VIRAL METAGENOMIC ANALYSIS OF SWEET POTATO USING HIGH THROUGHPUT DEEP SEQUENCING

Thulile Faith Nhlapo

A thesis submitted to the Faculty of Science, University of the Witwatersrand, Johannesburg, in fulfilment of the requirements for the degree of Doctor of Philosophy

Johannesburg, 2019

Supervisor
Prof. M.E.C Rey

Co-Supervisor
Dr. D.J.G Rees

Advisor
Dr. D.A. Odeny

DECLARATION

I, **THULILE FAITH NHLAPO (0604106A)**, am a student registered for the degree of Doctor of Philosophy in the academic year 2019.

I hereby declare the following:

- I am aware that plagiarism (the use of someone else's work without their permission and/or without acknowledging the original source) is wrong.
- I confirm that the work submitted for assessment for the above degree is my own unaided work except where explicitly indicated otherwise and acknowledged.
- I have not submitted this work before for any other degree or examination at this or any other University.
- The information used in the Thesis **HAS NOT** been obtained by me while employed by, or working under the aegis of, any person or organisation other than the University of the Witwatersrand and the Agricultural Research Council.
- I have followed the required conventions in referencing the thoughts and ideas of others.
- I understand that the University of the Witwatersrand may take disciplinary action against me if there is a belief that this is not my own unaided work or that I have failed to acknowledge the source of the ideas or words in my writing.

Signature  25th day of January 2019

ABSTRACT

Sweet potato is an orphan crop with significant importance in developing countries due to its high β-carotene content and ability to address food insecurity in rural communities. Sweet potatoes also serve as an economically viable crop to resource poor farmers. Production constraints in the form of viral diseases negatively impact crop quality and yield, which often results in catastrophic economic losses. Detailed genomic characterisation of sweet potato viruses in East and Southern Africa is still lacking. The development of virus detection, identification and characterisation techniques is essential for viral disease control. The aim of this study was to identify and characterise the diversity of sweet potato viruses causing sweet potato virus disease (SPVD) in the Eastern and Western Cape provinces of South Africa. Viral metagenomics and high-throughput deep sequencing revealed the presence of six previously detected viruses and two novel badnaviruses. Sweet potato chlorotic stunt virus (SPCSV) was identified as a major role player and causal agent of SPVD. SPCSV was detected in mixed infections with sweet potato feathery mottle virus (SPFMV) and begomoviruses. These viruses are known to act in synergy to exacerbate viral disease symptoms. The reference-guided and *de novo* assembly of next generation sequencing (NGS) data achieved over 70% genome coverage for all viruses. The use of deep sequencing of nucleic acids is therefore a reliable diagnostic tool for virus detection as well as for differentiating between diverse viral strains. Small RNA profiles in the NASPOT 1 (resistant) and Blesbok (susceptible) leaves were also investigated by analysing the expression patterns of virus derived small interfering RNAs (vsiRNAs) and endogenous micro RNAs (miRNAs) in response to mixed viral infection at 60 days post infection (dpi). MiRNAs are small non-coding RNAs involved in the regulation of important biological processes such as plant development, biotic and abiotic stress response and pathogen defense. Over 55 miRNA families were identified collectively from the two cultivars. Two miRNAs (miR160 and miR6300) were downregulated in both resistant and susceptible cultivars, while miR393, miR398, miR168, miR162 and miR482) were upregulated in both cultivars, after infection with viruses. These miRNAs could play a key role in pathogen defense responses, as they are known to target mRNAs that encode major genes, enzymes and proteins, which are involved in the plant defense mechanisms. This study lays a firm foundation for understanding host-pathogen interactions in sweet potato.

DEDICATION

I dedicate this work to my parents, Oumaki Sanny and Madala Petrus Nhlapo who have encouraged, inspired, and supported me in all areas of my life.

In memory of my grandmothers

Priscilla Gadifele Moitsheki

1939 – 2010

and

Maleponisa Paulina Nhlapo
1923 – 2013

ACKNOWLEDGEMENTS

All glory and honour to God, Father of my Lord and Saviour Jesus Christ. I have seen Him do exceedingly and abundantly above all I could ask, think or imagine. I have witnessed His power, favour, wisdom and love for me through this work.

