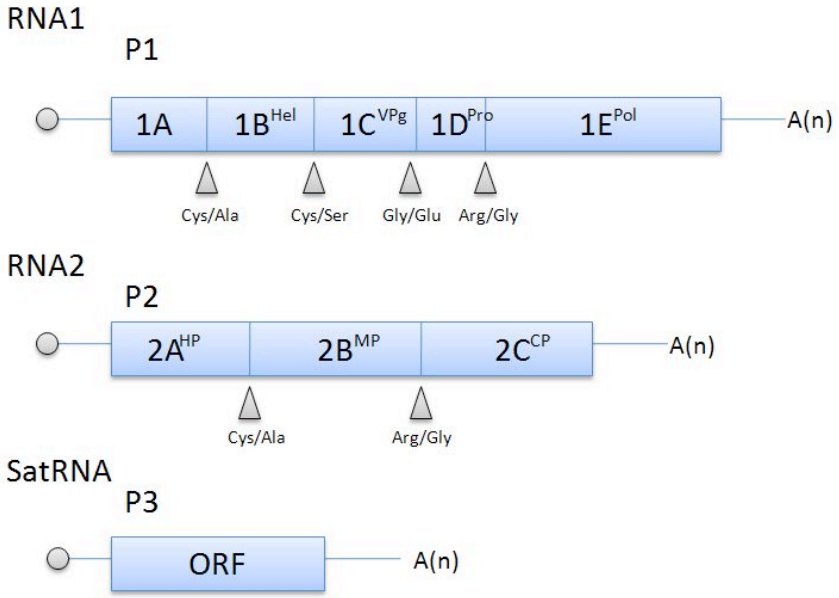


Figure 2.4: Diagrammatic representation of the bipartite viral genome and of the satRNA of GFLV-F13. The open reading frames (ORF) are represented by the open boxes and the VPg by the grey circles. The 5' and 3' noncoding regions are represented by the horizontal lines. The triangles represent the cleavage sites. Hel, helicase; VPg, viral genome linked protein; Pro, proteinase; Pol, RNA-dependent-RNA polymerase; Hp, homing protein; MP, movement protein; CP, coat protein; A(n), poly(A) tail (Adapted from Vigne et al., 1993).

Some nepovirus isolates have been shown to support the replication of satellite RNAs (satRNA). Satellite RNAs are RNA molecules that are smaller than 1500 nt and depend on helper viruses for replication, encapsidation, movement, and transmission. Most satRNAs share little or no sequence homology with the helper viruses (Murant and Mayo, 1982). Until recently, the satRNA of the French isolate F13 (GFLV-F13) was the only GFLV-associated satRNA characterised and remained the only described GFLV-associated satRNA for more than 20 years (Pinck et al., 1988). Two new GFLV-associated satRNAs have recently been described, associated with Californian isolates GFLV-R2 and GFLV-R6 (Gottula et al., 2013). The GFLV satRNAs are encapsidated and are 1114-1140 nt in length with a single ORF (codes for P3). It was demonstrated that the GFLV-F13 satRNA-encoded P3 is required for the replication of the satRNA (Pinck et al., 1988; Hans et al., 1992, 1993). The structure of the GFLV-F13 satRNA and ArMV satRNAs resemble the structure of the GFLV genomic RNA, in having a 5'-terminal genome linked protein (VPg) and a 3'-terminal poly(A) tail (Figure 2.4). The VPgs of the GFLV satRNA are indistinguishable from those attached to the helper

viruses, and are therefore probably encoded by the helpervirus genome (Mayo et al., 1979; Liu et al., 1991; Fritsch et al., 1993). Nepoviral satRNAs have a relatively short 5' untranslated region (UTR) and a similar sequence is present in the 5' leader sequence of all nepoviruses. No other sequences are similar between nepoviral satellites and the helper virus (Fritsch et al., 1993). The presence of satRNAs may enhance, attenuate or have no affect on the symptoms expressed by the helpervirus. It was recently demonstrated that the presence of a satRNA had no significant effect on GFLV symptom expression in *C. quinoa* (Gottula et al., 2013). It is however at this stage still unclear what the effect of satRNA presence may have on symptom expression in *V. vinifera*. A study that determined the occurrence and distribution of natural GFLV satRNAs in vineyards (Saldarelli et al., 1993) found that satRNAs were present in only 14.7% of 34 GFLV-infected grapevines tested. This is the only study to determine the occurrence and distribution of natural GFLV satRNA associations to date. Similar results (17%) were found for the natural occurrence of ArMV satRNAs (Wetzel et al., 2006).

2.2.3. Replication and cell-to-cell movement

It is proposed that GFLV replication takes place in the perinuclear area of a cell because of the accumulation of double stranded replicative forms, newly synthesized viral RNA and precursors and/or mature VPg proteins in this area (Ritzenthaler et al., 2002; Andret-Link et al., 2004). Cells infected with GFLV form a perinuclear complex, or "viral compartments" in the nuclear periphery, that are formed from the condensation and redistribution of endoplasmic reticulum (ER) membranes (Gaire et al., 1999). The viral compartments are formed by the aggregation of nonstructural RNA1-encoded proteins and RNA1-derived viral components. Replication is dependent on *de novo* lipid synthesis, and the ER-derived membranes are probably recruited by a coat protein complex vesicular trafficking mechanism. Therefore the RNA1-encoded polyprotein P1 is required for the viral RNA replication and is responsible for the formation of the viral compartments (Ritzenthaler et al., 2002).

From the viral compartments, GFLV particles move to the cell periphery and through the plasmodesmata to infect adjacent uninfected cells (Andret-Link et al., 2004). RNA2 codes for products involved in cell-to-cell movement. The cell-to-cell movement is separated in two steps: (1) the intracellular movement from the perinuclear area of RNA synthesis and virus assembly to the cell periphery; and (2) the intercellular movement across the cell wall into adjacent uninfected cells (Andret-Link et al., 2004). RNA2 is replicated *in trans* by the RNA1-encoded replication machinery and protein 2A is necessary, but not sufficient, for RNA2 replication. It is proposed that protein 2A possibly mediates the transport of P2-RNA2 complexes from their initial location (in the cytosol) to the perinuclear replication sites (Gaire et al., 1999). Both the MP and CP are needed for GFLV movement, but not for replication. As a result the MP and CP are only occasionally detected in the vicinity of the replication compartment (Ritzenthaler et al., 2002; Andret-Link et al., 2004). The CP may be involved in the intracellular transport mechanism of virions and/or MP. Once the MP is released from the precursor polyprotein, it is transported to the cell periphery (unknown mechanism). When the MP is in immediate proximity of the plasmodesmata, the MP self-assembles into tubules that allow protrusion into the cytoplasm of the neighbouring cells (Ritzenthaler et al., 2002; Laporte et al., 2003; Andret-Link et al., 2004). Less information is however known on the transport of the GFLV particles through the tubules into neighbouring cells. However, based on the similarities in virus spread between GFLV and Cowpea mosaic virus, the cell-to-cell movement of GFLV probably also involves MP-CP or MP-virion interactions during the intracellular and intratubular transport processes (Ritzenthaler et al., 1995, 2002; Belin et al., 1999; Carvalho et al., 2003; Andret-Link et al., 2004).

2.2.4. Transmission by vectors

Grapevine fanleaf virus is spread by the ectoparasitic dagger nematode *Xiphinema index*, family *Longidoridae*, and order *Dorylaimida* (Hewitt et al., 1958; Raski et al., 1983; Brown et al., 1995). The observation that *X. index* transmits GFLV was the first record of virus transmission by a nematode (Hewitt *et al.*, 1958). In South Africa, the Breede River area is infested with *X. index* and the vector is mainly spread by irrigation

water (Malan & Hugo, 2003). *Xiphinema index* is the sole vector of GFLV in the vineyard. The nematodes acquire the virus after feeding (in less than 15 min) on the growing plant root tips and they retain the ability to transmit the virus for several months, but not after moulting (Taylor and Robertson, 1975). The viruses do not multiply in their vectors and the virus is not transmitted to the egg. Specifically with GFLV, the virus particles are associated with the cuticular lining of the oesophagus (Hewit et al., 1958; Taylor and Robertson, 1975; Raski et al., 1983). The molecular determinants involved in the transmission process have been mapped to the RNA2-encoded CP and is the sole viral determinant responsible for the spread of GFLV by *X. Index* (Harrison et al., 1974; Harrison and Murant, 1977; Andret-Link et al., 2004). Recently, three-dimensional homology structure models of virions and CP subunits of ArMV and GFLV were constructed to further determine the transmission determinants on the CP. The researchers examined the models to predict amino acids that were exposed at the virion surface, which could possibly interact with specific vector factors. Five short amino acid stretches were identified, and were substituted in a GFLV RNA2 cDNA clone. After nematode transmission assays of the recombinant viruses the researchers concluded that a stretch of 11 residues (amino acids 188 to 198) in the βB-βC loop near the icosahedral 3-fold axis in the CP is essential for GFLV transmission by *X. index* (Schellenberger et al., 2010).

2.2.5. Genetic variability and recombination

Nucleotide substitutions and recombination are the most common forms of mutation in viral sequences that lead to genetic variability among virus isolates. Nucleotide substitutions occur during erroneous replication as a result of the non-proofreading RdRp, while recombination is the process in which gene segments are exchanged between two related viruses during co-infection of a host cell. As a result, RNA virus populations do not exist as a single member of a defined nucleotide sequence, but as a group of viruses related by a similar mutation or mutations, known as quasispecies (Drake et al., 1998, Garcia-Arenal et al., 2003).

GFLV infects its natural host for long periods of time and has theoretically a great potential for genetic variation (Naraghi-Arani et al., 2001; Andret-Link et al., 2004). Over 250 GFLV MP and CP sequences are available on GenBank and variability in these areas, especially the CP, has been studied extensively. Genetic variation studies have been conducted on isolates from France (Vigne et al., 2004), Germany (Wetzel et al., 2001), Iran (Sokhandan-Bashir et al., 2007; Pourrahim et al., 2007; Sokhandan-Bashir and Melcher, 2012), Slovenia (Pompe-Novak et al., 2007), South Africa (Liebenberg et al., 2009), Tunisia (Fattouch et al., 2005; Boulila, 2007) and the United States (Naraghi-Arani et al., 2001; Mekuria et al., 2009). Divergence in the CP was seen at levels up to 19% and 9% at nucleotide and amino acid level, respectively. In a previous SA study, the coat protein nucleotide sequences of twelve SA GFLV isolates were sequenced to investigate the variability between these variants as well as between the isolates available in GenBank (Liebenberg et al., 2009). Sequence identities between clones from different GFLV isolates from SA were between 86-99% and 94-99% at the nucleotide and amino acid levels, respectively. Phylogenetic analysis based on the CP revealed that all the isolates segregated into two distinct clades, clade A and B. It was discovered that the virus variants corresponding to clade B had 6 extra nt in the 3' non-coding region of RNA 2 when compared to the other group of variants. Clade A, whose variants did not show the extra nt, grouped with isolates from France, Germany and Chile, while clade B clustered with isolates from France and Slovenia (Liebenberg et al., 2009).

Recently, the nucleotide divergence was investigated in the full coding region of RNA2 as well as partial $1E^{Pol}$. This study was performed on Californian (USA) isolates and was the first study that investigated nucleotide divergence in RNA1 from a single geographic region. Analysis of the partial nucleotide sequences and deduced amino acid sequences of $1E^{Pol}$ indicated an average divergence of 11.5% and 9.9% respectively. Similarly, analysis of the coding RNA2 nucleotide sequences and deduced amino acid sequences among Californian isolates showed an average divergence between isolates was 10.8% and 6.0% respectively (Olivier et al., 2010). The lowest level of divergence on nucleotide and amino acid level was found in the CP gene, followed by the $1E^{Pol}$, MP

and HP. Phylogenetic analysis indicated that, based on full-length RNA2 sequences, isolates can be segregated into two distinct groups, as was seen with previous studies based on the CP from isolates from France (Vigne et al., 2004, 2009), SA (Liebenberg et al., 2009) and Tunisia (Fattouch et al., 2005). As mentioned before, the CP determines vector transmission and therefore amino acid divergence in this gene is likely the lowest because of possible strong evolutionary constraints (Oliver et al., 2010).

Recombination in viruses plays an important role in virus evolution (Oliver et al., 2010). Recombination within a viral species or among related viral species, known as intra- and interspecies recombination, may result in dramatic changes on the biology of viral population levels, and may lead to the emergence of new viruses and eventually new diseases (Garcia-Arenal et al., 2003). Inter- and intraspecies recombination in RNA2 between GFLV and ArMV isolates have been described extensively (Mekuria et al., 2009; Oliver et al., 2010; Pompe-Novak et al., 2007; Vigne et al., 2004; Vigne et al., 2005; Vigne et al., 2008). The first intraspecies recombination was described by Vigne et al. (2008) between GFLV-Ghu and ArMV-Ta. They found recombination events with multiple crossover sites mapping to the 5' UTR between nt 137 and 160, in gene HP between nts 880 and 921 and in gene MP between nt 1969 and 2031. No interspecies recombination was found in the CP (Vigne et al., 2008). Interspecies recombination in the HP and MP was also described between Californian GFLV isolates and ArMV (Oliver et al., 2010), and another research group found interspecies recombination events in the HP and MP of GFLV, ArMV and GDefV, but not once was interspecies recombination reported in the CP (Mekuria et al., 2009; Oliver et al., 2010). This corresponds with the findings of the initial study by Vigne et al (2008), which also found no interspecies recombination in the CP. The fact that no interspecies recombination in the CP was observed for any nepovirus can once again be attributed to the high evolutionary constraints on this gene (Oliver et al., 2010). Based on these reports, it seems that the HP and MP are "hotspots" for sequence variation and interspecies genetic exchange between closely related grapevine viruses (Mekuria et al., 2009).

2.3. The construction and uses of plant virus infectious clones

2.3.1. Background

Plant viruses are obligate parasites that are able to redirect the host cell's replication machinery for their own replication, translation and transcription. Since the discovery of the first plant virus, Tobacco mosaic virus (TMV), plant viruses have served as important model systems and are now often used as tools in biomedicine and plant biotechnology. RNA viruses are the most common viruses of plants, and since the discovery of the reverse transcriptase enzyme (Baltimore, 1970), the study of RNA viruses were particularly accelerated as researchers were now able to synthesize and clone cDNAs. Full-length cDNA infectious clones of plant RNA viruses could now be constructed by assembling RT-PCR generated cDNA fragments into bacterial plasmids and brought under the control of a chosen promoter. The availability of these full-length clones could be used for the study of viral replication, assembly and entry by modifying the genomes by mutagenesis. The small size of plant viruses make them suitable to be assembled into full length clones and are thus relatively easy to manipulate according to the researcher's needs for a particular study.

Full-length viral sequences are not necessarily needed for a fully functional infectious clone. Depending on the study and/or the application, restricted genome parts of the plant viruses can be used to construct viral vectors from only particular viral sequences. These viral vectors are commonly used to insert viral genes in plant genomes to prepare virus-resistant transgenic plants. Partial or restricted viral vectors are also used to exploit certain regulatory elements that are present in plant viruses, such as the Cauliflower mosaic virus (CaMV) 35S promoter (Guilley et al., 1982) and the translational enhancer of TMV (Gallie and Kado, 1989).

2.3.2. Full-length infectious clones of plant viruses

Transcripts from full-length infectious cDNA clones are generated either *in vitro* or *in vivo*, depending on the promoter that was used. Earlier infectious clones were transcribed *in vitro* by the use of *E. coli* promoters such as P_M or P_R, or bacteriophage promoters such as the SP6, T3 or T7. With this method, infectious viral RNA has to be transcribed with a bacterial RNA polymerase prior to inoculating host plants. However, this approach is expensive as it mostly necessitates transcription kits as well as expensive biolistic equipment for the transfection of host plants. Also, RNA produced by *in vitro* transcription is very sensitive to RNA degradation by RNAses and careful precautions are needed when working with RNA. An advantage of *in vitro* transcription is that the produced RNA transcripts are translation ready mRNA and therefore does not have to be introduced into the nucleus of the plant cell (Nagyova and Subr, 2007). *In vitro* transcripts can be mechanically inoculated into plant cells by rubbing the nucleic acids onto abrasive-injured leaves, however, this method is restricted to herbaceous and solanaceous species (Hull, 2002). When mechanical inoculation is not possible, other transfecting methods such as particle bombardment (Bilang et al., 1993) and microinjection (Klein et al., 1987) are used.

The CaMV 35S promoter drives *in vivo* transcription of transcripts and it circumvents the expensive synthesis of RNA transcripts using *in vitro* promoters. Because transcription is driven *in vivo*, the cDNA clones may be directly infectious and no extra precautions have to be taken to overcome RNA degradation. Also, cDNA clones driven with a CaMV 35S promoter are infectious by manual inoculation. However, some studies have shown that certain clones were not able to infect their natural hosts, and it is thought that the lack of infection by manual inoculation may be due to unsuccessful delivery into the plant cell nucleus (Dagless et al., 1997). Particle bombardments were originally used in these cases to deliver the DNA in the nucleus; however this method was also expensive. Agro-infiltration is a popular method of DNA delivery and it doesn't need expensive biolistic equipment.

Agro-inoculation exploits the ability of the crown gall bacterium *Agrobacterium tumefaciens* to infect plants and to transfer a segment of DNA, named transfer-DNA or T-DNA, to host plants (Joos et al., 1983). Agrobacterium introduces the T-DNA into plant genome at unpredictable sites, and as a result, viral cDNA cloned into the T-DNA are expressed in the plant cells. Several binary vector systems are widely used for the delivering of viral vectors into plants as they are able to transfer DNA to various plant host species (Ryu et al., 2004). Binary vectors are molecular tools used for higher plant transformations and are essentially disarmed tumor-inducing plasmids, since its tumor-inducing genes located in the T-DNA have been removed. Binary vectors are designed to include the T-DNA borders, multiple cloning sites and replication origins for *E. coli* and *A. tumefaciens* (Komari et al., 2006). Genes placed between these T-DNA borders are transferred to the plant cell and will relocate to the nucleus where they are then transcribed. Agro-inflitration is a very popular method of agro-inoculation, especially in *Nicotiana* species. Agro-inflitration is a method in which at least two true leaves are infiltrated with a needle-less syringe filled with an agrobacterium suspension that contains the recombinant viral vectors. Agrodrench is another agro-inoculation method, which involves the drenching of the plant rhizosphere, or the crown part of the plant, with *A. tumefaciens* containing the viral vector. Agrodrench has shown success in plant species that are otherwise difficult to agro-infiltrate, such as the *Solanaceae* species, as well as very young seedlings (Ryu et al., 2004).

Certain considerations need to be taken in account when cDNA clones of RNA viruses are being constructed, as serious limiting factors exist that may render full-length clones non-infectious, whether the transcripts are produced *in vitro* or *in vivo*. For high infectivity of transcripts it is important that the 5’ end of the transcript is near identical to the wild-type viral sequence and the presence of non-viral nucleotide (nt) between the promoter and the 5’ end of the viral sequence should be avoided. However, non-viral nt at the 3’ end are more easily tolerated. The presence of point mutations may also cause the clones to be non-replicating. Point mutations are generally caused by inaccurate

synthesis of cDNA fragments due to the reverse transcription and amplification by polymerases. Point mutations and deletions may also arise as a result of the instability of cloned cDNA in *E. coli*, as particular viral sequences may be toxic to the bacterial cells. This can however be eliminated by using a different strain or a different vector system (Boyer and Haenni, 1994; Nagyova and Subr, 2007).

2.3.3. The use of plant RNA virus infectious clones as viral vectors

Full-length infectious cDNA clones are usually used to study the life cycle of viruses in plant hosts. Viral vectors can be constructed from existing virus infectious cDNA clones, by inserting a multiple cloning site in the viral sequences of the infectious clone. Viral vectors are usually tested by incorporating reporter genes, such as the green fluorescent protein (GFP). Visible detection of the reporter genes confirms that the viral vector is still infectious and functional. After confirmation, viral vectors are modified to contain transgenes to be introduced into plant cells; either to express heterologous proteins in the plants, or to silence endogenous genes of the host plant. In the following sections the use of viral vectors for transient heterologous expression and transient silencing of endogenous genes are discussed.

2.3.3.1. Viral expression vectors

Plant virus-based vectors for expressing heterologous proteins in plants present promising biotechnological tools, as plants have been used as “green bioreactors” to express many different classes of recombinant proteins, such as therapeutic proteins (antibodies, vaccine allergens, etc), industrial proteins or enzymes (hydrolases, proteases, etc.), and biopolymers such as collagen (Xu et al., 2011). Existing methods require the use of either stable transgenic plants or transient expression systems using plant virus-based vectors. Transient gene expression provides many advantages over the traditional use of stably transformed transgenic plants, as stable transgenic plants are time- and resource- consuming. In the transient system, virus infection establishes within days of infection and the expression can be evaluated in days after the infection. The expressed proteins may also accumulate in high concentrations throughout the

plant. Therefore, transient gene expression by viral vectors provides a cost effective method for the overproduction of proteins in plants. Several RNA plant viruses have been converted into expression vectors and depending on the structure of the virus and expression needs, these viruses were used for different applications, including gene exchange or replacement, gene insertion, epitope presentation or complementation systems (Zhang and Ghabrial, 2006; Nagyova and Subr, 2007). Plant-based expression vectors based on TMV, potato virus X (PVX), Alfalfa mosaic virus, Cucumber mosaic virus, and Cowpea mosaic virus have been extensively used for the production of proteins (Danson et al., 1991; Chapman et al., 1992; Canizares et al., 2000; Gopinath et al., 2000; Sanchez-Navarro et al., 2001; Musiychuk et al., 2007; Fujiki et al., 2008).

Initial engineered plant virus-based expression vectors were based on full-length viruses, which were modified to contain heterologous genes to be expressed. The proteins of interest were either fused to the coat protein to be expressed, or were expressed from viral subgenomic promoters. These so-called first generation plant virus vectors based on full-length viral sequences, such as TMV and PVX, generally resulted in over 1g recombinant protein per kg of fresh leaf biomass. However, protein yields based on other plant virus vectors varied greatly, depending on the protein that was expressed. The initial expression vectors based on full-length viruses had some undesirable features, such as the size limitation of inserts that could be tolerated without affecting the replication of the virus. These constructs were also unstable and many produced low protein yields. Researchers addressed these problems by modifying the vector sequences to develop “deconstructed” viral vectors in which only the viral elements required for efficient expression of the sequence of interest were maintained, and the missing functions were provided using non-viral components. The “deconstructed” vector strategy, together with agrobacterium-mediated delivery, allowed for a more efficient, controlled and safe process of heterologous expression (Gleba et al., 2007; Fijiki et al., 2008).

Low expression of target proteins may also be a result of post-transcriptional gene silencing (PTGS) in agrobacterium-mediated transient expression. A previous study has

shown that by adding a suppressor of gene silencing, p19 of Tomato bushy stunt virus, to a GFP expression system, GFP expression levels increased by more than 50-fold in the presence of p19 compared to the plants that were only infiltrated with the virus-based expression vector. The p19 prevented the onset of PTGS in the infiltrated areas and therefore allowed for higher levels of transient expression. The GFP expression also persisted for much longer; GFP could be detected up to 12 days post inoculation (dpi), as opposed to only 5 dpi without p19 (Voinnet et al., 2003). In another noteworthy study, authors have agro-infiltrated a GFP-expressing TMV vector together with the silencing suppressor p19. This expression system produced approximately 600 to 1200 µg of recombinant GFP per gram of agroinfiltrated tissue, which was 10-25 fold more than previously available 35S promoter-driven transient expression systems (Lindo, 2007). It is known that there are still some limitations in transient heterologous protein expression using virus-based vectors; however this exciting field has made a tremendous progress since its inception in the last two decades (Gleba et al., 2007).

2.3.3.2. Virus induced gene silencing vectors

Viral vectors have also been converted to virus-induced gene silencing (VIGS) vectors for analysis of gene function in host species. With all the new sequencing technologies being developed and the wealth of sequencing data generated, the next challenge for researchers was to determine the primary function of the predicted genes. VIGS technology is an approach to explore the functionality of expressed sequences. This technology exploits the natural PTGS pathway of the plant's host defence mechanism against foreign RNAs (Lindbo et al., 1993; Ratcliff et al., 1997; Soosaar et al., 2005). One of the most important roles of RNA silencing in plants, apart from gene regulation, is to protect against viruses. Foreign RNA gets degraded in a sequence-specific manner through a pathway known as PTGS in plants, and is also known as RNA interference in animal cells (Baulcombe et al., 2004). For the purposes of this literature review, only PTGS in plants will be discussed.

Post-transcriptional gene silencing has two main stages, initiation and maintenance. RNA silencing is triggered by invading double stranded RNA (dsRNA) intermediates

(Figure 2.5). It is widely viewed that PTGS evolved to protect the plant from viruses (and other invading DNAs and RNAs) as in most cases RNA viruses undergo a dsRNA replicative form (RF) during its replication cycle. It was also initially thought that the dsRNA RF of viruses are the only templates that are able to trigger RNA silencing, but it was subsequently shown that PTGS is also triggered by highly structured genomic regions of viruses (Molnar et al., 2005). Once foreign dsRNA (from viruses) is detected in the cells they are processed into RNA duplexes of 21-24 nt called short interfering RNAs (siRNA) by a ribonuclease III-like enzyme known as Dicer (Fire et al., 1998; Bernstein et al., 2001). The siRNAs are denatured and one strand is incorporated into a multi-subunit nuclease complex called the RNA-induced silencing complex (RISC). The RISC complex consists of several interacting proteins that bind target sequences and small RNAs, including the Argonaute protein. Once associated with the RISC, the siRNA recognizes and base-pairs with the homologous RNA that eventually leads to the degradation of the target RNA by an RNase H cleavage mechanism. The silencing effect is maintained once it was initiated, by passing the silencing signal throughout the plant. Cells that have received this signal would then be primed to recognize the viral RNA which leads to subsequent degradation of the virus throughout the rest of the plant (Padmanabhan et al., 2008).

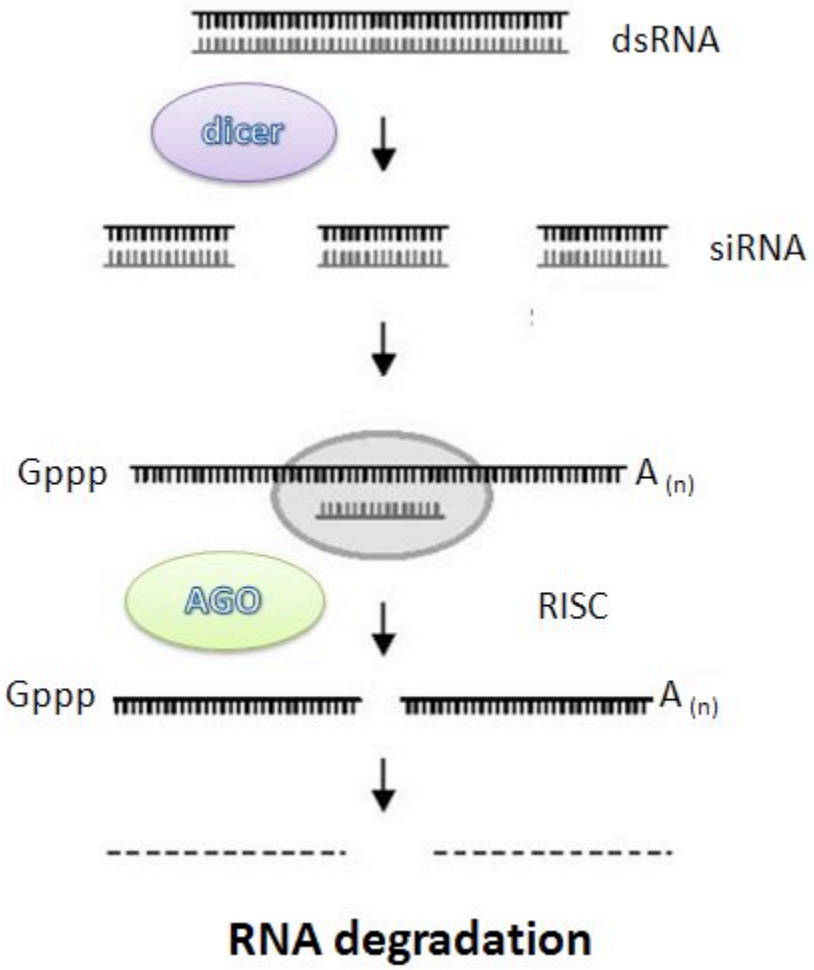


Figure 2.5: A brief overview of PTGS in plants. The RF of viruses provides the dsRNA that triggers PTGS. The dsRNA are processed into siRNAs by the endonuclease Dicer. The siRNA is loaded into the RISC complex that contains AGO and guides the RISC complex to the mRNA by base-pairing. The RISC complex cleaves the mRNA and is subsequently degraded. dsRNA, double-stranded RNA; siRNA, short interfering RNA; RISC, RNA-induced silencing complex; AGO, Argonaute.

VIGS vectors are infectious clones of plant viruses with an introduced multiple cloning site to insert partial sequences of an endogenous gene to be silenced in the host plant of interest, and takes advantage of the plant's defence mechanism to silence the inserted endogenous gene, as the PTGS would target and degrade both the viral and the corresponding endogenous mRNA (Baulcombe, 1999; Nagyova and Subr 2007). Factors that play an important role in the efficiency of VIGS vectors are the size and the

orientation of the insert. It was initially thought that a fragment of a minimum of 23 nt with a 100% identity to the targeted transgene is adequate for silencing (Thomas et al., 2001). However it was later established that longer (between 200 bp and 1300 bp) fragments should be used (Ekengren et al., 2003; Liu and Page, 2008). In terms of their efficiency to trigger the silencing effect, inverted repeats were shown to be the best, followed by inserts in the antisense orientation, and lastly inserts that are inserted in the sense orientation (Lacomme et al., 2003; Hein et al., 2005). Engineered vectors carrying host-derived nucleotide sections are delivered into the plants by either *in vitro* or *in vivo* methods. During infection caused by the recombinant virus, the RNA silencing system is initiated and directed both against the viral genome and the host transcript derived from the inserted DNA fragment. The virus and target transcripts are degraded by silencing that ultimately leads to a silenced phenotype. Therefore, the engineered vector initiates the plant's PTGS pathway and "knocks down" the expressed genes of choice (Small, 2007). Virus-induced gene silencing is a time- and cost-efficient *in vivo* tool for studying gene functionality in host plants and does not require plant transformation (Burch-Smith et al., 2004).

Several plant RNA and DNA viruses have been converted to VIGS vectors. The first VIGS vector based on an RNA virus, was based on TMV. Transcripts of recombinant TMV, carrying a sequence encoding for the Phytoene desaturase (PDS) enzyme, were produced *in vitro* and inoculated to *N. benthamiana* plants. Systemic infection produced the silencing phenotype of photobleaching of the leaves, indicating successful silencing of PDS (Kumagai et al., 1995). Since then, many plant RNA viruses have been engineered into VIGS vectors, which were used to elucidate the function of a number of plant genes that are involved in defense responses (Liu et al., 2002, 2004b), programmed cell death (Downward, 2003; Brinbaun, 2004; Gabriels et al., 2004; Liu et al., 2005; An et al., 2008), plant growth and development (Burger et al., 2003; Liu et al., 2004a; Ahn et al., 2005; Yang et al., 2007) and abiotic stress (Senthil-Kumar et al., 2006, 2007).

Recently, satellite virus-based vectors were also used for efficient gene silencing in plants co-inoculated with the respective helper viruses. The satellite virus of tobacco mosaic virus (STMV) was used as a vector for the delivery of inhibitory RNA into tobacco plants, together with TMV as helpervirus and this two-component system was termed as satellite virus-induced silencing system (SVISS) (Gosselé et al., 2002). The authors argued that using a satellite VIGS system may have several advantages over using VIGS vectors based on genomic RNA or DNA viruses. Satellite and helpervirus systems are a natural two-component (or deconstructed) system, which separates the replication and movement of the virus from the silencing component (the satellite), and which may lead to a more stable system. Infectious clones of satellites are also easier to construct and manipulate *in vitro* due to their small size. Also, because of the small size of satellites, they can accumulate in higher numbers in the cytoplasm because of higher replication efficiencies, which may lead to a more pronounced silencing effect (Gosselé et al., 2002). To date, the satellite viruses of the Begomoviruses, Geminivirus alpha satellite, Tomato yellow leaf curl virus and Tobacco curly shoot virus have all been successfully used as SVISS vectors in *Nicotiana* and *Solanaceae* species (Tao and Zhou, 2004; Huang et al., 2009, 2011; Zhou and Huang, 2012).

VIGS and expression vectors have mainly been used to silence or over-express transgenes in herbaceous hosts. As a result these important tools are limited for perennial and woody plants, such as citrus, plums, apples, pears, berries and grapevine, and are delaying research efforts for these important crops. To date only two VIGS vectors exist for grapevine, the first is based on GVA (Muruganantham et al., 2009) and the second on GLRaV-2 (Kurth et al., 2012). The GVA vector was the first vector that was able to silence PDS in both *N. benthamiana* and in *V. vinifera,* although the photobleaching was only observed in vein and inter-vein regions due to virus's phloem limitation (Muruganantham et al., 2009). The GLRaV-2 vector was used as a GFP expression vector as well as a VIGS vector in both *N. benthamiana* and *V. vinifera.* The authors used the GFP-GLRaV-2 vector to visualise, for the first time, the spread of the virus in its natural host. Even though the virus is phloem limited, after four weeks it was observed that the virus entered the petioles, leaves and berries via the veins and

after six months the virus was detected in the roots. This vector has contributed immensely by elucidating how the virus colonizes the vine during the seasons. The GLRaV-2 vector was also engineered to simultaneously express GFP while silencing the PDS gene in grapevine. The dual-purpose vector was used to monitor the spread of the virus as the PDS were silenced in the leaves. Initially bleaching was observed in leaf cells surrounding the virus infected GFP-expressing cells and later it was observed that the bleaching had expanded along the vascular system into areas that were not yet infected with the virus (Kurth et al., 2012). This GLRaV-2 derived gene expression and VIGS vector is the first of its kind for grapevine research and provides a platform for multiple applications in grapevine functional genomics and biotechnology. It has also paved the way for generating more vectors for grapevine and other woody plants, as only two VIGS vectors for grapevine is unfortunately not enough for the wealth of research that needs to be conducted. A virus that is not phloem-limited, such as Grapevine fanleaf virus, would be an ideal candidate to convert into a VIGS vector for functional genomics in grapevine, as systemic silencing of endogenous genes may be observed in a shorter period of time compared to the phloem limited viruses.

2.4. Conclusion

Grapevines constitute one of the most economically important fruit species in both SA and worldwide. Grapevine is subject to several plant pathogens that have a negative impact on the economy of wine-producing countries. Grapevine fanleaf virus is one such virus - it is the most globally dispersed of all grapevine-infecting viruses. Because of its notoriety, various aspects of the virus have been studied extensively. However, very limited information exists for GFLV isolates present in SA. In the chapters that follow, the characterisation of two isolates of GFLV found in SA, and the discovery of a new satRNA, is described. The two isolates were characterized on a molecular level and novel features are discussed. Also, with the availability of the SA GFLV full-length sequences, we made progress in an attempt to construct a GFLV infectious clone for functional analysis in grapevine, as these molecular tools for this valuable crop are limited at present.

With this PhD thesis we aimed to address the following research questions:

- How would GFLV isolates from SA differ on a nucleptide sequence level to the other known GFLV isolates? How much difference would exist between SA isolates at this level?
- Are there GFLV satRNAs present in SA? If so, how do they compare to the other characterised GFLV satRNAs? What are the effects of satRNA on the host plants? Can sequence variation between GFLV isolates reveal why certain GFLV and ArMV isolates able to replicate satRNAs and others not?
- With the availability of a full-length infectious cDNA clone based on SA GFLV isolates, can they be converted into a VIGS or expression viral vector system for *V. vinifera*? Can the GFLV satRNA be utilized as the VIGS or expression vector component of the system? And how would it compare to the traditional VIGS or expression systems?

2.5. References

1. Ahn CS, Lee JH, Pai HS (2005). Silencing of NbNAP1 encoding a plastidic SufB-like protein affects chloroplast development in Nicotiana benthamiana. Mol. Cell, 31, 112–118.
2. An SH, Sohn KH, Choi HW, Hwang IS, Lee SC, Hwang BK (2008). Pepper pectin methylesterase inhibitor protein CaPMEI1 is required for antifungal activity, basal disease resistance and abiotic stress tolerance. Planta, 228, 61–78.
3. Andret-Link P, Laporte C, Valet L, Ritzenthaler C, Demangeat G, Vigne E, Laval V, Pfeiffer P, Stussi-Garaud C, Fuchs M (2004). Grapevine fanleaf virus: still a major threat to the grapevine industry. J of Plant Path, 86, 183-195.
4. Arenal F, Fraile A, Malpica J (2001). Variability and genetics structure of plant viruses populations. Annual Review of Phytopathology 39, 157–189.
5. Baltimore, D. (1970) Viral RNA dependent DNA polymerase in virions of RNA tumor viuses. Nature, 226, 1209–1211.
6. Baulcombe DC (1999). Fast forward genetics based on virus-induced gene silencing. Current Opinion in Plant Biology, 2,109-113.
7. Baulcombe DC (2004). RNA silencing in plants. Nature, 431 (7006), 356-363.
8. Belin C, Schmitt C, Gaire F, Walter B, Demangeat G, Pinck L (1999). The nine C-terminal residues of the grapevine fanleaf nepovirus movement protein are critical for systemic virus spread. J Gen Virol 80, 1347-1356.

9. Bernstein E, Caudy AA, Hammond SC, Hannon GJ (2001). Role for a bidentate ribonuclease in the initiation step of RNA interference. Nature, 409 (6818), 363-366.
10. Bilang R, Zhang S, Leduc N, Iglesias VA, Gisel A, Simmonds J, Potrykus I, Sautter C (1993). Transient gene expression in vegetative shoot apical meristems of wheat after ballistic microtargeting. The Plant Journal 4, 735-744.
11. Boulila M (2007). Phylogeny and genetic recombination of Grapevine fanleaf virus isolates from naturally infected vineyards in Tunisia. Phytopath Mediterr, 46, 285-294.
12. Brown DJF, Robertson WM, Trudgill DL (1995). Transmission of viruses by plant nematodes. Ann Rev Phytopathology, 33, 223-249.
13. Burch-Smith TM, Anderson JC, Martin GB, Dinesh-Kumar SP (2004). Applications and advantages of virus-induced gene silencing for gene function studies in plants. Plant J 39, 734–746.
14. Burger C, Rondet S, Benveniste P, Schaller HJ (2003). Virus-induced silencing of sterol biosynthetic genes: identification of a Nicotiana tabacum L. obtusifoliol- 4alpha-demethylase (CYP51) by genetic manipulation of the sterol biosynthetic pathway in Nicotiana benthamiana. Exp Bot, 54, 1675–1683.
15. Cañizares MC, Li L, Yolande P, Estratios T, George PL (2006). A bipartite system for the constitutive and inducible expression of high levels of foreign proteins in plants. Plant Biotechnol J, 4,183-193.
16. Carvalho CM, Wellink J, Ribeiro SG, Goldbach RW, van Lent JWM (2003). The C-terminal region of the movement protein of Cowpea mosaic virus is involved in binding to the large but not to the small coat protein. J Gen Virol, 84, 2271-2277.
17. Chapman S, Kavanagh T, Baulcombe D (1992). Potato virus X as a vector for gene expression in plants. Plant J, 2, 549-557.
18. Digiaro M, Elbeaino T, Martelli GP (2007). Development of degenerate and species-specific primers for the differential and simultaneous RT-PCR detection of grapevine-infecting nepoviruses of subgroups A, B and C. J of Virological Methods, 141, 34–40
19. Downward J (2003). Metabolism meets death. Nature, 424, 896–897.
20. Drake JW, Charlesworth B, Charlesworth D, Crow JF (1998) Rates of spontaneous mutation. Genetics, 148, 1667–1686.
21. Ekengren SK, Liu Y, Schiff M, Dinesh-Kumar SP, Martin GB (2003). Two MAPK cascades, NPR1, and TGA transcription factors play a role in Pto-mediated disease resistance in tomato. Plant J, 36, 905–917.
22. Fattouch S, Acheche H, M'hirsi S, Mellouli L, Bejar S, Marrackchi M, Marzouki N (2005). RT-PCR-RFLP for genetic diversity analysis of Tunisian Grapevine fanleaf virus isolates in their natural host plants. J Virol Methods, 127, 126-132.
23. Fire A, Xu S, Montgomery MK, Kostas SA, Driver SE, Mello CC (1998). Potent and specific genetic interference by double-stranded RNA in Caenorhabditis elegans. Nature, 6669, 806–811.