I would like to extend my gratitude to my supervisor and mentor Dr. Jasper Rees. Thank you for the many opportunities you've afforded me and for believing in me from the very first day. Thank you for the academic, personal and financial support. Thank you for encouraging and challenging me to do and be more in all areas of my life. You have shown me that scientists can be great leaders. I admire you and most certainly want to be like you one day.

To my dearest parents, Oumaki and Madala Nhlapo, thank you for your endless love, support and sacrifices. I am a reflection of your hard work and selfless efforts. To my sisters Ntombizodwa and Kholwa Nhlapo, I am honoured to be your sibling, knowing that you look up to me has always been great motivation. Thank you for your support. I love you. Omolemo and Portia Moitsheki, thank you for the constant love and support. To the rest of my extended family, thank you for supporting me.

To Julia Mulabisana, thank you for your support and encouragement. Thank you for making a great contribution to this work. I was very honoured to work with you and share this difficult yet rewarding journey with you.

To Prof. Chrissie Rey and Dr. Damaris Odeny, thank for your financial and academic support.

To my friends Petunia Malatji, Molemi Rauwane, Lesego Tau, Tsebo Malape, Lesedi Milanzi and Keabetswe Ncube; thank you for your love, encouragement and support.

Thank you to the Agricultural Research Council – Professional Development Programme (ARC-PDP), National Research Foundation (NRF), and Department of Science and Technology (DST) for financial support throughout my studies.

CONTENTS PAGE

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS

aa	Amino acid
AGO	Argonaute
ARC	Agricultural Research Council
ARF	Auxin Responsive Factor
BLAST	Basic Local Alignment Search Tool
bp	Base Pair
BTP	Biotechnology Platform
°C	Degree Celsius
cDNA	Complementary Deoxyribonucleic Acid
CP	Coat Protein
CPm	Minor Coat Protein
DCL	Dicer-like
DNA	Deoxyribonucleic Acid
dsDNA	Double-stranded DNA
dsRNA	Double-stranded RNA
EST	Expressed Sequence Tag
FAO	Food and Agriculture Organisation
gDNA	Genomic DNA
h	Hours
I. batatas	*Ipomoea batatas*
iba-miRNA	Sweet Potato MicroRNA
kb	Kilo Basepairs
kDa	Kilodalton
min	Minutes
miRBase	MicroRNA Registry Database
miRNA	MicroRNA
MP	Movement Protein
NB-LRR	Nucleotide Binding Site Leucine Rich Repeat
NGS	Next Generation Sequencing
nm	Nanometer
NRF	National Research Foundation
nt	Nucleotide

ORF	Open Reading Frame
PCR	Polymerase Chain Reaction
pre-miRNA	MicroRNA Precursor
pri-miRNA	Primary MicroRNA Transcript
psRNATarget	Plant Small RNA Target Analysis Server
PTGS	Post-Transcriptional Gene Silencing
RCA	Rolling Circle Amplification
RdRp	RNA-dependent RNA Polymerase
RISC	RNA-Induced Silencing Complex
RNA	Ribonucleic Acid
RT-PCR	Reverse Transcription PCR
s	Seconds
sgRNA	Subgenomic RNA
siRNA	Small Interfering RNA
SPCSV	Sweet Potato Chlorotic Stunt Virus
SPFMV	Sweet Potato Feathery Mottle Virus
SPVD	Sweet Potato Virus Disease
sRNA	Small RNA
ssDNA	Single-stranded DNA
ssRNA	Single-stranded RNA
TGS	Transcriptional Gene Silencing
µl	Microliter
vsiRNA	Virus-derived Small Interfering RNA
VSR	Viral Suppressor of RNA Silencing