24. Fujiki M, Kaczmarczyk JF, Yusibov V, Rabindran S (2008). Development of a new cucumber mosaic virus-based plant expression vector with truncated 3a movement protein. Virology, 381, 136-142.
25. Gabriels SH, Takken FL, Vossen JH, de Jong CF, Liu Q, Turk SC, Wachowski LK, Peters J, Witsenboer JH, de Wit PJ, Joosten MH (2006). CDNAAFLP combined with functional analysis reveals novel genes involved in the hypersensitive response. Mol Plant Microbe Interact, 19, 567–576.
26. Gaire F, Schmitt C, Stussi-Garaud C, Pinck L, Ritzenthaler C (1999). Protein 2A of grapevine fanleaf nepovirus is implicated in RNA2 replication and colocalizes to the replication site. Virology, 264, 25-36.
27. Gaire F, Schmitt C, Stussi-Garaud C, Pinck L, Ritzenthaler C (1999). Protein 2A of grapevine fanleaf nepovirus is implicated in RNA2 replication and colocalizes to the replication site. Virology, 264(1), 25-36.
28. Gallie DR, Kado CI (1989). A translational enhancer derived from tobacco mosaic virus is functionally equivalent to a Shine-Dalgarno sequence.. Proc Natl Acad Sci 86, 129-132.
29. Garcia-Arenal F, Fraile A, Malpica JM (2003). Variation and evolution of plant virus populations. Int Microbiol, 6, 225-232.
30. Gleba Y, Klimyuk V, Marillonnet S (2007). Viral vectors for the expression of proteins in plants. Curr. Opin. Biotechnol, 18, 134-141.
31. Gopinath K, Wellink J, Porta C, Taylor KM, Lomonossoff GP, van Kammen A (2000). Engineering Cowpea Mosaic Virus RNA-2 into a Vector to Express Heterologous Proteins in Plants. Virology, 267, 159-17.
32. Gosselé V, Faché I, Meulewater F, Cornelissen M, Metzlaff M (2002). SVISS – a novel transient gene silencing system for gene function discovery and validation in tobacco plants. The Plant Journal, 32, 859-866.
33. Gottula JW, Lapato D, Cantilina KK, Saito S, Bartlett B, Fuchs M (2013). Genetic variability, evolution and biological effects of Grapevine fanleaf virus satellite RNAs. Phytopathology, http://dx.doi.org/10.1094/PHYTO-11-12-0310-R.
34. Guilley H, Dudley RK, Jonard G, Balàzs E, Richards KE (1982). Transcription of Cauliflower mosaic virus DNA: detection of promoter sequences, and characterization of transcripts. Cell 30(3), 763-73.
35. Hans F, Fuchs M, Pinck L (1992). Replication of the grapevine fanleaf virus satellite RNA transcripts in Chenopodium quinoa protoplasts. J Gen Virol, 73, 2517-2523.
36. Hans F, Pinck M, Pinck L (1993). Location of the replication determinants of the satellite RNA associated with grapevine fanleaf virus (strain F13). Biochimie, 75, 597-603.
37. Harrison BD and Murant AF (1977). Nematode transmissibility of pseudorecombinant isolates of tomato black ring virus. Annals of App Biol, 86, 209-212.

38. Harrison BD, Murant AF, Mayo MA, Roberts IM (1974). Distribution of determinants for symptom production, host range and nematode transmissibility between the two RNA components of raspberry ringspot virus. J Gen Virol, 22, 233-247.
39. Hein I, Barciszewska-Pacak M, Hrubikova K, Williamson S, Dinesen M, Soenderby IE, Sundar S, Jarmolowski A, Shirasu K, Lacomme C (2005). Virus-induced gene silencing-based functional characterization of genes associated with powdery mildew resistance in barley. Plant Physiol, 138, 2155–2164.
40. Hewitt WB, Raski DJ, Goheen AC (1958). Nematode vector of soil-borne fanleaf virus of grapevines. Phytopathology, 48, 586-595.
41. Huang C, Xie Y, Zhou X (2009). Efficient virus-induced gene silencing in plants using a modified geminivirus DNA1 component. Plant Biotechnol J, 7(3), 254-65.
42. Huang CJ, Zhang T, Li FF, Zhang XY, Zhou X (2011). Development and application of an efficient virus-induced gene silencing system in *Nicotiana tabacum* using geminivirus alphasatellite. J Zhejiang Univ Sci B, 12(2), 83–92.
43. Hull R, 2002. Matthews' Plant Virology. Academic Press, New York.
44. Joos H, Timmerman B, Van Montagu M, Schell J (1983). Genetic analysis of transfer and stabilization of *Agrobacterium* DNA in plant cells. EMBO J, 2(12), 2151–2160.
45. Klein TM, Wolf ED, Wu R, Sanford JC (1987). Highvelocity microprojectiles for delivering nucleic acids into living cells. Nature, 327, 70-73.
46. Komari T, Takakura Y, Ueki J, Kato N, Ishida Y, Hiei Y (2006). Binary Vectors and Super-binary Vectors. Methods in Molecular Biology, 343, 15-42.
47. Kumagai MH, Donson J, Della-Cioppa G, Harvey D, Hanley K, Grill LK (1995). Cytoplasmic inhibition of carotenoid biosynthesis with virus-derived RNA. Proc Natl Acad Sci USA 92, 1679–1683.
48. Kurth EG, Peremyslov VV, Prokhnevsky AI, Kasschau KD, Miller M, Carrington JC, Dolja VV (2012). Virus-derived gene expression and RNAi vector for grapevine. J Virol, 86, 6002-6009.
49. Lacomme C, Hrubikova K, Hein I (2003). Enhancement of virus-induced gene silencing through viral-based production of inverted-repeats. Plant J, 34, 543–553.
50. Laporte C, Vetter G, Loudes AM, Robinson GD, Hillmer S, Stussi-Garaud C, Ritzenthaler C (2003). Involvement of the secretory pathways and the cytoskeleton in intracellular targeting and tubule assembly of Grapevine fanleaf virus movement in tobacco BY-2 cells. The Plant Cell, 15, 2058-2075.
51. Liebenberg A, Freeborough MJ, Visser CJ, Bellstedt DU, Burger JT (2009). Genetic variability within the coat protein gene of Grapevine fanleaf virus isolates from South Africa and the evaluation of RT-PCR, DAS-ELISA and immunostrips as virus diagnostic assays. Virus Res, 142, 28-35.
52. Lindbo JA (2007). High-efficiency protein expression in plants from agroinfection-compatible *Tobacco mosaic virus* expression vectors. Plant Physiol, 145, doi 10.1186/1472-6750-7-52.

53. Lindbo JA, Silva-Rosales L, Proebsting WM, Dougherty WG (1993). Induction of a highly specific antiviral state in transgenic plants: implications for regulation of gene expression and virus resistance. Plant Cell, 5, 1749–1759.
54. Liu E and Page JE (2008).Optimized cDNA libraries for virus-induced gene silencing (VIGS) using tobacco rattle virus. Plant Methods, 4(5), doi 10.1186/1746-4811-4-5.
55. Liu Y, Nakayama N, Schiff M, Litt A, Irish VF, Dinesh-Kumar SP (2004). Virus induced gene silencing of a DEFICIENS ortholog in Nicotiana benthamiana. Plant Mol. Biol. 54, 701–711.
56. Liu Y, Schiff M, Czymmek K, Talloczy T, Levine B, Dinesh-Kumar SP (2005). Autophagy regulates programmed cell death during the plant innate immune response. Cell 121, 567–577.
57. Liu Y, Schiff M, Dinesh-Kumar SP (2004). Involvement of MEK1 MAPKK, NTF6 MAPK, WRKY/MYB transcription factors, COI1 and CTR1 in N-mediated resistance to tobacco mosaic virus. Plant J. 38, 800–809.
58. Liu Y, Schiff M, Marathe R, Dinesh-Kumar SP (2002). Tobacco Rar1, EDS1 and NPR1/NIM1 like genes are required for N-mediated resistance to tobacco mosaic virus, Plant J, 30(20), 415–429.
59. Malan AP and Hugo HJ (2003). Present status of grapevine nepoviruses and their nematode vectors in South Africa. Wynland, 14, 120-123.
60. Margis R, Ritzenthaler C, Reinbolt J, Pinck M, Pinck L (1993) Genome organization of grapevine fanleaf nepovirus RNA2 deduced from the 122K polyprotein P2 in vitro cleavage products. J Gen Virol, 74, 1919-1926.
61. Martelli GP (2012). Grapevine Virology Highlights: 2010-2012. Proceedings of the 17th Congress of ICVG, Davis, California, USA. October 7-14, 2012.
62. McGovern PE, (2004). Ancient wine: the search for the origins of viniculture. Princeton University Press.
63. Mekuria TA, Gutha LR, Martin RR, Naidu RA (2009). Genome diversity and intra- and interspecies recombination events in Grapevine fanleaf virus. Phytopathology, 99, 1394-1402.
64. Molnar A, Csorba T, Lakatos L, Varallyay E, Lacomme CC, Burgyan J (2005). Plant virus-derived small interfering RNAs originate predominantly from highly structured single-stranded viral RNAs. J Virol, 79, 7812–7818.
65. Musiychuk K, Stephenson N, Bi H, Farrance CE, Orozovic G, Brodelius M, Brodelius P, Horsey A, Ugulava N, Shamloul AM, Mett V, Rabindran S, Streatfield SJ, Yusibov V (2007). A launch vector for the production of vaccine antigens in plants. Influ Other Respir Viruses, 1, 19-25.
66. Nagyova A, Šubr Z (2007). Infectious full-length clones of plant viruses and their use for construction of viral vectors. Acta Virologica, 51, 223-237.
67. Naraghi-Arani P, Daubert S, Rowhani A (2001). Quasispecies nature of the genome of Grapevine fanleaf virus. J Gen Virol, 82, 1791-1795.
68. Pinck M, Reinbolt J, Loudes AM, Le Ret M, Pinck L (1991). Primary structure and location of the genome-linked protein (VPg) of Grapevine Fanleaf Nepovirus. FEBS Letters, 284, 117-119.

69. Pink L, Fuchs M, Pinck M, Ravelonandroi M, Walter B (1988). A Satellite RNA in Grapevine Fanleaf Virus Strain F13. J Gen Virol, 69, 233-239.
70. Pompe-Novak M, Gutierrez-Aguirre I, Vojvoda J, Blas M, Tomazic I, Vigne E, Fuchs M, Ravnikar M, Petrovic N (2007). Genetic variability within RNA2 of Grapevine fanleaf virus. Eur J Plant Path, 117, 307-312.
71. Pourrahim R, Farzadfar SH, Golnaraghi AR, Ahoonmanesh A (2007). Partial molecular characterization of some Grapevine fanleaf virus isolates from North-east of Iran. J Phytopathol, 155, 754-757.
72. Qiu W, Avery JD Jr, Lunden S (2007). Characterization of a severe virus-like disease in Chardonnay grapevines in Missouri. Online. Plant Health Progress, doi 10.1094/PHP-2007-1119-01-BR.
73. Quacquarelli A, Gallitelli D, Savino V, Martelli GP (1979). Properties of Grapevine fanleaf virus. J Gen Vir, 32, 349-360.
74. Raski DJ, Goheen AC, Lider LA, Meredith CP (1983). Strategies against grapevine fanleaf virus and its nematode vector. Plant Disease, 67, 335-339.
75. Ratcliff F, Harrison BD, Baulcombe DC (1997). A similarity between viral defence and gene silencing in plants. Science, 276, 1558–1560.
76. Rathay E (1882). Die Gabler- oder Zwiewipflerreben, eine vorläufige Mittheilung. Österr Bot Z, 32, 316-320.
77. Rathay E (1883). Über die in Nieder-Österreich als “Gabler“ oder “Zwiewipfler” bekannten Reben. Verlag der k.k. önol. und pomol. Lehranstalt, Klosterneuburg.
78. Ritzenthaler C, Laporte C, Gaire F, Dunoyer P, Schmitt C, Duval S, Piéquet A, Loudes AM, Rohfritsch O, Stussi-Garaud C, Pfeiffer P (2002). Grapevine Fanleaf Virus Replication Occurs on Endoplasmic Reticulum-Derived Membranes. J Virol, 76, 8808–8819.
79. Ritzenthaler C, Schmit AC, Michler P, Stussi-Garaud C, Pinck L (1995). Grapevine fanleaf nepovirus P38 putative movement protein is located on tubules in vivo. Mol Plant-Microbe Interact, 8, 379-387.
80. Ritzenthaler C, Viry M, Pinck M, Margis R, Fuchs M, Pinck L (1991). Complete nucleotide sequence and genetic organization of grapevine fanleaf nepovirus RNA1. J Gen Virol, 72, 2357-2365.
81. Ryu CM, Anand A, Kang L, Mysore KS (2004). Agrodrench: a novel and effective agroinoculation method for virus induced gene silencing in roots and diverse Solanaceous species. The Plant Journal, 40, 322-331.
82. Saldarelli P, Minafra A, Walter B (1993). A survey of grapevine fanleaf nepovirus isolates for the presence of satellite RNA. Vitis, 32, 99-102.
83. Sanchez-Navarro J, Miglino R, Ragozzino A, Bol JF (2001). Engineering of Alfalfa mosaic virus RNA 3 into an expression vector. Arch Virol, 146, 923-939.

84. Sanfacon H, Wellink J, Le Gall O, Karasev A, van der Vlugt R, Wetzel T (2009). Secoviridae: a proposed family of plant viruses within the order *Picornavirales* that combines the families *Sequiviridae* and *Comoviridae*, the unassigned genera *Cheravirus* and *Sadwavirus*, and the proposed genus Torradovirus. Arch Virol, 154(5), 899-907.

85. Schellenberger P, Andret-Link P, Schmitt-Keichinger C, Bergdoll M, Marmonier A, Vigne E, LemaireO , Fuchs M, Demangeat G, Ritzenthaler C (2010). A stretch of 11 amino acids in the βB-βC loop of the coat protein of *Grapevine fanleaf virus* is essential for transmission by the nematode *Xiphinema index*. J Virol, 84, 7924–7933.

86. Senthil-Kumar M, Govind G, Kang L, Mysore KS, Udayakumar M (2007). Functional characterization of Nicotiana benthamiana homologs of peanut water deficit-induced genes by virus-induced gene silencing. Planta, 225, 523–539.

87. Senthil-Kumar M, Udayakumar M (2006). High-throughput virus-induced gene-silencing approach to assess the functional relevance of a moisture stress-induced cDNA homologous to lea4. J Exp Bot, 57, 2291–2302.

88. Small I (2007). RNAi for revealing and engineering plant gene functions. Current Opinion in Biothechnology, 18, 148-153.

89. Sokhandan-Bashir N and Melcher U (2012). Population genetic analysis of grapevine fanleaf virus. Archives of Virology. DOI 10.1007/s00705-012-1381-0.

90. Sokhandan-Bashir NS, Zarghani SN, Hejazi MS (2007). Diversity of Grapevine fanleaf virus isolates from Iran. Virus Res, 128, 144-148.

91. Soosaar JL, Burch-Smith TM, Dinesh-Kumar SP (2005). Mechanisms of plant resistance to viruses. Nat Rev Microbiol, 3, 789–798.

92. Tao X, Zhou X (2004). A modified viral satellite DNA that suppresses gene expression in plants. Plant J, 38, 850–860.

93. Taylor CE and Robertson WM (1975). Acquisition, retention and transmission of viruses by nematodes. In Nematode Vectors of Plant Viruses. Plenum Press, London, 253-276.

94. Thomas CL, Jones LDC, Maule AJ (2001). Size constraints for targeting posttranscriptional gene silencing and for using RNA-directed methylation in N. benthamiana using a potato virus X vector. Plant J, 25, 417–425.

95. Vigne E, Bergdoll M, Guyader S, Fuchs M (2004). Population structure and genetic variability within isolates of Grapevine fanleaf virus from naturally infected vineayard in France: evidence for mixed infection and recombination. J Gen Virol, 85, 2435-2445.

96. Vigne E, Demangeat G, Komar V, Fuchs M (2005) Characterization of a naturally occurring Grapevine fanleaf virus recombinant isolate. Arch Virol, 150, 2241–2255.

97. Vigne E, Marmonier A, Fuchs M (2008) Multiple interspecies recombination events within RNA2 of Grapevine fanleaf virus and Arabis mosaic virus. Arch Virol, 153, 1771-1776.

98. Vigne E, Marmonier A, Komar V, Lemaire O, Fuchs M (2009). Genetic structure and variability of virus populations in cross-protected grapevines superinfected by Grapevine fanleaf virus. Virus Res, 144, 154-162.
99. Viry M, Serghini MA, 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.
100. Voinnet O, Rivas S, Mestre P, Baulcombe D (2003). An enhanced transient expression system in plants based on suppression of gene silencing by the p19 protein of tomato bushy stunt virus. Plant Journal, 33, 949-956.
101. Wetzel T, Meunier L, Jaeger U, Reustle GM, Krczal G (2001). Complete nucleotide sequences of the RNAS 2 of German isolates of Grapevine fanleaf and Arabis mosaic nepoviruses. Virus Res, 75, 139-145.
102. Wetzel T., Bassler A., Amren MA, Krczal G (2006). A RT/PCR-partial restriction enzymatic mapping (PREM) method for the molecular characterisation of the large satellite RNAs of Arabis mosaic virus isolates. J Virol Methods, 132, 97-103.
103. Xu J, Dolan MC, Medrano G, Cramer CL, Weathers PJ (2011). Green factory: Plants as bioproduction platforms for recombinant proteins. Biotechnology Advances, 30, 1171–1184.
104. Yang KS, Kim HS, Jin UH, Lee SS, Park JA, Lim YP, Pai HS (2007). Silencing of NbBTF3 results in developmental defects and disturbed gene expression in chloroplasts and mitochondria of higher plants. Planta, 225, 1459–1469.
105. Zhang C and Ghabrial (2006). Development of *Bean pod mottle virus*-based vectors for stable protein expression and sequence-specific virus-induced gene silencing in soybean. Virology, 344 (2), 401–411.
106. Zhou X, Huang C (2012). Virus-induced gene silencing using begomovirus satellite molecules. Methods Mol Biol, 2894, 57-67.

2.6. Electronic references

http://www.dpvweb.net/dpv/showdpv.php?dpvno=385
http://www.sahistory.org.za/people/johan-anthoniszoon-jan-van-riebeeck
http://www.sawis.co.za/info/stats_wine_grape_vines.php
http://www.sawis.co.za/info/download/Wynbedryfinligting_mei_2013_Eng.pdf
http://www.wosa.co.za/sa/stats_worldwide.php

Chapter 3

Complete nucleotide sequence of a South African isolate of Grapevine fanleaf virus

The work represented in this chapter was published in a condensed form in Virus Genes.

Lamprecht RL, Maree HJ, Stephan D, Burger JT (2012). Complete nucleotide sequence of a South African isolate of Grapevine fanleaf virus. Virus Genes, 45, 406-410.

3.1. Abstract

The complete sequences of RNA1 and RNA2 have been determined for a South African isolate of Grapevine fanleaf virus (GFLV-SAPCS3). The two RNAs are respectively 7342 and 3817 nucleotides in length, excluding their poly (A) tails. RNA1 has a large open reading frame (ORF) of 6852 nucleotides, and a 5'- and 3' UTR of 243 and 244 nucleotides, respectively. The full length nucleotide sequence of GFLV-SAPCS3 RNA1 had the highest nucleotide identity (86.5%) to the French isolate GFLV-F13. RNA2 comprises of a single ORF of 3330 nucleotides, and has the highest nucleotide identity (90.4%) with GFLV-F13. The 5'- and 3' UTRs of GFLV-SAPCS3 RNA2 are 272 and 212 nucleotides (nt) in length, respectively. The GFLV-SAPCS3 RNA2 5' UTR is 32-53 nt longer when compared to other GFLV isolates. This 5' UTR is also more closely related to GFLV-GHu and Arabis mosaic virus (ArMV) isolates than to other GFLV isolates. Putative intra- and interspecies recombination events between GFLV and ArMV isolates involving GFLV-SAPCS3 RNA1 and RNA2 were investigated. Recombination analysis

software indicated that the GFLV-SAPCS3 5'UTR might have evolved from a recombination event between a GFLV-F13-type and an (ArMV) Ta-type isolate.

3.2. Introduction

Grapevine fanleaf virus (GFLV) is one of the most devastating grapevine viruses worldwide. It causes a wide range of symptoms on grapevine including degeneration and malformation of canes and berries, and yellowing of the leaves (Andret-Link et al., 2004; Hewit, 1969; Raski et al., 1983). Grapevine fanleaf virus has a positive-sense, single-stranded bipartite RNA genome and is classified in the genus *Nepovirus*, family *Secoviridae* (Sanfacon et al., 2009). Each RNA molecule contains a single ORF encoding for the polyproteins, P1 and P2, which are proteolytically cleaved into functional proteins by the RNA1-encoded viral protease. Both RNAs have a 5' covalently attached viral genome-linked protein (VPg) and are polyadenylated at their 3' end. RNA1 is approximately 7.3 Kb in length and P1 is proteolytically cleaved into the proteinase co-factor (also known as protein 1A), a helicase ($1B^{Hel}$), the viral genome-linked protein ($1C^{VPg}$), a proteinase ($1D^{Pro}$) and the RNA-dependant RNA polymerase ($1E^{Pol}$). RNA2 is approximately 3.8 Kb in length and the protein products of P2 are the homing protein ($2A^{HP}$), the movement protein ($2B^{MP}$) and the coat protein ($2C^{CP}$) (Andret-Link et al., 2004; Margis et al., 1993; Pinck et al., 1988; Quacquarelli et al., 1976; Ritzenthaler et al., 1991). To date, the full genome sequences of RNA1 and RNA2 are only available for three strains; isolate GFLV-F13 from France and two Washington isolates, GFLV-WAPN173 and GFLV-WAPN6132 (Serghini et al., 1990; Ritzenthaler et al., 1991; Mekuria et al., 2009). Only partial GFLV $2B^{MP}$ and $2C^{CP}$ sequences are available from the African countries Tunisia and South Africa (SA) (Fattouch et al., 2005; Boulila 2007; Liebenberg et al., 2009). In this chapter the first full-length GFLV sequences for RNA1 and RNA2 from the SA isolate GFLV-SAPCS3 are reported and putative inter- and intraspecies recombination events in RNA1 and RNA2 were investigated, using recombination analysis software.

3.3. Materials and Methods

3.3.1. Virus source, detection and maintenance

GFLV-SAPCS3 was sampled in 2006 from a grapevine plant (*Vitis vinifera* cv Cabernet Sauvignon) collected in the Paarl-Wellington wine growing region of SA. Canes were cut, rooted and maintained in a greenhouse. Virus was propagated in the greenhouse in *Chenopodium quinoa* (6-8 leaf stage) by grinding 1g young grapevine leaves in 4 ml freshly made-up inoculation buffer (30 mM K_2HPO_4, 50 mM Glycine, 1% Celite, 1% Bentonite, pH 9.2) on ice, followed by rub-inoculating the plant sap onto the two top leaves of each plant. Resulting infections were verified 14-18 days post inoculation (dpi), by one step RT-PCR. Crude RNA extractions were made by macerating 0.3 g of newly expanded, systemically infected *C. quinoa* leaves in 3 ml grinding buffer (0.1 M Na_2CO_3-$NaHCO_3$ pH 9.6, 2% (w/v) Polyvinyl Pyrrolidone (PVP) 40, 0.2% (w/v) Bovine Serum Albumin (BSA), 0.05% (v/v) Tween 20, 1% (w/v) $Na_2S_2O_5$). Four microlitres of the extract was added to 1 x GES buffer (100 mM glycine-NaOH pH 8.0, 50 mM NaCl, 1 mM EDTA, 0.5% (v/v) Triton X100) to give a final volume of 30 µl. The sample was denatured for 10 min at 95°C followed by direct cooling on ice for 5 min. Two microlitres of the denatured RNA was directly added to a one step RT-PCR mix (final concentration of 1x Kapa Buffer A, 1x Cresol, 5 mM Dithiothreitol (DTT), 0.4 µM of each primer GFLV1_02b_2772_F and GFLV1_03b_3521_R (Table 3.1), 200 µM dNTPs, 0.5 U AMV RT (Fermentas), 0.5 U Kapa Taq (Kapa Biosystems) and adjusted to a final volume of 25 µl with ddH_20. The cycling conditions were as follows: a reverse transcription step at 45°C for 45 min, one denaturation step at 94°C for 5 min, followed by 30 cycles of 94°C for 30 sec, 55°C for 30 sec, 72°C for 45 sec, and a final elongation step at 72°C for 7 min. PCR products were visualized on a 1% agarose gel stained with Ethidium Bromide (EtBr). A PCR fragment corresponding to 760 bp indicated the successful transmission of GFLV to the herbaceous hosts. For virus transmission to *Nicotiana benthamiana*, plant sap from GFLV-positive *C. quinoa* plants were used to inoculate 6-8 leaf stage *N. benthamiana* plants as described above and resulting infection detection were also

performed as described as above. All mechanical inoculations, and subsequent maintenance of infected herbaceous hosts, were performed in an insect free greenhouse facility with a natural light cycle. The greenhouse facility is controlled with temperatures between 18°C and 28°C and a relative humidity of approximately 70%. To exclude the possibility that the co-infection of ArMV had an influence on the GFLV-SAPCS3 sequencing results, the grapevine plant containing GFLV-SAPCS3, and subsequent infected herbaceous hosts, were additionally screened for the presence of ArMV by DAS-ELISA (Bioreba) following the manufacturer's instructions.

3.3.2. RNA extraction and genome sequencing

3.3.2.1. RNA extraction

For the genome sequencing of GFLV-SAPCS3, three methods were used for RNA extraction. For the amplification of most of the genome, total RNA was extracted from *C. quinoa* leaves using a CTAB method (White et al., 2008). To determine the 5'- and 3' ends, high quality total RNA was extracted from *N. benthamiana* using the Plant RNA Reagent™ kit (Invitrogen), and double stranded RNA (dsRNA) extracted from *N. benthamiana* using a double stranded RNA extraction protocol adapted from Jordan et al., (1983), Morris and Dodds (1979) and Valverde et al. (1990). For GFLV-SAPCS3 dsRNA extraction, 10 g of GFLV-SAPCS3-infected *N. benthamiana* leaves were ground in liquid nitrogen and 100 ml extraction buffer that consisted of 45 ml 2xSTE (0.2 M NaCl, 0.1 M Tris, 2 mM EDTA, pH 6.8), 15 ml 10%SDS, 50 ml phenol-chloroform mixture (1:1 v/v), 1.2 ml 40mg/ml bentonite and 1ml 2-mercapthoethanol) was added to the thawed powder and allowed to shake at room temperature for 30 min. The suspension was centrifuged for 15 min at 10 000 rpm at 20°C, and the supernatant collected and adjusted to 16% EtOH. Three grams of CF-11 cellulose powder (Whatman) were added and mixed at room temperature for 45 min. The sample was loaded onto a column and allowed to run through, and washed with 200ml 1x STE-EtOH buffer (0.1 M NaCl, 50 mM Tris, 1 mM EDTA adjusted to 16% EtOH). The dsRNA was eluted with 50ml 1xSTE buffer and again adjusted to 16% EtOH. The dsRNA was further purified by another round of cellulose chromatography by adding 5g of CF-11

cellulose to the eluate and shaken at room temperature for 30min. The final elution was done by loading the suspension on a new column and run through, washed with 50ml 1xSTE-EtOH (16%) buffer, followed by eluting the dsRNA in 9ml STE buffer. The dsRNA was kept overnight at -20°C after the addition of 0.9 ml 3M sodium acetate and 25ml absolute EtOH. Double stranded RNA was centrifuged at 10 000 rpm for 60min at 4°C, the pellet washed with 70% EtOH and then resuspended in 50µl ddH_20 and viewed on a 1% agarose gel. The dsRNA was stored at -80°C.

3.3.2.2. Primers

Primers for cDNA synthesis and PCR were initially designed from GFLV-F13 sequences available on GenBank (RNA1 accession no. NC003615 and RNA2 accession no. NC003623), and subsequently from newly generated GFLV-SAPCS3 sequences (Table 3.1).

Table 3.1: Primers used in this study to determine the full-length sequences of GFLV SAPCS3 RNA1 and RNA2. Positions of primers as well as the isolate used for primer design are indicated.

Primer Name	Sequence	Designed from isolate	Primer position
RNA1			
5'GFLV_54_RNA1	AAGAGTTTAAGAAACTCAAGA	GFLV-F13	54-74
GFLV1_01_2086_R	ATTGCATTATCAAGTCCCTG	GFLV-F13	2067-2086
GFLV1_02_1801_F	TCGCTTTAGGTGCTATGGTT	GFLV-F13	1801-1820
GFLV1_02b_2772_F	GCGAGTTCTATGATTGATG	GFLV-SAPCS3	2773-2791
GFLV1_02_3321_R	AAGAGAATCCATCATCCTTG	GFLV-F13	3302-3321
GFLV1_03_5583_R	AGATAGAGAGGCGTCCACAT	GFLV-F13	5565-5583
GFLV1_03b_3521_R	CTACCTTGCTTTGTCCT	GFLV-SAPCS3	3522-3538
GFLV1_04_5201_F	AGAAACCCCAAAAGACGAGA	GFLV-F13	5201-5220
GFLV1_04a_5457_F	GATGGCTTGATTACGGACC	GFLV-SAPCS3	5458-5476
GFLV1_04_6912_R	CTGTTTCCGAGTACATGGCT	GFLV-F13	6894-6913
GFLV1_05_6706_F	ATGCCAAAACAAACTCTTAT	GFLV-F13	6706-6725
GFLV1_outer _SMART	GCACCAAAGGCGGATAAACA	GFLV-SAPCS3	570-589
GFLV1_inner _SMART	CCTTTGGATTTTGGGCTGAC	GFLV-SAPCS3	469-488
GFLV_RNA1_RACE_GSP1	TTCACACCTTTGCTTTAG	GFLV-SAPCS3	514-531
GFLV_RNA1_RACE_GSP2	CGCTATCCTGCTCTTCCTTA	GFLV-SAPCS3	383-402
GFLV_RNA1_RACE_GSP3	AGTTCCTTAGCCTCCGCATT	GFLV-SAPCS3	298-317
RNA2			
GFVL2_1_01_F	ATGGGCAARTTTTATTAYCC	GFLV-F13	233-252
GFLV2_1_858_R	GATCCATTGTGCCCTACCG	GFLV-F13	1087-1105
GFLV2_2_548_F	GGAGGAAGTGGTATGACACTAGTG	GFLV-F13	777-800
GFLV2_2_1724_R	CGTGAACGGCTTCTGATACAG	GFLV-F13	1954-1974
GFLV2_3_1573_F	GCHACTATGATGGGTAAYAC	GFLV-F13	1082-1821
GFLV2_3_2915_R	CTATGCAATCCATGGCARGC	GFLV-F13	3218-3237
GFVL2_4_2848_F	GAAGATGATTATTTYGTGTG	GFLV-F13	3077-3096
GFLV2_4_3333_R	CTAGACTGGGAAGCTGG	GFLV-F13	3547-6-3562
GFLV2_outer_SMART	TGGAGAAGGGGGTGGTGACG	GFLV-SAPCS3	601-620
GFLV2_innerR_SMART	TGGCGGAAGGAGGAATCAGT	GFLV-SAPCS3	375-394
GFLV_RNA2_RACE_GSP1	AAGAAAACATCCGAACAG	GFLV-SAPCS3	439-456
GFLV_RNA2_RACE_GSP2	ACAGTTTGGCGGAAGGAGGA	GFLV-SAPCS3	381-400
GFLV_RNA2_RACE_GSP3	AGACAAATAAGACGCCCGCT	GFLV-SAPCS3	309-327
Universal primers			
		Procedure	**Reference**
dT(17)	TACGATGGCTGCAGT(17)	3' end sequence determination	Meng et al., 2005
TSrG 5' end	AAGCAGTGGTATCAACGCAGAGTA CGCGGG	SMART	Matz et al., 1999
HeelA (carrier)	GTAATACGACTCACTATAGGGCAAG CAGTGGTATCAACGCAGAGGTCCC	SMART	Matz et al., 1999
HeelB (specific)	GTAATACGACTCACTATAGGGCA	SMART	Matz et al., 1999
Abridged Anchor Primer	GGCCACGCGTCGACTAGTACGGGII GGGIIGGGIIG	5' RACE System	Invitrogen

3.3.2.3. cDNA synthesis and PCR

High fidelity enzymes for cDNA synthesis (Superscript III Reverse Transcriptase, Invitrogen) and PCR (Ex Taq DNA Polymerase, Takara) were used to avoid introducing nucleotide errors during amplification. For cDNA synthesis, approximately 300 ng total RNA extracted from GFLV-SAPCS3-infected *N. benthamiana*, 1 μM antisense primer and 0.5 μM dNTPs were mixed and made up to a 13 μl total volume with ddH_2O, and incubated at 65°C for 5 min, followed by immediate cooling on ice for 1 min. The RNA and primer mixture was added to 10 μl cDNA synthesis mix (1x First Strand Synthesis Buffer, 5 mM DTT, 2 U RNaseOUT, 10 U Superscript III RT). The reactions were incubated at 50-55°C for 1 hour, and inactivated at 85°C for 5 min. Five microlitres of the cDNA was used for PCR reactions.

For PCR amplification, different combinations of the sense and antisense primers were used to obtain amplicons that would span the entire genome (Figure 3.1). The annealing temperatures and the extension times for each PCR reaction were dependent on the primer combinations used. For PCR amplification, 5 µl cDNA was mixed with 1x Ex Taq Buffer, 1x Cresol, 0.2 µM of each sense and antisense primer, 200 µM dNTPs and 1.25 U Ex Taq to a final volume of 50 µl. The cycling conditions were as follows: 94°C for 5 min, 30 cycles of 94°C for 30 sec, specific annealing temperature for 1 min, 72°C for specific extension time, followed by final elongation at 72°C for 7 min. The PCR reactions were held at 4°C and stored at -20°C. PCR products were separated and viewed on a 1% agarose gel stained with EtBr.

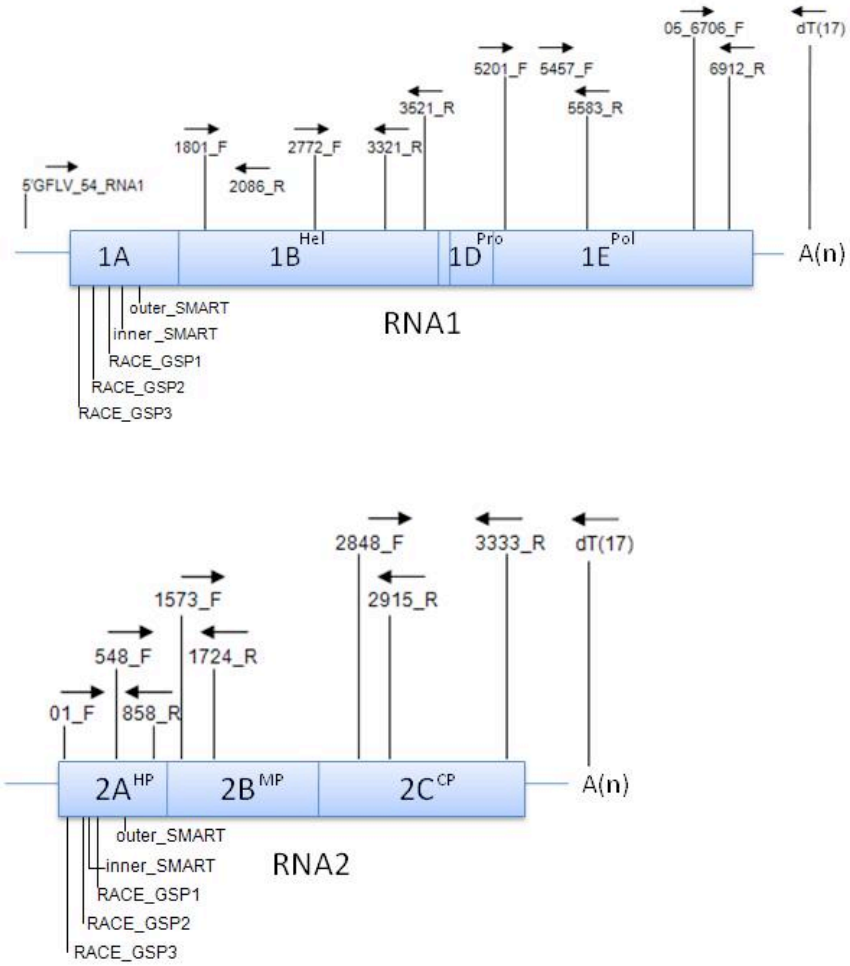


Figure 3.1: Sequencing strategy of GFLV-SAPCS3 RNA1 and RNA2. The blue blocks represent the genes of RNA1 and RNA2 and the blue horizontal lines represent the 5' and 3' UTRs. The arrows represent the primers and their orientation while vertical black lines lines represent the primer positions (not drawn to scale). Sense and antisense primers were used in different combinations to obtain PCR fragments. The primers below the horizontal blue boxes are the GFLV-SAPCS3 specific antisense primers used for the 5' UTR determination. The poly(A) tail is represented by A(n). Details of the primers are listed in Table 3.1.

3.3.2.4. Cloning

PCR products with the correct size were excised and purified using a Zymoclean Gel DNA Recovery kit (Zymo Research). Purified amplicons were either sequenced directly, or sequenced bidirectionally after being cloned into a TA cloning vector (pGEM-T Easy, Promega). About 20 ng of each gel-purified PCR product was ligated into 50 ng of the TA cloning plasmid pGEM-T-Easy (Promega) in a 10 µl reaction. The ligation mix was prepared according to the manufacturer's instructions and kept at 4°C overnight. For transformation, 100 µl of chemically competent *E. coli* JM109 cells was added to the

ligation reaction, mixed gently and incubated on ice for 20 min. The cells were heat shocked for 45 sec in a 42°C water bath and immediately cooled on ice for 2 min. Nine hundred microlitres of sterile Luria Broth (LB, Merck) was added to the transformation reaction and shaken at 150 rpm for 90 min at 37°C. One hundred microlitres was plated onto LB bacteriological agar (Merck) plates containing 100 µg/ml Ampicillin (Amp), and for blue-white screening, 40 µg/ml 5-bromo-4-chloro-3-indolyl-β-D-galactoside (X-Gal, Fermentas) and 0.2 mM Isopropyl-β-D-thiogalactoside (IPTG, Fermentas) were added. Recombinant colonies were picked with a sterile toothpick and used to inoculate LB medium, containing Amp, to grow overnight shaking at 150rpm at 37°C. Plasmids were extracted using the GeneJet Plasmid Miniprep kit (Thermo Scientific) according to the manufacturer's instructions and plasmids isolated were confirmed by control restriction digests. Plasmids with the correct insert sizes were sequenced. The sequences generated from the overlapping amplicons were used to build a contiguous sequence using CLC Main Workbench version 6.5 (CLC Bio).

3.3.2.5. 5' and 3' end sequence determination

The 5' end sequence of GFLV-SAPCS3 was determined by SMART (switching mechanism at the 5' end of the RNA transcript) (Matz et al., 1999; Zhu et al., 2001) and the 5' RACE System (Invitrogen). For the SMART procedure, first strand cDNA was synthesized using Superscript III RT on dsRNA as template with antisense GFLV-SAPCS3 primers GFLV1_outer_SMART or GFLV2_outer_SMART, and the "switching" mechanism was made with the SMART primer TSrG_5' end. The 5' ends of RNA1 and RNA2 were amplified using Ex Taq (Takara) with the HeelA primer and GFLV-SAPCS3 primers GFLV1_outer_SMART or GFLV2_outer_SMART, and then re-amplified with the HeelB primer and nested primers GFLV1_inner_SMART or GFLV2_inner_SMART (see Table 3.1 for primer details). PCR amplicons representing the 5' end of each RNA from the SMART procedure were gel purified, cloned into pGEM-T-Easy and four cDNA clones of each were sequenced.

For the 5' RACE System (Invitrogen), gene-specific reverse primers (RNA1_RACE_GSP1 and RNA2_RACE_GSP1) and the Abridged Anchor primers were

used or the first strand cDNA synthesis, following the manufacturer's instructions. GFLV-SAPCS3-specific nested primers for RNA1 (GFLV_RNA1_RACE_GSP2 and GFLV_RNA1_RACE_GSP3) and for RNA2 (GFLV_RNA2_RACE_GSP2 and GFLV_RNA2_RACE_GSP3) were used to generate amplicons, which were then cloned. Six cDNA clones for each RNA segment were sequenced, from which a consensus sequence was derived.

Primer $dT_{(17)}$ (Meng et al., 2005) was used for cDNA synthesis of RNA1 and RNA2 for the determination of the 3' terminal sequence using Superscript III RT (Invitrogen). For PCR reactions, GFLV-F13 sequence-specific primers, complementary to RNA1 (GFLV1_05_6706_F) and RNA2 (GFVL2_4_2848_F), were used with primer $dT_{(17)}$. PCR products were cloned into pGEM-T-Easy and three clones of each were sequenced.

3.3.2.6. Phylogenetic and recombination analysis

Full-length sequences of GFLV-SAPCS3 RNA1 and RNA2 were compiled by assembling all newly generated sequences. Multiple sequence alignments of RNA1 and RNA2 of GFLV and ArMV isolates were performed using the ClustalW (Thompson et al., 1994) default settings. The nucleotide and protein sequence identity, phylogenetic analysis and pairwise distance calculations were performed using the Mega5 analysis package (Tamura et al., 2011). Sequences of Tobacco ringspot virus (TRSV, Accession numbers U50869 and AY363727) were used as an outgroup for phylogenetic tree analysis. For recombination analysis, Recombination Detection Program version 3.44 (RDP3) (Martin et al., 2010) and Simplot version 3.2 (Lole et al., 1999) were used with other full-length GFLV and ArMV sequences to find possible recombination events between GFLV-SAPCS3 and the different isolates. Default settings were used.

3.4. Results

3.4.1. Virus detection and maintenance

Plant sap of GFLV-SAPCS3 was routinely mechanically inoculated to either *C. quinoa* or *N. benthamiana*. The grapevine from which GFLV-SAPCS3 was isolated (and subsequent GFLV-SAPCS3-infected herbaceous hosts) was not infected with ArMV as tested by DAS-ELISA (results not shown). The first transmission experiment was done by rub-inoculating GFLV-SAPCS3 plant sap onto 15 *C. quinoa* (6-8 leaf stage) plants, including negative controls inoculated with buffer alone. At 14 dpi, 11 out of the 15 plants showed yellow mottling and deformation of the leaves, as well as yellow line patterns on the leaves (Figure 3.2). The other 4 plants showed no symptoms while the healthy and buffer controls also showed no symptoms. The 11 symptomatic plants were shown to be infected with GFLV when the expected 760 bp amplicon was produced by RT-PCR on newly expanded, systemically-infected leaves. Once GFLV-SAPCS3 was transferred to *C. quinoa*, the virus was maintained in *C. quinoa* or *N. benthamiana* by mechanically inoculating new plants every 4-6 weeks. When transferred from *C. quinoa* to *N. benthamiana*, the virus caused little or no symptoms on *N. benthamiana* under the same greenhouse conditions, although RT-PCR confirmed GFLV infection (Figure 3.3). Whenever virus transmission was between herbaceous hosts, the resulting infection rate was always between 95-100% as screened by RT-PCR.

Figure 3.2: The symptoms expressed by GFLV-SAPCS3 on *C. quinoa* (14dpi).

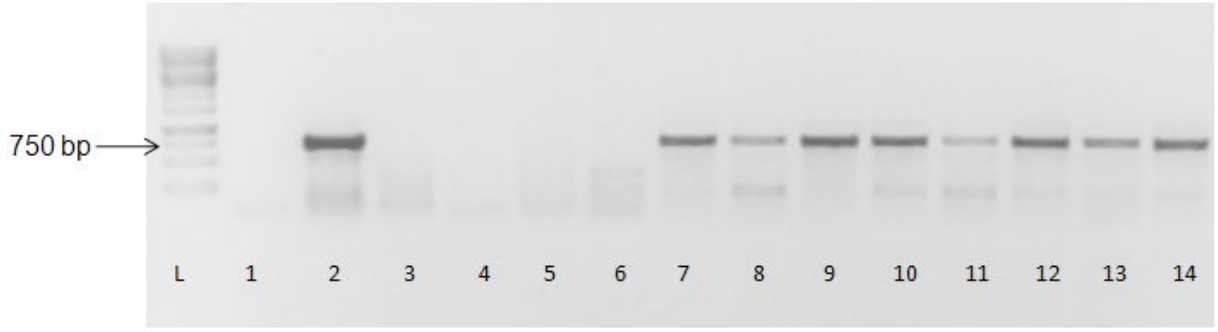


Figure 3.3: PCR products for GFLV detection in systemically infected leaves in the herbaceous host *N. benthamiana*. A 760 bp fragment indicated successful virus transmission and systemic infection. (L) GeneRuler 1kb ladder (Fermentas), (1) no template PCR control, (2) GFLV-SAPCS3 total RNA positive PCR control, (3) GES buffer no template PCR control, (4) buffer-inoculated plant, no virus (5) healthy plant, (6-14) GFLV-SAPCS3 transmitted *N. benthamiana plants* (14 dpi). In this round of transmission, only one plant (lane 6) was not infected with GFLV-SAPCS3. The RT-PCR results show GFLV infection, however these plants showed no symptoms.

3.4.2. RNA extraction and genome sequencing

The CTAB method (White et al., 2008) and the Plant RNA Reagent™ kit (Invitrogen), were used for extracting total RNA from *C. quinoa* and *N. benthamiana*, and both methods yielded good total RNA. For both methods, absorbance levels of 1.9-2.1 (measured at A260/280) and 1.8-2.2 (measured at A260/230 absorbance ratio) were obtained. Double stranded RNA was also extracted from *N. benthamiana* to determine the 5' end sequence of GFLV-SAPCS3 RNA1 and RNA2. The dsRNA was visualized on an agarose gel to inspect its quality (Figure 3.4) and the two bands at >7 kb and >3.5 kb that represented RNA1 and RNA2 could be seen. As the dsRNA preparation did not include an RNAse step, ribosomal RNA could also be detected on the gel, indicated by the asterisks (Figure 3.4).

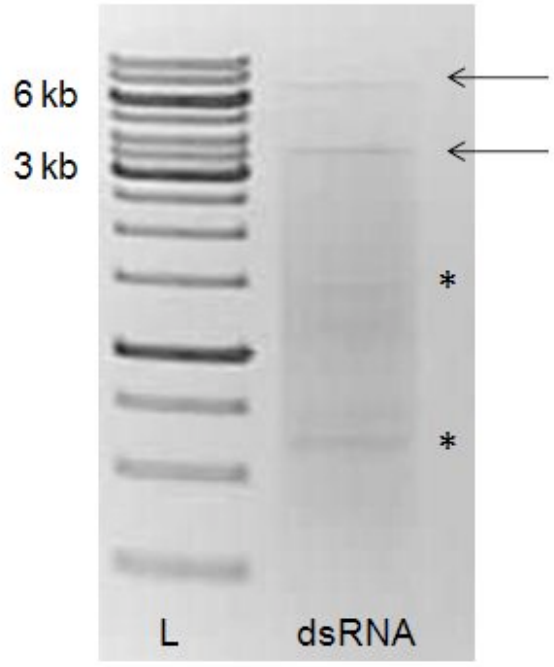


Figure 3.4: Double stranded RNA extracted from GFLV-SAPCS3-infected *N. benthamiana*. The two upper bands indicated by the arrows are the RNA1 and RNA2 segments of GFLV. The bands indicated by the asterisks are most likely ribosomal RNA (40S and 60S rRNA), as the dsRNA preparation did not include

an RNAse step. The dsRNA was used as template for the determination of the 5' ends of RNA1 and RNA2 of GFLV-SAPCS3.

For genome sequencing of GFLV-SAPCS3, six and five overlapping amplicons spanning most of RNA1 and RNA2, respectively, were cloned and sequenced and a consensus sequence was constructed that excluded the primer sequences. Both the SMART procedure (Matz et al., 1999; Zhu et al., 2001) and the 5' RACE System (Invitrogen) were used to determine the 5' end sequences. Initially, the sequences generated by SMART were verified by including them in an alignment of different GFLV isolates. The alignment indicated that the sequences generated by the SMART procedure were incomplete at the 5' end for both GFLV-SAPCS3 RNA1 and RNA2 (short by 10 and 13 nt respectively). To complete the 5' UTR sequences of RNA1 and RNA2 the 5' RACE System (Invitrogen) was used. Six clones of each were sequenced and the sequencing results were all identical. Three clones representing the 3' ends of GFLV-SAPCS3 RNA1 and RNA2 were sequenced. All the sequences generated were compiled and a contiguous sequence representing GFLV-SAPCS3 RNA1 and RNA2 were generated using CLC Main Workbench version 6.5 (CLC bio).

Sequencing revealed that the RNA1 of GFLV-SAPCS3 was 7342 nt in length, excluding the poly(A) tail. The 5' UTR of RNA1 is 243 nt and the 3' UTR is 244 nt in length. The ORF of RNA1 is 6852 nt in length. RNA2 of GFLV-SAPCS3 is 3817 nt in length, excluding its poly(A) tail. The 5' and 3' UTRs are 272 nt and 212 nt in length, respectively. The 5' UTR of SAPCS3 is 32-53 nt longer than those of other GFLV isolates, except for the Hungarian isolate GFLV-GHu (acc EF426852), which also has a long 5' UTR (264 nt). The ORF located on RNA2 coding for P2 is 3330 nt in length. The complete sequences of RNA1 and RNA2 have been deposited in the GenBank database with the accession numbers JF968120 and JF968121.

3.4.3. Phylogenetic and recombination analysis

Multiple full-length nucleotide and amino acid sequence alignments of GFLV-SAPCS3 RNA1 and RNA2, and other full-length GFLV and ArMV isolates were performed. The

full-length nucleotide and amino acid (aa) sequences of GFLV-SAPCS3 RNA1 and RNA2 were the closest to the French isolate GFLV-F13, showing a nucleotide identity of 86.5% (P1 aa 94%) and 90.4% (P2 aa 98%), respectively. The RNA1 full-length sequence of GFLV SAPCS3 was more distantly related to GFLV Washington isolate WAPN6132 with a nucleotide identity of 85% (P1 aa 92%). Similarly, the GFLV-SAPCS3 RNA2 full-length nucleotide sequence was more distantly related to GFLV-GHu with a nucleotide identity of 81.9% (P2 aa 86%) followed by the GFLV Washington isolates WAPN165 and WAPN6132 (both nucleotide identities 83.8% and P2 aa 91%).

Multiple nucleotide sequence alignments of ORFs encoding P1 and P2 of GFLV-SAPCS3 and all other available full-length sequences of GFLV and ArMV isolates indicated that GFLV-SAPCS3 grouped separately between GFLV and ArMV isolates. The phylogenetic tree based on P1 sequences suggested that GFLV SAPCS3 grouped separately (Figure 3.5). More GFLV RNA1 full-length sequences are however needed to clarify the phylogenetic relationship between GFLV isolates.

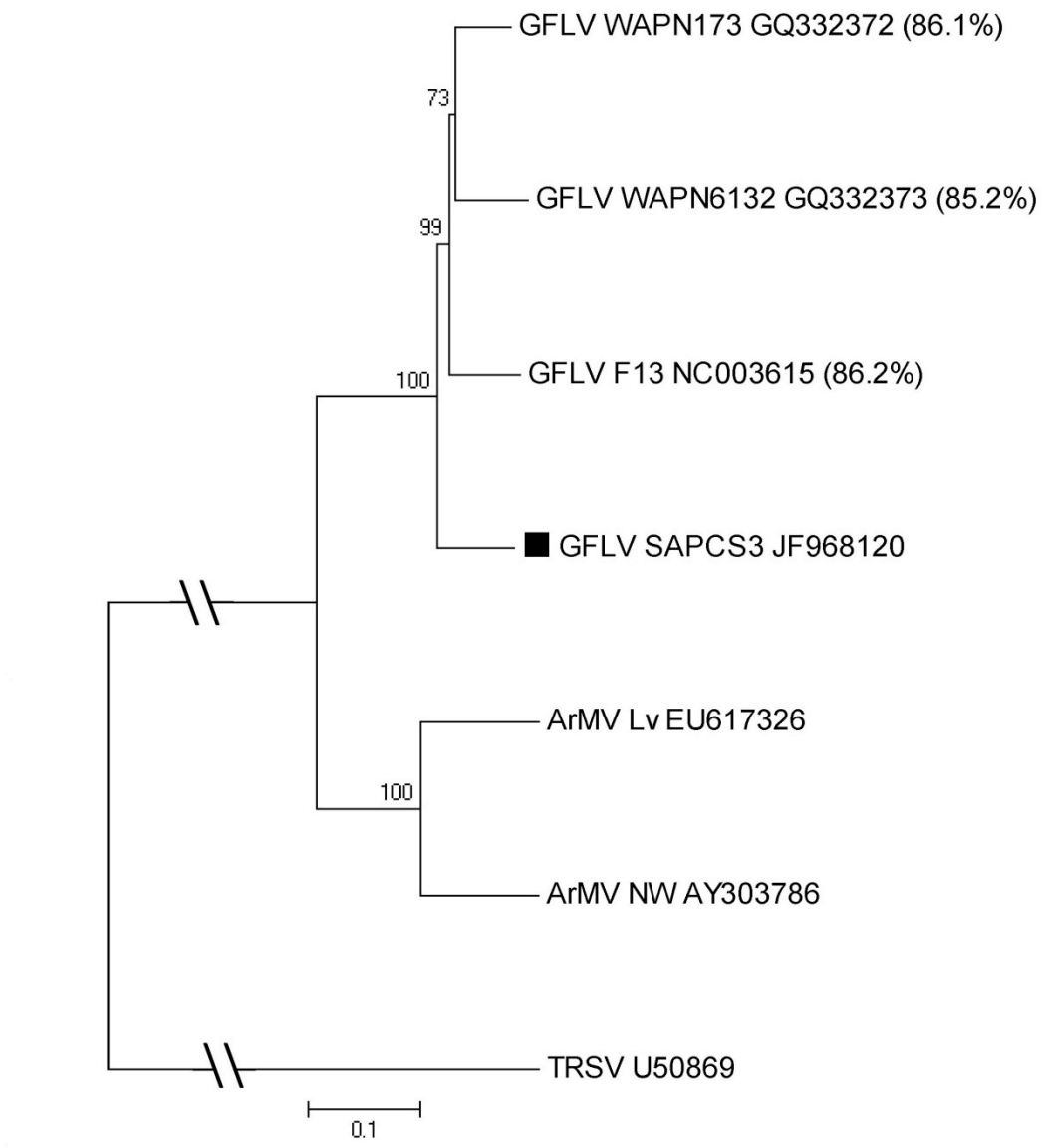


Figure 3.5: Phylogenetic tree based on nucleotide sequences of P1 of GFLV and ArMV isolates. The nucleotide identities between GFLV SAPCS3 and other GFLV isolates are indicated in brackets. The accession numbers of the isolates are indicated next to the isolate names. GFLV-SAPCS3 is indicated on the tree as a solid block. All the phylogenetic trees were constructed using the Neighbour-Joining method. The percentage of replicate trees in which the associated taxa clustered together in the bootstrap test (1000 replicates) is shown next to the branches. The evolutionary distances were computed using the Maximum Composite Likelihood method. Phylogenetic analysis was conducted in MEGA5 (Tamura et al., 2011)

The phylogenetic tree based on the P2 sequences of various GFLV isolates indicated that GFLV-SAPCS3 grouped separately to the grouping of GFLV-F13, the Californian strains CACSC1, CAZINA4, CACSB5, as well as to the grouping of Washington GFLV isolates (WAPN165, WAPN6132, WACH911, WAPN8133, WAPN173 and WAPN57), while the ArMV isolates grouped together in different clades and GFLV-NW, GFLV-WACF2142 and GFLV-GHu also all grouped separately (Figure 3.6).

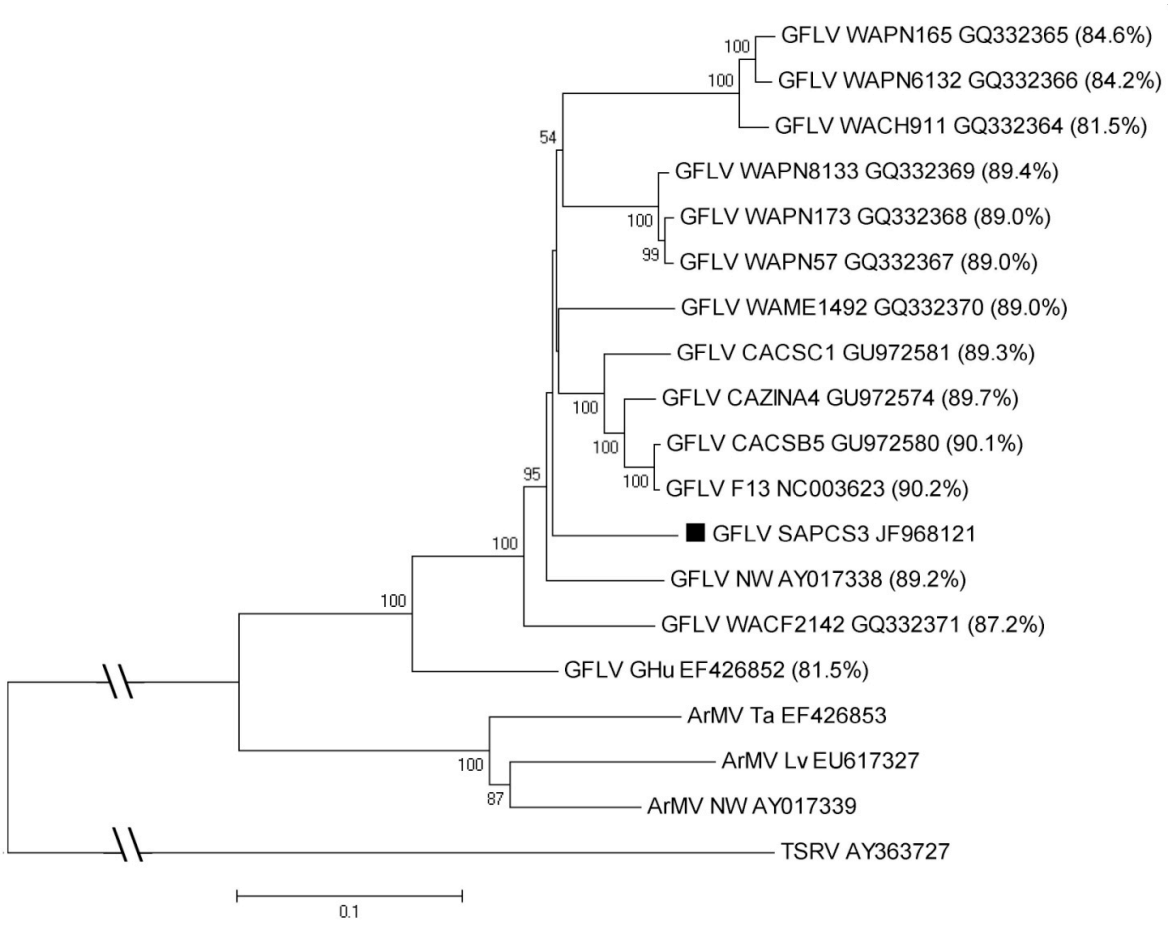


Figure 3.6: Phylogenetic tree based on nucleotide sequences of P2 of GFLV and ArMV isolates. The nucleotide identities between GFLV SAPCS3 and other GFLV isolates are indicated in brackets. The accession numbers of the isolates are indicated next to the isolate names. GFLV-SAPCS3 is indicated on the tree as a solid block. All the phylogenetic trees were constructed using the Neighbour-Joining method. The percentage of replicate trees in which the associated taxa clustered together in the bootstrap test (1000 replicates) is shown next to the branches. The evolutionary distances were computed using the Maximum Composite Likelihood method. Phylogenetic analysis was conducted in MEGA5 (Tamura et al., 2011)

The availability of a higher number of $2C^{CP}$ sequences allowed the construction of a phylogenetic tree incorporating significantly more GFLV isolates (Figure 3.7). This $2C^{CP}$-based tree confirmed a previous study showing that the South African isolates could be divided in two clades (Liebenberg et al., 2009). GFLV-SAPCS3 clustered with the South African isolates in Clade A together with GFLV isolates mostly from the USA, Slovenia and South America, whereas Clade B includes the other South African isolates and GFLV isolates mostly from France and USA.

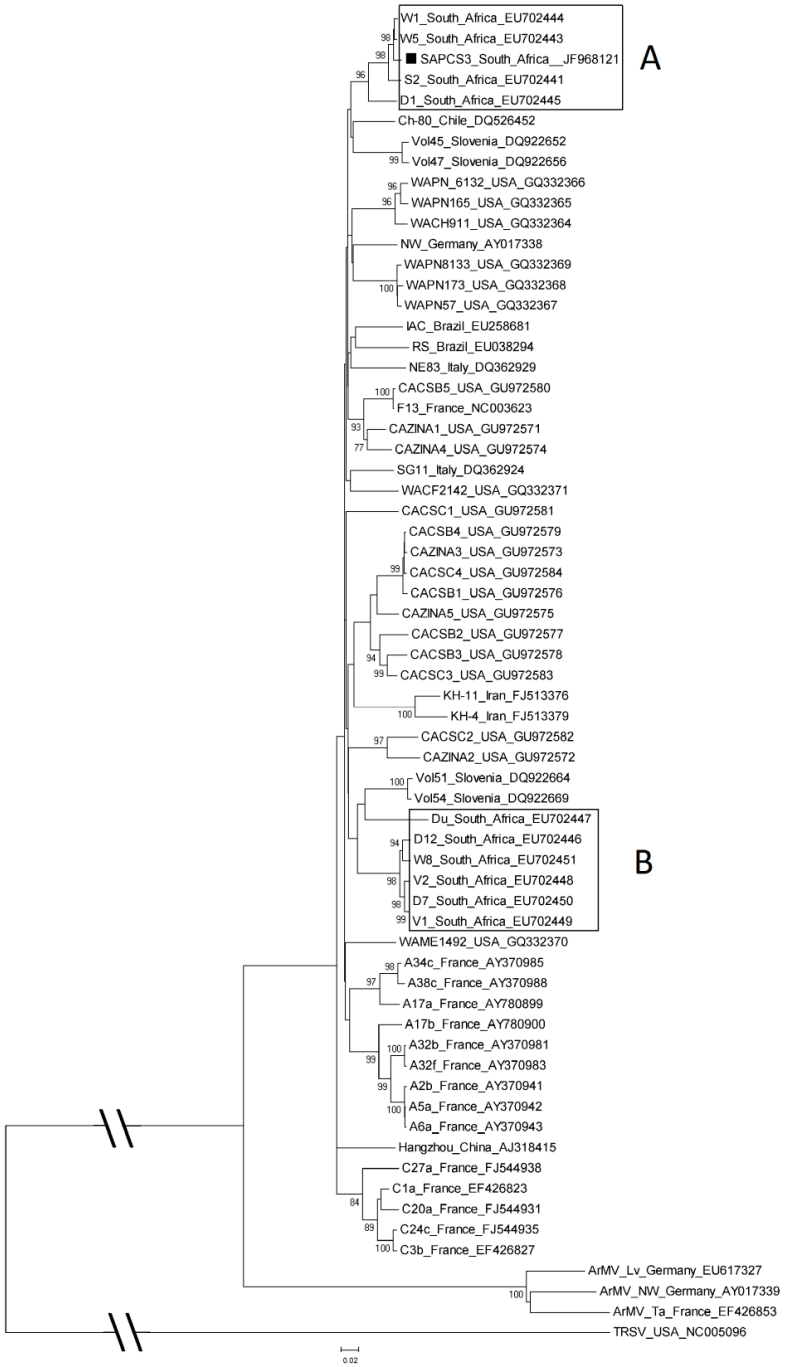


Figure 3.7: Phylogenetic tree based on the CP nucleotide sequences of GFLV and ArMV isolates. The accession numbers of the isolates are indicated next to the isolate names. GFLV-SAPCS3 is indicated on the tree as a solid block. The South African isolates are blocked to represent the two clades, A and B. All the phylogenetic trees were constructed using the Neighbour-Joining method. The percentage of replicate trees in which the associated taxa clustered together in the bootstrap test (1000 replicates) is shown next to the branches. The evolutionary distances were computed using the Maximum Composite Likelihood method. Phylogenetic analysis was conducted in MEGA5 (Tamura et al., 2011).

Inter- and intraspecies recombination in RNA2 between GFLV and ArMV isolates have been described extensively (Mekuria et al., 2009; Oliver et al., 2010; Pompe-Novak et al., 2007; Vigne et al., 2004; Vigne et al., 2005; Vigne et al., 2008; Sokhandan-Bashir and Melcher, 2002) but there are little reports for such events for GFLV RNA1. Putative recombination events for RNA1 and RNA2 of GFLV-SAPCS3 were assessed using RDP3 (Martin et al., 2010) and Simplot (Lole et al., 1999). Putative intraspecies recombination events in RNA1 were found with a GFLV-F13-type as the recombinant and a GFLV-WAPN173-type and a GFLV-SAPCS3-type as parental lineages (major and minor parents, respectively) with a RDP p-value of $3.876x10^{-4}$. The detected recombination breakpoint is located within the $1E^{Pol}$ gene starting at 6821 nt towards the end of 3' end (no breakpoint end determined). A Simplot analysis of aligned RNA1 sequences from recombinant GFLV-F13-type with GFLV-SAPCS3-type and GFLV-WAPN173-type confirmed this recombination event within $1E^{Pol}$ (Figure 3.8). The putative recombination event in $1E^{Pol}$ was further demonstrated by calculating the pairwise-distances between the three isolates before and after position 6821. Pairwise distances calculated from position 1-6821 nt indicated that GFLV-F13 had the highest nucleotide identity to GFLV-WAPN173 (89.8%) compared to GFLV-SAPCS3 (87.3%). However, from position 6812-7341, GFLV-F13 had the highest nucleotide identity to GFLV-SAPCS3 (88.7%) compared to GFLV-WAPN173 (78.1%). A phylogenetic tree based on the 3' UTR of RNA1 of the different isolates further demonstrated this putative recombination event, showing GFLV-F13 and GFLV-SAPCS3 grouping together while GFLV isolates WAPN173 and WAPN6132 grouped together (Figure 3.9). The other predicted recombination event in RNA1 involved a GFLV-WAPN173-type, with parental lineages a GFLV-F13-type and a GFLV-WAPN6132-type in the $1A\text{-}1B^{Hel}$ gene region with a RDP p-value of $4.251876x10^{-4}$. This intraspecies recombination event in GFLV-WAPN173 was also predicted in a previous study (Mekuria et al., 2009) and the similar results obtained in our study and from the previous study (Mekuria et al., 2009) thus validates the putative recombination event in $1E^{Pol}$ described in this chapter. No statistically significant interspecies recombination events in RNA1 were found between any of the other GFLV and ArMV isolates in this study, or any other previous study

(Oliver et al., 2010). More full-length sequences of GFLV and ArMV RNA1 are needed to clarify the phylogenetic relationship between GFLV and ArMV RNA1.

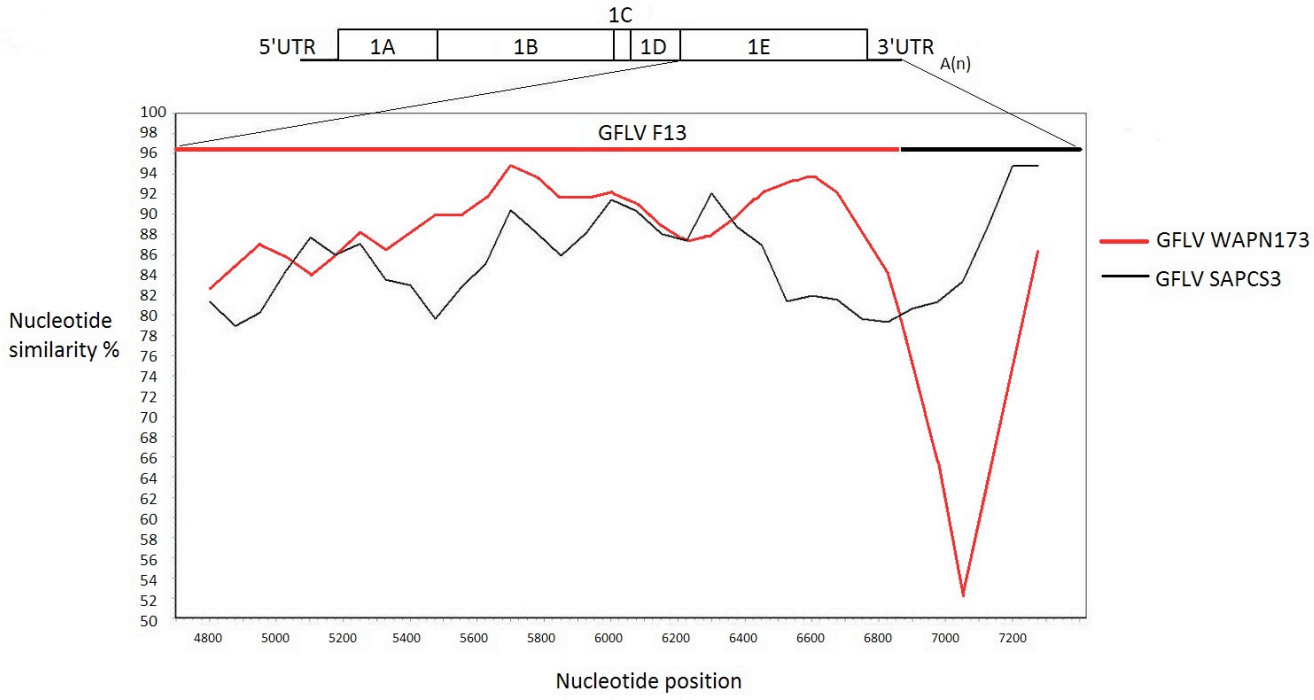


Figure 3.8: Graphic representation of Simplot analysis within $1E^{Pol}$. The $1E^{pol}$ and 3'UTR of GFLV-F13 RNA1 (nt 4623 – 7342) is compared to the two putative parent isolates GFLV-WAPN173 (red) and GFLV-SAPCS3 (black) with the recombination breakpoint at 6821 nt with no end breakpoint determined. Window size: 200 nt; step size: 75 nt. The ORFs are represented by the wide boxes on top of the figure, and the 5' and 3' UTRs are represented by the bold black lines attached to the wide boxes and the poly(A) tail is represented by A(n).

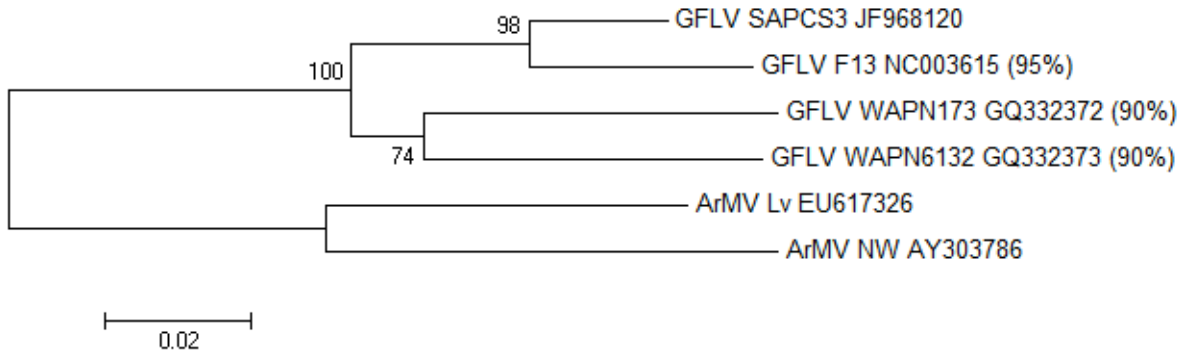


Figure 3.9: A phylogenetic tree based on the RNA1 3' UTRs of available GFLV isolates. This tree confirms the high nucleotide identity in this region between GFLV-F13 and GFLV-SAPCS3, due to a putative recombination event in GFLV-F13 (GFLV-WAPN173 and GFLV-SAPCS3 as the major and minor parents, respectively). The nucleotide identities in the 3' UTR between GFLV SAPCS3 and other GFLV isolates are indicated in brackets. The accession numbers of the isolates are indicated next to the isolate names. All the phylogenetic trees were constructed using the Neighbour-Joining method. The percentage of replicate trees in which the associated taxa clustered together in the bootstrap test (1000 replicates) is shown next to the branches. The evolutionary distances were computed using the Maximum Composite Likelihood method. Phylogenetic analysis was conducted in MEGA5 (Tamura et al., 2011).

Recombination events within GFLV RNA2 have been described, including interspecies recombination events between GFLV-GHu and ArMV-Ta (acc EF426853) (Vigne et al., 2008). Both GFLV-GHu and GFLV-SAPCS3 isolates have long 5' UTRs with lengths comparable to those of ArMV isolates. Simplot analysis (Figure 3.10) on the 5' UTR alignment of GFLV-SAPCS3 and GFLV-F13 and ArMV isolates indicated that GFLV-SAPCS3 is a recombinant with parental lineages of a GFLV-F13-type (major parent) and a ArMV-Ta-type (minor parent) in the first 171 nt of the 5' UTR (no start breakpoint determined). Since the grapevine plant from which GFLV-SAPCS3 was isolated was negative for ArMV infection, this possible recombination event had to occur in a different plant at an undetermined time. This putative recombination event is supported by the high 5' UTR sequence identity found between GFLV-SAPCS3 and GFLV-Ghu and the other three available ArMV isolates (68-77.3%) in the first 171 nt. An insertion of 22-25 nt, which includes the conserved sequence stretch AA/GTCCGTT/CA, is present in

GFLV-SAPCS3 (position 73-98), GFLV-GHu, ArMV-Ta, ArMV-NW and ArMV-Lv, but cannot be found in any other GFLV isolate.

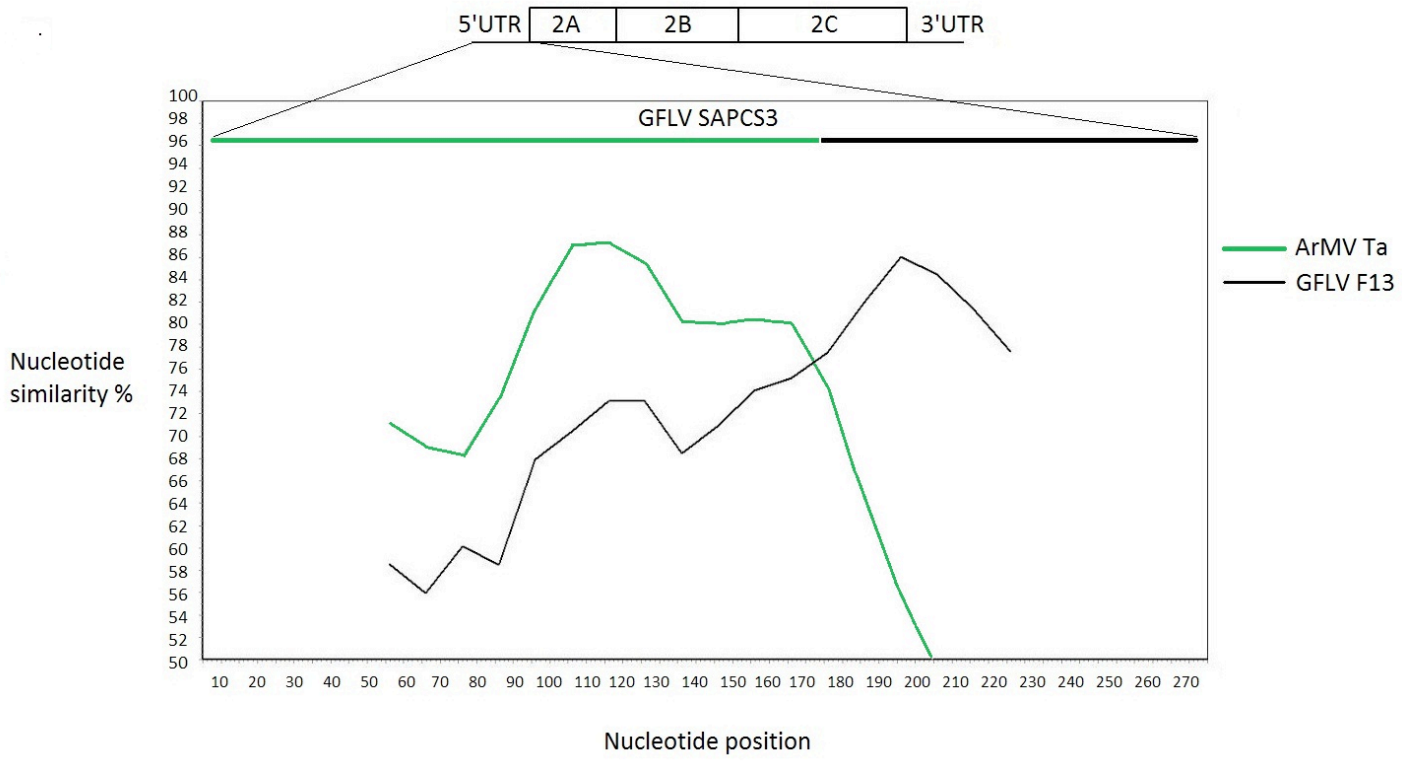


Figure 3.10: Graphic representation of Simplot analysis within RNA2 5'UTR. The RNA2 5'UTR of GFLV-SAPCS3 is compared to the putative recombination parental isolates GFLV-F13 (black) and ArMV-Ta (green). The recombination breakpoint in the 5'UTR is at position 171 nt with no start breakpoint determined. Window size: 100 nt; step size: 10 nt. The ORFs are represented by the wide boxes on top of the figure, and the 5' and 3' UTRs are represented by the bold black lines attached to the wide boxes and the poly(A) tail is represented by A(n).

3.5. Discussion

Isolate GFLV-SAPCS3 is the first GFLV isolate from SA to be characterised on a biological and molecular level. Symptomatology associated with GFLV-SAPCS3 when transmitted to *C. quinoa* included deformation of the leaves as well as mosaic and line patterns on new systemically-infected leaves. The mechanisms of symptom development in plants, following virus infection, remains poorly understood. For nepoviruses, back in the 1970s it was shown that the RNA2 of Raspberry ringspot virus

caused yellowing in *Petunia hybrida* while RNA1 was responsible for the severity of these symptoms in *C. quinoa* (Harrison et al., 1972). Two more recent reports indicated the potential symptom determinants of nepoviruses. For Grapevine chrome mosaic virus, it was shown that the 5' noncoding region of RNA2 triggers a necrotic response in *N. benthamiana*, by using a potato virus X (PVX) viral vector. However, the necrotic symptoms induced from the vector were not the typical symptoms produced by the native virus, suggesting that this sequence may not be the dominant symptom determinant (Fernandez et al., 1999). However, these results were obtained using a PVX-based vector, therefore the heterologous system may have induced the atypical symptoms. The use of homologous viral vector systems should produce conclusive results. With regards to GFLV, previous studies have found no association between symptomatology and genetic variation in GFLV populations. However, these studies have focussed on the genetic variability in RNA2 among different isolates expressing different symptomatologies (Pompe-Novak et al., 2007, Liebenberg et al., 2009). Recently, the symptom determinants in GFLV were mapped to the $1E^{Pol}$ of RNA1 by the use of full-length infectious GFLV cDNA clones. The authors mapped the symptom determinants by using the infectious cDNA clones of GFLV-F13 and GFLV-Ghu that displayed distinct symptom expressions on *N. benthamiana*. A gain-of-symptom approach was employed by using different assortants and also developing chimeras of the clones representing the two isolates, and analysing the observed symptoms on *N. benthamiana.* It was unambiguously concluded that the GFLV-Ghu symptoms could be mapped to RNA1, and later shown that the coding region responsible for symptom expression is in the $1E^{Pol}$ (Vigne et al., 2012). The putative recombination event in the $1E^{Pol}$ shown in this study may be significant with regards to symptom development. The symptomatology associated with GFLV-SAPCS3 in *C. quinoa* is rather unique, and it would be interesting to know how these symptoms compare to the GFLV-F13 symptomatology on *C. quinoa*, as these two isolates share similar sequences in the $1E^{Pol}$. More GFLV full-length infectious clones of different isolates eliciting different symptoms are needed to further investigate the role of $1E^{Pol}$, and maybe other regions, on symptom expressions in herbaceous hosts and grapevine.

Isolate GFLV-SAPCS3 RNA1 and RNA2 were sequenced using basic RT-PCR based techniques, and the genomic full-length sequences were deposited in GenBank with accession numbers JF968120 and JF968121, respectively. Phylogenetically, GFLV-SAPCS3 RNA1 and RNA2 grouped separately from other GFLV isolates, indicating that this is a new variant. The RNA2 5' UTR of GFLV-SAPCS3 is the longest of the GFLV 5' UTRs reported to date. The RNA2 5' UTRs of GFLV-SAPCS3 and GFLV-Ghu are very similar in size and nucleotide identity to the 5' UTRs of ArMV isolates. In a previous study, interspecies recombination was detected in the 5' UTR of GFLV-Ghu RNA2, involving recombination events between GFLV and ArMV. This was the first report of interspecies recombination between GFLV and ArMV (Vigne et al., 2008). GFLV-SAPCS3 and GFLV-Ghu have similar RNA2 5' UTR structures and sizes and this prompted a further investigation into possible recombination events between GFLV-SAPCS3 and ArMV in the 5' UTR, as well as the rest of the genome.

In this chapter, putative inter- and intraspecies recombination that involves GFLV-SAPCS3 RNA1 and RNA2 are described. In RNA1, we found a putative recombination event within $1E^{Pol}$, involving GFLV-F13 (recombinant) with parental lineages of a GFLV-WAPN173-type (major parent) and a GFLV-SAPCS3- type (minor parent) isolates. This putative recombination event was further demonstrated by comparing the pairwise-distances of before and after the predicted breakpoint. A phylogenetic tree also indicated that the GFLV-F13 RNA1 3' UTR is phylogenetically closer related to the GFLV-SAPCS3 3' UTR than the GFLV-WAPN173 3' UTR. In previous studies, intraspecies recombination was found in the $1A\text{-}1B^{Hel}$ in one isolate (Mekuria et al., 2009) and in the $1E^{Pol}$ region of nine Californian isolates (Vigne et al., 2010). However, in the latter study, only partial $1E^{Pol}$ sequences (1140 nt) were used, and although this may give an indication on recombination in this particular area, it will not provide a deeper understanding of recombination events in areas surrounding $1E^{Pol}$. The study reported putative recombination events in the $1E^{Pol}$, with determined start and end breakpoints, in different regions than what was reported in this chapter. It is likely that recombination events in the rest of the $1E^{Pol}$ (extending to the 3' UTR) were missed due to the use of partial $1E^{Pol}$ sequences. The different recombination breakpoints between

isolates may also indicate that sequence exchanges among GFLV isolates may occur randomly (Mekuria et al, 2009). The putative recombination towards the end of the $1E^{Pol}$ that included the 3' UTR discussed in this chapter is the first report of recombination in this region. It is likely that similar studies will follow as soon as more complete $1E^{Pol}$ and full-length RNA1 sequences become available.