CHAPTER 1

Motivation for Study

1.1 General Introduction

Sweet potato is an important root crop with nutritional and economical value in South Africa. The crop is used to address food insecurity by reducing rural poverty through the promotion of sweet potato cultivation and improvement of vitamin A intake by women and children. Sweet potato production also serves as income for rural poor subsistence farmers. It has been documented that production constraints such as insect pests, viral, bacterial and fungal diseases have caused catastrophic economic losses by substantially reducing crop quality and yield. Viral diseases can reduce crop yield by up to 100%, depending on the cultivar (variety) and the type of virus infecting the plant. Extensive studies have been done to elucidate the etiology of sweet potato virus diseases from all sweet potato growing regions including Korea (Kwak et al., 2015), Peru (Gutierrez et al., 2003) and Uganda (Gibson et al., 2004). Nevertheless, detailed genomic characterisation of sweet potato viruses in East and Southern Africa is still lacking. Development of virus detection and identification techniques is important for viral disease control. After viruses are detected and identified, rapid indexing procedures can be developed, resistant cultivars can be recognised and scientists can further devise other disease control strategies. There are however difficulties that arise in the detection of sweet potato viruses. Sometimes problems arise due to low virus titers, inhibitors, or ineffective diagnostic assays. Limitations that may be encountered in the diagnosis of sweet potato viruses include the occurrence of mixed infections, diverse viral strains, and uneven virus distribution within the plant. The presence of certain viruses such as sweet potato feathery mottle virus (SPFMV), have been known to mask the presence of other viruses in sweet potato, thus hindering efforts to successfully isolate and identify all the viruses infecting the host. A traditional method that has been reliably used for virus detection in the past is biological indexing using the indicator plant, *Ipomoea setosa*. Previously it was believed that this plant was a host for all viruses infecting sweet potato but some viruses, like tomato spotted wilt virus (TSWV), do not cause visible symptoms in the host and therefore more reliable methods of virus detection are needed.

A metagenomic approach and the use of next generation sequencing (NGS) techniques will circumvent the limitations stated above. The deep sequencing technology will be able to

detect viral pathogens occurring at very low titers as well as detect mixed infections and diverse strains present in a particular host thereby enabling the detection of known and novel viruses and their strains, if present. The use of the Illumina Sequencing platforms is expected to provide high coverage of detected viruses and is likely to result in the generation of complete viral genomes.

MicroRNAs (miRNAs) are an important class of small RNAs that are involved in the regulation of cellular functions including cell differentiation, pathogen defense and growth. Studies focusing on the identification of miRNAs in *Ipomoea batatas* (sweet potato) have not been done and availability of sweet potato miRNA sequences in the major small RNA repositories is lacking. Identification of miRNA sequences that could be useful in crop improvement; for traits such as virus disease resistance, pest or drought resistance, heat tolerance, and high vitamin content; through the use of artificial microRNA technology is a very important endeavour. In order to employ the use of microRNA technology in breeding programmes, it is imperative to investigate and understand the underlying defense mechanisms in virus resistant and susceptible sweet potato varieties. In South Africa sweet potato cultivar Blesbok is a high yielding and popular variety in the informal and formal fresh produce markets (Laurie, 2001, Laurie et al., 2015), even though it is very susceptible to viral infection (Domola, 2004). The NASPOT 1 cultivar is resistant to sweet potato virus disease (Domola, 2004). The disease defense mechanisms of these two cultivars will be investigated. The availability of sequence data (genomic resources) will not only aid in the identification of viral pathogens infecting sweet potato but also help researchers to understand the effect of the identified viruses on South African cultivars (varieties). Therefore NGS strategies can be employed together with traditional techniques for crop improvement in order to address food insecurity.

1.2 Aims

This study aimed to achieve the following:

a. To carry out a viral metagenomic analysis of sweet potato viruses in the Eastern Cape and Western Cape provinces in order to determine the identity and distribution of sweet potato viruses in these two provinces.
b. To sequence the full genomes of SPFMV and SPCSV isolates identified from the survey.

c. To identify the small RNAs (siRNAs and miRNAs) involved in host-pathogen interactions.
d. To investigate the miRNA expression patterns of the NASPOT 1 (resistant) and Blesbok (susceptible) cultivars in response to multiple virus infection.