An interspecies recombination event was also described in this chapter, with GFLV-SAPCS3 as the recombinant, resulting from the recombination between a GFLV-F13-type as the major parent and an ArMV-Ta-type as the minor parent within the 5' UTR of RNA2. The fact that the 5' UTR in GFLV-SAPCS3 is similar in nucleotide similarity and length of the 5' UTR GFLV-Ghu, and therefore also the same as the 5' UTRs of most ArMV isolates, supports this putative recombination event. Also, the conserved sequence stretch AA/GTCCGTT/CA was found in the 5' UTR of GFLV-SAPCS3, GFLV-Ghu and three other ArMV isolates, but is absent in all the other GFLV isolates sequenced to date. The significance of this insertion is unknown and should be investigated. The insertion in the RNA2 5' UTR of both GFLV-SAPCS3 and GFLV-Ghu also suggests that both these isolates are descendants from a single recombination event. Several studies have described the interspecies recombination between GFLV and ArMV, and most of the recombination sites were detected in the 5' UTR, $2A^{HP}$ or the $2B^{MP}$, but never in the $2C^{CP}$ (Mekuria et al., 2009; Oliver et al., 2010). The $2C^{CP}$ is responsible for subunit assembly and vector transmission and therefore genetic exchange within the $2C^{CP}$ is less likely to generate viable viruses due to the high structural constraints associated with the CP (Vigne et al., 2008). It seems that the RNA2 5' UTR, $2A^{HP}$ and $2B^{MP}$ may be recombination "hotspots" for the genetic exchange between closely related grapevine-infecting nepoviruses (Mekuria et al., 2009). The higher recombination rates in $2A^{HP}$ and $2B^{MP}$ among different grapevine-infecting nepoviruses may play a role in host adaptation and specialisation (Oliver et al., 2010). However, the so-called recombination "hot-spots" in RNA1 cannot yet be investigated due to the limited number of full-length sequences available for both GFLV and ArMV. The interspecies recombination of GFLV and ArMV RNA2 described in this

chapter and in previous studies are important to study as these chimeras could lead to increased genetic diversity and may result in the emergence of new virus variants.

3.6. Conclusion

In this chapter the first full-length sequence of a South Africa GFLV isolate, GFLV-SAPCS3, was determined. This study has contributed to the ever-growing sequence database of nepoviruses that may aid in understanding nepovirus diversity and naturally occurring GFLV recombinants. Putative recombination events in GFLV-SAPCS3 RNA1 and RNA2 were demonstrated. A putative recombination event was detected towards the 3' end of RNA1 of a GFLV-SAPCS3-type and a GFLV-WAPN173-type to form the recombinant GFLV-F13-type, and this is the first report of a recombination event in this region. An interspecies recombination event was also described in this chapter, suggesting that the GFLV-SAPCS3 RNA2 5' UTR may have originated from a recombination event between ArMV and GFLV isolates. The recombination event in the RNA2 5'UTR was further confirmed by the presence of a conserved insert that was only found in GFLV-Ghu and all ArMV isolates sequenced to date. The significance of this insertion is however unknown and should be investigated. Comprehensive investigations regarding GFLV and ArMV recombination, which may lead to the rise of novel nepovirus variants, will only be able to continue when more full-length sequences become available.

3.7. References

1. Andret-Link P, Laporte C, Valat L, et al (2004). Grapevine fanleaf virus: still a major threat to the grapevine industry. J Plant Pathol, 86,183-195.
2. Boulila M (2007). Phylogeny and genetic recombination of Grapevine fanleaf virus isolates from naturally infected vineyards in Tunisia. Phytopathol Mediterr, 46,285-294.
3. Fattouch S, Acheche H, M'hirsi S, et al (2005). RT-PCR-RFLP for genetic diversity analysis of Tunisian Grapevine fanleaf virus isolates in their natural host plants. J Virol Methods, 127,126-132.
4. Fernandez IT, Candresse T, Le Gall O, Dunez J (1999). The 5' noncoding of grapevine chrome mosaic nepovirus RNA2 triggers a necrotic response on three Nicotiana spp. MPMI, 12, 337-344.

5. Harrison BD, Murrant AF, Mayo MA, Roberts IM (1974). Distribution of determinants for symptom production, host range and nematode transmissibility between two RNA components of raspberry ringspot virus. J Gen Virol, 22, 233-247.
6. Hewitt WB (1968). Viruses and virus diseases of the grapevine. Rev of App Mycology, 47, 433-455.
7. Jordan RL, Dodds JA, Ohr HD (1983). Evidence for virus-like agents in avocado. Phytopathology, 73, 1130-1135.
8. Liebenberg A, Freeborough MJ, Visser CJ, Bellstedt DU, Burger JT (2009). Genetic variability within the coat protein gene of Grapevine fanleaf virus isolates from South Africa and the evaluation of RT-PCR, DAS-ELISA and ImmunoStrips as virus diagnostic assays. Virus Res, 142, 28-35.
9. Lole KS, Bollinger RC, Paranjape RS, Gadkari K, Kulkami SS, Novak NG, Ingersoll R, Sheppard HW, Ray SC (1999). Full length human immunodeficiency virus type 1 genomes from subtype C-infected seroconverters in India, with evidence of intersubtype recombination. J Virol, 73, 152-160.
10. Margis R, Ritzenthaler C, Reinbolt J, Pinck M, Pinck L (1993). Genome organization of grapevine fanleaf nepovirus RNA2 deduced from the 122K polyprotein P2 in vitro cleavage products. J Gen Virol, 74, 1919-1926.
11. Martin DP, Lemey P, Lott M, Moulton V, Posada D, Lefeuvre P (2010). RDP3: a flexible and fast computer program for analyzing recombination. Bioinformatics, 26, 2462-2463.
12. Matz M, Shagin D, Bogdanova E, Britanova O, Lukyanov S, Diatschenko L, Chenchik A (1999). Amplification of cDNA ends based on template-swichting effect and step-out PCR. Nucleic Acids Research, 27, 1558-1560.
13. Mekuria TA, Gutha LR, Martin RR, Naidu RA (2009). Genome Diversity and Intra-and Interspecies Recombination Events in Grapevine fanleaf virus. Phytopathology, 99, 1394-1402.
14. Meng B, Li C, Wang W, Goszczynski D, Gonsalves D (2005). Complete genome sequences of two new variants of Grapevine rupestris stem pitting-associated virus and comparative analyses. J Gen Virol, 86, 1555-1560.
15. Morris TJ, Dodds JA (1979). Isolation and analysis of double-stranded RNA from virus-infected plant and fungal tissue. Phytopathology, 69, 858-858.
16. Oliver JE, Vigne E, Fuchs M (2010). Genetic structure and molecular variability of Grapevine fanleaf virus populations. Virus Res, 152, 30-40.
17. Pinck L, Fuchs M, Pinck M, Ravelonandro M, Walter BA (1988). Satellite RNA in grapevine fanleaf virus strain F13. J Gen Virol, 69, 233-239.
18. Pompe-Novak M, Gutierrez-Aguirre I, Vojvoda J, Blas M, Tomazic I, Vigne E, Fuchs M, Ravnikar M, Petrovic N (2007). Genetic variability within RNA2 of Grapevine fanleaf virus. Eur J Plant Pathol, 117, 307–312.
19. Quacquarelli A, Gallitelli D, Savino V, Martelli GP (1976). Properties of grapevine fanleaf virus 1976. J Gen Virol, 32, 349-360.

20. Raski DJ, Goheen AC, Lider LA, Meredith CP (1983). Strategies against grapevine fanleaf virus and its nematode vector. Plant Dis, 67, 335-338.
21. Ritzenthaler C, Viry M, Pinck M, Margis R, Fuchs M, Pinck L (1991). Complete nucleotide sequence and genetic organization of grapevine fanleaf nepovirus RNA1. J Gen Virol, 72, 2357-2365.
22. Sanfaçon H, Wellink H, Le Gall O, Karasev A, van der Vlugt A, Wetzel T (2009). Secoviridae: a proposed family of plant viruses within the order Picornavirales that combines the families Sequiviridae and Comoviridae, the unassigned genera Cheravirus and Sadwavirus, and the proposed genus Torradovirus. Arch Virol, 154 (5), 899-907.
23. Serghini MA, Fuchs M, Pinck M, Reinbolt J, Walter B, Pinck L (1990). RNA2 of grapevine fanleaf virus, sequence analysis and coat protein cistron location. J Gen Virol, 71, 1433-1441.
24. Sokhandan-Bashir N and Melcher U (2012). Population genetic analysis of grapevine fanleaf virus. Archives of Virology. DOI 10.1007/s00705-012-1381-0.
25. Tamura K, Peterson D, Peterson N, Stecher G, Nei M, and Kumar S (2011). MEGA5: Molecular Evolutionary Genetics Analysis using Maximum Likelihood, Evolutionary Distance, and Maximum Parsimony Methods. Molecular Biology and Evolution, 28, 2731-2739.
26. Thompson JD, Higgins DG, Gibson TJ (1994). CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. Nucleic Acids Res, 22, 4673-4680.
27. Valverde RA, Nameth ST, Jordan RL (1990). Analysis of double-stranded RNA for plant virus diagnosis. Plant Dis, 74, 255-258.
28. Vigne E, Bergdoll M, Guyader S, Fuchs M (2004). Population structure and genetic diversity within *Grapevine fanleaf virus* isolates from a naturally infected vineyard: evidence for mixed infection and recombination. J Gen Virol, 85, 2435–2445.
29. Vigne E, Demangeat G, Komar V, Fuchs M (2005). Characterization of a naturally occurring Grapevine fanleaf virus recombinant isolate. Arch Virol, 150, 2241–2255.
30. Vigne E, Marmonier A, Fuchs M (2008). Multiple interspecies recombination events within RNA2 of Grapevine fanleaf virus and Arabis mosaic virus. Arch Virol, 153, 1771-1776.
31. Vigne E, Schmitt-Keichener C, Komar V, Rakotomalala L, Lemaire O, Ritzenthaler C, Fuchs M (2012). Symptom determinants of Grapevine fanleaf virus in Nicotiana species. Proceedings of the 17th Congress of ICVG, Davis, California, USA, October 7-14, 2012.
32. White EJ, Venter M, Hiten NF, Burger JT (2008). Modified Cetyltrimethylammonium bromide method improves robustness and versatility: The benchmark for plant RNA extraction. Biotechnol J, 3, 1424–1428.
33. Zhu YY, Machleder EM, Chenchik A, Li R, Siebert PD (2001). Reverse transcriptase template switching, A SMART approach for full-length cDNA library construction. BioTechniques, 30, 892-897.

Chapter 4

A new satellite RNA associated with a Grapevine fanleaf virus isolate found in South Africa

Scientific outputs:

The full-length genomic and satellite RNA sequences of GFLV-SACH44, as well as the construction of the full-length infectious cDNA clone of the satellite RNA, was published in a condensed form in Viruses.

RL Lamprecht, M Spaltman, D Stephan, T Wetzel, JT Burger (2013). Complete nucleotide sequence of a South African isolate of Grapevine fanleaf virus and its associated satellite RNA. Viruses, 5 (7), 1815-1823.

The full-length sequence and the phylogenetic analysis of GFLV-SACH44 satellite RNA, together with the sequence analysis of GFLV-SAPCS3 RNA1 and RNA2 (Chapter 3), were collectively presented at the 17th Congress of the International Council for the Study of Virus and Virus-like Diseases of the Grapevine (ICVG), on 7 October 2012, held at UC Davis, California, USA.

Lamprecht RL, Maree HJ, Stephan D, Wetzel T, Burger JT. Molecular Characterization of South African isolates of Grapevine Fanleaf virus and an Associated New Satellite RNA. Proceedings of the 17th Congress of ICVG, Davis, California, USA, October 7-14 2012, p 36.

***Note**: The work presented in the published article and thesis chapter was part of a larger study that included Ms RL Lamprecht (primary investigator for this PhD study), Ms M Spaltman (BSc (Hons) student, Stellenbosch University), Ms A Baßler and Dr T*

Wetzel (both at AlPlanta Plant Research Institute, Germany). Ms RL Lamprecht discovered and characterised the new South African GFLV satRNAs and developed full-length infectious cDNA clones of the GFLV-SACH44 satRNA. Ms A Baßler and Dr T Wetzel were involved in confirming the infectivity of the GFLV-SACH44 satRNA clones in herbaceous hosts by Northern Blot analysis at AlPlanta. Parts of the GFLV-SACH44 genomic sequence was determined by Ms M Spaltman in 2012 as part of her BSc Honours Degree, entitled "Molecular characterization of a Grapevine fanleaf virus isolate in South Africa associated with a satellite RNA" (unpublished). The aim of the BSc (Hons) project was to find additional GFLV-associated satRNAs in a South African vineyard, and to compare the genomic sequence data for the 1E^{Pol} gene of GFLV satRNA isolates with other GFLV isolates without satRNAs, in order to determine genomic regions that might be responsible for satRNA association and replication. However, no satRNAs were found and Ms M Spaltman started sequencing parts of the GFLV-SACH44 genome (plasmids containing GFLV-SACH44 fragments provided by Ms RL Lamprecht). The BSc (Hons) study was designed and co-supervised by Ms RL Lamprecht, together with Prof JT Burger and Dr D Stephan. The BSc (Hons) Degree was completed by Ms M Spaltman in December 2012. Ms RL Lamprecht continued in determining the rest of the GFLV-SACH44 genomic sequence (including the 5' and 3' ends of both RNA1 and RNA2) and performed the phylogenetic analyses.

Also, part of the research described in this chapter was conducted by Ms RL Lamprecht at the AlPlanta Plant Research Institute (Neustadt an der Weinstraße, Germany) as part of a bilateral project between Stellenbosch University (Western Cape, South Africa) and AlPlanta. Therefore some of the detection procedures (i.e. primers used and RT-PCR chemistries) were different than the detection procedures used throughout this thesis.

4.1 Abstract

In this chapter, newly discovered satellite RNAs (satRNA) associated with South African isolates of Grapevine fanleaf virus (GFLV) are described, the satRNAs of GFLV-SACH44 and GFLV-SACH47. Both the entire genome and satRNA of GFLV-SACH44

satRNA were completely sequenced. The full-length sequence of the satRNA was found to be more closely related to the large satRNAs of three ArMV isolates (86.7-87.8% nucleotide (nt) identity) than to the GFLV-F13 satRNA (81.8% nt identity). Full-length infectious clones of GFLV-SACH44 satRNA were constructed. The infectivity of the clones was tested using three nepovirus subgroup A isolates, GFLV-NW, ArMV-NW and GFLV-SAPCS3, as helper viruses. The clones were mechanically inoculated on *Chenopodium quinoa* and *Nicotiana benthamiana* and were infectious when co-inoculated with the two GFLV variants, but not when co-inoculated with ArMV-NW.

4.2 Introduction

Small molecular parasites are known to be associated with several plant viruses, *i.e.* satRNAs or satellite viruses, collectively designated as satellites. Satellites are completely dependent on helper viruses for replication. Due to their ability to use the replication machineries of helper viruses, research interests in satellites have increased in recent years (Hu et al., 2009). The major difference between satRNAs and satellite viruses is that the latter encode proteins responsible for their own encapsidation, for example Tobacco mosaic satellite virus (Dodds, 1998), while satRNAs do not code for any known functional proteins, and are therefore dependent on the helper virus for encapsidation as well as replication (Murant and Mayo, 1982; Francki, 1985; Frisch and Mayo, 1989; Hu et al., 2009). In this chapter, only satRNAs will be discussed.

Satellite RNAs are RNA molecules that are smaller than 1500 nucleotides (nt) and that depend on their cognate helper viruses for replication, encapsidation, movement, and transmission. Most satRNAs share little or no sequence homology to the helper viruses (Murant and Mayo, 1982). Satellite RNAs can be either double stranded (ds) or single stranded (ss) (Figure 4.1), and can be sub-divided into three subgroups; subgroup 1, the large linear ss satRNAs; subgroup 2, the short linear ss satRNAs; and subgroup 3, the circular ss satRNAs (Mayo et al., 2005).

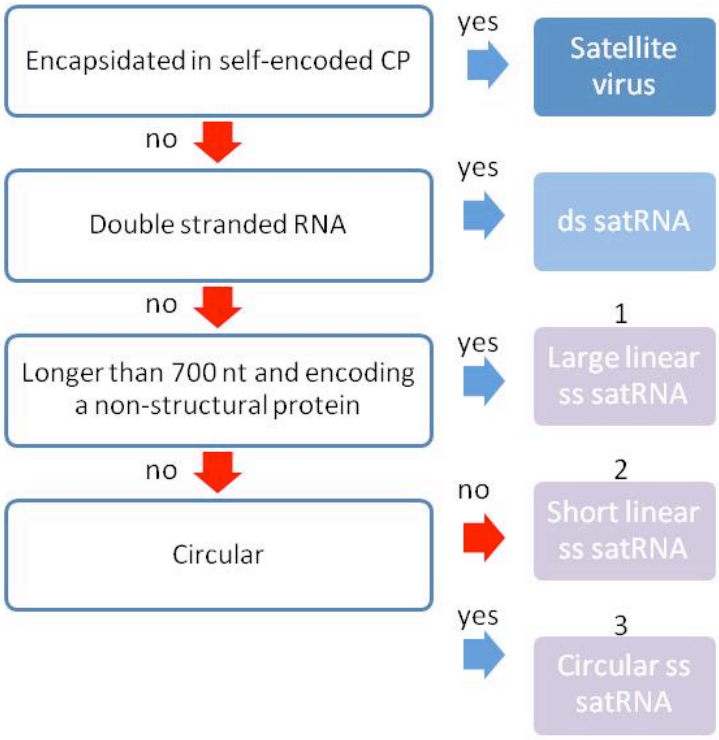


Figure 4.1: A flow-diagram representing satRNA classification. The blocks on the left indicate the most prominent feature of each group. The blue arrows indicate the subgroup the satellite belongs to, if the prominent feature holds true (red arrow for false). The satRNAs can be classified into the either double-stranded (ds) satRNAs or single-stranded (ss) satRNAs. The ss satRNAs that can be subdivided into the (1) large linear ss satRNAs, (2) the short linear ss satRNAs and (3) the circular satRNAs.

The large satRNAs of subgroup 1 are between 0.7 to 1.5 kb and codes for at least one nonstructural protein. Satellite RNAs in this group are often referred to as messenger-type satRNAs (Mayo et al., 2005). The large satRNAs have *cis*-acting elements (either short stretches of nt sequences or ordered-structures) that interact with translation factors of their helper viruses for translation and replication (Huang et al., 2010). Species in subgroup 1 are for example the ArMV large satRNA (Liu et al., 1990), GFLV satRNA (Pinck et al., 1989), Bamboo mosaic virus satRNA (Lin et al., 1994), Chicory yellow mottle virus (ChYMV) large satRNA (Rubino et al., 1990 b), Tomato black ring

virus (ToBRV) satRNA, etc (Meyer et al., 1984). The best studied satellites of this subgroup are those associated with nepoviruses. The satellites of nepoviruses have a 5' terminal VPg protein and a 3′ end poly (A) tail that both resemble the corresponding regions of the helper virus genome (Mayo et al., 1979; Mayo et al., 1982). The biological function of the putative proteins encoded by satRNAs in this subgroup is mostly unknown. However, it was shown that the proteins encoded by the satRNAs of ToBRV, ArMV and GFLV are needed for replication of the satRNAs (Hemmer et al., 1993; Hans et al., 1993; Liu and Cooper, 1993). When the AUG at the start of protein P3 (encoded by the ORF of the GFLV satRNA) was substituted in a GFLV-F13 satRNA mutated clone, replication of the satRNA was abolished. The authors concluded that P3 (together with the 5' and 3' non-coding regions) is important for the satRNA replication. However, the conformational changes of the RNA that may have been created by the substitutions are also a possible reason for replication abolishment (Hans et al., 1993).

The short linear satRNAs of subgroup 2 are smaller than 700 nt and do not have any coding functionality, and are thus also sometimes referred to as the non-coding satRNAs (Mayo et al., 2005; Hu et al., 2009). Contrary to the messenger-type satRNAs, the non-coding satRNAs rely on their structural organisation (*e.g.* hairpins and pseudoknots) for their replication, rather than *cis*-acting elements that interact with translation factors as required for the messenger-type satRNAs (Huang et al., 2010). Examples of small satRNAs include Cucumber mosaic virus satRNA (Kaper and Tousignant, 1997), ChYMV satRNA (Rubino et al., 1990 a), Pea enation mosaic virus satRNA (Demler and De Zoeten, 1989), Panicum mosaic virus small satRNA (Buzen et al., 1984), Peanut stunt virus satRNA (Whitmer Collmer et al., 1985), Turnip crinkle virus satRNA (Simon and Howell, 1986), Tomato bushy stunt virus (ToBSV) satRNAs B1 and B10 (Celix et al., 1997).

The small circular satellites in subgroup 3 are sometimes referred to as virosoids because they resemble the structural nature of viroids. The circular satRNAs do not encode any ORFs and they tend to be highly structured and replicate by rolling-circle mechanisms. This group includes ArMV small satRNA (Kaper et al., 1986), Cereal

yellow dwarf virus-RPV satRNA (Miller et al., 1991), ChYMV satRNA (Rubino et al., 1990 b), Lucerne transient streak virus satRNA (Sehgal et al., 1993) and Tobacco ringspot virus satRNA (Buzayan et al., 1986).

Satellites that were previously isolated from grapevine are only those associated with the nepoviruses; GFLV, ArMV and Grapevine Bulgarian Latent virus (GBLV). No other satellite species are known to naturally infect grapevine. Large satRNAs were found to be associated with both GFLV (isolates GFLV-F13, GFLV-R2 and GFLV-R6) and GBLV, while some ArMV-infected plants had the presence of large satRNAs and/or small circular satRNAs (Davis and Clark, 1983; Liu et al., 1990). The first GFLV satRNA described was the satRNA of the French isolate GFLV-F13. The GFLV-F13 satRNA is 1114 nt in length and requires the presence of a helper virus to replicate. The GFLV-F13 satRNA sequence has no significant sequence homology to its helper virus, except for the short (10 nt) consensus sequence present in the 5' UTR of RNA1 and RNA2 (Pinck et al., 1988; Fuchs et al 1989). The GFLV-F13 satRNA was the only described GFLV-associated satRNA for over 20 years. Recently, however, two new GFLV-associated satRNAs were described, found to be associated with Californian isolates GFLV-R2 and GFLV-R6 (Gottula et al., 2013). Both these satRNAs are 1140 nt in length and have a 5'- and 3' UTR of 24 nt and 78 nt in length, respectively. The ORFs of both GFLV-R2 and GFLV-R6 satRNAs correspond to a translated product of 346 aa, compared with 341 aa of the GFLV-F13 satRNA. Several satRNAs from different ArMV isolates have been described. The large satRNAs from isolates ArMV-hop, ArMV-lilac, ArMV-p119, ArMV-NW, ArMV-P116 and ArMV-J86 are between 1092-1139 nt in length (Liu et al., 1990; Wetzel et al., 2006). One small circular satRNA (300 nt in length) was also found to be associated with a hop ArMV isolate (Kaper et al., 1988).

Here, the discovery and characterisation of the GFLV-SACH44 satRNA is reported. This satRNA was found to be associated with a GFLV isolate isolated from grapevine material collected in Robertson (Western Cape, South Africa). Interestingly, the GFLV-SACH44 satRNA is more closely related to the ArMV satRNAs (on a nt and aa level) than to the GFLV satRNAs. In this chapter both the genome and the satRNA of GFLV-

SACH44 were sequenced in order to determine if variation exists between GFLV isolates that are able to replicate satRNAs and those that are not. A full-length cDNA infectious clone of the GFLV-SACH44 satRNA was constructed and shown to be capable of replication in herbaceous hosts when mechanically co-inoculated with the helper virus isolates GFLV-NW and GFLV-SAPCS3, but not with ArMV-NW.

4.3 Materials and Methods

4.3.1 Satellite RNA detection in grapevine

Leaf material from grapevine plants displaying typical symptoms of GFLV infection were collected from a vineyard in Robertson (Western Cape, South Africa) during autumn in 2010. In total, 29 plants were screened for the presence of a satRNA with RT-PCR using the satRNA diagnostic primers, satRNA_dgF, satRNA_174F and satRNA_dgR (Table 4.1). These diagnostic primers were designed from conserved areas obtained from alignments of GFLV-F13 satRNA (acc NC003203), ArMV-NW satRNAs (acc DQ187317 and DQ187315) full-length nt sequences. Initially, the grapevine plants were screened by RT-PCR for satRNA with satRNA_dgF and satRNA_R with an expected amplicon size of 385 bp. Primers satRNA_174F and satRNA_R were later also used as a diagnostic set for the detection of GFLV satRNAs in grapevine with the same one-step RT-PCR protocol. For satRNA detection in grapevine, total RNA was extracted from grapevine leaves as described by White et al. (2008). The total RNA was added to a one step RT-PCR cocktail (section 3.3.1) and the cycling conditions were as follows: a reverse transcription step at 45°C for 45 min, one denaturation step at 94°C for 5 min, followed by 30 cycles of 94°C for 30 sec, 55°C for 30 sec, 72°C for 45 sec, and a final elongation step at 72°C for 7 min. PCR products were visualized on a 1% agarose gel stained with EtBr. A PCR amplicon corresponding to 749 bp indicated the presence of satRNA in the sample.

4.3.2 Virus maintenance

Grapevine plants found that had the presence of satRNAs were cut and rooted in the greenhouse facility. Leaves were harvested when needed. The greenhouse is controlled with temperatures between 18°C and 28°C and a relative humidity of approximately 70%. As described in Chapter 3 (section 3.3.1), these grapevine plants were screened for the presence of ArMV by DAS-ELISA (Bioreba) according to the manufacturer's instructions. This was done to exclude the possibility that co-infection with ArMV may have had an effect on the sequencing and phylogenetic results.

4.3.3 Sequencing of the GFLV-SACH44 genome and satRNA

Grapevine plant SACH44 was selected for GFLV genomic and satRNA sequencing. As part of a BSc (Hons) study, parts of the GFLV-SACH44 genome have already been sequenced with primers that were designed from the GFLV-SAPCS3 sequence (Chapter 3). Total RNA was extracted from a grapevine plant (SACH44) using the CTAB method (White et al., 2008). High fidelity enzymes for cDNA synthesis (Superscript III Reverse Transcriptase, Invitrogen) and PCR (Ex Taq DNA Polymerase, Takara) were used for the amplification of the GFLV-SACH44 genome and satRNA as described previously (section 3.3.2.3). Primers that were designed from existing GFLV-SACH44 RNA1 and RNA2 sequences for completion of the remainder of the genome are listed in Table 4.1. Also listed in Table 4.1 are the primers designed for determination of GFLV-SACH44 satRNA full-length sequences, as well as the construction of a full-length infectious clone.

Table 4.1: Primers used in this study for the detection and sequencing of the GFLV-SACH44 satRNA. Primers used for the determination of the genomic and satRNA full-length sequences are also listed.

Primer Name	Primer sequence	Designed from	Purpose	Position
RNA1				
CH44_RNA1_seq_703_F	TAACACCAAGGGAAAAGTCGTC	GFLV-SACH44	Sequencing	777-798
CH44_RNA1_2131_F	CTGGGAGCTTTTCATATTA	GFLV-SACH44	PCR	2131-2150
CH44_RNA1_seq_2701_F	TGGAGTGGTTATGCCAGGCA	GFLV-SACH44	Sequencing	2701-2720
CH44_RNA1_seq_4282_R	ACTTCTCATACTTAATGGGTTGCG	GFLV-SACH44	Sequencing	4259-4282
CH44_RNA1_4912_R	CAACTTCATCAAGCAGTTTC	GFLV-SACH44	PCR	4893-4912
CH44_RNA1_block2_seq_F	GTGCCAGGTAGGCATCAATC	GFLV-SACH44	Sequencing	5355-5374
CH44_RNA1_block3_seq_F	CTGGTGCTTATAAAGAGTTG	GFLV-SACH44	Sequencing	6128-6147
CH44_RNA1_3'UTR_6982_F	TGTGCAGGGGTCGCAACTAA	GFLV-SACH44	PCR	6982-7001
RNA2				
CH44_RNA2_134_F	CCAAAGCGAAGAGTTTAAGA	GFLV-SACH44	PCR	134-153
CH44_RNA2_686_R	CACAGTGGCCCGTATAAACC	GFLV-SACH44	PCR	667-685
CH44_RNA2_seq_gap_3	TGAGCAAGGCCTACCGCT	GFLV-SACH44	Sequencing	1270-1287
CH44_RNA2_block3_seq_F	GAGGCTGAACCCAGATTGAG	GFLV-SACH44	Sequencing	2057-2076
CH44_RNA2_3'UTR_3557_F	GCCTCGTCCAGGTTTCAGTT	GFLV-SACH44	PCR	3556-3575
SatRNA				
GFLV_satRNA_RACE_GSP4	CTTTTCAGCAGGAGCCCAGA	GFLV-SACH44	5' RACE	813-832
GFLV_satRNA_RACE_GSP5	AGATAGAAGTGAGGGTGAAA	GFLV-SACH44	5' RACE	538-557
GFLV_satRNA_RACE_GSP6	CTGCTGTTTGTGTCCCTTCG	GFLV-SACH44	5' RACE	252-271
GFLV_satRNA_174F	GTAAGCAAACGGACCT	GFLV and ArMV	Diagnostic/ Sequencing	173-189
GFLV_satRNA_dgR	ACCTTACGCAACATCCG	GFLV and ArMV	Diagnostic/ Sequencing	905-921
GFLV_satRNA_dgF	TTCACCCTCACTGCTATC	GFLV and ArMV	Diagnostic/ Sequencing	539-556
A/G_sat1_F (Wetzel et al., 2006)	ATTGTCGTGTAAGCACCGTG	ArMV-NW	Diagnostic	133-152
A/G_sat4_R (Wetzel et al., 2006)	ACGCGTGCGAGTGTCACCAC	ArMV-NW	Diagnostic	872-891
GFLV_ic_satRNA_1F_AscI	AA**GGCGCGCC**ATGAAAAATTTC TATGGGTTCTCGT	GFLV-SACH44	Infectious clone	1-25
GFLV_ic_satRNA_pR2_Bsp120I	AA**GGGCCC**(Tx30)GAGTTGGCT AATGAGCAACC	GFLV-SACH44	Infectious clone	1085-1104

4.3.3.1 GFLV-SACH44 RNA1 and RNA2 full-length sequence completion

Parts of GFLV-SACH44 RNA1 and RNA2 have been sequenced (64% of GFLV-SACH44 RNA1 and 86 % for GFLV-SACH44 RNA2) in a BSc (Hons) study, using primers that were used for the GFLV-SAPCS3 sequence determination. In order to complete the genome sequence of GFLV-SACH44, sequence-specific primers were designed (Table 4.1) to amplify and sequence incomplete regions. The resulting PCR products were gel-purified using the Zymoclean Gel-DNA Recovery kit (Zymo Research), cloned into pGEM-T-Easy (Promega), and at least three clones from each of the PCR products sequenced in both directions (methods are the same as described in sections 3.3.2.3-3.3.2.4). The nt sequence of the 5' ends of GFLV-SACH44 RNA1 and RNA2 were determined using a 5'-RACE System (Invitrogen) using GFLV-SAPCS3-based 5' RACE primers (Table 3.1) and following the manufacturer's instructions (method described in section 3.3.2.5). Primer dT(17) (Meng et al., 2005) was used for cDNA synthesis for the determination of the 3' terminal sequence of GFLV-SACH44 RNA1 and RNA2, and used together with primers CH44_RNA1_3'UTR_6982_F and CH44_RNA2_3'UTR_3557_F, respectively for amplification and sequencing (method described in section 3.3.2.5). All the sequences generated from the overlapping amplicons were used to build a contiguous sequence using CLC Main Workbench version 6.5 (CLC Bio). Primer sequences were excluded during assembly.

4.3.3.2 GFLV-SACH44 satRNA full-length sequence determination

The molecular structure of GFLV-F13 satRNA and a graphical representation of the sequencing strategy of the GFLV-SACH44 satRNA is shown in Figure 4.2 (a and b). The first part of the genome that was sequenced was the 749 bp amplicon that was produced by using the diagnostic primers satRNA_174F and satRNA_dgR. This amplicon was gel-purified using the Zymoclean Gel-DNA Recovery kit (Zymo Research), and sequenced directly from both directions. The 5' RACE System (Invitrogen) was used for determining the nt sequence of the 5' UTR of GFLV-SACH44 satRNA. For this, high quality total RNA was extracted from grapevine using the Plant RNA Reagent™ (Invitrogen) and first strand cDNA was synthesized using the gene-

specific reverse primer GFLV_satRNA_RACE_GSP4, following the manufacturer's instructions. GFLV-SACH44 satRNA-specific nested primers (GFLV_satRNA_RACE_GSP5, GFLV_satRNA_RACE_GSP6), together with the Abridged Anchor primer and the Abridged Universal Amplification Primer of the 5' RACE System, were used to generate amplicons which were cloned in pGEM-T-Easy (section 3.3.2.4) and sequenced. cDNA clones representing the 5' terminal region of GFLV-SACH44 satRNA were sequenced and a consensus sequence was constructed using CLC Main Workbench version 6.5 (CLC Bio). In order to determine the nt sequence of the 3' UTR of the GFLV-SACH44 satRNA, primer $dT_{(17)}$ (Meng et al., 2005) was used for cDNA synthesis followed by a PCR reaction using primers satRNA_dgF and primer $dT_{(17)}$ for amplification of cDNA. The amplicon was cloned into pGEM-T-Easy and cDNA clones representing the 3' terminal region of GFLV-SACH44 satRNA were sequenced. The full-length sequence of GFLV-SACH44 satRNA was compiled by assembling all newly generated sequences using CLC Main Workbench version 6.5 (CLC bio).

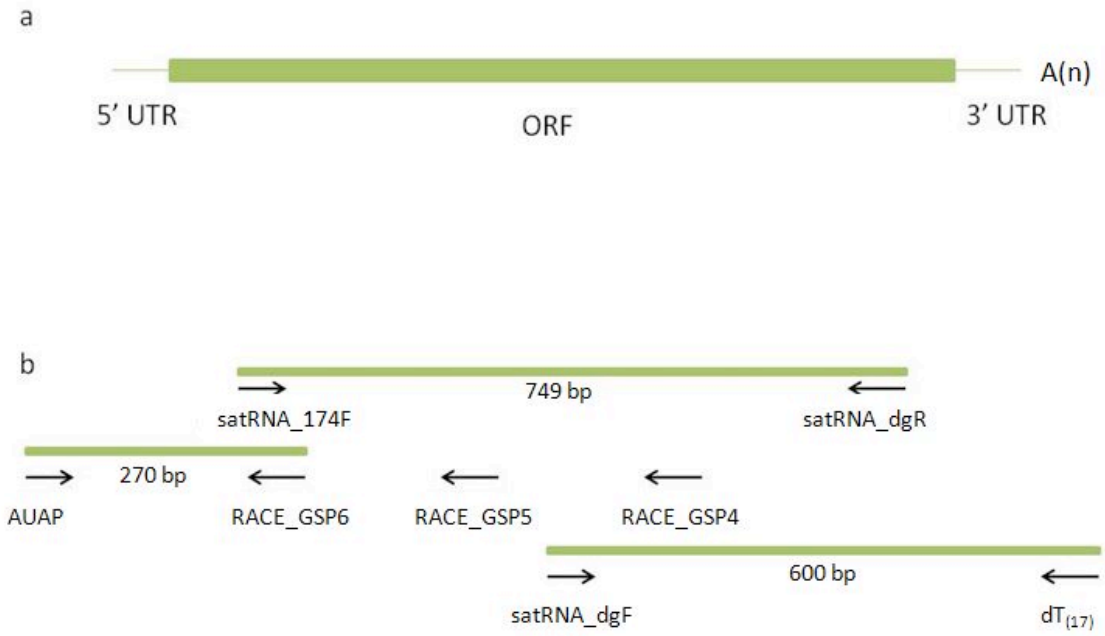


Figure 4.2: (a) A graphic representation of the molecular organisation of the GFLV-F13 satRNA. The thin lines represent the 5' and 3' UTRs, the block represents the single ORF of the large satRNAs, and the poly(A) tail is represented by A(n). (b) The sequencing strategy for GFLV-SACH44 satRNA. The green lines represent the amplicons that were sequenced, and the arrows represent the primers used (Table 5.1). The size of each amplicon is also indicated. Each area was sequenced 3-6 times, and consensus sequences were compiled to form the full-length sequence for the GFLV-SACH44 satRNA.

4.3.4 Phylogenetic analysis

The full-length nt and aa sequences of GFLV-SACH44 RNA1, RNA2 and the satRNA were compared to the full-length sequences of other GFLV and ArMV isolates by performing multiple sequence alignments using ClustalW (Thompson et al., 1994). The shared nt and aa identities, pairwise distance calculations and phylogenetic analyses were performed using the MEGA 5 analysis package (Tamura et al., 2011).

4.3.5 Construction of a GFLV-SACH44 satRNA full-length infectious clone

Using the sequence information obtained by determining the 5' and 3' ends, the entire GFLV-SACH44 satRNA was amplified as one RT-PCR amplicon. For that, primer GFLV_RNA3_pR2_Bsp120I (Table 4.1) was used for cDNA synthesis, with Superscript III RT (Invitrogen) according to the manufacturer's instructions (section 3.3.2.3). The cDNA was PCR-amplified (section 3.3.2.3) using Ex Taq DNA Polymerase (Takara) and primers GFLV_ic_satRNA_1F_AscI and GFLV_RNA3_pR2_Bsp120I, therefore adding 5' terminal *Asc*I and 3' terminal *Bsp*120I restriction enzyme sites to the amplicon for cloning purposes. The entire satRNA fragment was cloned into pGEM-T-Easy, resulting in pGEM-T-Easy-satfrag. Following a population cloning strategy (Yu and Wong, 1998), recombinant colonies were picked and screened with restriction enzymes to confirm the satRNA insertion. The restriction enzyme *Bsp*120I is cytosine methylation sensitive, and therefore all the pGEM-T-Easy-satfrag plasmids, as well as the L140 vector (Figure 4.3; Stephan, 2005) were re-transformed (section 3.3.2.4) in *E. coli* strain JM110 (Fisher Scientific) that lacks *dam* and *dcm* methylation activity.

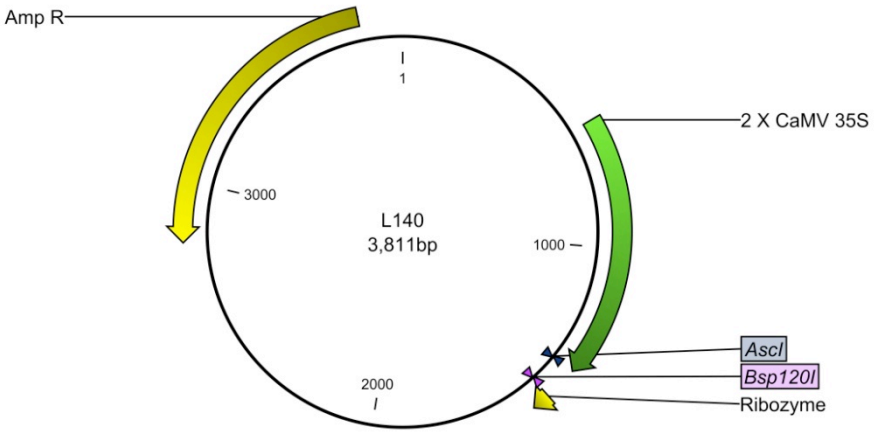


Figure 4.3: Schematic representation of the L140 vector (Stephan, 2005) that contains a double Cauliflower mosaic virus (CaMV) 35S promoter (Töpfer et al., 1987), hammerhead ribozyme termination sequence (Hofacker et al., 2003) and ampicillin resistance gene.

Plasmid DNA of ten confirmed clones were spectrophotometrically quantified and mixed in equal amounts to generate a population of clones. The pGEMT-T-Easy-satfrag population was digested using *Asc*I and *Bsp*120I restriction enzymes and the fragment representing the satRNA was ligated into L140 vector (Figure 4.3) digested with the same enzymes. For the ligation of the insert into the digested L140 vector, insert DNA, digested vector (3:1 molar ratio over vector), 1x T4 DNA Ligase Buffer (Fermentas) and 5 U T4 DNA Ligase (Fermentas) were mixed and the ligation mixture was made up to 20 µl. The ligation reaction was left at 4°C overnight. All ligation reactions were transformed in *E. coli* JM109 (section 3.3.2.4), the transformants were grown overnight in Luria broth and ampicilin, and plasmids were extracted using the GeneJet Plasmid Miniprep kit (Thermo Scientific). Plasmids isolated were confirmed by control restriction digests and sequencing.

4.3.6 Mechanical inoculation of L140-GFLV-satRNAfl in herbaceous hosts

DNA of L140-GFLV-satRNAfl clones (construct 3 and 12, 1 µg each) were mixed with 10 µl plant sap of ArMV-NW, GFLV-NW or GFLV-SAPCS3 that was macerated in an inoculation buffer (30mM K_2HPO_4, 50mM Glycine, 1% Celite, 1% Bentonite, pH 9.2). The DNA and plant sap mixture was mechanically inoculated by finger-rubbing onto the top two leaves of *C. quinoa* plants (6-8 leaf-stage), one plant per clone. The *C. quinoa* plants were left in a darkroom for four hours prior to inoculation. For controls, the plants were mechanically inoculated with plant sap of ArMV-NW, GFLV-NW or GFLV-SAPCS3 excluding plasmid DNA, or just inoculation buffer.

Table 4.2: Experimental design for the initial infectivity testing of the two L140-GFLV-satRNAfl constructs. The two constructs were co-inoculated with ArMV-NW, GFLV-NW or GFLV-SAPCS3 as helper viruses.

Experiment	Clone	2 µg plasmid	Helper virus (in 10 µl plant sap)	Herbaceous host
1	L140-GFLV-satRNAfl 3	10 µl	GFLV-NW	*C. quinoa*
2	L140-GFLV-satRNAfl 3	10 µl	ArMV-NW	*C. quinoa*
3	L140-GFLV-satRNAfl 12	10 µl	GFLV-NW	*C. quinoa*
4	L140-GFLV-satRNAfl 12	10 µl	ArMV-NW	*C. quinoa*
5	Transmission control	-	GFLV-NW	*C. quinoa*
6	Transmission control	-	ArMV-NW	*C. quinoa*
7	Healthy control	-	-	*C. quinoa*
8	L140-GFLV-satRNAfl 12	10 µl	GFLV-SAPCS3	*C. quinoa*
9	Transmission control	-	GFLV-SAPCS3	*C. quinoa*
10	Healthy control	-	-	*C. quinoa*

4.3.7 Detection of satRNA in herbaceous hosts

4.3.7.1 Detection by RT-PCR

GFLV and ArMV DAS-ELISA (Bioreba) was performed on the top systemically-infected leaves of the inoculated *C. quinoa* plants, ten days post-inoculation (dpi), to confirm successful virus transmission. Total RNA was extracted from the top newly formed systemically-infected leaves from the DAS-ELISA positive plants, using the CTAB method (White et al., 2008). The presence of RNA derived from L140-GFLV-satRNAfl in the total RNA was tested by either using the SuperScript One-Step RT-PCR with Platinum Taq (Invitrogen) system (according to the manufacturer's instructions) with primers A/G_sat1_F and A/G_sat4_R; or using the one-step RT-PCR protocol described in section 4.3.1 with the full-length infectious clone primers, GFLV_ic_satRNA_1F_AscI and GFLV_RNA3_pR2_Bsp120I. A PCR amplicon with the length of 759 bp or 1.1 kb (depending on the primer set used) was expected when RNA derived from GFLV-SACH44-satRNAfl was present. PCR products were visualized on an EtBr stained 1% agarose gel with a UV light. To confirm that satRNA amplification from the systemic leaves was from the systemic spread of RNA, and not from plasmid DNA contamination, the RT-PCR was repeated without reverse transcriptase.

4.3.7.2 Detection by Northern Blot analysis

For the detection of the satRNA, 1 µg of total RNA isolated from infected *C. quinoa* was separated on a 1% agarose-formaldehyde gel, blotted by capillary transfer to a nylon membrane (Hybond N+, Amersham) in 20x SSC buffer (1.5M NaCl, 0.15 M trisodium citrate, pH 7.0), and cross-linked under UV light (310 nm). For the preparation of the probe, 25 ng of a purified PCR-amplified fragment (a 300 bp fragment from the coding region of GFLV-SACH44 satRNA) was labelled with [α-^{32}P]dCTP (3,000 Ci/mmol, Perkin-Elmer) using the Decalabel DNA labelling kit (Fermentas) following the manufacturer's instructions. The membranes were hybridised at 65°C for 16 hours in PerfectHyb Plus buffer (Sigma). Post-hybridisation washings were performed at 65°C in

2% SDS, 2x SSC (20 minutes), 1% SDS, 1x SSC (20 minutes), and 0.1% SDS, 0.5xSSC (10 minutes). Stripping of the membranes (for hybridisation with another probe) were performed at 85°C in 0.1% SDS, 0.1xSSC (2x 30 minutes), and 1 minute in 2xSSC at room temperature. Fuji screens and a scan phosphoimager Pharaos FxPlus molecular imager (Biorad) were used to visualize the hybridisation signals.

4.4 Results

4.4.1 Satellite RNA detection in grapevine

Grapevines displaying typical GFLV symptoms were screened for the presence of a GFLV-associated satRNA. Out of 29 plants which were shown to be GFLV infected by DAS-ELISA, two plants (*Vitis vinifera* cv Chardonnay, SACH44 and SACH47) showed the expected 385 bp amplicon after RT-PCR using the diagnostic primers satRNA_dgF and satRNA_dgR (Figure 4.4, lanes 16 and 17), which first indicated the potential presence of satRNAs. However the diagnostic PCR protocol using primers satRNA_dgF and satRNA_dgR produced a lot of non-specific bands close to the 500 bp region (for example lane 10 and 11) and was therefore not a reliable diagnostic primer set. When the same one-step RT-PCR protocol was used with primers satRNA_174F and satRNA_dgR, it produced less non-specific bands (Figure 4.5).

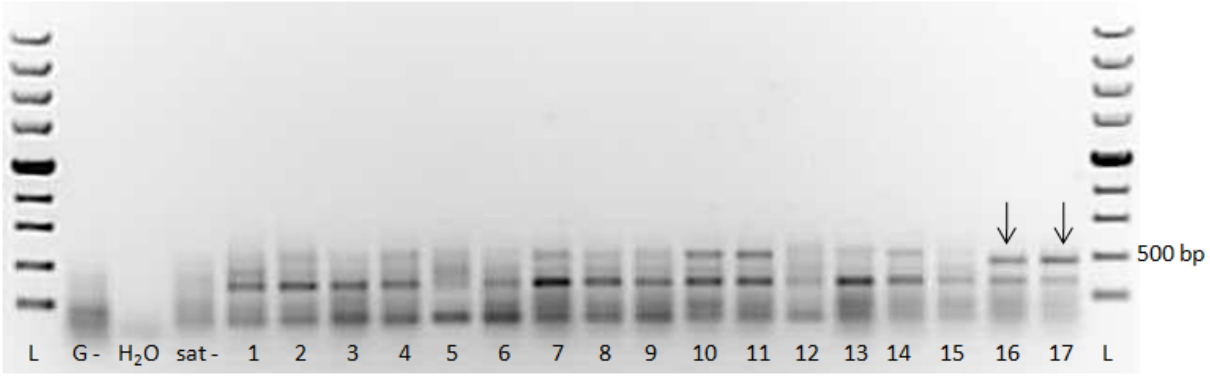


Figure 4.4: Initial satRNA detection by RT-PCR using primers satRNA_dgF and satRNA_dgR that produced a 385 bp fragment. The arrows indicate the two satRNA positive plants (lanes 16 and 17, from grapevine plants SACH44 and SACH47, respectively) with PCR bands just below the 500 bp mark. (L) GeneRuler 1kb ladder (Fermentas); (G-) GFLV-free grapevine control; (H2O) no template PCR control; (sat-) GFLV-SAPCS3 no satRNA control; (1-17) some of the samples used for the detection of GFLV satRNA.

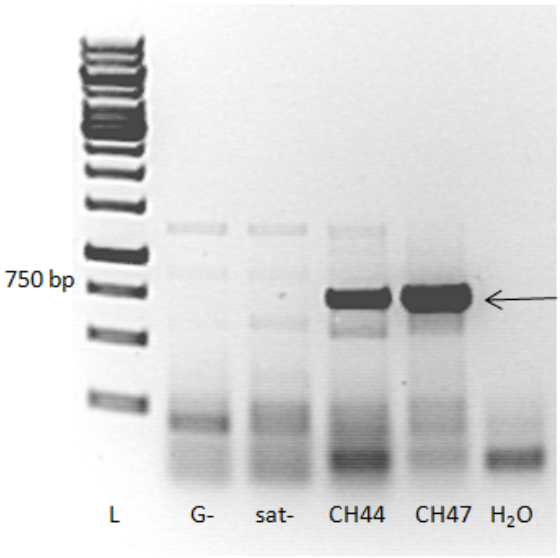


Figure 4.5: PCR amplicons resulting from the one-step RT-PCR using primers satRNA_174F and satRNA_dgR for the detection of satRNAs in grapevine plants. A 750 bp PCR amplicon is an indication of satRNA presence in the grapevine sample. The two fragments indicated by the arrow were gel-purified and sent in for sequencing. (L) GeneRuler 1kb ladder (Fermentas); (G-) GFLV-free grapevine control; (sat-) GFLV-SAPCS3 no satRNA; (CH44) GFLV-SACH44 satRNA; (CH47) GFLV-SACH47 satRNA; control; (H2O) no template PCR control.

The larger PCR fragment that was produced using the primers satRNA_174F and satRNA_dgR for the two satRNA positive samples were excised from the gel, purified using the Zymoclean Gel DNA Recovery kit (Zymo Research) and directly sequenced. The first sequenced fragments of the GFLV-SACH44 and GFLV-SACH47 were shown to be closely related to other satRNAs when it was analysed using the BLAST algorithm (http://www.ncbi.nlm.nih.gov). Initial sequencing results indicated that both satRNAs are more closely related to the ArMV-P119 satRNA, with Expect (E) values of $2.09E^{-140}$ and $1.5E^{-150}$, respectively. The BLAST E-values of GFLV-SACH44 and GFLV-SACH47 satRNAs were $3.12E^{-62}$ and $1.85E^{-57}$ respectively when compared against the GFLV-F13 satRNA. The two GFLV-satRNA infected Chardonnay plants SACH44 and SACH47 were rooted and maintained in a greenhouse facility. The grapevine plants were tested repeatedly to confirm the presence of the satRNA by RT-PCR. Both grapevine plants SACH44 and SACH47 were also confirmed not to be co-infected with ArMV by DAS-ELISA.

4.4.2 Sequencing and phylogenetic analysis

The entire genome of South African isolate GFLV-SACH44, including its associated satRNA, has been sequenced. The complete sequences of RNA1, RNA2 and the satRNA were deposited in the GenBank database with the accession numbers KC900162, KC900163 and KC900164.

The two RNAs of GFLV-SACH44 are 7341 nt and 3816 nt in length, respectively, and its satRNA is 1104 nt in length, all excluding the poly (A) tail. The RNA1 5'- and 3' UTRs are 243 and 246 nt in length, respectively. The open reading frame (ORF), coding for P1, of GFLV-SACH44 RNA1 is 6852 nt (2284 aa) in length, from nt positions 244-7095. The 5' and 3' UTRs of GFLV-SACH44 RNA2 are 271 and 215 nt in length, respectively. The 5' UTR of GFLV-SACH44 RNA2 is one nt short than the RNA2 5' UTR of GFLV-SAPCS3, which has the longest 5' UTR to date (Lamprecht et al., 2012). The 5' UTR of GFLV-SACH44 RNA2 also has the insertion AA/GTCCGTT/CA at position 73–98 that is present in GFLV-SAPCS3, GFLV-Ghu and other ArMV isolates, but not present in any

other GFLV isolate sequenced to date. The ORF of RNA2, coding for P2 is 3330 nt (1110 aa) in length, from nt positions 272-3601. The entire GFLV-SACH44 satRNA was 1104 nt in length, excluding the poly(A) tail. The sequence analysis of the satRNA revealed a single ORF of 1014 nt in length (338 aa), and its 5' and 3' UTRs are 14 and 74 nt in length, respectively. Nucleotide positions 1-17 were identical to the same region of three ArMV satRNA isolates, while the first 15 nt were identical to the same region of GFLV-F13 satRNA. Only one area in the satRNA fragment was found that was identical to the GFLV and ArMV genomes, in nt positions 1-7.

To determine the pairwise distances of GFLV-SACH44 RNA1 and RNA2, multiple full-length nt and aa sequence alignments were performed with other full-length GFLV isolates. Similarly, to determine the pairwise distances of the satRNAs, full-length nt and aa sequence multiple alignments were performed with the satRNA of GFLV-SACH44, GFLV- and ArMV satRNAs. The full-length nt and aa sequences of GFLV-SACH44 RNA1 and RNA2 were the closest related to the South African isolate GFLV-SAPCS3, showing a nt identity of 98.2% (P1 aa identity 99%) and 98.6% (P2 aa 99.1%), respectively, followed by GFLV-F13 with a nt identity of 87.3% (P1 aa identity 93.8%) and 90.1% (P2 aa 96%), respectively. The RNA1 full-length sequence of GFLV-SACH44 was most distantly related to GFLV Washington isolate WAPN6132 with a nt identity of 86.1% (P1 aa 91.8%) and the GFLV-SACH44 RNA2 full-length nt sequence was most distantly related to GFLV-GHu with a nt identity of 84.4% (P2 aa 90.2%). Interestingly, the full-length sequence of GFLV-SACH44 satRNA is more closely related to the ArMV satRNA isolated from lilac (Liu et al., 1990) with a nt identity of 87.8%, followed by ArMV-P119 satRNA, ArMV-P116 satRNA and GFLV-F13 satRNA with nt identities of 85 %, 83.5 %, and 82.6 % respectively. The GFLV-SACH44 satRNA was most distantly related to the GFLV-R2 satRNA (71.0%). Pairwise distances based on the full-length aa sequences between the GFLV and ArMV satRNAs indicated that the GFLV-SACH44 satRNA was closer related to ArMV-P119 satRNA (80.1%) than to the lilac isolate of ArMV (77.3%). GFLV-SACH44 satRNA shared aa identities of 76%, 65%, and 66% with GFLV-F13, GFLV-R2 and GFLV-R6 satRNAs, respectively.

The phylogenetic tree based on full-length RNA1 sequences showed that GFLV-SACH44 and GFLV-SAPCS3 grouped together (Figure 4.6 a) to the other four full-length GFLV sequences. A new GFLV isolate (GFLV-1050-02, accession number JX513889) has been completely sequenced since our previous study (Lamprecht et al., 2012) and was also included in the phylogenetic tree. Therefore, since 2012, two more GFLV RNA1 full-length sequences were added to Genbank, one from our study as well as isolate GFLV-1050-02. However, more GFLV RNA1 full-length sequences are still needed to clarify the phylogenetic relationship between GFLV isolates.

In the meantime, four new full-length RNA2 sequences of Iranian GFLV isolates were also reported (Zarghani et al., 2013), which were also included in our phylogenetic analysis. A phylogenetic tree based on the full-length RNA2 sequences of GFLV-SACH44 and 19 other isolates revealed that all isolates, except for GFLV-Ghu, were grouped in three main clades (Figure 4.6 b) that seem to be linked to geographic origin. Clade 1 mainly consists of Washington and Californian isolates and can be further divided into 3 sub-clades; of which GFLV-SACH44 and GFLV-SAPCS3 are placed separately in subgroup C (Figure 4.6 b). Clade 2 includes three Washington isolates, whereas clade 3 contains the Iranian isolates (Zarghani et al., 2013).

A phylogenetic tree based on the satRNA full-length sequences (Figure 4.6 c) was constructed and revealed that there are two distinct clades (clade 1 and 2) of the GFLV and ArMV satRNAs. Clade 1 included the GFLV-SACH44 satRNA, the lilac ArMV satRNA isolate, ArMV-P116 satRNA, ArMV-P119 satRNA and GFLV-F13 satRNA. The other clade, clade 2, included the satRNAs of the hop ArMV isolate, ArMV-J86, ArMV-NW, GFLV-R2 and GFLV-R6. The grouping of the 2 clades cannot be attributed to geographical origin, since the lilac ArMV satRNA isolate originated from the United Kingdom (Liu et al., 1991) and the other satRNAs in clade 1 were isolated from grapevine in different areas in Germany (Wetzel et al., 2006), France (Pinck et al., 1989) and South Africa (this study). In clade 2 the satRNAs were isolated from hop in the United Kingdom (Wetzel et al., 2002), ArMV-J86 with unknown origin (Wetzel et al.,

2006), ArMV-NW from Germany (Wetzel et al., 2001) and the satRNAs of GFLV-R2 and GFLV-R6 from California, USA (Gottula et al., 2013).

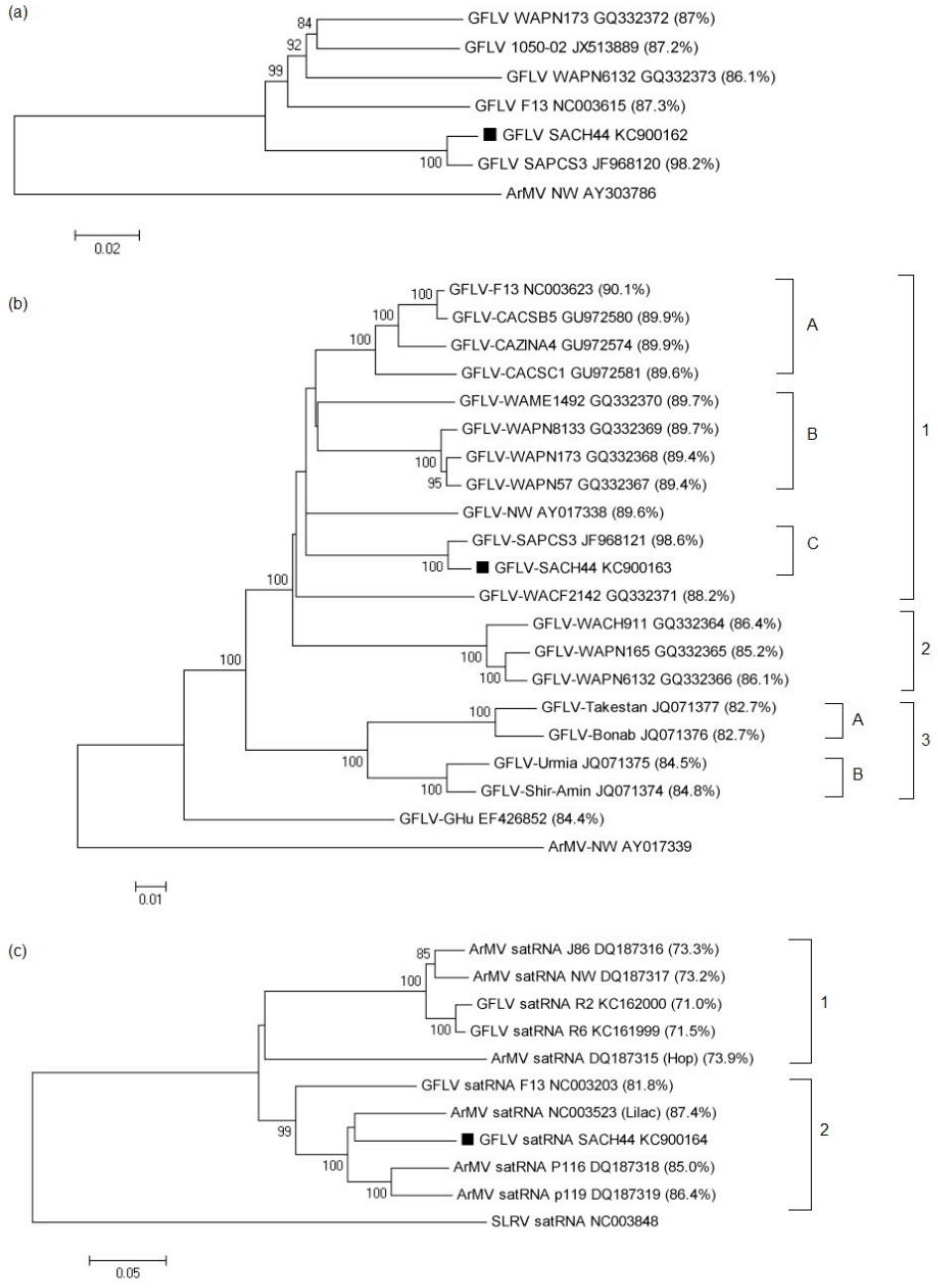


Figure 4.6: Phylogenetic trees based on full-length nt sequences of (a) RNA1 of GFLV isolates; (b) RNA2 of GFLV isolates and (c) satRNAs of GFLV and ArMV isolates. The nt identity between GFLV-SACH44 and other GFLV or ArMV isolates are indicated in brackets. The accession numbers of the isolates are indicated next to the isolate names. GFLV-SACH44 is indicated on the tree as solid block. For

phylogenetic analysis of RNA1 and RNA2, ArMV-NW was used as an outgroup (acc AY303786 and AY017338, respectively) and Strawberry latent ringspot virus (SLRSV) satRNA (acc NC003848) was used as an outgroup for satRNA phylogenetic analysis. All the phylogenetic trees were constructed using the Neighbour-Joining method. The percentage of replicate trees in which the associated taxa clustered together in the bootstrap test (1000 replicates) is shown next to the branches. Phylogenetic analysis was conducted in MEGA5 (Tamura et al., 2011).

4.4.3 Infectivity screening of L140-GFLV-satRNAfl

The entire GFLV-SACH44 satRNA was amplified as one RT-PCR fragment and cloned into the L140 expression vector (Stephan, 2005). This led to L140-GFLV-satRNAfl (Figure 4.7), a vector in which the GFLV-satRNA was cloned 6 nt downstream of the transcription start of an enhanced Cauliflower mosaic virus (CaMV) 35S promoter (Töpfer et al., 1987) and upstream of a self-processing hammerhead ribozyme sequence (Hofacker, 2003). Successful self-processing would lead to potentially three foreign nt at the 3' end of the primary transcripts derived from L140-GFLV-satRNAfl.

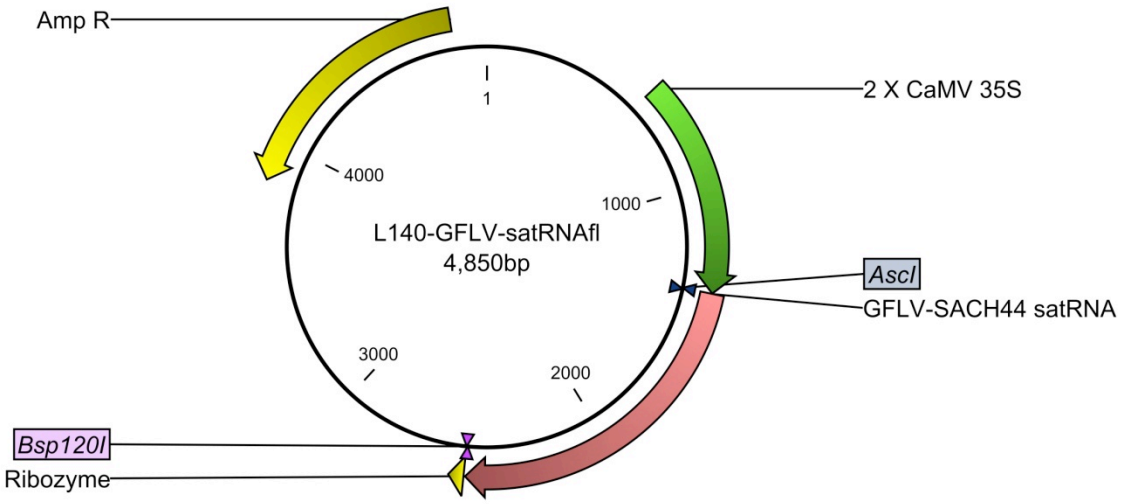


Figure 4.7: Graphic representation of clone L140-GFLV-satRNAfl. The pGEM-T-Easy satfrag population mix was digested with *Asc*I and *Bsp*120I restriction enzymes, and the resulting fragment 1.1 kb was inserted in vector L140 that was opened with the same restriction enzymes. The resulting plasmid contained a double CaMV 35S promoter (2x35S) (Töpfer et al., 1987), the entire GFLV-SACH44 satRNA insert (satfrag), a self-processing hammerhead ribozyme sequence (Ribozyme) (Hofacker, 2003) and Ampicilin resistance (Amp R) gene.

Only two clones (clones 3 and 12), from a total of 6 clones, of the population of L140-GFLV-satRNAfl showed the correct restriction digest patterns (Figure 4.8, 2.8 kb + 2 kb) when digested with *Eco*RI and *Hind*III (Fermentas). The digest patterns of the other four clones showed a band smaller than the expected 2kb band. The clones were sequenced and it was shown that the intact satRNA was cloned between the double CaMV 35S promoter and the Ribozyme only in clones 3 and 12. Therefore only clones 3 and 12 were mechanically inoculated onto *C. quinoa.*

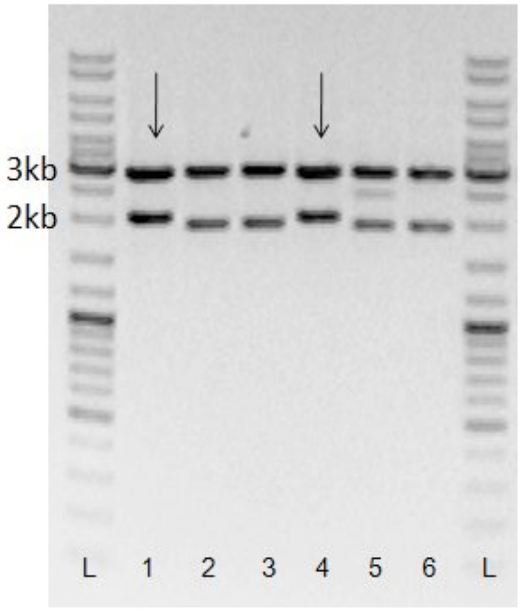


Figure 4.8: Control restriction digests on the individual L140-GFLV-SACH44-satRNA clones that were obtained by a population cloning strategy. Lanes 1 and 4 (clones 3 and 12, respectively), indicated by the arrows, produced the expected restriction digest patterns (2.8 kb + 2 kb) when digested with *Eco*RI and *Hind*III and were used for mechanical inoculations in *C. quinoa*. (L) GeneRuler™ DNA Ladder Mix (Fermentas).

Initially, each of the clones was mechanically co-inoculated with plant sap taken from singly ArMV- or GFLV-infected *C. quinoa* plants. Four out of seven inoculated plants (see Table 4.3) were shown by DAS-ELISA to be infected with either ArMV or GFLV. Nevertheless, all co-inoculated plants and the respective negative controls were used to screen for the presence of GFLV-SACH44 satRNA by RT-PCR.

Table 4.3: Infectivity results of L140-GFLV-satRNAfl as determined by RT-PCR. *Chenopodium quinoa* plants were initially co-inouculated with the satRNA constructs and ArMV-NW or GFLV-NW as helper viruses, and later repeated with GFLV-SAPCS3. Sections in bold are in which cases systemic spread of the satRNA were found. (+) indicates a positive ELISA or RT-PCR result; (-) a negative ELISA or RT-PCR result; nd not determined.

Experiment	Helper virus	GFLV ELISA result	ArMV ELISA result	satRNA RT-PCR
L140-GFLV-satRNAfl, clone 3	ArMV-NW	-	+	-
L140-GFLV-satRNAfl, clone 12	ArMV-NW	-	-	-
L140-GFLV-satRNAfl, clone 3	GFLV-NW	-	-	-
L140-GFLV-satRNAfl, clone 12	**GFLV-NW**	**+**	**-**	**+**
Transmission control	ArMV-NW	-	+	-
Transmission control	GFLV-NW	+	-	-
Healthy control	-	-	-	-
L140-GFLV-satRNAfl, clone 12	**GFLV-SAPCS3**	**+** 2/3	nd	**+** 1/2
Transmission control	GFLV-SAPCS3	+	-	-
Transmission control	GFLV-NW	-	-	nd
Healthy control	-	-	-	-

In the initial infectivity assay, one plant clearly showed the expected 760 bp amplicon for satRNA (Figure 4.6 a, lane 6) and this particular plant was co-inoculated with GFLV-NW and L140-GFLV-satRNAfl clone 12. The plant that was successfully infected with ArMV-NW and co-inoculated with L140-GFLV-satRNAfl clone 3 showed no indication of systemic spread of the satRNA by RT-PCR. The plants that were co-inoculated with GFLV-NW and L140-GFLV-satRNAfl clone 3, or ArMV and L140-GFLV-satRNAfl clone 12, had no RT-PCR products. This was expected due to the unsuccessful transmission of either GFLV or ArMV to these particular plants. None of the four plants showed any

symptoms of virus infection and there were no differences between the infected and control plants. Therefore the effect of the satRNA on either GFLV or ArMV symptomatology in *C. quinoa* could not be determined.

The infectivity of L140-GFLV-satRNAfl clone 12 was repeated and confirmed in a separate experiment by mechanical co-inoculation of the L140-GFLV-satRNAfl plasmid DNA with plant sap from GFLV-SAPCS3-infected *C. quinoa* (Table 4.3). The plasmid/plant sap mixes were rub-inoculated on three *C. quinoa* plants. Two of the three plants were successfully inoculated with GFLV-SAPCS3 plant sap, however only one of the two infected plant tested positive for the systemic spread of satRNA by RT-PCR (Figure 4.6 b).

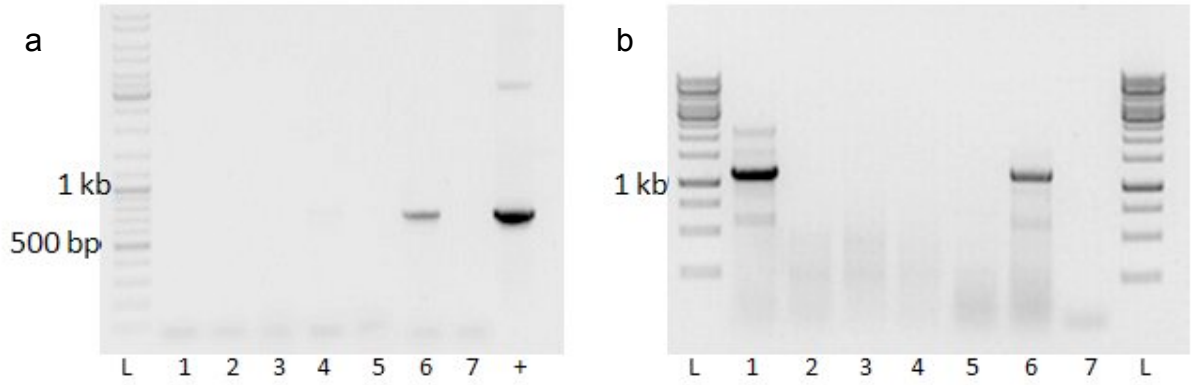


Figure 4.6: Infectivity testing of the GFLV-SACH44-satRNA full-length cDNA clones in *C. quinoa.* Total RNA was screened for the presence of satRNA systemically by RT-PCR with two primer sets; (a) A/G_sat1_F and A/G_sat4_R (expected PCR fragment 760bp) and (b) GFLV-satRNA_174F and GFLV_satRNA_dgR (expected PCR fragment 1.1 kb). (a) Initial infectivity screening of L140-GFLV-satRNAfl clone 3 and 12 co-inoculated with GFLV-NW and ArMV-NW. (L) GeneRuler™ DNA Ladder Mix (Fermentas), (1) No template control, (2) *C. quinoa* healthy control, (3) ArMV transmission control, (4) L140-GFLV-satRNAfl clone 3 with GFLV-NW; (5) L140-GFLV-satRNAfl clone 3 with ArMV-NW and L140; (6) L140-GFLV-satRNAfl clone 12 with GFLV-NW; (7) L140-GFLV-satRNAfl clone 12 with ArMV-NW and (+) plasmid satRNA clone control. (b) Infectivity testing L140-GFLV-satRNAfl clone 12 co-infected with GFLV-SAPC3 plant sap in *C. quinoa*. (L) GeneRuler 1kb ladder (Fermentas), (1, 2) L140-GFLV-satRNAfl clone 12 with GFLV-SAPCS3, (2) plant 2 with GFLV-SAPCS3 and L140-GFLV-satRNAfl clone 12, (3) *C. quinoa* healthy control, (4) GFLV-NW transmission control, (5) GFLV-SAPCS3 transmission control (no satRNA), (6) GFLV-SACH44 total RNA control, extracted from grapevine and (7) no template control.

Furthermore, Northern blot analysis was used to confirm the infectivity of the full-length satRNA cDNA clones in *C. quinoa* plants. The full-length cDNA clones (both L140-GFLV satRNA clones 3 and 12) were mechanically co-inoculated with plant sap from either ArMV-NW- or GFLV-NW-infected *C. quinoa*. The Northern blot analysis showed that both full-length cDNA clones were infectious in *C. quinoa* with co-inoculated with plant sap derived from GFLV-NW-infected *C. quinoa*, but not with ArMV-NW (Figure 4.7). Both L140-GFLV-satRNAfl clones produced a signal with the probe when it was co-inoculated with GFLV-NW, but not with ArMV-NW. With the controls, there was no signal with ArMV-NW (no satRNA), and no signal with the healthy control nor with GFLV-NW control without satRNA.

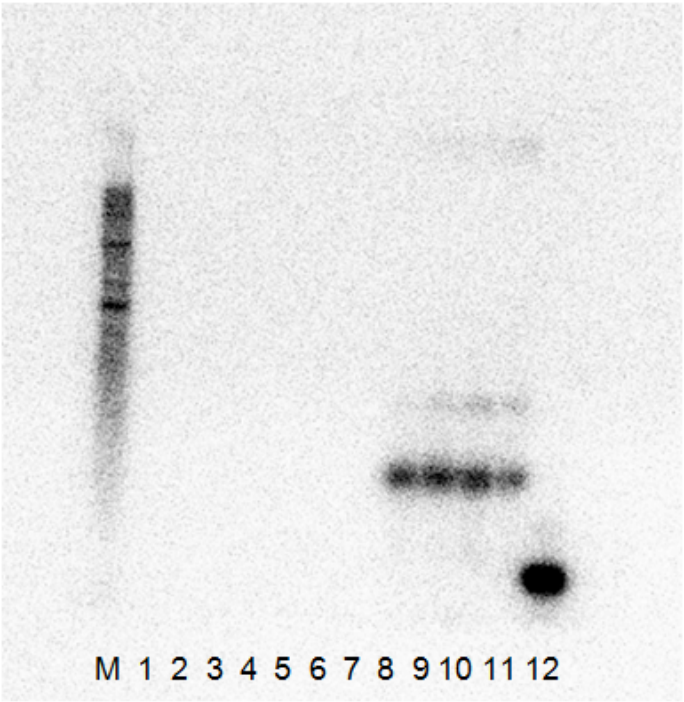


Figure 4.7: The Northern blot analysis results of the two L140-GFLV-satRNAfl clones co-inoculated with ArMV-NW and GFLV-NW. (M) RNA millenium (Ambion) marker, (1) ArMV-NW no satRNA control, (2-3) L140-GFLV-satRNAfl clone 3 with ArMV-NW, (4-5) L140-GFLV-satRNAfl clone 12 with ArMV-NW, (6) healthy *C. quinoa* control, (7) GFLV-NW no satRNA control, (8-9) L140-GFLV-satRNAfl clone 3 with GFLV-NW, (10-11) L140-GFLV-satRNAfl clone 12 with GFLV-NW and (12) positive PCR product control (probe ~ 400 bp).

4.5 Discussion

In this chapter, new GFLV-associated satRNAs naturally found in a SA vineyard are described. Twenty-nine GFLV-infected field plants were screened for the presence of satRNAs, of which two were found to be associated with satRNA. In a BSc (Hons) study, conducted by Ms M Spaltman (Stellenbosch University, Western Cape, South Africa), an additional 80 grapevine samples from the same area (De Wetshof, Robertson, South Africa) were also screened with GFLV satRNA-specific primers for the presence of satRNAs. No satRNAs were detected in any of these plants, even after using additional primer combinations (unpublished). Therefore, in total, more than a 100 plants from the same area were screened, and only two plants were found to contain satRNAs. In similar surveys, satRNAs could only be detected in 5 of 34 (14.7%) tested GFLV-infected plants (Saldarelli et al., 1993), and 8 out of 47 (17%) tested ArMV-infected plants (Wetzel et al., 2006). Both these previous studies included grapevine samples from a wider geographic region which may have led to the higher satRNA incidence rate than what was found in this present study, suggesting that a wider geographic region in South Africa should be screened. The initial sequencing results of the two satRNAs (GFLV-SACH44 and GFLV-SACH47) indicated that both were more closely related to the satRNAs of various ArMV isolates, rather than to the other completely sequenced GFLV satRNAs. The sequences of the GFLV-SACH44 and GFLV–SACH47 satRNAs were highly similar on a nt level (only six differences of the ~800 nt sequenced). It was therefore decided to determine the genomic and satRNA sequences of GFLV-SACH44 to compare with other GFLV isolates that do not contain satRNAs. Of all the ArMV and GFLV isolates that are reported to have satRNAs, GFLV-F13 and ArMV-NW are the only isolates that have their full genomes and satRNAs completely sequenced (Serghini et al., 1990; Ritzenthaler et al., 1991; Fuchs et al., 1989; Wetzel et al., 2001, 2004, 2006). This study has therefore contributed towards the body of knowledge regarding GFLV satRNAs by characterising another “helper virus and satRNA sequence set”. The full-length GFLV-SACH44 RNA1, RNA2 and satRNA

were deposited in Genbank with accession numbers KC900162, KC900163 and KC900164.

The two RNAs of GFLV-SACH44 are 7341 nt and 3816 nt in length and its satRNA is 1104 nt in length, all excluding the poly (A) tail. The 5' UTR of GFLV-SACH44 RNA2 is one nt shorter than that of GFLV-SAPCS3, which has the longest 5'UTR to date. The GFLV-SACH44 RNA2 also has the AA/GTCCGTT/CA motif at positions 73–98, which is present in GFLV-SAPCS3, GFLV-Ghu and other ArMV isolates, but not in other GFLV isolates sequenced to date. This suggests that GFLV-SACH44, like GFLV-SAPCS3, may have arisen from the same ancestor that might have originated from an interspecies recombination event in the 5' UTR region between a GFLV-F13 type and an ArMV-Ta type isolates (Lamprecht et al., 2012). It would be interesting to know how often this ArMV-based insertion occurs in South African GFLV isolates and elsewhere, as it may reveal where these recombinant viruses might have originated from. As expected, the GFLV-SACH44 genome was closely related to the GFLV-SAPCS3 isolate, and with very little significant variation. Therefore, the natural association of satRNAs to certain GFLV isolates seems not to be dependent on nt sequence alone.

For more than two decades, the only GFLV-associated satRNA that was completely sequenced was the satRNA of isolate GFLV-F13. Recently two new GFLV-associated satRNAs were described, roughly during the same period that our GFLV-SACH44 satRNA was characterised. The GFLV-SACH44 satRNA described in this chapter is 1104 nt long, with 5'- and 3' UTRs 14 and 74 nt in length, respectively. When compared to other full-length satRNA sequences that were available on GenBank, the GFLV-SACH44 satRNA was phylogenetically more closely related to three ArMV satRNAs than to the other three GFLV satRNA. When a phylogenetic tree was constructed based on all the available full-length ArMV and GFLV satRNA sequences, the satRNAs separated into two distinct clades. The four GFLV satRNAs did not group together as expected, and instead, grouped in two different clades; GFLV-F13 and GFLV-SACH44 grouped in one clade, while GFLV-R2 and GFLV-R6 grouped closely together in the other clade. Even with the newly described Californian GFLV satRNAs, the GFLV-

SACH44 satRNA still seems to be distinct, in sharing nt identity of only 71.0-81.8% with the GFLV-satRNAs. The high nt identity between GFLV-SACH44 and ArMV-lilac satRNAs (and other ArMV satRNA isolates) might be indicative of the presence of ArMV as a helper virus in the grapevine plant SACH44. However, this was shown by ArMV-specific DAS-ELISA not to be the case – in fact no plants from this particular vineyard were infected by ArMV. Therefore, GFLV-SACH44 satRNA was replicating in this plant with a GFLV isolate as a helper virus and not with an ArMV isolate even if the sequence information suggests otherwise. It may just be that there is a lack of GFLV satRNA full-length sequences that may lead one to assume that the GFLV-SACH44 satRNA is from an ArMV origin. This was also the case with the GFLV satRNAs of GFLV-R2 and GFLV-R6 that also shared higher nt identities with other ArMV isolates (ArMV-J86 and ArMV-NW) than to the GFLV-F13 satRNA. It is clear that more full-length sequences of GFLV- and ArMV-associated satRNAs are needed to gain a clearer picture regarding their phylogenetic relationships.

Full-length infectious cDNA clones of the GFLV-SACH44 satRNA were obtained by cloning the entire GFLV-SACH44 satRNA downstream of the transcription start of an enhanced CaMV 35S promoter (Töpfer et al., 1987) and upstream of a self-processing hammerhead ribozyme sequence (Hofacker, 2003). A population cloning strategy was used to maximize the probability of obtaining infectious cDNA clones, as described by Yu and Wong (1998). Only two clones from this population were used to inoculate *C. quinoa* plants. The L140-GFLV-satRNAfl clone 12 was shown, by RT-PCR, to be infectious by mechanical inoculation in *C. quinoa* with both GFLV-NW and GFLV-SAPCS3. As expected, since the GFLV-NW virus transmission to this plant was unsuccessful, L140-GFLV-satRNAfl clone 3 did not spread systemically in *C. quinoa* in the initial infectivity experiment. The infectivity testing of this clone was not repeated with GFLV-SAPCS3. Neither L140-GFLV-satRNAfl clone 3 nor clone 12 was infectious when co-inoculated with ArMV-NW. Infectivity of L140-GFLV-satRNAfl clone 12 was further demonstrated by an independent Northern blot analysis. Surprisingly, L140-GFLV-satRNAfl clone 3 was also shown to be infectious in *C. quinoa* co-inoculated with GFLV-NW in Northern blot analysis. The infectivity of L140-GFLV-satRNAfl clone 3 was

initially missed, as the GFLV-NW helper virus transmission to that particular plant was unsuccessful and led to a negative RT-PCR result. The Northern blot also confirmed that neither of the clones were infectious when they were co-inoculated with ArMV-NW, even though the GFLV-SACH44 satRNA is more similar to ArMV isolates than GFLV satRNAs on a nt level. Therefore it seems at this stage that the clones are infectious when co-inoculated with a homologous helper virus. One can only speculate why the GFLV-SACH44 satRNA only replicates along with GFLV isolates as a helper virus system, and not with this particular ArMV isolate, even though they are phylogenetically more closely related to the latter. It is possible that these clones may replicate with other ArMV isolates, but not with ArMV-NW. As an example, the infectious clone of the GFLV-F13 satRNA was able to replicate in *C. quinoa* protoplasts when co-infected with an ArMV isolate (ArMV-S) as a helper virus (Hans et al., 1993). Why certain ArMV or GFLV isolates may act as a helper virus for a particular satRNA, and the possible determinants in the genome responsible for their association, is not known at present. More full-length sequences and full-length infectious cDNA clones are needed for both ArMV and GFLV satRNAs, because at this stage it is difficult to predict from sequence analysis alone why some isolates of ArMV or GFLV are helpers for the replication of a specific satellite, and other isolates not (Liu et al., 1993, Wetzel et al., 2006).