1.3 Research Questions and Hypotheses

The major questions that this study seeks to answer include:

a). Are there new sweet potato virus introductions into SA since the last survey?

b). Is NGS technology a more effective technique for sweet potato virus detection?

The above research questions will be addressed based on the following hypotheses:

a). That a survey of two provinces (eight locations) will be sufficient to detect some of the sweet potato viruses present in the country.

b). The sequencing techniques will be sufficient to identify previously undetected sweet potato viruses in South Africa.

1.4 Chapter Summary

The thesis is comprised of six chapters. In chapter one the general introduction, aims and objectives and research outputs to the study are outlined. In chapter two the literature review on the background topics (plant virus diagnostics, next generation sequencing (NGS) and RNA silencing) are thoroughly explained. Chapter three is the first research chapter, which describes the identification and diversity study of sweet potato RNA viruses in two South African provinces. Chapter four describes the identification of sweet potato geminiviruses using rolling circle amplification (RCA) and NGS. The first report of sweet potato badnaviruses in South Africa is also described in chapter four. Chapter five describes the identification of pathogen related microRNAs (miRNAs) in sweet potato. Chapter six is the general discussion, conclusion and recommendations for future studies.

1.5 Research Outputs

Conference presentations

Nhlapo, T.F., Mulabisana, J., Odeny, D., Rey, M.E.C., Rees, D.J.G. Viral metagenomic analysis of sweet potato using high-throughput deep sequencing. Oral presentation delivered at the African Centre for Gene Technologies meeting, June 2014, Pretoria, South Africa.

Thulile F. Nhlapo, Julia Mulabisana, Chrissie M.E. Rey, Damaris A. Odeny, Lindy Esterhuizen, and D. Jasper G. Rees. Viral metagenomic analysis of sweet potato using high-throughput deep sequencing. Poster presentation delivered at the International Plant and Animal Genome Conference, January 2015 San Diego, California, USA.

Nhlapo, T.F. Sweet spud sequencing for the nation: solving food insecurity one step at a time. Oral presentation delivered at the Africa's Women in Agriculture, Entrepreneurship and Youth Development Summit, May 2015, Development Bank of Southern Africa, Midrand, Johannesburg, South Africa.

Nhlapo, T.F. Sweet spud sequencing for the nation: solving food insecurity one step at a time. Oral presentation delivered at the third Global Conference on Agricultural Research for Development (GCARD3), March 2016, Pretoria, South Africa.

Nhlapo, T.F., Mulabisana, J., Odeny, D., Rey, M.E.C., Rees, D.J.G. The status of sweet potato viruses in South Africa: using viral metagenomics for the detection and characterisation of potyviruses, criniviruses, geminiviruses and badnaviruses. Oral presentation delivered by D.J.G Rees on behalf of T.F. Nhlapo, at Sweet potato and Yam Genomics Workshop, International Plant and Animal Genome Conference, January 2017, San Diego, California, USA.

Nhlapo, T.F. Next-Generation Sequencing Saves the Sweet Potato. Illumina iCommunity Newsletter May, 2015. Link: nhlapo-miseq-sweetpotato-article-1370-2015-002.pdf

Peer-reviewed publications

Nhlapo, T.F., Mulabisana, J., Odeny, D., Rey, M.E.C., Rees, D.J.G. 2018. First report of Sweet potato badnavirus A and Sweet potato badnavirus B in South Africa. *Plant Disease* https://doi.org/10.1094/PDIS-08-17-1235-PDN.

Nhlapo, T.F., Rees, D.J.G, Odeny, D., Mulabisana, J., Rey, M.E.C. 2018. Viral metagenomics reveals sweet potato virus diversity in the Eastern and Western Cape provinces of South Africa. *South African Journal of Botany,* 117, 256-267.

Works in progress

Nhlapo, T.F., Rauwane, M.E., Rey, M.E.C., Rees, D.J.G. Identification of miRNAs associated with mixed viral infections in susceptible and resistant sweet potato cultivars. Manuscript writing in progress and will be submitted to *Virus Research* in April 2019.

References

DOMOLA, M. J. 2004. *Survey and characterisation of sweet potato viruses in South Africa.* University of Pretoria.