4.6 Conclusion

In this chapter the genomic and satRNA full-length sequences of a GFLV from South Africa is reported. No clear sequence variation could be detected between the GFLV isolates that are naturally associated (GFLV-SACH44) with a satRNA to those that are not (GFLV-SAPCS3), indicating that recognition of the satellite by the replication machinery is more complex than it appears. Interestingly, the GFLV-SACH44 satRNA was more closely related in nt identity and sequence length to the ArMV satRNAs than to the other described GFLV satRNAs. It was shown that the grapevine plant from which the GFLV-SACH44 satRNA was isolated was not infected with ArMV, and therefore the satRNA described in this chapter, even though it shows high similarity to ArMV satRNAs, is clearly associated with a GFLV isolate.

Furthermore, full-length infectious cDNA clones of GFLV-SACH44 satRNA were constructed that were able to replicate with two different GFLV isolates (GFLV-NW and GFLV-SAPCS3) after mechanical inoculation. There was no change in symptom expression in the presence of the satRNAs; however, the small number of plants used to confirm the infectivity of selected clones does not allow for an overall assumption on symptomatology. L140-GFLV-satRNA clone 12, in combination with two GFLV isolates as the helper virus, was shown to be infectious in newly formed systemically-infected *C. quinoa* leaves in three independent experiments. The infectivity of this clone was also confirmed using two different techniques, by RT-PCR and Northern blot analysis. This clone was selected to be cloned into a binary vector system to allow agroinfiltration experiments in *N. benthamiana* for further analysis (Chapter 5).

4.7 References

1. Buzayan JM, Gerlach WL and Bruening G (1986). Satellite tobacco ringspot virus RNA: A subset of the RNA sequence is sufficient for autolytic processing. Proc Nati Acad Sci, 83, 8859-8862.
2. Buzen FG, Niblett CL, Hooper GJ, Hubbard J and Newman MA (1984). Further characterization of panicum mosaic virus and its associated satellite virus. Phytopathology, 74, 313–318.
3. Celix A, Rodriguez-Cerezo E, Garcia-Arenal F (1997). New Satellite RNAs, but No DI RNAs, Are Found in Natural Populations of Tomato Bushy Stunt Tombusvirus. Virology, 239 (2), 277–284.
4. Davies DL and Clark MF (1983). A satellite-like nucleic acid of arabis mosaic virus associated with hop nettlehead disease. Annals of App Biol, 103, 439–448.
5. Dodds JA (1998). Satellite tobacco mosaic virus. Annu Rev Phytopathol, 36, 295-310.
6. Demler SA and De Zoeten GA (1989). Characterization of a Satellite RNA Associated with Pea Enation Mosaic Virus. J Gen Virol, 70 (5), 1075-1084.
7. Francki RIB (1985). Plant virus satellites. Annu Rev Microbiol, 39, 151-174.
8. Fritsh C, Mayo MA (1989). Satellites of plant viruses. In C. L. Mandehar (ed.), Plant Viruses, vol I, Structure and replication, p. 289-321. CRC Press, Inc., Boca Raton, Fla.
9. Fuchs M, Pinck M, Serghini MA, Ravelonandro M, Walter B, Pinck L (1989). The nucleotide sequence of satellite RNA in grapevine fanleaf virus, strain F13. J Gen Virol, 170, 955-962.
10. Gottula JW, Lapato D, Cantilina KK, Saito S, Bartlett B, Fuchs M (2013). Genetic variability, evolution and biological effects of Grapevine fanleaf virus satellite RNAs. Phytopathology, http://dx.doi.org/10.1094/PHYTO-11-12-0310-R.

11. Hans F, Pinck M, Pinck L (1993). Location of the replication determinants of the satellite RNA associated with grapevine fanleaf nepovirus (strain FI3). Biochimie, 75, 597-603.
12. Hemmer O, Oncino C, Fritsch C (1993). Efficient replication of the *in vitro* transcripts from cloned cDNA of tomato black ring virus satellite RNA requires the 48K satellite RNA-encoded protein. Virol, 194, 800-806.
13. Hofacker IL (2003). Vienna RNA secondary structure server. Nucleic Acids Res, 31, 3429-3431.
14. Hu CC, Hsu YH, Lin NS (2009). Satellite RNAs and Satellite Viruses of Plants. Viruses, 1, 1325-1350.
15. Huang YW, Hu CC, Lin NS, Hsu YZ (2010). Mimicry of molecular pretenders. RNA Biol, 7, 162-171.
16. Kaper JM and Tousignant ME (1977). Cucumber mosaic virus-associated RNA 5. I. Role of host plant and helper strain in determining amount of associated RNA 5 with virions. Virology, 80, 186–195
17. Kaper JM, Tousignant ME and Steger G (1988). Nucleotide sequence predicts circularity and self-cleavage of 300-ribonucleotide satellite of arabis mosaic virus. Biochem Biophys Res Commun, 154 (1), 318-325.
18. Kaper JM., Tousignant ME, Steger G (1988). Nucleotide sequence predicts circularity and self-cleavage of 300-ribonucleotide satellite of arabis mosaic virus. Biochem Biophys Res Commun, 154, 318-325.
19. Lin NS and Hsu YH (1994). A satellite RNA associated with bamboo mosaic potexvirus. Virology, 202 (2), 707-14.
20. Liu YY and Cooper JI (1993). The multiplication in plants of arabis mosaic satellite RNA requires the protein. J Gen Virol, 74, 1471-1474.
21. Liu YY, Cooper JI, Edwards ML and Hellen CUT (1991). A satellite RNA of arabis mosaic nepovirus and its pathological impact. Annals of App Biol, 118, 577-587.
22. Liu YY, Hellen CU, Cooper JI, Bertioli DJ, Coates D, Bauer G (1990). The nucleotide sequence of a satellite RNA associated with arabis mosaic nepovirus. J Gen Virol, 71, 1259-1263.
23. Liu YY, Hellen CU, Cooper JI, Bertioli DJ, Coates D, Bauer G (1990). The nucleotide sequence of a satellite RNA associated with arabis mosaic nepovirus. J Gen Virol, 71, 1259-1263.
24. Mayo MA, Barker H, Harrison BD (1979). Polyadenylate in the RNA of five nepoviruses. J Gen Virol, 43, 603-610.
25. Mayo MA, Barker H, Harrison BD (1982). Specificity and properties of the genome-linked proteins of nepoviruses. J Gen Virol, 59, 149-162.
26. Mayo MA, Leibowitz MJ, Palukaitis P, Scholthof KBG, Simon AE, Stanley J, Taliansky M (2005). In Virus Taxonomy: VIIIth Report of the International Committee on Taxonomy of Viruses; Faquet CM, Mayo MA, Maniloff J, Desselberger U, Ball IAE. Elsevier Academic Press, New York, USA, pp. 1163-1169.

27. Meng B, Li C, Wang W, Goszczynski D, Gonsalves D (2005). Complete genome sequences of two new variants of Grapevine rupestris stem pitting-associated virus and comparative analyses. J Gen Virol, 86,1555-1560.

28. Meyer M, Hemmer O and Fritsch C (1984). Complete Nucleotide Sequence of a Satellite RNA of Tomato Black Ring Virus. J Gen Virol, 65, 1575-1583.

29. Miller WA, Hercus T, Waterhouse PM, Gerlach WL (1991). A satellite RNA of barley yellow dwarf virus contains a novel hammerhead structure in the self-cleavage domain. Virology, 183, 711–720.

30. Murant AF; Mayo MA (1982). Satellites of plant viruses. Annu. Rev. Phytopathol. 20, 49-70.

31. Pink L, Fuchs M, Pinck M, Ravelonandroi M, Walter B (1988). A Satellite RNA in Grapevine Fanleaf Virus Strain F13. J Gen Virol, 69, 233-239.

32. Ritzenthaler C, Viry M, Pinck M, Margis R, Fuchs M, Pinck L (1991) Complete nucleotide sequence and genetic organization of grapevine fanleaf nepovirus RNA1. J Gen Virol, 72, 2357-2365.

33. Rubino L, Tousignant ME, Steger G and Kaper JM (1990). Nucleotide sequence and structural analysis of two satellite RNAs associated with chicory yellow mottle virus. J Gen Virol, 71(9), 1897-1903.

34. Rubino L, Burgyan J, Grieco F and Russo M (1990). Sequence Analysis of Cymbidium Ringspot Virus Satellite and Defective Interfering RNAs. J Gen Virol, 71 (8), 1655-1660.

35. Sehgal OP, Sinha RC, Gellatly DL, Ivanov I and AbouHaidar MG (1993). Replication and encapsidation of the viroid-like satellite RNA of lucerne transient streak virus are supported in divergent hosts by cocksfoot mottle virus and turnip rosette virus. J Gen Virol, 74, 785-788.

36. Serghini MA, Fuchs M, Pinck M, Reinbolt J, Walter B, Pinck L (1990) RNA2 of grapevine fanleaf virus, sequence analysis and coat protein cistron location. J Gen Virol, 71, 1433-1441.

37. Shi BJ, Ding SW, Symons RH (1997). Plasmid vector for cloning infectious cDNAs from plant RNA viruses: high infectivity of cDNA clones of tomato aspermy cucumovirus. J Gen Virol, 78, 1181-1185.

38. Simon AE and Howell SH (1986). The virulent satellite RNA of turnip crinkle virus has a major domain homologous to the 3' end of the helper virus genome. EMBO J, 5(13), 3423–3428.

39. Stephan D (2005). Molekulare Charakterisierung von *Beet mild yellowing virus* (BMYV) und *Beet chlorosis virus* (BChV) sowie Selektion von BMYV *Amplicon*-transgenen *Nicotiana benthamiana*. PhD Dissertation, Hannover University, Germany.

40. Tamura K, Peterson D, Peterson N, Stecher G, Nei M, and Kumar S (2011). MEGA5: Molecular Evolutionary Genetics Analysis using Maximum Likelihood, Evolutionary Distance, and Maximum Parsimony Methods. Molecular Biology and Evolution, 28, 2731-2739.

41. Thompson JD, Higgins DG, Gibson TJ (1994). CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. Nucleic Acids Res, 22, 4673-4680.

42. Töpfer R, Matzeit V, Gronenborn B, Schell J, Steinbiss HH, (1987). A set of plant expression vectors for transcriptional and translational fusions. Nucleic Acids Res, 15, 5890.

43. Wetzel T, Fuchs M, Bobko M, Krczal G (2002). Size and sequence variability of the *Arabis mosaic virus* protein 2A. Arch Virol, 147, 1643–1653.

44. Wetzel T, Meunier L, Jaeger U, Reustle GM, Krczal G (2001). Complete nucleotide sequences of the RNAs 2 of german isolates of grapevine fanleaf and arabis mosaic nepoviruses. Virus Res, 75 (2), 139–145.

45. Wetzel T., Bassler A., Amren MA, Krczal G (2006). A RT/PCR-partial restriction enzymatic mapping (PREM) method for the molecular characterisation of the large satellite RNAs of Arabis mosaic virus isolates. J Virol Methods, 132, 97-103.

46. Wetzel,T., Beck,A., Wegener,U. and Krczal,G (2004). Complete nucleotide sequence of the RNA 1 of a grapevine isolate of Arabis mosaic virus. Arch Virol, 149 (5), 989-995.

47. Whitmer Collmer C, Hadidi A and Kaper JM (1985). Nucleotide sequence of the satellite of peanut stunt virus reveals structural homologies with viroids and certain nuclear and mitochondrial introns. Proc Nati Acad Sci, 82, 3110-3114.

48. Xiang C, Han P, Lutziger I, Wang K and Oliver DJ (1999). A mini binary vector series for plant transformation. Plant Molecular Biology, 40, 711-717.

49. Yu HH and Wong SM (1998). Synthesis of biologically active cDNA clones of cymbidium mosaic potexvirus using a population cloning strategy. Arch Virol, 143 (8), 1617-1620.

50. Zarghani SN, Shams-Bakhsh M, Bashir NS and Wetzel T (2013). Molecular Characterization of Whole Genomic RNA2 From Iranian Isolates of Grapevine Fanleaf Virus. J Phytopathology, doi: 10.1111/jph.12089.

Chapter 5

Construction of a full-length cDNA clone of a South African isolate of Grapevine fanleaf virus and a satellite RNA

5.1. Abstract

Infectious cDNA clones of plant viruses provide an excellent tool for the research of viral gene functions and virus-host interactions. However, only a few infectious clones based on grapevine viruses are available. In this chapter, the first attempt at constructing full-length infectious clones based on a South African GFLV isolate, including a satellite RNA, is described. Full-length viral sequences were cloned downstream of a 35S Cauliflower mosaic virus (CaMV) promoter and transferred to a binary vector. A population cloning strategy was followed and several constructs were agro-infiltrated in *N. benthamiana.* After three agro-infiltrations were attempted, none of the clones seemed to be infectious. When investigated, significant mutations in the GFLV RNA1 full-length clone were found. Possible causes and proposed methods to overcome these mutations are discussed.

5.2. Introduction

A variety of techniques have been used to examine plant responses induced by virus infection and plant–virus interactions, and as a result full-length genomic cDNA clones have been produced for a number of plant viruses (Lamprecht & Jelkman 1997). Construction of an infectious clone is generally the first step for determining the function of sequences of RNA viruses by reverse genetics (Chapman, 2008). An infectious clone

is essentially a full-length cDNA copy corresponding to the RNA sequence of the virus. This approach allows for the modification of genes in the virus to gain insight into the genomic organisation and gene function (Boyer and Haenni, 1994). One of the first plant RNA viruses that were converted into an infectious clone was Brome mosaic virus (Ahlquist et al., 1984). Since then, many more RNA viruses have been converted into infectious clones. Generally cDNA fragments are linked together and the full-length cDNAs are brought under the control of a chosen promoter. Earlier infectious clones were under the control of a Pm or a T7 promoter, thus allowing the production of infectious RNAs by *in vitro* transcription. More recently infectious clones were fused to the Cauliflower mosaic virus (CaMV) 35S promoter (Töpfer et al., 1987) to drive *in vivo* transcription. The CaMV 35S promoter is a strong constitutive promoter and serial arrangements of several copies of the promoter are often used in vectors (Kay et al., 1987). A number of 35S-cDNA infectious clones were able to infect plant hosts by manual inoculation. However, the 35S-cDNAs that were able to infect some plants following manual inoculation were not all able to infect the normal host range of the corresponding virus (Weber et al., 1992; Neeleman et al., 1993; Ding et al., 1995; Dagless et al., 1997). Also, lack of infection via manual inoculations may be due to unsuccessful delivery into the plant cell nucleus (Dagless et al, 1997). Some viruses are also phloem-limited, and may not be able to transmit via manual inoculations. Because of this, several infectious clones are cloned into binary vectors that allow for agro-infiltration with the use of *Agrobacterium tumefaciens*. With this method, the viral genome will be transcribed *in vivo* and the virus will replicate and move systemically throughout the plant. Infectious clones can further be modified to expression vectors to express foreign proteins at high concentrations. Furthermore, infectious clones can also be converted into virus-induced gene silencing (VIGS) vectors, which exploit the RNA-mediated post-transcriptional gene-silencing mechanisms in plants and are commonly used for functional genomic studies of the host plant.

Viry et al. (1993) reported the first full-length GFLV cDNA infectious clone, based on isolate GFLV-F13. The cDNA clones of RNA1 and RNA2 were each cloned downstream from a T7 promoter. However, the transcripts were only infectious in *Chenopodium*

quinoa protoplasts when inoculated by electroporation. Other full-length infectious cDNA clones that are based on grapevine viruses include infectious clones of ArMV-NW (Dupuis et al., 2009), Grapevine virus A (GVA) (Muruganantham et al., 2009), Grapevine rupestris stem pitting-associated virus (GRSPaV) (Meng et al., 2012) and Grapevine Leafroll-associated virus-2 (GLRaV-2) (Kurth et al., 2012). However, of these clones, only the GVA and GLRaV-2 clones are infectious in the natural host *V. vinifera*. Both these clones have been successfully converted into VIGS vectors. The major drawback of the GVA-based VIGS vector is that the virus is naturally phloem-restricted in grapevine, and as a result the GVA-based vector therefore has limitations in causing systemic silencing. The GVA-derived VIGS vector was restricted only to the leaf veins and inter-leaf veins in *V. vinifera* and this is not ideal for a VIGS system. The GLRaV-2 VIGS vector was able to silence *V. vinifera* endogenous PDS and express GFP systemically, even though this filamentous virus is also phloem-limited. More grapevine virus-based infectious clones are needed as they will greatly assist in developing a functional genomics tool for grapevine, as these molecular tools are desperately needed for the development of grapevine functional genomics. In this chapter we report the first attempt at construction of full-length cDNA clones (genomic and satellite RNA) based on South African GFLV isolates. The modification of the satellite RNA into an expression and VIGS vector is also described.

5.3. Materials and Methods

5.3.1. Virus source and total RNA extraction

Isolate GFLV-SAPCS3 was routinely maintained in *C. quinoa* and/or *N. benthamiana* in an insect-free greenhouse as described in section 3.3.1. The greenhouse facility is controlled with temperatures between 18°C and 28°C and relative humidity of approximately 70%. For the construction of the full-length cDNA clones, total RNA was extracted from GFLV-SAPCS3 infected *N. benthamiana* using the CTAB method (White et al., 2008).

5.3.2. Primers

Primers for the construction of the full-length cDNA clones were designed from full-length RNA1 and RNA2 GFLV-SAPCS3 sequences (Chapter 3, Accession numbers JF968120 and JF968121). Primers used for the cloning of ArMV-NW full-length RNA1 into the binary vector system are listed (section 5.3.4). The primers were designed so that the resulting PCR fragments would overlap and that each amplicon contains unique restriction enzyme sites. Some of the primers were designed to add an extra restriction enzyme site to the fragments to facilitate subsequent cloning steps. The primers for the construction of the infectious clones are listed in Table 5.1

Table 5.1: Primers used in this study to construct full-length cDNA clones for both GFLV-SAPCS3 and ArMV-NW

Primer pairs	Sequence (5'-3')	Additional	Product size
GFLV RNA1			
GFLV SAPCS3 RNA1 5'F	ATGAAAAATTTTTGTGAGTTCTTAC		Frag1 317 bp
GFLV RNA1 RACE GSP3	AGTTCCTTAGCCTCCGCATT		
GFLV_ic_RNA1_pF2	TCCAGCGAAGAGTTTGAGAA		Frag2 2132 bp
GFLV _ic_RNA1_pR2_SacI	AA**GAGCTC**AGTAAGTTCGCCTATCGCCC	*Sac*I flap	
GFLV_ic_RNA1_pF3	GCATGCGTGTTGTAAAAGCG		Frag3 3130 bp
GFLV_ic_RNA1_pR3	TTCCCCAACAAAAAATCTCG		
GFLV_ic_Frag4_F	ATTCCTTGAAGATATTTC		Frag4 2772 bp
GFLV_ic_RNA1_pR4_dT30_EcoRI	AA**GAATTC**T(30)ATTAAATGCAAAACAGTAAAAAG	*Eco*RI flap	
pCASS2_e35S_F_XmaI	AA**CCCGGG**CATGGTGGAGCACGACACTCT	*Xma*I flap	PCR1 4646 bp
GFLV_RNA1_PCR1_R_SalI	AA**GTCGAC**CAGGCGCTGTTAACAGCCGTCA	*Sal*I flap	
GFLV_RNA1_PCR2_HpaI F	GCTGTGACGGCTGTTAACAGCGCCTCTATA		PCR2 3932 bp
pCASS2_35ST_R_SalI (GFLV)	AA**GTCGAC**CAGCTGGCACGACAGGTTTC	*Sal*I flap	
GFLV RNA2			
GFLV_ic_RNA2_pF3	CTGAACATACGTGATATGAT		RNA2 frag 3 1611 bp
GFLV_ic_RNA2_pR3_HindIII	AA**AAGCTT**(Tx34)ATAAATTTGCAAAACAGTAAAAAG	*Hind*III flap	
GFLV_SAPCS3_RNA2_5'F	ATGAAAAATCTGCTGGGGTTTTCAT		Partial RNA2
GFLV2_3743R	ACAACACACTGTCGCCACTAAAAGC		

			3743 bp
GFLV2_EcoRI_3462 F	GCAAACCC**GAATTC**TCATTTCTCCC	*Eco*RI mutagenesis	RNA2 3'end 412 bp
GFLV_ic_RNA2_pR3	AA (Tx34)ATAAATTTGCAAAACAGTAAAAAG		
pCASS2_e35S_F_ClaI	AA**ATCGAT**CATGGTGGAGCACGACACTCT	*Cla*I flap	2x35S-RNA2-35ST 5042 bp
pCASS2_35ST_R SalI (GFLV)	AA**GTCGAC**CAGCTGGCACGACAGGTTTC	*Sal*I flap	
GFLV satRNA			
GFLV_satRNA_ic_F	ATGAAAAATTTCTATGGGTTCTCGT		satRNA 1129 bp
GFLV_satRNA_ic_R_EcoRI	AA**GAATTC**(Tx17)GAGTTGACTTATGAGCAA	*Eco*RI flap	
GFLV_sat 3'end_F	GTTCTCCGGATGTTGCGTCA		Sat 3' end 254 bp
pCB301_ClaI_R	CGACGGTATCGATAAGCTTG		
P2A_F_BamHI	AA**GGATCC**AACTTTGACCTTCTTAAG	*Bam*HI	P2A-GFP 784 bp
GFP_SnaBI_R	AA**TACGTA**GTACAGCTCGTCCATGCCGA	*Sna*BI	
GFP100_F_BamHI	AA**GGATCC**GAGTACAACTACAACAGC	*Bam*HI	GFP-100 116 bp
GFP100_R_SnaBI	AA**TACGTA**TGCCGTCCTCGATGTTGT	*Sna*BI	
GFP150_F_BamHI	AA**GGATCC**TACGTCCAGGAGCGCACC	*Bam*HI	GFP-150 166 bp
GFP150_R_SnaBI	AA**TACGTA**CAGCTTGTGCCCCAGGAT	*Sna*BI	
GFP200_F_BamHI	AA**GGATCC**CACCACCGGCAAGCTGCC	*Bam*HI	GFP-200 216 bp
GFP200_R_SnaBI	AA**TACGTA**CGAACTTCACCTCGGCGC	*Sna*BI	
Sat_mut_exp_F_BamHI-SnaBI	GTTGCTTTGGCGACTTGAGG**GGATCC**GGG**TACGTA** GGAAGGCTGTTGTGGGACAT	*Bam*HI-*Sna*BI	3281 bp
Sat_mut_exp_R_BamHI-SnaBI	ATGTCCCACAACAGCCTTCC**TACGTA**CCC**GGATC** CCCTCAAGTCGCCAAAGCAAC	*Bam*HI-*Sna*BI	
Sat_mut_sil_F_BamHI-SnaBI	GCTTTGGCGACTTGAGGTAG**GGATCC**GGG**TACGTA** GGAAGGCTGTTGTGGGACAT	*Bam*HI-*Sna*BI	3284 bp
Sat_mut_sil_R_BamHI-SnaBI	ATGTCCCACAACAGCCTTCC**TACGTA**CCC**GGATC** CCTACCTCAAGTCGCCAAAGC	*Bam*HI-*Sna*BI	
ArMV RNA1			
pCASS2_e35S_F_XmaI	AA**CCCGGG**CATGGTGGAGCACGACACTCT	*Xma*I flap	PCR1 4101 bp
ArMV_RNA1_PCR1_R_SalI	AA**GTCGAC**GCACATTGTACTTGTACATTTT	*Sal*I flap	
ArMV_RNA1_PCR2_BsrGI_F	GAAAAATGTACAAGTACAATGTGCTTGGTA		PCR2 4455 bp
pCASS2_35ST_R_SalI (ArMV RNA1)	AA**GTCGAC**CCAGCTGGCGAAAGGGGG	*Sal*I flap	

5.3.3. Construction of full-length cDNA clones based on GFLV-SAPCS3

CLC Bio Main Workbench was used for the *in silico* design of the cloning strategies. Unique restriction enzymes spanning the entire fragment were identified and primers flanking these regions were designed (Table 5.1). Each sub-cloning step was verified by

control restriction digests. Sequencing after every single assembly step was not possible as it was not cost and time efficient. PCR fragments were first cloned into a TA vector (pGEM-T-Easy, Promega), then sub-cloned directly between the double enhanced CaMV 35S promoter (Töpfer et al., 1987) and the 35S terminator sequence in the vector pCASS2 (Shi et al., 1997). The blunt cutter *Stu*I restriction enzyme site located at the transcription start of pCASS2 vector was utilized to ligate PCR fragments directly to the promoter, as extra non-viral nucleotides at the 5' end of the viral genome may significantly reduce the infectivity of constructs (Nagyova and Subr, 2007). The full-length cDNA flanked by the double enhanced CaMV 35S promoter and the 35S terminator sequence was then cloned into the pCB301 binary vector (Xian et al., 1999) between the right and left border sequences. Following ligation, vectors were transformed (section 3.3.2.4) in *E. coli* JM109. However, some plasmids were also re-transformed in *E. coli* JM110 (*dam*/*dcm*-) when methylation sensitive restriction enzymes (*Stu*I, *Xba*I and *Bsp120I*) were used.

A population cloning strategy (Yu and Wong, 1998) was followed in constructing the full-length cDNA clones. This strategy allows for simultaneous cloning of a large population of independent full-length clones. At each sub-cloning step, between 8 and 10 independent plasmids were pooled together and used for subsequent cloning steps. This approach maximizes the chances of obtaining infectious full-length cDNA clones (Yu and Wong, 1998). The cloning steps of full-length cDNAs of GFLV-SAPCS33 RNA1, RNA2, and the GFLV-SACH44 satRNA cloned into the pCB310 binary vector are described in the following sections. The modification of the satRNA clone into an expression and VIGS vector is also described.

5.3.3.1 Construction of full-length cDNA clones of GFLV-SAPCS3 RNA1

The cloning strategy of GFLV-SAPCS3 RNA1 involved the identification of unique restriction enzymes spanning the RNA1 sequence. Once the restriction sites were identified, four primer sets were designed for the amplification of the four overlapping fragments, designated RNA1 fragments 1- 4 (Figure 5.1). High fidelity enzymes were used for the cDNA synthesis (Superscript III Reverse Transcriptase, Invitrogen) and

PCR (Ex Taq DNA Polymerase, Takara) of the four fragments (methods described in section 3.3.2.3). The PCR products were gel purified using the Zymoclean Gel DNA Recovery kit (Zymo Research), followed by the ligation into the pGEM-T-Easy (Promega) vector. The four cloned overlapping GFLV-SAPCS3 RNA1 fragments were named pGEM-RNA1-frag1, pGEM-RNA1-frag2, pGEM-RNA1-frag3 and pGEM-RNA1-frag4. The ligation reactions were transformed as in section 3.3.2.4 and the plasmids were recovered using GeneJET Plasmid Miniprep kit (Thermo Scientific) according to manufacturer's instructions. The plasmids were verified by control restriction enzyme digests. The DNA of correct plasmids were spectrophotometrically quantified and mixed in equal amounts to gain a population of clones.

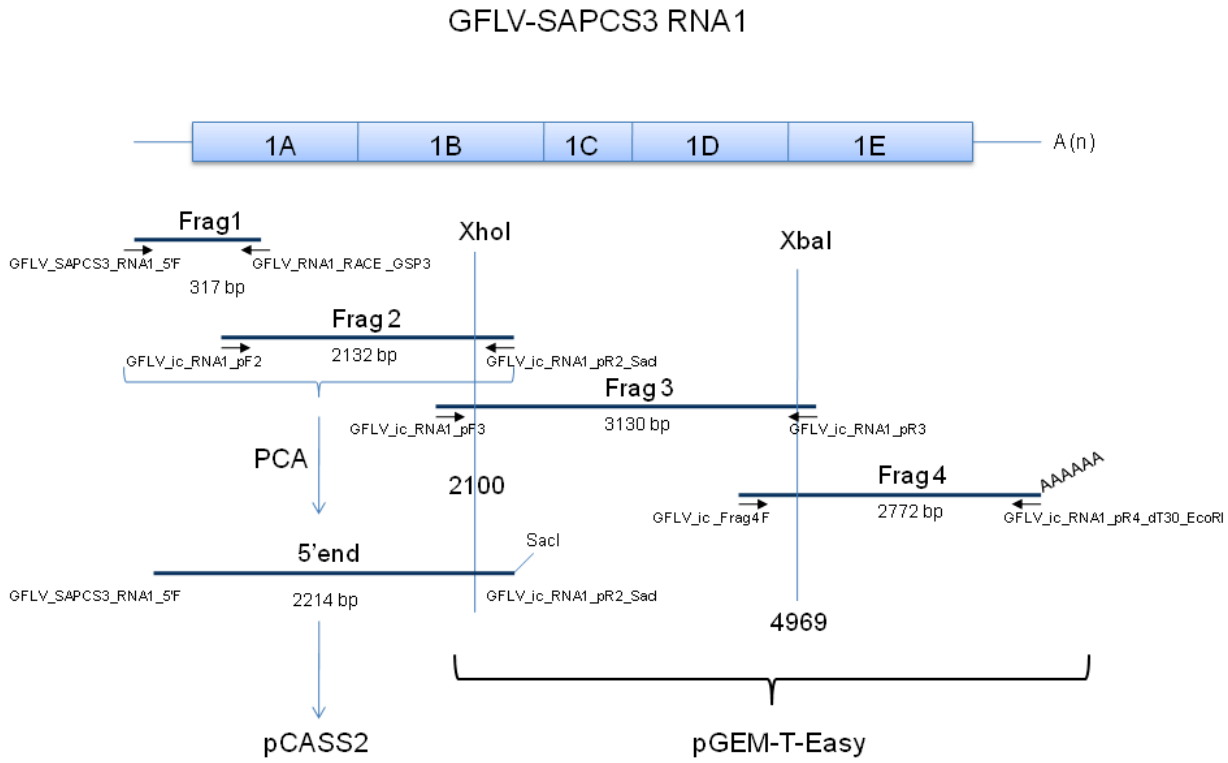


Figure 5.1: Diagrammatic representation showing the full-length cDNA assembly strategy of GFLV-SAPCS3 RNA1. Four overlapping PCR fragments each containing unique restriction enzyme sites were cloned into either pCASS2 (Shi et al., 1997) or pGEM-T-Easy (Promega) vectors. The genome organisation of GFLV RNA1 is represented by the top horizontal blue block. The horizontal blue lines are the 4 overlapping PCR fragments and their respective fragment sizes are indicated below the lines. The arrows represent the primers that were used to generate the PCR products. The vertical lines represent the positions of the unique restriction enzyme sites that were used to assemble the fragments, the enzymes are shown on top of the vertical lines and the position of the cut site is indicated at the bottom of the vertical line. The 5' end of RNA1 was generated by polymerase cyclic assembly and was directly cloned into the pCASS2 vector (Shi et al., 1997). Fragments 3 and 4 were cloned into pGEM-T-Easy (Promega).

For assembling the full-length GFLV-SAPC3 RNA1 under the control of a double 35S CaMV promoter, RNA1 fragments 1 and 2 were joined by polymerase cyclic assembly (PCA) to form the 5'end part of GFLV-SAPCS3 RNA1. Fragments 1 and 2 were amplified from their respective plasmids with the primers that were used to generate them. The amplicons were viewed on an agarose gel, and the expected 317 bp and 2.1 kb fragments were excised, and purified using Zymoclean Gel DNA Recovery kit (Zymo Research), according to the manufacturer's instructions. To assemble fragments 1 and 2 as one fragment, the purified amplicons were mixed in equal amounts and added to a PCR mixture of 1x HF Buffer (Finnzymes), 200µM dNTPs, Phusion Polymerase (Finnzymes) and made up to 20 µl with H_20 without any primers. Cycling conditions were as follows: 98°C for 30 sec, 10 cycles of 98°C for 10sec, 55°C 30 sec, and 72°C 2.5min, and a final extension step of 72°C for 7 min. The PCA product was subjected to PCR, using Phusion Polymerase (Finnzymes) as described above with the external primers GFLV_SAPCS3_RNA1_5'F and GFLV _ic_RNA1_pR2_SacI that resulted in a 2.4 kb amplicon including a SacI flap at the 3' end. The 2.4 kb PCR product was digested with *Sac*I (Fermentas), followed by phosphorylation (T4 Polynucleotide Kinase, Promega) according to the manufacturer's instructions. The *Sac*I digested fragment was ligated into a *Stu*I and *Sac*I (Fermentas) digested pCASS2 vector (transformed with *E. coli* JM110) as described in section 4.3.4, followed by transformation in *E. coli* JM109. The resulting pCASS2-GFLV-RNA1-5'-end plasmids were screened with restriction enzyme digestion and the DNA of correct pCASS2-GFLV-RNA1-5'-end plasmids were quantified using a Nanodrop spectrophotometer (Thermo Fisher). These plasmids were mixed in equal amounts to serve as a vector plasmid mix for the next cloning step. A population mix of pGEM-T-Easy-RNA1-frag4 plasmids was digested with *Eco*RI and separated on an agarose gel. The 2.8 kb digested fragment was ligated in the *Eco*RI digested pCASS2-GFLV-RNA1-5'-end vector population mix. The resulting pCASS2-GFLV-RNA1-5'+3'end plasmids were screened by restriction enzyme digests and the DNA of correct plasmids were quantified and mixed in equal amounts. A population mix of pGEM-T-Easy-RNA1-frag3 (transformed in *E. coli* JM110) was digested with *Xba*I and *Xho*I and the resulting fragment was cloned into the population mix of pCASS2-

GFLV-RNA1-5'+3'end (transformed in *E. coli* JM110) digested with the same restriction enzymes. The resulting clone, pCASS2-GFLV-SAPCS3-RNA1, therefore had the full-length GFLV-SAPCS3 RNA1 cDNA sequence between the double 35S CaMV promoter and the 35S termination sequence. The full-length pCASS2-GFLV-RNA1 was confirmed by colony PCR and control restriction digests.

The pCASS2-GFLV-RNA1 population mix was used as template for PCR to continue with the cloning into the pCB301 binary vector. To clone the entire GFLV-SAPCS3 RNA1 cDNA, together with the double 35S CaMV promoter and the 35S termination sequence, two fragments (PCR1 and PCR2) were amplified from the population mix and cloned into pGEM-T-Easy (named pGEM-2x35S-RNA1-PCR1 and pGEM-RNA1-35ST-PCR2, respectively). For PCR1, primers pCASS_e35S_F_XmaI and GFLV_RNA1_PCR1_SalI were used. This fragment included the entire double 35S CaMV promoter as well as a 3893 bp section of GFLV-SAPCS3 RNA1, flanked by the inserted *Xma*I and *Sal*I restriction enzyme flaps. The second fragment, PCR2, was amplified from pCASS2-GFLV-RNA1 with primers GFLV_RNA1_PCR2_HpaI_F and pCASS2_35ST_R_SalI (GFLV). This fragment included the rest of the GFLV-SAPCS3 RNA1 (3484 bp), the poly (A) tail, and the 35S termination sequence with an internal existing *Hpa*I site and flanked by the *Sal*I site at the 3' end. Population mixes were mixed for both pGEM-2x35S-RNA1-PCR1 and pGEM-RNA1-35ST-PCR2. The population mix of pGEM-2x35S-RNA1-PCR1 was digested with *Xma*I and *Sal*I (Fermentas) and the resulting fragment 4636 bp fragment was ligated into pCB301, digested with the same restriction enzymes. The resulting plasmid, pCB301-2x35S-RNA1-PCR1, was confirmed with control restriction digests and confirmed plasmids mixed for population cloning. When needed, rolling circle amplification (RCA), using phi29 DNA Polymerase (Thermo Scientific) was used, according to the manufacturer's instructions, to increase plasmid DNA concentrations for subsequent cloning steps. A population mix of pGEM-RNA1-35ST-PCR2 was digested with *Hpa*I and *Sal*I (Fermentas) and the 3947 bp fragment that included the rest of the RNA1 section and the 35S termination sequence was cloned into a population mix of pCB301-2x35S-RNA1-PCR1 digested with the same restriction enzymes. This resulted in pCB301-

2x35S-GFLVRNA1-35ST, with the double enhanced CaMV 35S promoter, GFLV-SAPCS3 RNA1 and 35S termination sequence, cloned between the T-DNA Left and Right borders of pCB301 binary vector.

5.3.3.2 Construction of full-length infectious clone GFLV-SAPCS3 RNA2

Several vectors and cloning strategies were assessed in the assembly of the GFLV-SAPCS3 RNA2 cDNA clone. These included dividing the genome in 3 fragments and using the internal unique restriction sites to assemble the fragments (as with GFLV-SAPCS3 RNA1), or to assemble the three fragments by circular polymerase extension cloning (CPEC) (Quan and Tian, 2009) (not discussed). In the end however, only one of the aforementioned fragments was used for the assembly of GFLV-SAPC3 RNA2 full-length clone. Fragment 3 was obtained by amplifying with primers GFLV_ic_RNA2_pF3 and GFLV_ic_RNA2_pR3 and cloning the resulting purified PCR fragment (1661 bp) in pGEM-T-Easy (Promega), resulting in pGEM-T-Easy-RNA2-frag3. This fragment included the 3' terminal part and a poly (A) tail of GFLV-SAPCS3 RNA2. The rest of the GFLV-SAPCS3 RNA2 was amplified as one PCR fragment and used for further sub-cloning steps (see below).

A population mix of pGEM-T-Easy-RNA2-frag3 (10 plasmids) was used as template to amplify the 3' end of GFLV-SAPCS3, using primers GFLV2_CPEC_3'F and GFLV2_ EcoRI_3462F (all 10 plasmids were in reversed orientation). Primer GFLV2_CPEC_3'F was designed to anneal to the pGEM-T-Easy backbone at the *Eco*RI site, while primer GFLV2_ EcoRI_3462F was designed to produce a silent mutation at position 3468-3476 of RNA2, changing the sequence ACTGTTAGG to AGAATTCGG. This fragment therefore included two *Eco*RI sites, one mutated internally at the 5' end of the fragment without affecting the amino acid sequence, and another downstream the poly (A) tail of the 3' end. The resulting 412 bp PCR product was gel purified, digested with *Eco*RI, and ligated into a *Eco*RI digested pCASS2 vector, resulting in pCASS2-RNA2-3'end. Colony PCR was used to confirm the correct orientation of the insert. Plasmids were recovered using GeneJET Plasmid Miniprep kit (Thermo Scientific), according to manufacturer's instructions, and were sequenced for confirmation. Plasmids were retransformed in *E.*

coli JM110. The cDNA of the remainder of the GFLV-SAPCS3 RNA2 genome was synthesized from total RNA using primer dT(17) (Meng et al., 2005), followed by PCR with primers GFLV_SAPCS3_RNA2_5'F and GFLV2_3743R. The resulting 3743 bp PCR fragment was gel purified, digested with *Xba*I (Fermentas), phosphorylated with T4 Polynucleotide Kinase (Promega) and ligated into pCASS2-RNA2-3'end (transformed in *E. coli* JM110) that was digested with *Stu*I and *Xba*I (Fermentas). This resulted in the full-length pCASS-GFLV-SAPCS3-RNA2, which included the full-length GFLV-SAPCS3 RNA2 flanked by the double enhanced CaMV 35S promoter and 35S termination sequences. The full-length clone was confirmed by restriction digest and sequencing.

To clone the full-length GFLV-SAPCS3 RNA2 into the binary vector pCB301, the entire region (including the double enhanced CaMV 35S promoter and 35S termination sequence) was amplified with high-fidelity Ex Taq DNA Polymerase (Takara) with primers pCASS2_e35S_F_ClaI and pCASS2_35S_T_R_SalI (GFLV), resulting in a fragment that was flanked by *Cla*I and *Sal*I restriction sites. pCASS-GFLV-SAPCS3-RNA2 was used as template. The resulting 5042 bp fragment was cloned into pGEM-T-Easy (Promega), resulting in pGEM-T-easy-2x35S-GFLV-RNA2-35ST. A control restriction digest showed that the insert was inserted in a reverse orientation upon cloning into pGEM-T-Easy (Promega). Therefore the entire fragment was flanked by *Sal*I on either side (as a result of the primer insertion and the internal *Sal*I site of the vector backbone) and digestion with *Cla*I was not necessary. The entire fragment including RNA2, promoter and termination sequence was excised from the TA vector using only *Sal*I, and ligated into pCB301 that was also digested with *Sal*I. This resulted in pCB301-2x35S-GFLVRNA2-35ST, with the double enhanced CaMV 35S promoter, RNA2 and 35S termination sequence, cloned between the T-DNA Left and Right borders of pCB301 binary vector.

5.3.3.3 Construction and modification of full-length infectious clone GFLV-SACH44 satRNA into expression and silencing vectors.

The full-length infectious cDNA clone L140-GFLV-satRNAfl clone 12 (Chapter 4) was used to construct a new satRNA infectious clone, one that had the same vector

backbones (pCASS2 and pCB301) as the GFLV-SAPCS3 RNA1 and RNA2 full-length clones. The strategy also ensured that the new satRNA clone would have no non-viral nucleotides between the promoter and the viral start site, unlike L140-GFLV-satRNAfl.

L140-GFLV-satRNAfl clone 12 was used as template for PCR using primers GFLV_satRNA_ic_F and GFLV_satRNA_ic_R_EcoRI. The PCR product with the *Eco*RI flap on the 3' end was *Eco*RI digested, phosphorylated (T4 Polynucleotide Kinase, Promega) and ligated into pCASS2 (retransformed in *E. coli* JM110) that was digested with *Stu*I and *Eco*RI. This resulted in pCASS2-GFLV-satRNA in which the satRNA fragment was cloned directly downstream the double enhanced 35S CaMV promoter and upstream the 35S termination sequence.