GIBSON, R. W., ARITUA, V., BYAMUKAMA, E., MPEMBE, I. & KAYONGO, J. 2004. Control strategies for sweet potato virus disease in Africa. *Virus Research,* 100**,** 115-122.

GUTIERREZ, D., FUENTES, S. & SALAZAR, L. 2003. Sweetpotato virus disease (SPVD): distribution, incidence, and effect on sweetpotato yield in Peru. *Plant Disease,* 87**,** 297-302.

KWAK, H.-R., KIM, J., KIM, M.-K., SEO, J.-K., JUNG, M.-N., KIM, J.-S., LEE, S. & CHOI, H.-S. 2015. Molecular characterization of five potyviruses infecting Korean sweet potatoes based on analyses of complete genome sequences. *The Plant Pathology Journal,* 31**,** 388.

LAURIE, S. Overview of breeding and evaluation of orange fleshed sweet potato in South Africa. Proc. of VITAA project regional workshop, Nairobi, Kenya, 2001.

LAURIE, S., MAJA, M., NGOBENI, H. & DU PLOOY, C. 2015. Effect of different types of mulching and plant spacing on weed control, canopy cover and yield of sweet potato (Ipomoea batatas (L.) Lam). *American Journal of Experimental Agriculture,* 5**,** 450.

CHAPTER 2

LITERATURE REVIEW

2.1 Sweet potato origin

Sweet potato (*Ipomoea batatas* (L.) Lam.) is a dicotyledonous perennial plant that produces starchy tuberous roots that are consumed as a vegetable. Sweet potato belongs to the Morning Glory family, *Convolvulaceae,* and within this family there are 55 genera (Kreuze, 2002). More than 500 species are found in the *Ipomoea* genus with ploidy levels ranging from 2x to 6x (Kreuze, 2002, Manamela, 2009). It is suggested that sweet potato originated from a cross between *I. trifida* and other wild *Ipomoea* species in central or northern South America 5000 years ago (O'Brien, 1972, Kreuze, 2002, Tairo et al., 2006) and it is now cultivated and consumed all over the world in tropical and sub-tropical regions (Kreuze, 2002, Valverde et al., 2007). Sweet potato is the only *Ipomoea* species known to have economic and nutritional importance (Kreuze, 2002).

2.2 Nutritional and economic importance

Sweet potato is ranked as the seventh most important staple crop and the third most important root crop (Kays, 2004, Valverde et al., 2007). The crop is particularly beneficial to farmers with limited resources as it is very easy to grow. It performs well in nutrient poor soils; it is drought tolerant, crowds out weeds and has a short growing period (Kreuze, 2002, Martin, 1998). When compared to other major starch staple crops such as potato, it has the largest rate of production and ranks number one of all vegetables in terms of nutrition (Centre for Science in the Public Interest (CSPI, 2009). Sweet potatoes are rich in vitamin A, vitamin C, vitamin B6, potassium, folate, and essential mineral salts. The tubers are high in carbohydrates and dietary fiber, and the leaves are also used as greens (Kreuze, 2002, Loebenstein et al., 2003). Yellow and orange-fleshed sweet potatoes have high beta-carotene content and they are used to alleviate vitamin A deficiency (Kreuze, 2002), especially in developing countries such as Zambia, South Africa, Uganda, Kenya, Nigeria, and Tanzania (Clark et al., 2012). Statistics show that 43 million African children under the age of five are threatened by vitamin A deficiency, a condition causing blindness, disease, premature death, maternal death, death from measles and diarrhoea in children, reduced resistance to infections, and delayed recovery from illness (van Jaarsveld et al., 2005). It is therefore

important to maintain crop production for food security and good health status in developing countries.