To clone the entire promoter, the satRNA and the 35S termination sequences into the pCB301 binary vector, pCASS2-GFLV-satRNA was digested with *Hind*III and *Pvu*II (Fermentas) and the 2137 bp fragment ligated into pCB301 that was digested with *Hind*III and *Sma*I (Fermentas). The resulting full-length satRNA clone, pCB301-2x35S-GFLV-satRNA-35ST, contained a double enhanced 35S CaMV promoter, full-length GFLV-SACH44 satRNA and a 35S termination sequence between the T-DNA Left and Right Borders. The clones were confirmed with control restriction digests.

To modify the full-length pCB301-2x35S-GFLV-satRNA-35ST clones into a GFP expression vector and a VIGS vector for silencing endogenous GFP in transgenic *N. benthamiana* 16c plants, *Bam*HI and *Sna*BI restriction enzyme sites were incorporated at the end of the ORF of the satRNA sequence for insertion of GFP sequences. All primers used for site-directed mutagenesis and amplification of the various fragments for gene expression or gene silencing are also listed in Table 5.1. The 3' end of the satRNA that included the end of the ORF, the added poly (A) tail and partial pCB301 vector backbone (at the *Cla*I site) were amplified with primers GFLV_sat_3'end_F and pCB301_ClaI_R and the resulting fragment (254 bp) was cloned into pGEM-T-Easy (Promega) resulting in pGEM-sat-3'end-mut. This plasmid would serve as the template for site-directed mutagenesis to incorporate of *Bam*HI and *Sna*BI restriction sites at the end of the satRNA ORF. Inserts for expression or silencing for the modified satRNA

vectors were obtained by PCR from a construct (Stephan, 2005) that consisted of the self-cleaving 2A peptide (P2A) from foot and mouth disease (Donnely et al., 2001) fused to GFP. For expression of GFP, the entire P2A-GFP fragment flanked by added *Bam*HI and *Sna*BI restriction enzyme sites was amplified and cloned into pGEM-T-Easy (Promega) leading to pGEM-P2A-GFP. For silencing of GFP, three different lengths (100bp, 150bp and 200bp) of GFP were amplified from the same plasmid that contained the P2A-GFP fusion peptide. *Bam*HI and *Sna*BI restriction sites were added to the GFP partial sequences during amplification and were cloned into pGEM-T-Easy (Promega), to yield pGEM-GFP100, pGEM-GFP150 and pGEM-GFP200.

For the GFP satRNA expression vector, the stop codon of the satRNA ORF was replaced by *Bam*HI and *Sna*BI sites to allow for the insertion of the P2A-GFP fusion gene at this site. pGEM-sat-3'end-mut was used as template for PCR amplification with the primers sat_mut_exp_F_BamHI-SnaBI and sat_mut_exp_R_BamHI-SnaBI, followed by *Dpn*I digest to cleave the methylated sites in the template plasmid, to ensure that only the mutated plasmid would remain. The mutated plasmids with the incorporated restriction sites were transformed in *E. coli* JM109. This resulted in plasmid pGEM-sat-3'end-exp-BamHI-SnaBI-mut. The entire P2A-GFP was amplified from a construct that was available (Stephan, 2005) with primers that added a 5' *Bam*HI and a 3' *Sna*BI restriction site to the PCR product and cloned into pGEM-T-Easy, resulting in pGEM-P2A-GFP. The P2A-GFP insert was cut from pGEM-P2A-GFP with *Bam*HI and *Sna*BI, and cloned into pGEM-sat-3'end-exp-BamHI-SnaBI-mut, digested with the same enzymes. The resulting plasmid, pGEM-sat-3'end-P2A-GFP, contained the satRNA 3' end, with the incorporated P2A-GFP fusion gene directly after the satRNA ORF. From here the pGEM-sat-3'end-P2A-GFP was digested with *Bsp*EI and *Cla*I, and the resulting 1015 bp fragment was cloned into pCB301-2x35S-GFLV-satRNA-35ST, digested with the same enzymes. This resulted in pCB301-GFLV-satRNA-P2A-GFP, which included the entire satRNA sequence, with the P2A-GFP fusion gene inserted directly after the satRNA ORF followed by the 3' UTR of the satRNA.

Similarly, to modify the satRNA clone into a VIGS vector, pGEM-sat-3'end-mut was used to insert the *Bam*HI and *Sna*BI restriction sites directly after the stop codon of the satRNA ORF, by PCR using primers sat_mut_sil_F_BamHI-SnaBI. The amplification was followed by a *Dpn*I digest and transformation of the PCR products. This resulted in plasmid pGEM-sat-3'end-sil-BamHI-SnaBI, and this plasmid was used as the vector to clone in the three different lengths of GFP (GFP-100, GFP-150 and GFP-200). The resulting plasmids (pGEM-sat-si-GFP100, pGEM-sat-si-GFP150 and pGEM-sat-si-GFP200) were used to excise the fragments that consisted of the satRNA 3' end and partial GFP sequences. These fragments were cloned into the pCB301-2x35S-GFLV-satRNA-35ST, by using the flanking *Bsp*EI and *Cla*I sites. This resulted in three constructs, pCB301-GFLV-satRNA-sil-GFP100, pCB301-GFLV-satRNA-sil-GFP150 and pCB301-GFLV-satRNA-sil-GFP200, which included the entire satRNA sequence, with the partial GFP sequences inserted directly after the satRNA ORF stop codon followed by the 3' UTR and poly (A) tail of the satRNA.

5.3.4 ArMV-NW full-length clones as positive controls

Full-length infectious cDNA clones of ArMV-NW1 RNA1, RNA2 and satRNA were provided by Dr T Wetzel (AlPlanta Institute for Plant Research, Germany). All three constructs were already cloned into the pCASS2 vector (named pCASS2-ArMV-NW1, pCASS2-ArMV-NW2 and pCASS2-ArMV-NWsat) and were infectious in *C. quinoa* by mechanical inoculation. Therefore, to be used in this study as positive controls, these constructs had to be cloned into pCB301 to be comparable to the GFLV-SAPSC3 full-length clones.

Similarly to the pCB301-2x35S-GFLV-RNA1-35ST, the cloning of the full-length ArMV-NW RNA1, together with the double enhanced 35S CaMV promoter and termination sequence, two separate fragments were amplified from pCASS2-ArMV-NW1 as template. Primers that were used for cloning are listed in Table 5.1. For PCR 1, primers pCASS2_e35S_F_ClaI and ArMV_RNA1_PCR1_R_SalI were used to produce a 4101 bp fragment, which included the double 35S promoter and the first 3347 bp of ArMV-

NW RNA1. The purified PCR product was cloned into pGEM-T-Easy (Promega) to produce pGEM-2x35S-ArMV-RNA1-PCR1. This plasmid was digested with *Cla*I and *Sal*I (Fermentas) and ligated into pCB301 digested with the same restriction enzymes. This produced pCB301-2x35S-ArMV-RNA1-PCR1. For the second PCR, primers ArMV_RNA1_PCR2_BsrGI_F and pCASS2_35ST_R_SalI (ArMV) were used to amplify a fragment of 4394 bp, that included the remaining 4015 bp of the ArMV-NW RNA1 and the 35S termination sequence, from pCASS2-ArMV-NW1 as template. The resulting purified PCR product was cloned into pGEM-T-Easy (Promega), named pGEM-ArMV-RNA1-35ST-PCR2. This plasmid was digested with *Bsr*GI and *Sal*I (Fermentas) and ligated into pCB301-2x35S-ArMV-RNA1-PCR1 that was digested with the same restriction enzymes. This resulted in pCB301-2x35S-ArMV-RNA1-35ST. Both pCASS2-ArMV-NW2 and pCASS2-ArMV-NWsat were digested with *Hind*III and *Pvu*II (Fermentas), and the resulting fragments (5039 bp and 2343 bp, respectively) were cloned into pCB301 digested with *Hind*III and *Sma*I (Fermentas). This produced the two full-length clones of pCB301-2x35S-ArMV-RNA2-35ST and pCB301-2x35S-ArMV-satRNA-35ST. Therefore, the constructs pCB301-2x35S-ArMV-RNA1-35ST, pCB301-2x35S-ArMV-RNA2-35ST and pCB301-2x35S-ArMV-satRNA-35ST all had the full-length viral sequence cloned directly downstream enhanced 35S CaMV promoter and upstream from the 35S termination sequence, between the T-DNA Left and Right borders.

5.3.5 Electroporation and agro-infiltration

All full-length constructs (GFLV and ArMV) were electroporated into electrocompetent *A. tumefaciens* C58CI cells, containing the helper plasmid pCH32 to enhance T-DNA transfer (Hamilton et al., 1996). Settings were as follows: capacitance at 25 µF, resistance at 200 Ω and voltage at 1.5 kV. Transformants were grown on Kanamycin (Kan) and Tetracyclin (Tet) LB agar plates for 48 hours at 28°C.

5.3.6 Agro-infiltration of *N. benthamiana* plants

Agro-infiltration was performed as described by Bazzini et al. (2007). One fresh colony of plated *A. tumefaciens* strain C58CI cells (with helper plasmid pCH32) carrying the full-length GFLV and ArMV constructs was used to inoculate 40 ml of LB containing Kan and Tet and grown overnight at 28°C with shaking at 200 rpm. The bacteria were centrifuged at 4000 rpm for 5 min, and the pellet was resuspended in infiltration buffer (10 mM $MgCl_2$, 10 mM MES and 100 μM acetosyringone). The bacterial solution concentration was measured at OD 600nm and adjusted to the optic density of 0.5 cells/ml. Equal volumes of the cultures carrying pCB301-2x35S-GFLVRNA1-35ST (six final constructs) and pCB301-2x35S-GFLVRNA2-35ST (one final construct) were mixed for GFLV. Therefore, for GFLV infiltrations, six bacterial mixes were made, each one representing one pCB301-2x35S-GFLVRNA1-35ST construct from the population cloning strategy, mixed with an equal volume of pCB301-2x35S-GFLVRNA2-35ST.

For the ArMV infiltrations two bacterial mixes were made; equal volumes of pCB301-2x35S-ArMV-RNA1-35ST and pCB301-2x35S-ArMV-RNA2-35ST were mixed to represent ArMV alone infiltration, the second mix was made up with pCB301-2x35S-ArMV-RNA1-35ST, pCB301-2x35S-ArMV-RNA2-35ST and pCB301-2x35S-ArMV-satRNA-35ST to represent ArMV and satRNA infiltration. See Table 5.2 for the construct combinations used and experimental design

Table 5.2: Experimental design for the agro-infiltration of GFLV and ArMV full-length cDNA clones in *N. benthamiana.* Two plants were infiltrated per experiment, and repeated three times

Experiment	Virus	pCB full-length clones/controls
1	GFLV	RNA1 construct 1 + RNA2
2		RNA1 construct 2 + RNA2
3		RNA1 construct 3 + RNA2
4		RNA1 construct 4 + RNA2
5		RNA1 construct 5 + RNA2
6		RNA1 construct 6 + RNA2
7		GFLV-SAPCS3 plant sap
8	ArMV	RNA1 + RNA2
9		RNA1 + RNA2 + satRNA
10		ArMV-NW plant sap
11	No virus	Empty pCB301 vector
12	No virus	Infiltration buffer alone
13	No virus	Healthy control

The bacterial suspensions were left at room temperature for at least 3 hours. The bacterial suspensions were drawn up in a single needle-less syringe for each bacterial mix. For infiltration into *N. benthamiana*, the tip of the syringe was pressed against the underside of a leaf while simultaneously applying gentle pressure from the other side of the leaf. For each experiment, two young *N. benthamiana* plants (named a and b) were infiltrated, using two fully developed upper leaves. Entire leaves were infiltrated. For negative controls, healthy *N. benthamiana* plants were used and other *N. benthamiana* plants were infiltrated with either empty pCB301 vector or infiltration buffer alone. For virus transmission and DAS-ELISA controls, plant sap derived from GFLV-SAPCS3 or ArMV-NW infected plants were used for mechanical transmission to new plants. For each experiment, two plants were infiltrated/ inoculated, and all the experiments were repeated three times.

5.3.7 Detection of GFLV and ArMV in infiltrated *N. benthamiana*

All infiltrated plants were screened for virus infection derived from the full-length cDNA clones by performing either GFLV- or ArMV-specific DAS-ELISA (Bioreba) on newly formed systemically-infected leaves 10-14 days post inoculation (dpi), according to the manufacturer's instructions. GFLV and ArMV positive controls supplied with the DAS-ELISA kits were also included.

For the GFLV-infiltrated plants, RT-PCR was performed on infiltrated and systemically-infected leaves in order to detect RNA transcripts derived from the full-length cDNA clones. For this, total RNA was extracted from infiltrated and systemically-infected leaves using the SV Total RNA isolation system (Promega), according to the manufacturer's instructions. The one-step RT-PCR protocol described in section 3.3.1 was used for GFLV detection, using primers GFLV1_02b_2772_F and GFLV1_03b_3521_R.

5.3.8 Sequencing of the GFLV-SAPCS3 full-length cDNA clones

Plasmid DNA of both pCB301-2x35S-GFLVRNA1-35ST and pCB301-2x35S-GFLVRNA2-35ST were sequenced. Primers used in Chapter 3 (Table 3.1) were used to sequence the entire RNA1 and RNA2. New primers were designed to sequence the regions flanking the viral sequences; *i.e.* the pCASS2 double enhanced 35S promoter and 35S termination sequence, as well as the T-DNA Left and Right borders of pCB301. The sequences generated were assembled using CLC Main Workbench 6.5 (CLC Bio).

5.4 Results and Discussion

5.4.1 Construction of full-length cDNA clones based on South African GFLV isolates

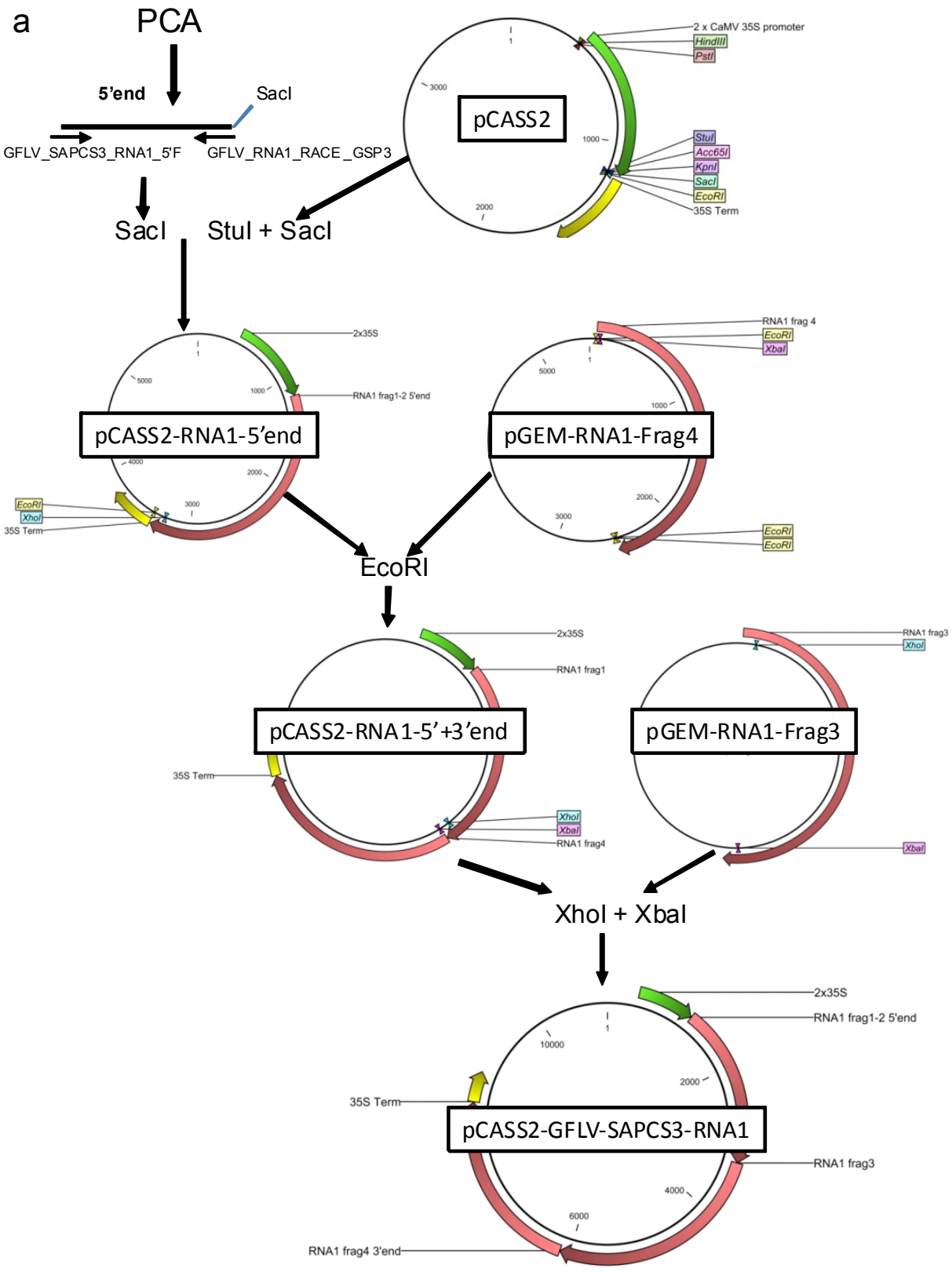
Several vectors and cloning strategies were assessed for the cloning of full-length GFLV viral sequences under the control of a CaMV 35S promoter and then into a binary vector. In the meantime, existing full-length infectious cDNA clones based on ArMV-NW were made available to be used as positive controls in this study (Courtesy of Dr T Wetzel, AlPlanta Institute for Plant Research, Germany). These full-length ArMV-NW RNA1, RNA2 and satRNA constructs were previously cloned in the pCASS2 vector backbone (Shi et al., 1997) and it was therefore decided to clone the GFLV full-length viral sequences (genome and satRNA) in the same backbone for comparison purposes.

Overlapping PCR fragments of GFLV-SAPCS3 RNA1 and RNA2 and GFLV-SACH44 satRNA, with added 3' end poly (A) tails, were cloned in pCASS2 vector (Shi et al., 1997) directly downstream the double enhanced 35S promoter (with no extra non-viral nucleotides between the promoter and the viral 5' UTR start site) using the *Stu*I site, and upstream of the 35S termination sequence. The GFLV pCASS2 full-length clones were repeatedly confirmed by different control restriction digests before carrying on with subsequent cloning into pCB301 (Xiang et al., 1999). The entire region that included the double enhanced CaMV 35S promoter, full-length viral sequences and termination sequences of the GFLV pCASS2 full-length clones were cloned between the T-DNA Left and Right borders of pCB301 (Xiang et al., 1999). Initial design of cloning strategies would have used different binary vectors but in the end pCB301 was chosen for the GFLV and ArMV full-length clones, since pCB301 is less than half the size (3.5kb) of the other binary vectors that are often used, *e.g.* the pBIN19 backbone. The use of fairly large binary vectors is known to make *in vitro* manipulation inconvenient (Xiang et al., 1999), and therefore using a smaller binary vector seemed to be more favorable.

5.4.1.1 Construction of full-length cDNA clones of GFLV-SAPCS3 RNA1

The cloning strategy used and described in section 5.3.3.1 was successful in obtaining full-length cDNA clones of GFLV-SAPCS3 RNA1 (pCASS2-GFLV-RNA1 and pCB301-2x35S-GFLVRNA1-35ST) and the cloning steps followed are depicted in Figure 5.2 a and b. To bring the entire GFLV-SAPCS3 RNA under the control of a CaMV 35S promoter (Figure 5.2 a), fragments 1 and 2 were joined together to make up the 5' end part of GFLV-SACH44 RNA1, and were cloned into pCASS2. Ten correct plasmids were mixed in equal amounts to make up a population mix of pCASS2-RNA1-5'end. A population mix (10 plasmids) of pGEM-RNA1-frag4 (including the poly (A) tail) was digested and cloned into the pCASS2-RNA1-5'end population mix that yielded pCASS2-RNA1-5'+3'end. Nine correct pCASS2-RNA1-5'+3'end plasmids were mixed in equal amounts to serve as the vector for the next cloning step, which involved the cloning of the insert from a population mix (10 plasmids) of pGEM-RNA1-frag3. Ten full-length GFLV-SAPCS3 RNA1 constructs under the control of a double enhanced CaMV 35S promoter were obtained. At each sub-cloning step the plasmids were thoroughly inspected by colony PCR and/or several control restriction digests. The final pCASS2-GFLV-RNA1 plasmids were not sequenced, as there were ten to be used in total, after using the population cloning strategy. For cloning of the GFLV-SAPCS3 RNA1 into pCB301, two large fragments were amplified from the pCASS2 full-length constructs, as appropriate restriction sites that were common between these constructs and the pCB301 could not be found. Appropriate restriction enzyme sites were added to the PCR products for cloning purposes. It cannot be excluded that mutations were incorporated during the amplification steps, even though high fidelity proof reading enzymes were used, such as Ex Taq (Takara) and Phusion Polymerase (Thermo Scientific). A population mix (10 plasmids) of pCASS2-GFLV-RNA1 was used as template for amplification of the two fragments. The first amplified fragment, PCR1, was cloned into pGEM-T-Easy and seven plasmids of pGEM-2x35S-RNA1-PCR1 were mixed into a population mix. The insert of the population mix was cloned into pCB301 vector and yielded pCB301-2x35S-RNA1-PCR1. Seven correct plasmids were obtained

and mixed in a population mix. Eight pGEM-T-Easy plasmids containing the second PCR fragment, PCR2, were made up to a population mix of pGEM-RNA1-35ST-PCR2, and the insert was cloned into the population mix of pCB301-2x35S-RNA1-PCR1. This resulted in pCB301-2x35S-GFLVRNA1-35ST, with the double enhanced CaMV 35S promoter, GFLV-SAPCS3 RNA1 and 35S termination sequence, cloned between the T-DNA Left and Right borders of pCB301 binary vector. Six full-length constructs from the population cloning strategy were obtained.



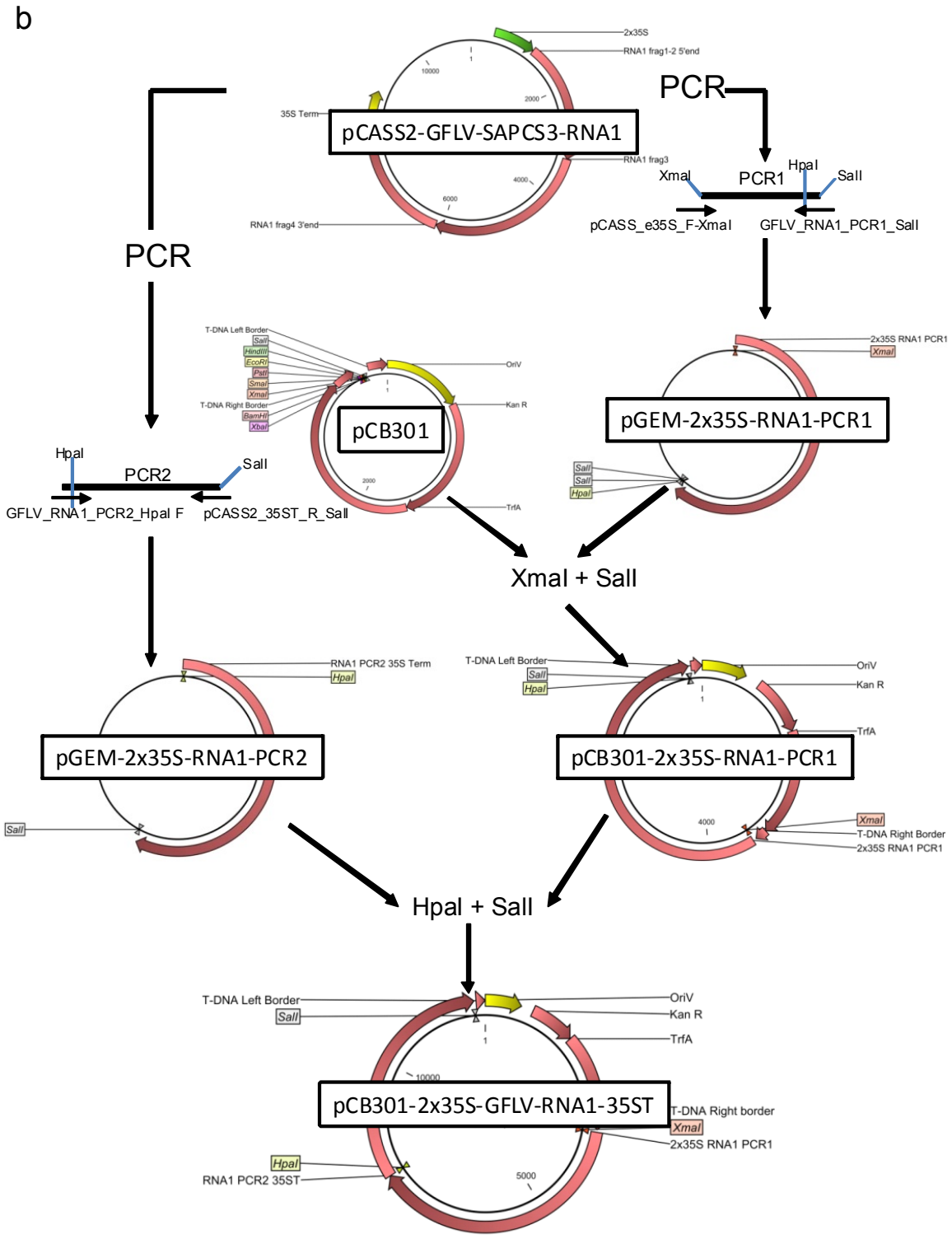


Figure 5.2: Steps that were followed to assemble the full-length cDNA clone of GFLV-SAPCS3 RNA1 in the (a) pCASS2 35S vector and (b) in the pCB301 binary vector. All plasmids and restriction enzymes involved are shown in the diagram. See text for full description. Primers used are represented by the small arrowheads and are labeled with the appropriate primer names; blue thin lines represent the approximate restriction enzyme positions (either internally or as flaps); thick arrowheads represent the sequential steps that were followed.

5.4.1.2 Construction of full-length cDNA clones of GFLV-SAPCS3 RNA2

The cloning strategy used and described in section 5.3.3.2 was successful in obtaining full-length cDNA clones of GFLV-SAPCS3 RNA2 (pCASS2-GFLV-RNA2 and pCB301-2x35S-GFLVRNA2-35ST) and are depicted in Figure 5.3. The 3' end including an poly (A) tail of GFLV-SAPCS3 RNA2 was first cloned in a TA vector and the resulting pCASS2-RNA2-3'end plasmids were verified by sequencing. Only one of the plasmids was used for subsequent cloning steps, as sequencing confirmed it was in the correct orientation. The remainder of the GFLV-SAPCS3 RNA2 was amplified as one PCR product and it was cloned into the pCASS2-RNA2-3'end vector. This resulted in pCASS2-GFLV-RNA2 was verified by control restriction digests and one plasmid was sequenced. Sequencing revealed that the entire RNA2 was intact, flanked by the promoter regions and termination sites. The RNA2 insert was compared to the GFLV-SAPCS3 RNA2 full-length sequence of Chapter 3, and it was found that only one silent mutation had occurred during cloning. Therefore, with sequencing between sub-cloning steps, a population mix of pCASS2-GFLV-RNA2 was not used and only the sequenced plasmid was used for further cloning into pCB301. For cloning of the GFLV-SAPCS3 RNA2 into pCB301, the entire region that included the RNA2 and the promoter regions were amplified as one PCR fragment, flanked by appropriate restriction enzyme sites and cloned directly into pCB301, resulting in pCB301-2x35S-GFLVRNA2-35ST. One plasmid was obtained and this full-length pCB301-2x35S-GFLVRNA2-35ST plasmid was to be mixed with six plasmids of pCB301-2x35S-GFLVRNA1-35ST to be used for agro-infiltration.

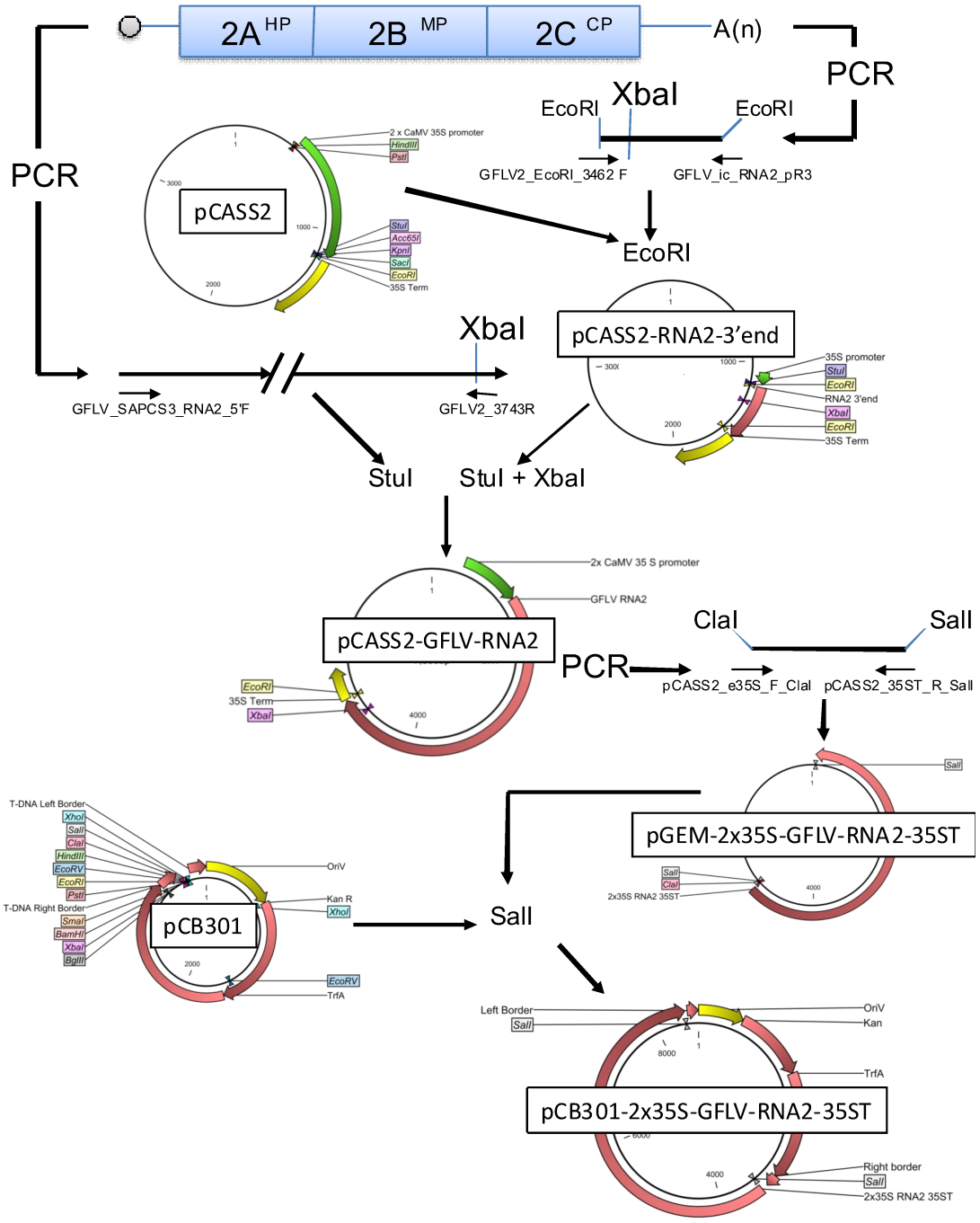


Figure 5.3: Steps that were followed to assemble the full-length cDNA clone of GFLV-SAPCS3 RNA2 in the pCASS2 35S vector and finally in the pCB301 binary vector. All plasmids and restriction enzymes involved are shown in the diagram. See text for full description. Primers used are represented by the small arrowheads and are labeled with the appropriate primer names; blue thin lines represent the approximate restriction enzyme positions (either internally or as flaps); thick arrowheads represent the sequential steps that were followed.

5.4.1.3 The modification of full-length GFLV-SACH44 satRNA infectious cDNA clone into an expression and silencing vector

The cloning strategy described in section 5.3.3.3 that yielded the full-length pCB301-2x35S-GFLV-satRNA-35ST is depicted in Figure 5.4. The full length infectious cDNA L140-GFLV-satRNAfl clone (Chapter 4) was used as template for PCR to clone the satRNA insert into pCASS2, so that all the constructed clones in this study have the exact same backbones for comparison reasons. Once pCASS2-GFLV-satRNA was obtained, the entire region that included the satRNA sequence, the promoter and termination sequence were cloned into pCB301 in one ligation step. The resulting full-length satRNA clone, pCB301-2x35S-GFLV-satRNA-35ST, contained a double enhanced 35S CaMV promoter, full-length GFLV-SACH44 satRNA and a 35S termination sequence between the T-DNA Left and Right Borders. Control restriction digests were used to verify the final plasmids.

The full-length satRNA clones were modified to be used as vectors for expression and/or silencing of foreign genes. To be modified into an expression vector, the P2A-GFP peptide was fused to the ORF of the satRNA. The P2A-GFP peptide, as opposed to just using GFVP alone, was used in this vector, as P2A is a self-cleaving peptide that allows for the expression of multiple (in this case two) proteins from a single ORF. The P2A peptide is cloned in between genes to allow for efficient production of protein products within one vector through a cleavage event within the P2A peptide sequence (Szymczak-Workman et al., 2012). Successful GFP expression would result in fluorescence under UV light in agro-infiltrated *N. benthamiana*. To use the satRNA as a silencing vector, partial GFP sequences of different lengths (100 bp, 150 bp and 200 bp) were inserted directly downstream the satRNA ORF. Silencing will be observed as a loss of GFP phenotype in transgenic 16c *N. benthamiana*. The cloning steps that were followed in succesfull cloning of these modified satRNA clones are depicted in Figure 5.5.

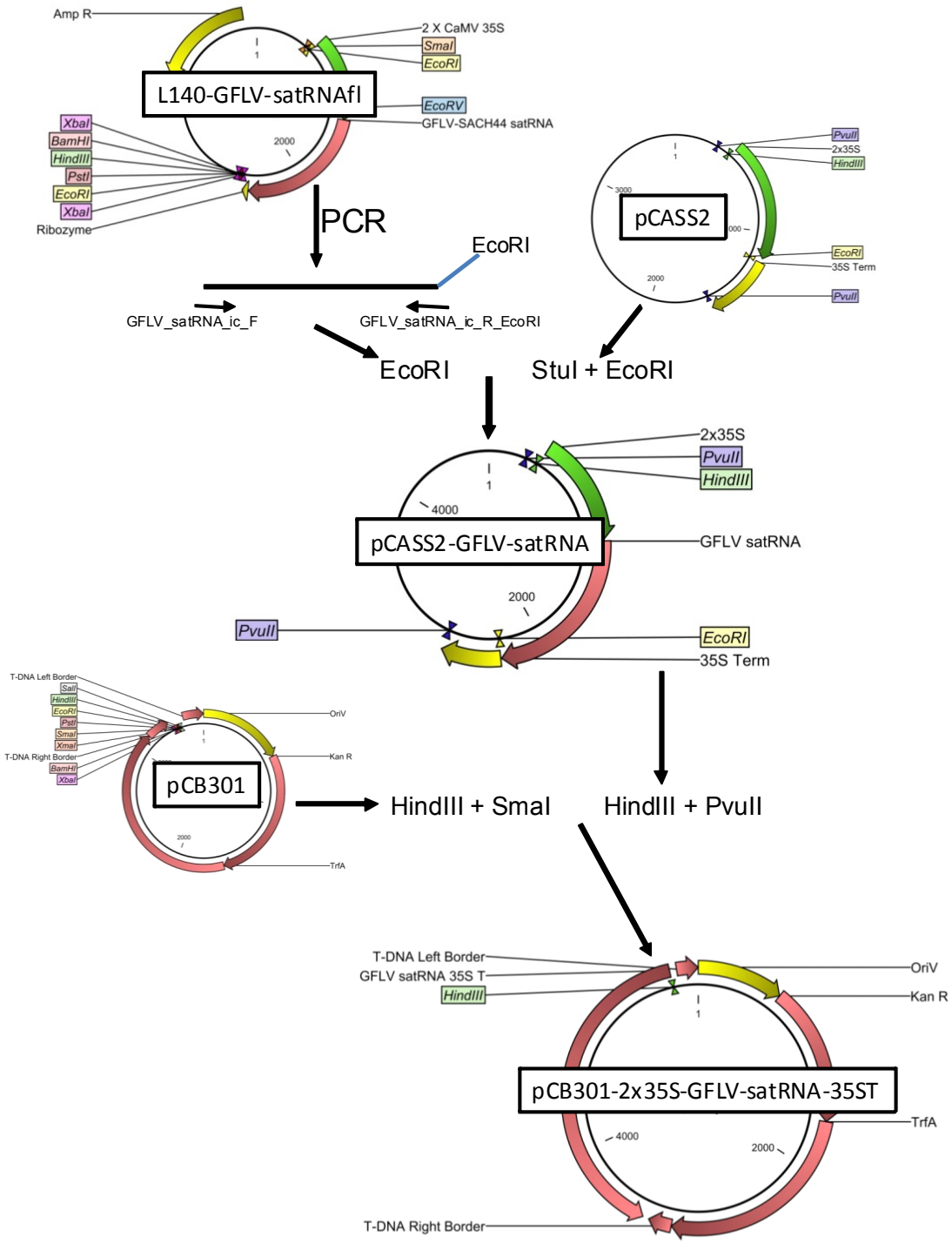


Figure 5.4: The steps that were followed to clone the GFLV-SACH44 satRNA into the pCB301 binary vector. L140-GFLV-satRNAfl clone12 (Chapter 4) was used as template to construct another full-length cDNA clone that would use the same vector backbones used for the GFLV-SAPCS3 RNA1 and RNA2 full-length clones. All plasmids and restriction enzymes involved are shown in the diagram. See text for full description. Primers used are represented by the small arrowheads and are labeled with the appropriate primer names; blue thin lines represent the approximate restriction enzyme positions (either internally or as flaps); thick arrowheads represent the sequential steps that were followed.

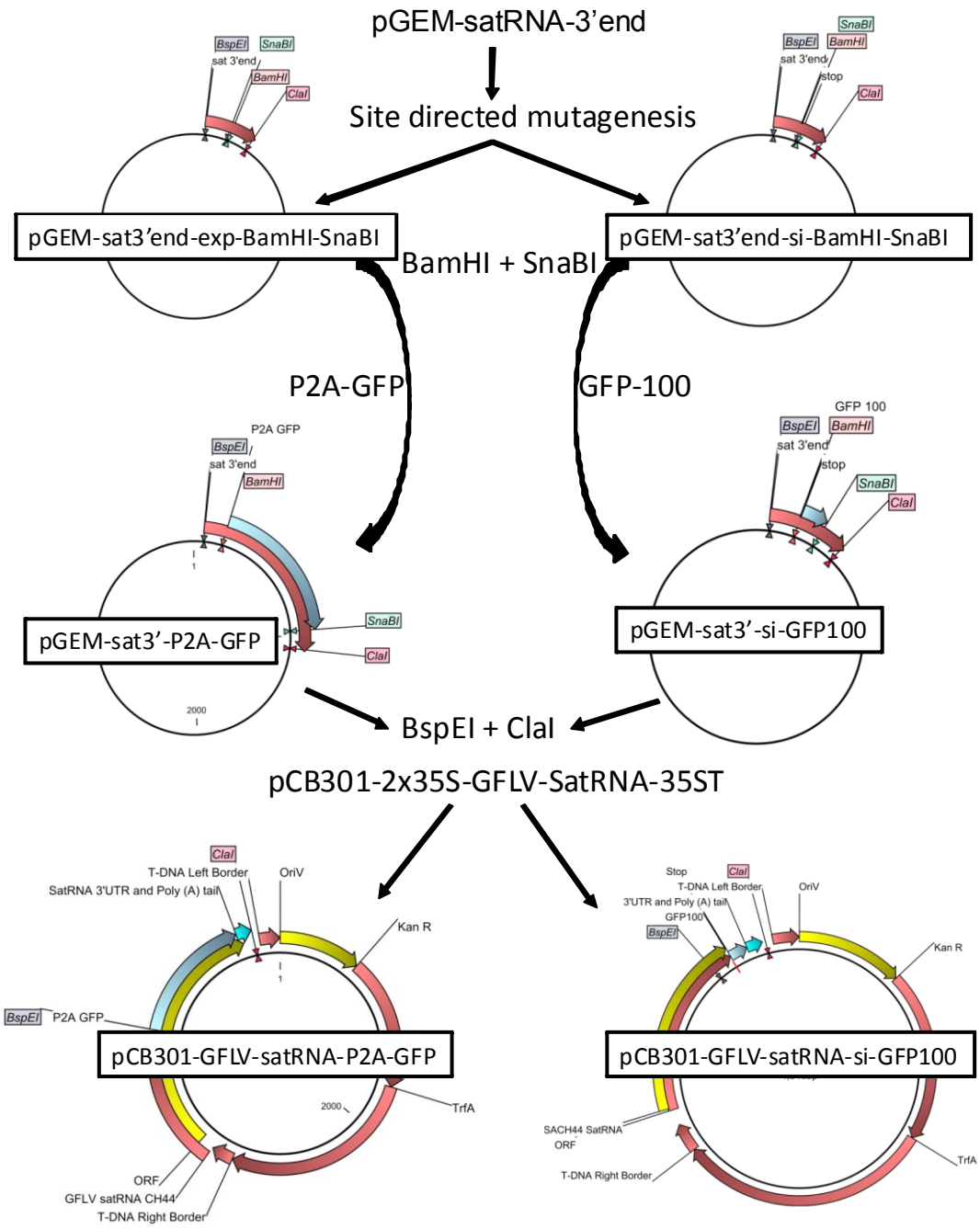


Figure 5.5: The modification of full-length pCB301-2x35S-GFLV-satRNA-35ST clone into a GFP-expression vector (left) and a GFP silencing vector (right). Restriction sites *Bam*HI and *Sna*BI were incorporated at the end of the satRNA ORF to allow for insertion of either full-length or partial GFP sequences. See text for details.