In South Africa, sweet potato addresses two national problems, namely malnutrition and food insecurity (South African Human Rights Commission, 2008). A survey conducted in the rural areas of KwaZulu-Natal, the Eastern Cape and the North West provinces showed that an estimated 64% of children between the ages of one to nine years are vitamin A deficient and sweet potato plays an important role in alleviating the deficiency (Manamela, 2009). Organizations such as the Medical Research Council (MRC) and the Agricultural Research Council (ARC) have developed and implemented projects aimed at improving vitamin A intake through the production and consumption of provitamin A-rich vegetables and fruit such as orange-fleshed sweet potato (van Jaarsveld et al., 2005). Furthermore, sweet potato is of considerable economic value within marketing chains, which are structured for the local and export markets (Department of Agriculture, Forestry and Fisheries, 2010). Sweet potato is also produced and sold by informal sectors, however the figures are not included in the official production records. In 2015 the gross value of sweet potato production in South Africa was R200 million (Department of Agriculture, Forestry and Fisheries, 2010). Statistics show that the sweet potato industry has been making a significant contribution to the gross value of agriculture (Department of Agriculture, Forestry and Fisheries, 2010). In terms of production, the annual production of sweet potato in 2014 was 62569 tons, with an average yield of 29300 hectograms/hectare (hg/ha) (FAOSTAT, 2017). Currently, sweet potato ranks among the 10 most important food crops globally based on the dry weight produced, yielding over 130 million metric tons per year and nine million hectares (Clark et al., 2012). The highest producers of sweet potato are China, followed by Nigeria and the United States of America (Table 2.1).

2.3 Production constraints

Sweet potato has high production yields and can withstand several production stresses such as high temperature and water shortages (Kays, 2004). However, production can be constrained by insects, weeds, rodents, bacterial, fungal and viral diseases (Kapinga and Carey, 2003, Nderitu et al., 2009, Souto et al., 2003, Kreuze et al., 2009, Valverde et al., 2007), which have a negative economic impact on the livelihoods of farmers and those depending on the crop for nutrition (Tesfaye et al., 2011). The sweet potato weevil species *Cylas brunneus* and *Cylas puncticolis* are known to cause the most devastating damage on

sweet potato (Aritua et al., 2007, Kreuze, 2002) but viral diseases are the second most devastating constraints (Ateka et al., 2004b).

Table 2.1: Sweet potato world production and highest producing countries in each continent for 2014 (FAOSTAT, 2017)

Continent	Selected Countries	Production (t)
Asia	Total	79213661
	China	71305000
Africa	Total	22630750
	Nigeria	3775425
Americas	Total	3828140
	United States of America	1341910
Oceania	Total	883150
	Papua New Guinea	671740
Europe	Total	45901
	Portugal	22591
World	Total	106601602

2.3.1 Viruses as production constraints

The sweet potato crop is sensitive to viral infection thus viruses can be major production constraints (Domola, 2006), (Tesfaye et al., 2011). Viral diseases may cause up to 100% yield loss, this depends on the cultivar type, cultivar susceptibility or whether the viral infection is singular or multiple, as seen in a virus complex (Gibson et al., 1997). The first extensive report to demonstrate the effect of viruses on sweet potato in South Africa showed that viruses significantly lower yield and quality of tubers (Domola, 2006). In the United States of America yield losses of 30-50% in famer's fields have been observed (Valverde et al., 2007), and losses of 80-90% have been recorded in areas affected by virus complexes that include sweet potato feathery mottle virus (SPFMV) and sweet potato chlorotic stunt virus (SPCSV) (Kreuze, 2002, Valverde et al., 2007, Mukasa et al., 2003). Viruses co-infecting a

plant can interact in several ways to cause severe disease. A common mechanism is where one virus assists the replication of a second co-infecting virus, leading to increased titers and more severe symptoms (Latham and Wilson, 2008). This process is known as viral synergism (Kreuze, 2002). A good example of synergism in sweet potato is the known interaction between SPCSV and SPFMV (Valverde et al., 2007), which causes sweet potato virus disease (SPVD). SPVD was first reported in 1940 in Uganda, Burundi, Rwanda and eastern Belgian Congo (Kreuze, 2002, Ateka et al., 2004b, Cuellar et al., 2008). The symptoms associated with this disease include chlorosis, deformed leaves and severe stunting (Ateka et al., 2004b, Kreuze, 2002) (Figure 2.1). SPVD can reduce yields by 80% - 90% (Ateka et al., 2004b, Kreuze, 2002) making it the most economically devastating disease affecting sweet potato (Cuellar et al., 2008, Kreuze, 2002). In the mid-1940s viral diseases were recognized in quality and yield deterioration of sweet potato in South Africa (Tesfaye et al., 2011). Although many of the viruses and their strains, which occur in South Africa, have been characterised molecularly and biologically, SPCSV has not been the focus of many studies and has therefore not been well characterised.

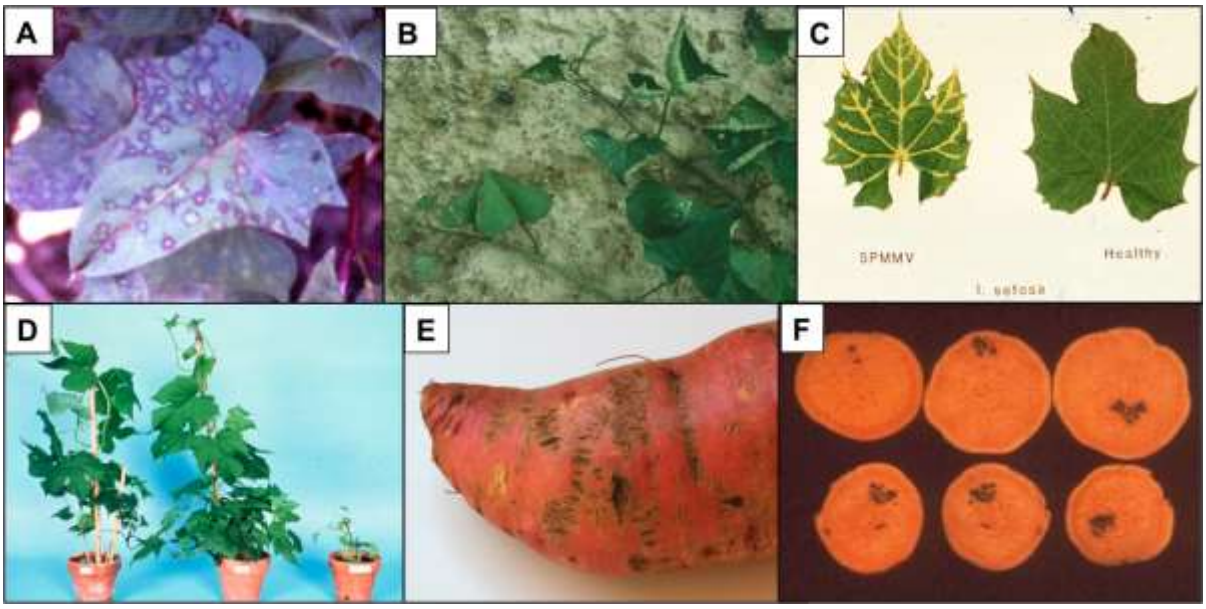

Figure 2.1: Symptoms associated with sweet potato viral infection. (A) Leaves displaying chlorotic spots with purple rings are usually associated with SPFMV. (B) Upward curling of leaves is often an indication of the presence of SPLCV. (C) Yellowing of leaf veins in the indicator plant (*I. setosa*) is a symptom caused by SPMMV. (D) Plants affected with SPVD, caused by a combination of SPCSV and SPFMV, exhibit severe stunting (far right plant) in comparison to mild symptoms observed in single infections of SPCSV and SPFMV (first two tall plants). Tubers display symptoms of viral infection such as external lesions (E) and internal root necrosis (F) often attributed to the russet crack strain of SPFMV. **Source**: http://keys.lucidcentral.org/keys/sweetpotato/.

2.4 Taxonomy and classification of plant viruses

2.4.1 Classification of viruses

By definition a virus is a set of one or more nucleic acid template molecules, normally encased in a protective coat or coats of protein or lipoprotein, that is able to organise its own replication, only within suitable host cells (Hull, 2009). The Baltimore system of virus classification places viruses into seven groups (Baltimore, 1971) according to the type of genome nucleic acid found encapsidated within a virus particle. The genome nucleic acid can be ssDNA, dsDNA, dsRNA, negative/ambisense sense ssRNA, positive/sense ssRNA, positive/sense ssRNA with a DNA intermediate and reverse transcribing viruses (dsDNA-RT) (Hull, 2009, Khan and Dijkstra, 2001). The ICTV also classifies viruses based on their chemical characteristics, diseases, vectors, and geographical distribution. Due to the nature of this study, the literature review will focus on describing viruses based on their genome characteristics.