5.4.2 ArMV-NW full-length clones as positive controls

The full-length pCASS2-ArMV-NW (RNA1, RNA2 and satRNA) were successfully cloned into binary vector pCB301. Similar to the cloning strategy that was used to construct pCB301-2x35S-GFLV-RNA1-35ST (Figure 5.2 b) no common restriction enzyme sites between pCB301 and pCASS2-ArMV-NW-RNA1 were present and two fragments were amplified from the full-length pCASS2 construct and sequentially ligated into pC301, and resulted in pCB301-2x35S-ArMV-RNA1-35ST. Appropriate common restriction sites between the ArMV-NW RNA2 and ArMV-NW satRNA pCASS2 constructs and pCB301 were however available and therefore the fragments were cloned into pCB301 by a single ligation reaction.

5.4.3 Transformation, electroporation and control restriction digests of final pCB301 constructs

All full-length pCB301 clones for GFLV and ArMV were transformed in *E. coli* JM109 and thereafter electroporated in *A. tumefaciens* C58CI (with helperplasmid pCH32). Several control restriction digests including different restriction enzymes were performed on all pCB301 full-length clones (Figure 5.6), isolated from both bacterial strains for verification. It is however important to note that restriction digest of these final plasmids would only indicate whether all expected fragments were present and if the inserts were cloned in the correct orientation. Restriction digests would however not show whether point mutations, such as insertions or deletions, occurred during the various cloning steps.

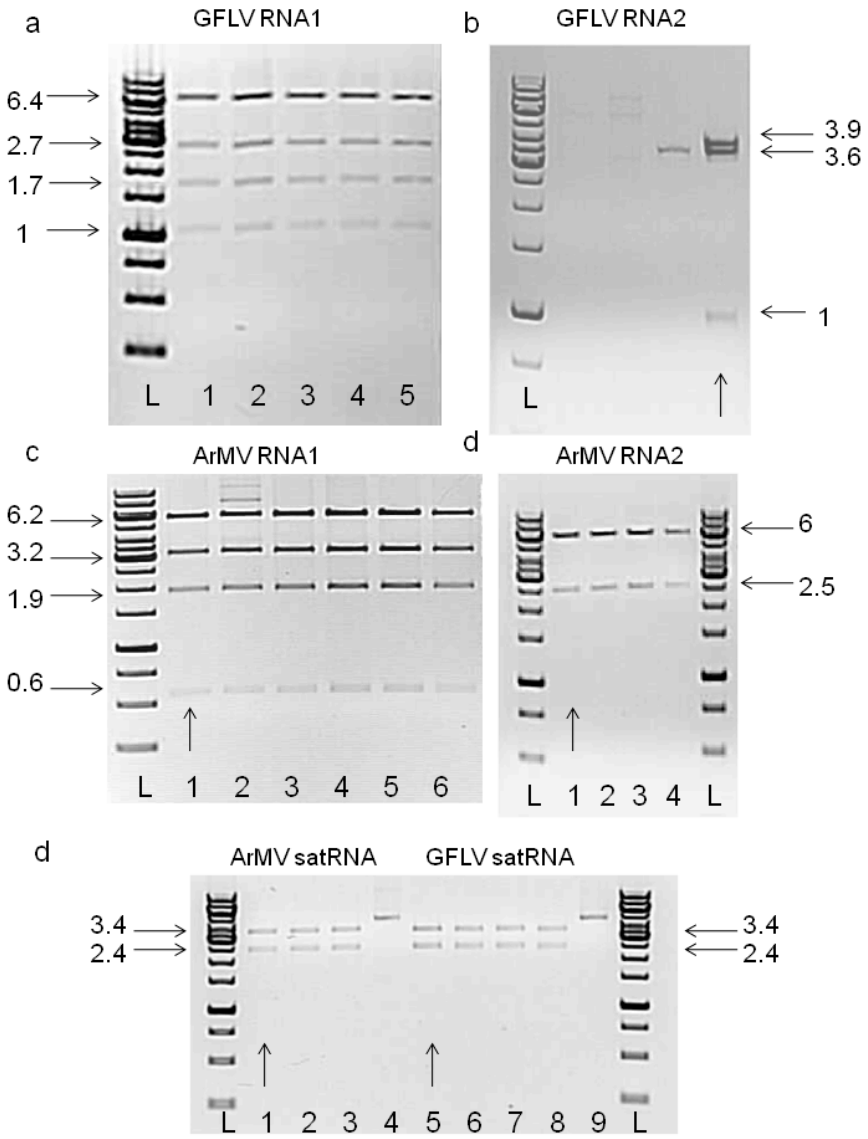


Figure 5.6: Expected fragment sizes after control restriction digests on pCB301 full-length GFLV and ArMV clones. Different restriction enzymes were used. (a) Restriction digests using *Hind*III on five plasmids of pCB301-2x35S-GFLVRNA1-35ST. Six plasmids in total were used for *A. tumefaciens* electroporation. (b) Restriction digests of pCB301-2x35S-GFLVRNA2-35ST using *Hind*III. Only one plasmid showed the correct fragments (indicated by the vertical arrow) and was used for *A. tumefaciens* electroporation. (c) Restriction digests of pCB301-2x35S-ArMV-RNA1-35ST using *Hind*III. (d) Double restriction digest of pCB301-2x35S-ArMV-RNA2-35ST using *Kpn*I and *Avr*II and (e) *Eco*RV restriction digests of both pCB301-2x35S-ArMV-satRNA-35ST and pCB301-2x35S-GFLV-satRNA-35ST. Horizontal arrows represent the correct fragments that were produced after the digests. Sizes of the fragments are indicated (kb). Vertical arrows represent the plasmids that were used for *A. tumefaciens* electroporation and *N. benthamiana* agro-infiltrations. (L) GeneRuler 1kb ladder (Fermentas).

5.4.4 Detection of GFLV and ArMV in infiltrated *N. benthamiana*

The agro-infiltration experiments of the full-length GFLV and ArMV clones were repeated three times. With the first round of infiltrations, two plants (8-10 leaf stage *N. benthamiana*) per experiment (Table 5.2) were agro-infiltrated. At 10 dpi GFLV and ArMV DAS-ELISA were performed on newly formed systemically-infected leaves, according to the manufacturer's instructions. Each sample was analyzed in two wells of the ELISA plate. The ELISA plate was inspected visually after two hours of adding the para-nitrophenyl-phospate to the wells. None of the plants that were infiltrated with either GFLV or ArMV full-length clones showed any colour development. The only samples that were positive were from the plants that had either GFLV or ArMV virus transmitted by mechanical inoculation with infected plant sap, and the wells that had the GFLV and ArMV positive controls that were supplied with the DAS-ELISA kit (Bioreba). The reasons for the negative ELISA results could have been one of the following: an infiltration error (incorrect buffers, infiltration method, etc), the plants may have been too old for infiltrations, the light intensity or the temperatures in the greenhouse might have not been optimal, and finally, the full-length cDNA clones were not infectious. The ArMV assembled clones were also not detected in the systemic leaves even though they were supposed to be the positive controls. A second agro-infiltration was performed by infiltrating younger *N. benthamiana* plants (6-8 leafstage). However this infiltration attempt yielded the same DAS-ELISA results described above.

In a recent study it was demonstrated that light represses the expression of flagella genes in *A. tumefaciens* that are important in the movement of the bacterium. Also, the reduction in flagella numbers caused by light intensity correlated with a reduction of *A. tumefaciens* infectivity (Oberpichler et al., 2008). To eliminate the strong natural light intensity experienced during the South African summer months, frames covered with shade nets (50%) were erected in the greenhouse around the plants. The plants were agro-infiltrated for a third time and the shades remained there for the duration of the experiment. After 12 dpi the plants were evaluated by GFLV- and ArMV-specific DAS-ELISA and the plates were investigated visually and spectrophotometrically measured

(xMark Microplate Absorbance Spectrophotometer, Biorad). Two plants per experiment were infiltrated, and duplicate ELISA readings were analyzed for each plant. The ELISA result readings are listed in Table 5.3. Unfortunately the same results were obtained as the previous two agro-infiltration attempts, even though the experiments were conducted under a shade net. Therefore the light intensity in the greenhouse could not be attributed to the negative results. The ArMV assembled clones was not detected in the systemic leaves in any of the agro-infiltration attempts. The greenhouse is a temperature controlled environment with temperatures between 18°C and 28°C. However, all agro-infiltrations were done between December and February 2013, the period when the Stellenbosch (Western Cape, South Africa) experiences the harshest summer temperatures (27 °C to 37 °C). Even with the automated air-conditioner facilities fitted in the greenhouse, temperature spikes were observed and these high temperatures may have affected the infectivity of the ArMV-NW clones. Also, similar amplification techniques that were used to clone the GFLV fragments into the binary vector were employed for the ArMV full-length clones, and these extra *in vitro* amplifications may have affected the infectivity of the new ArMV full-length clones. Further investigation (e.g sequencing of the constructs) into why the ArMV clones did not replicate was not explored.

Table 5.3: The ELISA absorbance values from the GFLV and ArMV infiltrated *N. benthamiana* plants after the third attempt. Absorbance values were measured at 405 nm and the average of the two readings are listed. Positive ELISA samples are listed in bold.

Experiment	Infiltrated/transmitted/control	Plant (x2)	Absorbance (405 nm)
1	GFLV constructs	a	0.094
1	GFLV constructs	b	0.112
2	GFLV constructs	a	0.097
2	GFLV constructs	b	0.1
3	GFLV constructs	a	0.088
3	GFLV constructs	b	0.102
4	GFLV constructs	a	0.099
4	GFLV constructs	b	0.096
5	GFLV constructs	a	0.091
5	GFLV constructs	b	0.094
6	GFLV constructs	a	0.092
6	GFLV constructs	b	0.099
7	**GFLV virus**	a	**2.528**
7	**GFLV virus**	b	**3.522**
8	ArMV constructs	a	0.179
8	ArMV constructs	b	0.159
9	ArMV and satRNA constructs	a	0.163
9	ArMV and satRNA constructs	b	0.152
10	**ArMV virus**	a	**0.555**
10	**ArMV virus**	b	**0.446**
11	Empty pCB301	a	0.085
11	Empty pCB301	b	0.081
12	Infiltration buffer	a	0.096
12	Infiltration buffer	b	0.161
13	Healthy	a	0.093
13	Healthy	b	0.146
14	**GFLV positive (kit)**	a	**1.107**
14	**GFLV positive (kit)**	b	**1.186**
15	**ArMV positive (kit)**	a	**0.975**
15	**ArMV positive (kit)**	b	**1.013**

After three failed attempts in detecting virus in the infiltrated *N. benthamiana* leaves it seemed likely that the assembled GFLV clones were not replicating efficiently, or not moving systemically throughout the infected plants. It was also possible that the full-length clones were expressing the virus (coat protein), but at titres too low for detection by ELISA. We therefore investigated by RT-PCR whether RNA transcripts produced from the GFLV clones could be detected in the systemic leaves. Total RNA extract from both infiltrated leaves (plants a and b pooled) and systemically-infected leaves (separate plant a and b samples), from two plants (a and b) per experiment, were subjected to the diagnostic RT-PCR. RT-PCR products (760 bp) were detected in the samples that represented the infiltrated leaves and the GFLV positive control, but no PCR products could be detected in the systemic leaves of the GFLV infiltrated plants (Fig 5.7 a). Therefore, no systemic RNA transcripts derived from GFLV full-length clones were found.

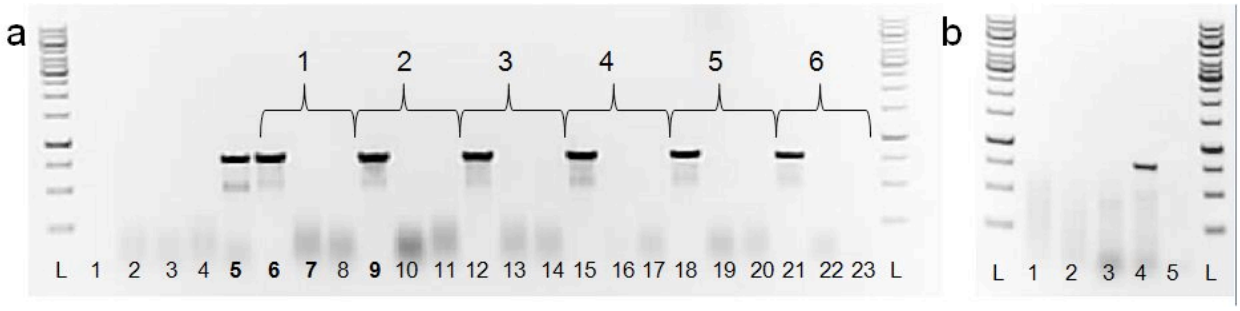


Figure 5.7: Results from RT-PCR to determine if RNA transcripts derived from the GFLV full-length clones were present in the infiltrated and systemic leaves. (a) RT-PCR on total RNA from *N. benthamiana* plants that were agro-infiltrated with GFLV full-length cDNA clones. The horizontal brackets represent each experiment, *i.e.* one of six pCB301-2x35S-GFLVRNA1-35ST constructs from the population cloning strategy, co-infiltrated with a single pCB301-2x35S-GFLVRNA2-35ST construct. For each experiment the infiltrated leaf areas (two plants pooled, in the first lane) and the systemic leaves (two separate samples, in the following two lanes) were screened for GFLV derived RNA by RT-PCR (760 bp expected fragment). (L) GeneRuler 1kb ladder (Fermentas); (1) no template control; (2) healthy control; (3) infiltration buffer control; (4) empty vector control; (5) GFLV positive control; (6, 9, 12, 15, 18, 21) infiltrated leaves; (7, 8, 10, 11, 13, 14, 16, 17, 19, 20, 22, 23) systemic leaves. (b) RT-PCR on total RNA extracted that were additionally treated with DNase. (1, 2) RT-PCR on infiltrated leaves from experiment 1 and 2 (lanes 6 and 9 from Fig 5.7 a) after an additional DNase treatment; (3) RT-PCR on systemic leaves

from experiment 1 (lane 7 from Fig 5.7 a) after an additional DNase treatment (4) RT-PCR on systemic leaves of the GFLV positive control plant, after an additional DNase treatment and (5) no template control.

However, the PCR products obtained from the infiltrated areas may have derived from plasmid DNA contamination due to the agro-infiltration procedure. The PCR products may have also been from RNA transcripts that were produced locally in the infiltrated leaves, but are unable to move systemically. To investigate this, total RNA extracted from the infiltrated areas of plant 1 and 2, as well as that from the GFLV positive control and the systemically-infected leaves of plant 1 a, were subjected to and extra DNase step (even though the SV Total RNA isolation system (Promega) already included a DNase step). This was done to possibly rule out that plasmid contamination produced the PCR products. The same one-step RT-PCR protocol was used on the additionally DNase treated total RNAs. Only the GFLV positive plant produced the expected 760 bp band (Fig 5.7 b), suggesting systemic spread of RNA from the natural virus. However, none of the infiltrated or systemically-infected leaves produced any PCR products. This suggests that the initial PCR products found in the infiltrated leaf samples (Fig 5.7 a) was from local plasmid DNA contamination resulting from the agro-infiltration procedure, and that no RNA transcripts from the full-length clones were produced either locally or systemically. The only explanation that remained was that none of the assembled clones were replicating. It was therefore decided to sequence the GFLV full-length clones to investigate where the problem may be.

5.4.5 Sequencing of the GFLV full-length cDNA clones

Both pCB301-2x35S-GFLVRNA1-35ST (construct 1 from the population cloning strategy) and pCB301-2x35S-GFLVRNA2-35ST were sequenced to determine if mutations (insertions and/or deletions) were responsible for rendering the GFLV clones not infectious. Primers were designed to sequence both the T-DNA borders of binary vector pCB301, as well as the 35S promoter and termination sequence. Various primers spanning the GFLV-SAPCS3 genome were used to sequence the RNA1 and RNA2 sequences of the assembled clones. In total 17 primers were used to sequence the

region of pCB301-2x35S-GFLVRNA1-35ST that included the T-DNA borders, the 35S promoter and termination sequence and the GFLV RNA1 (8880 bp). For sequencing of the pCB301-2x35S-GFLVRNA2-35ST clone, 11 primers were used that spanned the area that included the T-DNA borders, the 35S promoter and termination sequence (5372 bp). CLC Main Workbench version 6.5 (CLC Bio) was used to create contiguous sequences from the overlapping fragments. To determine at what positions mutations had occurred, the newly generated full-length sequences were compared to the *in silico* map that was created during the cloning strategy process for both RNA1 and RNA2. For both clones, the T-DNA borders and 35S promoter and termination sequences were intact and no mutations had occurred (data not shown). Also, the start of the viral 5' UTRs of RNA1 and RNA2 were inserted directly downstream the transcription start site (at GCC) of the double enhanced CaMV 35S promoter, with no extra non-viral nucleotides inserted, for both clones. The inserted poly(A) tails were present in both clones upstream from the 35S termination sequence.

In total, 36 mutations were detected in the GFLV-RNA1 viral sequence (Table 5.4 and Appendix A). The most significant of the mutations were four deletions that caused four frameshifts (Table 5.4) in the ORF sequence, which in turn caused that the expected single ORF was now divided in 5 ORFs (at positions 244-1707, 1665-2093, 2114-5836, 5974-6690 and 6789-7091). The first deletion (missing A at position 1648) caused the first significant mutation in gene 1B that lead to a frameshift of the ORF. A single frameshift mutation could render a full-length clone not infectious, especially if it occurred at the beginning of the ORF sequence. The other deletion mutations, that occurred in the RNA1, were at positions 2062 (in gene $1B^{Hel}$), 5828 and 6551, (both in the $1E^{Pol}$) by missing G, C and G, respectively, at these positions. The mutations in all four areas were confirmed by 2-3 overlapping sequences and the sequence chromatograms were also visually inspected. The single point mutations that occurred in the non-coding areas, as well as in the ORF sequence before position 1648, are listed in Table 5.4. The deletions that caused the frameshifts are also listed.

Table 5.4: Point mutations in pCB301-2x35S-GFLVRNA1-35ST. The most significant mutations that lead to at least 4 frameshifts are indicated in bold.

RNA1				
Position	GFLV-SAPCS3	Mutation	Change in aa sequence	Notes
140	A	G	-	5' UTR
274-277	TGTT	AACA	CC → NS	
368	C	T	A → V	
660	C	T	silent	
714	C	T	silent	
766	G	A	V → I	
1027	A	G	T → A	
1648	**A**	**-**	**Frameshift**	
2062	**G**	**-**	**Frameshift**	
5828	**C**	**-**	**Frameshift**	
6551	**G**	**-**	**Frameshift**	
7114	G	T	-	3' UTR
7251	T	-	-	3' UTR

The four deletions can be corrected by site-directed mutagenesis. *In silico* analysis revealed that if these deletions were corrected (by inserting an A, G, C and G at positions 1648, 2062, 5828 and 6551, respectively), the premature stop codons will be abolished and a single ORF of 6852 bp would be transcribed, corresponding to the ORF size of GFLV-SAPCS3 RNA1 (Lamprecht et al., 2012). If the deletions were corrected, 15 nonsynonymous mutations and 15 silent mutations would still exist in the GFLV RNA1 clone. Table 1 in the Appendix lists the predicted mutations that would remain in RNA1 if the four deletions were to be corrected. Even if the four deletions were to be corrected, it is at this stage unknown what the effect of the multiple nonsynonymous mutations in RNA1 might have on the infectivity of the GFLV clones. If the GFLV full-length clones are still not able to replicate after the deletions have been corrected, one might have to reconstruct the entire GFLV RNA1 full-length clone, as mapping the effects of each mutation might be laborious and impractical. A new cloning strategy

would have to be devised, one that involves less PCR amplification. It is also recommended to sequence after each sub-cloning step as often as possible.

Sequencing of pCB301-2x35S-GFLVRNA2-35ST revealed that no detrimental mutations had occurred in the single intact ORF (Table 5.5). Only 2 point mutations (at positions 1800 and 2059) lead to two amino acid changes. Fewer mutations were most likely introduced in this segment as less cloning steps, more specifically less PCR amplification steps, were required for assembling the GFLV RNA2 full-length clone. At this stage it be cannot determined what the effect of these two mutations may have on the infectivity of the GFLV RNA2 full-length clone, as it cannot at this stage be evaluated as the GFLV RNA1 full-length clone needs to be infectious for the system to work.

Table 5.5: Point mutations present in pCB301-2x35S-GFLVRNA2-35ST.

RNA2				
Position	GFLV-SAPCS3	Mutation	Change in aa sequence	Notes
127	A	G	-	In 5' UTR
258	T	C	-	In 5' UTR
632	T	C	-	
909	T	C	-	
1392	T	C	-	
1800	T	A	A → P	
1934	T	C	-	
1994	**T**	**C**	**-**	
2034	**C**	**A**	**-**	
2059	**A**	**G**	**E** → G	
2408	**T**	**C**	**-**	
3815	G	A	-	In 3' UTR

The incorporation of the mutations in the in the GFLV full-length clones may have been introduced by a number of ways. The nonsynonymous mutations in RNA1 and RNA2 may have occurred as a result of several serial passages of the natural GFLV-SAPCS3 virus in *C. quinoa* and *N. benthamiana* over a period of two years. In a previous study, the RFLP banding patterns of Californian GFLV isolates were simplified upon passage in *C. quinoa*, and the authors concluded that the genetic diversity among GFLV isolates

diminished upon serial passage in herbaceous hosts (Naraghi-Arani et al., 2001). It is therefore difficult to predict whether these mutations in the assembled GFLV full-length clones occurred during the serial passages in the herbaceous hosts due to host adaptation, or whether these mutations occurred as a result of the cloning strategy that was followed.

Viruses, especially RNA viruses, exist in a plant as a heterogeneous population, as a result of their error prone nature of the RNA dependent RNA polymerase (RdRp). It is therefore possible that the published GFLV-SAPCS3 sequences (Lamprecht et al., 2012) do not necessarily correspond to the infectious sequences. These heterogeneous populations may include lethal mutants in the plant cells that may have been incorporated in the final assembly, as all templates would have equal chance to be amplified by PCR. Furthermore, the proofreading enzymes used to amplify the viral fragments and the intermediate vectors could also have introduced point mutations, even if they are classified as high fidelity enzymes. The chances of erroneous incorporations increase especially with amplifying long viral genomes. However, in the past, PCR was applied successfully to obtain infectious clones (Hayes and Buck, 1990; Viry et al., 1993).

The cloning and propagation of the viral fragments in bacterial plasmids may also pose some complications. The instability of cDNA fragments in *E. coli* may lead to point mutations or deletions as particular viral sequences may be toxic for the bacteria (Boyer and Haenni, 1994) which may affect the ability of the clones to replicate in plants. This was observed with two flaviviruses, Yellow fever virus and Japanese encephalitis virus. Full-length cDNA clones were successfully obtained for these viruses, but multiplication of these clones in bacteria consistently introduced mutations in the viral sequences (Rice et al., 1989; Sumiyoshi et al., 1992).

Due to the inability of the GFLV full-length clones to replicate in *N. benthamiana*, the infectivity of the new pCB301-2x35S-GFLV-satRNA-35ST (derived from the GFLV-SACH44 satRNA full-length clone in Chapter 4) and the modified satRNA constructs

were not assessed. Modifications were applied to the pCB301-2x35S-GFLV-satRNA-35ST clones to possibly convert it into a VIGS and/or expression vectors. Different constructs serving different purposes were constructed. For expression of GFP, the stop codon of the GFLV-SACH44 satRNA clone was replaced with a construct comprising a self cleaving 2A peptide isolated from the foot-and-mouth disease virus (Donnely et al., 2001) fused to GFP. Also, the pCB301-2x35S-GFLV-satRNA-35ST clones were modified to be used as silencing vectors. Partial GFP fragments (100 bp, 150 bp and 200 bp fragments) were cloned directly downstream the GFLV satRNA stop codon. None of these constructs were evaluated *in planta* due to time constraints and the absence of a helper virus infectious clone. However, these chimaeric satRNA constructs are available for future studies.

5.5 Conclusion

To date, infectious clones based on grapevine viruses GFLV-F13 (Viry et al., 1993), ArMV-NW (Dupuis et al., 2009), GVA (Muruganantham et al., 2009), GRSPaV (Meng et al., 2012) and GLRaV-2 (Kurth et al., 2012) are available. However, only GVA- and GLRaV-2-based clones are able to replicate in *V. vinifera* (Muruganantham et al., 2009; Kurth et al., 2012). In this chapter we report the first attempts toward constructing GFLV genomic and satRNA full-length clones.

Full-length cDNA clones based on GFLV-SAPCS3 RNA1 and RNA2 constructed in this chapter were agro-infiltrated in *N. benthamiana*, but after several attempts were shown not to be replicating. It was established that the clones were not infectious in N. benthamiana due to lethal mutations that had occurred in RNA1. The continual serial passages of GFLV-SAPCS3 in herbaceous hosts, the potential viral replication errors caused by the RdRp, the different polymerases used to amplify the viral segments and the propagation of intermediate constructs in *E. coli* hosts (JM109 and JM110) may have all contributed to the multiple mutations that were observed in the GFLV full-length clones, especially in the GFLV-SAPCS3 RNA1 clone. It is clear that the GFLV full-

length clones need to be corrected by site-directed mutagenesis, or by constructing a new GFLV-SAPCS3 RNA1 clone.

The new satRNA clone and modified constructs were based on a fully-functional infectious clone (Chapter 4) and is ready to be evaluated *in planta*, either together with an infectious clone, or with native virus from infected plant sap as helper virus. The satRNA clone converted into an expression and silencing vector may be a valuable tool, even without a genomic infectious clone. Once proven to be infectious and capable of expression and silencing of heterologous genes, these constructs may prove a very much needed molecular tool for the grapevine research community, as functional genomics in grapevine is at present limited due to the lack of molecular and functional genomics tools.

5.6 References

1. Ahlquist, P., French, R., Janda, M., and Loesch-Fries, L. S. (1984). Multicomponent RNA plant virus infection derived from cloned viral cDNA. Proc Natl Acad Sci, USA 81, 7066–7070.
2. Boyer and Haenni. (1994) Infectious transcripts and cDNA clones of RNA viruses. Virology, 198, 145-426.
3. Chapman, S. (2008). Construction of Infectious Clones for RNA Viruses: TMV. Methods in Molecular Biology, 451, 477-490. DOI:10.1007/978-1-59745-102-4-32
4. Dagless EM, Shintaku MH, Nelson RS and Foster GD (1997). A CaMV 35S promoter driven cDNA clone of tobacco mosaic virus can infect host plant tissue despite being uninfectious when manually inoculated onto leaves. Arch Virol, 142, 183-191
5. Ding SW, Rathjen JP, Li WX, Swanson R, Healy H, Symons RH (1995). Efficient infection from cDNA clones of cucumber mosaic cucumovirus RNAs in a new plasmid vector. J Gen Virol, 76, 459-464
6. Dupuis L., C. Himber, P. Dunoyer, A. Bassler, M. Keller, T. Wetzel (2009). Symptom determinants on the *Arabis mosaic nepovirus* genome. Proceedings of the 16th Congress of ICVG, Dijon France, 31 October – 4 September 2009.
7. Gosselé V, Faché I, Meulewater F, Cornelissen M, Metzlaff M (2002). SVISS – a novel transient gene silencing system for gene function discovery and validation in tobacco plants. The Plant Journal, 32, 859-866.

8. H Sumiyoshi, C H Hoke, D W Trent (1992). Infectious Japanese encephalitis virus RNA can be synthesized from in vitro-ligated cDNA templates. J Virol, 66(9), 5425–5431.
9. Hamilton CM, Frary A, Lewis C, Tanksley SD (1996). Stable transfer of intact high molecular weight DNA into plant chromosomes. Proc Natl Acad Sci USA, 93, 9975–9979.
10. Hayes RJ, Buck KW, (1990). Infectious cucumber mosaic virus RNA transcribed *in vitro* from clones obtained from cDNA amplified using the polymerase chain reaction. J Gen Virol, 71, 2503-2508.
11. Kay R, Chan A, Daly M, McPherson J (1987). Duplication of CaMV-35S promoter sequences creates a strong enhancer for plant genes. Science, 236, 1299-1302.
12. Kurth EG, Peremyslov VV, Prokhnevsky AI, Kasschau KD, Miller M, Carrington JC Dolja V (2012). Virus-Derived Gene Expression and RNA Interference Vector for Grapevine. J Virol. 86(11), 6002-6009
13. Lamprecht S & Jelkman W. (1997). Infectious cDNA clone used to identify strawberry mild yellow edge-associated potexvirus as casual agent of the disease. J Gen Virol, 78, 2347-2353.
14. Lamprecht RL, Maree HJ, Stephan D, Burger JT (2012). Complete nucleotide sequence of a South African isolate of Grapevine fanleaf virus. Virus Genes, 45, 406-410.
15. Meng B, Venkataraman S, Li C, Wang W, Dayan-Glick C, Mawassi M. Infectivity Assays of Second Generation cDNA clones of Grapevine Ruperstris Stem Pitting-associated virus. Proceedings of the 17th Congress of ICVG, Davis, California, USA, October 7-14, 2012.
16. Murugananthham M, Moskovitz Y, Haviv S, Horesh T, Fenigstein A, du Preez J, Stephan D, Burger JT, Mawassi M (2009). Grapevine virus A-mediated gene silencing in Nicotiana benthamiana and Vitis vinifera. Journal of Virological Methods, 155 (2), 167-174.
17. Nagyova and Subr Z (2007). Infectious full-length clones of plant viruses and their use for construction of viral vectors. Acta Virologica, 51, 223-237.
18. Naraghi-Arani P, Daubert S, Rowhani A (2001). Quasispecies nature of the genome of Grapevine fanleaf virus. J Gen Virol, 82, 1791–1795.
19. Neeleman L, Van Der Vossen EAG, Bol JF (1993) Infection of tobacco with alfalfa mosaic virus cDNAs sheds light on the early function of the coat protein. Virology, 196, 883-887.
20. Oberpichler I, Rosen R, Rasouly A, Vugman M, Ron EZ, Lamparter T. (2008) Light affects motility and infectivity of Agrobacterium tumefaciens Environ Microbiol, 10(8), 2020-2029.
21. Quan J and Tian J (2009). Circular polymerase extension cloning for high throughput cloning of complex and combinatorial DNA libraries Nature Protocols, 6, 242–251.
22. Rice CM, Grakoui A, Galler R, Chambers TJ (1989). Transcription of infectious yellow fever RNA from full-length cDNA templates produced by in vitro ligation. New Biol, 1(3), 285–296.
23. Shi BJ, Ding SW and Symons RH (1997). Plasmid vector for cloning infectious cDNAs from plant RNA viruses: high infectivity of cDNA clones of tomato aspermycucumovirus. J Gen Virol, 78, 1181-1185.

24. Töpfer R, Matzeit V, Gronenborn B, Schell J, Steinbiss HH, 1987. A set of plant expression vectors for transcriptional and translational fusions. Nucleic Acids Res, 15, 5890.
25. Viry M., Serghini MA, 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.
26. Szymczak-Workman AL, Vignali KM, Vignali DAA (2012). Design and Construction of 2A Peptide-Linked Multicistronic Vectors. Cold Spring Harb Protoc doi:10.1101/pdb.prot067884.
27. Weber H, Haeckel P, Pfitzner AJP (1992) A cDNA clone of tomato mosaic virus is infectious in plants. J Virol, 66, 3909-3912.
28. Xiang C, Han P, Lutziger I, Wang K and Oliver DJ (1999). A mini binary vector series for plant transformation. Plant Molecular Biology, 40, 711-717.
29. Yu HH and Wong SM (1998).Synthesis of biologically active cDNA clones of cymbidium mosaic potexvirus using a population cloning strategy. Archives of Virology 143, 1617–1620.

Chapter 6

Concluding remarks and future prospects

The aim of this study was to determine the full-length sequences of South African (SA) Grapevine fanleaf virus (GFLV) isolates. Even though South Africa is consistently one of the top ten wine producers in the world, research on GFLV is limited and very little sequence information is available for SA GFLV isolates. Internationally, GFLV is one of the most widespread and devastating viruses affecting grapevine and has been studied extensively. Most of these studies have largely focused on the epidemiology of the virus and on the development of detection techniques. Other researchers have focussed their attention on determining partial sequences of RNA2 in order to study genetic variability among isolates. More full-length sequences of both the GFLV genomic components, RNA1 and RNA2, as well as biological characterisation of the virus, are needed in order to understand how GFLV evolves and to develop concomitant strategies to manage the disease. In this study, the full-length genomic sequences of two SA GFLV isolates (GFLV-SAPCS3 and GFLV-SACH44) are presented, including a full-length sequence of an associated satellite RNA (satRNA). Before this study, only three GFLV isolates have been completely sequenced (from France and the USA), even with the virus's worldwide notoriety. Also, a fully functional infectious clone based on a new GFLV satRNA was developed and is described.

The complete genome of GFLV-SAPCS3 was the first SA GFLV isolate to be fully sequenced. In addition to deformation of new systemically-infected leaves, this isolate also induced mosaic and line patterns in infected *Chenopodium quinoa* plants. GFLV-SAPCS3 RNA1 is 7342 nt in length and RNA2 is 3817 nt in length and phylogenetic analysis revealed that it was most closely related to GFLV-F13. The GFLV-SAPCS3 RNA2 5' UTR is the longest sequenced to date and was more closely related in

sequence and length to the 5' UTRs of Arabis mosaic virus (ArMV) isolates. This interesting observation prompted us to investigate whether recombination events in GFLV-SAPCS3 and ArMV could be detected in the 5' UTR, as well as the rest of the genome. Putative recombination events were detected in both GFLV-SAPCS3 RNA1 and RNA2 sequences. In RNA1, we found a putative recombination event within $1E^{Pol}$, involving GFLV-F13 (recombinant) with parental lineages of a GFLV-WAPN173-type (major parent) and a GFLV-SAPCS3- type (minor parent) isolates. To date this is only the second report of recombination events in the polymerase gene of GFLV. Also, an interspecies recombination event was detected in the 5' UTR of RNA2, with GFLV-SAPCS3 as the recombinant, resulting from the recombination between a GFLV-F13-type as the major parent and an ArMV-Ta-type as the minor parent. This recombination event was further illustrated by the presence of the conserved stretch, AA/GTCCGTT/CA, that is present in the 5' UTR of GFLV-SAPCS3, GFLV-Ghu (a known GFLV-ArMV recombinant) and three other ArMV isolates, but was absent in all the other GFLV isolates sequenced to date.

With the two SA studies combined (this study and a BSc Honours study of Ms M Spaltman), only two GFLV-infected plants out of more than a hundred plants screened tested positive for the presence of satRNA. It was decided to determine the full-length genomic and satRNA sequences of GFLV isolate from one of these plants, GFLV-SACH44, and to compare them to the full-length sequences of GFLV-SAPCS3 and other isolates have not naturally-associated satRNAs with them. The two RNAs of GFLV-SACH44 were 7341 nt and 3816 nt in length, respectively, and was most closely related to GFLV-SAPCS3 (98.2 % and 98.6 % nucleotide identities, respectively). The GFLV-SACH44 also had the same conserved sequence AA/GTCCGTT/CA at position 73–98 in the RNA2 5' UTR that was found in GFLV-SAPCS3, GFLV-Ghu and other ArMV isolates. This finding suggests that GFLV-SACH44, like GFLV-SAPCS3, may have arisen from a common ancestor that may have originated from an interspecies recombination event in the 5' UTR region between GFLV-F13 type and ArMV-Ta type isolates. An updated phylogenetic tree, including new GFLV full-length RNA1 and RNA2 sequences, indicated that GFLV-SAPCS3 and GFLV-SACH44 grouped together,

although distinct from the other sequenced isolates. The GFLV-SACH44 satRNA was 1104 nt in length and had a 5' UTR of 14 nt and a 3' UTR of 74 nt in length. Interestingly, the GFLV-SACH44 satRNA was more closely related in nucleotide identity and sequence length to the ArMV satRNAs (86.7-87.8% nucleotide identity) than to the other described GFLV satRNAs (71-81.8%). It was confirmed with a phylogenetic analysis based on full-length sequences of GFLV and ArMV satRNA that the GFLV-SACH44 satRNA grouped, although separately, with three ArMV satRNA isolates and the GFLV-F13 satRNA in one clade, while the Californian GFLV satRNAs and the other ArMV satRNA isolates grouped in another.

Full-length infectious cDNA clones of GFLV-SACH44 satRNA were constructed. The full-length sequence of the GFLV-SACH44 satRNA was cloned between a Cauliflower mosaic virus 35S promoter and upstream of a self-processing hammerhead ribozyme sequence. Two constructs were able to replicate with two different GFLV isolates, GFLV-NW (Germany) and GFLV-SAPCS3, after mechanical inoculation in *Chenopodium quinoa*. The infectivity of the satRNA clones were repeated in three independent experiments and were verified by RT-PCR and Northern blot analysis. Initial experiments showed no notable difference in symptom expression in the presence of the satRNA, however a small number of plants were used, and a larger number of plants are required to verify this result. The satRNA clone was also co-inoculated with ArMV-NW, however after several attempts, results have shown that the satRNA clone was not able to replicate with ArMV-NW as a helper virus. This was surprising as the GFLV-SACH44 is phylogenetically more closely related to ArMV satRNAs than to GFLV satRNAs. This finding illustrates that recognition of the satellite by the replication machinery of the helpervirus is more complex than it appears. In this study, in an attempt to find possible genomic areas responsible for satRNA replication, no clear sequence variation were shown between the GFLV isolates that have naturally-associated satRNAs (GFLV-SACH44) and those that have not (GFLV-SAPCS3). However, as was demonstrated later in this study, GFLV-SAPCS3 was able to replicate the satRNA, and therefore comparing the sequences of these two isolates was

superfluous. Additional sequences and infectious clones of GFLV satellites and their helper viruses will be needed to address this question.

In this study the foundation was laid towards the construction of a GFLV full-length infectious cDNA clone based on GFLV-SAPCS3. The satRNA of GFLV-SACH44 was only discovered well past the initial planning phases of construction of the GFLV-SAPCS3 full-length clone, and it therefore seemed to be a wasted effort to restart the process in order to construct genomic and satRNA infectious clones that represented a single isolate. We therefore continued with the construction of the GFLV-SAPCS3 full-length clone, and fortunately as demonstrated, GFLV-SAPCS3 was able to replicate the GFLV-SACH44 satRNA clone. This meant that the full-length cDNA clones representing the genomic and satRNA components could potentially be coupled together as a single system. With this in mind, the full-length satRNA clone was converted into a viral-based vector for the expression and silencing of transgenes. Unfortunately the GFLV-SAPCS3 genomic full-length clones were not infectious in *N. benthamiana*, likely due to lethal mutations that occurred in RNA1, and as a result, the modified satRNA constructs was not evaluated *in planta.* The modified satRNA clones are however ready and available to be inoculated along with a native helper virus, to be screened for its infectivity and efficiency as a virus-based gene expression and silencing vector.

In future studies the GFLV satRNA clone can be utilised to determine which GFLV or ArMV genomic regions are responsible for satRNA replication. This can be achieved by sequencing GFLV isolates that are not able to replicate the constructed GFLV satRNA clone in herbaceous hosts, and to compare those sequences to GFLV isolates that do support the replication of the satRNA, such as GFLV-SAPCS3 and GFLV-SACH44, which were sequenced in this study. Furthermore, the GFLV-satRNA clone can be utilised in determining the effects that satRNA presence may have on overall GFLV symptom expression. One could compare the symptoms expressed by herbaceous hosts infected with a symptom-inducing isolate, like GFLV-SAPCS3, in the presence of satRNA, to herbaceous hosts infected with the said isolate without satRNA. Lastly, the full-length GFLV-SAPCS3 cDNA clones may not currently be infectious due to deletion mutations in RNA1, but these mutations can be corrected by site-directed mutagenesis.

These full-length GFLV-SAPCS3 genomic clones can then be used in conjunction with the modified satRNA constructs for the expression and silencing of heterologous genes in *N. benthamiana* and possibly *V. vinifera*. Overall, the constructs assembled in this study may prove to be much-needed molecular tools for the grapevine research community.

This study contributed to the current knowledge of GFLV variability by determining the full-length sequences of two SA GFLV isolates. We have added two new full-length sequenced isolates to the existing three completely sequenced GFLV isolates. Furthermore we have characterised a new GFLV satRNA, both on a molecular level and biologically. The full-length GFLV satRNA infectious clone constructed in this study will prove to be an invaluable tool in future studies.

Appendix A

Table 1: The predicted mutations (nonsynonomous and silent) that will still exist in pCB301_2x35S_GFLVRNA1_35ST if the four deletions were to be corrected.

Position	SAPCS3	Mutation	Change aa
1745	A	G	E → V
2503	T	C	Silent
2510	A	G	E → G
2805	G	A	Silent
3000	T	C	Silent
3102	G	A	Silent
3115	A	G	S → G
3315	G	A	Silent
3361	C	T	Silent
3371	T	C	L → S
4204	T	C	F → L
5013	T	C	Silent
5211	A	C	Silent
5217	C	T	Silent
5235	C	T	Silent
5449	A	T	R → W
5577	C	T	Silent
5878	A	C	I → L
6006	T	C	Silent
6072	T	C	Silent
6515	A	G	K → R
6712	A	G	T → A
6833	C	T	A → V