I. **Single-stranded DNA (ssDNA):** This group consists of two families, *Geminiviridae* and *Nanoviridae*. *Geminiviridae* contains three genera (*Begomovirus*, *Mastrevirus* and *Curtovirus*) all transmitted by different vectors. These viruses have small circular ssDNA genomes having either one or two segments (monopartite or bipartite) (Hull, 2009, Khan and Dijkstra, 2001). Replication of geminiviruses occurs within the nucleus of the infected plant and ssDNA is converted into a double-stranded circular intermediate (Hull, 2009). Sweet potato viruses falling into this group include sweet potato leaf curl virus (SPLCV) and sweet potato mosaic virus (SPMaV), which are in the *Begomovirus* genus and sweet potato symptomless virus 1, (SPSMV-1) in the *Mastrevirus* genus (Clark et al., 2012, Kreuze et al., 2009, Mbanzibwa et al., 2016).

II. **Reverse-transcribing viruses (retroviruses):** Retroviruses are viruses that have dsDNA (family *Caulimoviridae*) or ssRNA genomes where replication is facilitated by reverse transcriptase (Khan and Dijkstra, 2001, Madigan et al., 1997). Viruses that have circular dsDNA genomes replicate through an RNA intermediate (Kashif, 2009) and those that have ssRNA genomes replicate through an ssDNA intermediate (Madigan et al., 1997). Many of these viruses integrate into their host genomes. In this group are badnaviruses, namely sweet potato badnavirus A (SPBVA) and sweet potato badnavirus B (SPBVB) (Kreuze et al., 2009).

III. **Double-stranded RNA (dsRNA):** Mycoviruses and some plant or fungal viruses (*Partiviridae* and *Reoviridae*) are assigned to this group. Double-stranded RNA viruses such as Reoviruses replicate in the cytoplasm of the host, and each strand of negative dsRNA produces many copies of positive strand RNA followed by synthesis of mRNA (Khan and Dijkstra, 2001, Madigan et al., 1997).

IV. **Negative sense single-stranded RNA (ssRNA):** These viruses have single stranded RNA genomes that are complementary to their messenger RNA (mRNA) (Madigan et al., 1997). In this group some or all of the genes are translated into proteins. Plant viruses in this group include *Rhabdoviridae* and *Fimoviridae* (Madigan et al., 1997).

V. **Positive sense single-stranded RNA (ssRNA +):** This category includes viruses that have the same orientation as the mRNA, therefore the viral genome serves directly as the mRNA. This mRNA encodes an RNA polymerase (RNA replicase) that makes complementary strands of RNA and then uses them as templates to make more plus strands. These strands can either be translated as mRNA or packed as the genome in the newly synthesized capsid (Madigan et al., 1997). Most plant viruses are in this group e.g. *Bromoviridae*, *Closteroviridae*, *Flexiviridae*, *Comoviridae*, *Potyviridae*, *Sequeviridae*, *Tombusviridae* and *Luteoviridae* (Khan and Dijkstra, 2001). The two major disease-causing viruses of sweet potato, SPFMV and SPCSV, fall into this category.

2.5 The taxonomy of sweet potato viruses

Compared to viruses of other agriculturally important crops, sweet potato viruses were poorly studied until recently (Kreuze and Fuentes, 2008). Over 30 viruses infecting sweet potato have been identified and assigned to 9 families namely: *Bromoviridae*, *Fimoviridae*, *Caulimoviridae*, *Closteroviridae*, *Comoviridae*, *Flexiviridae*, *Geminiviridae*, *Luteoviridae*, and *Potyviridae* (Clark et al., 2012). Most plant viruses infecting sweet potato have RNA genomes (double stranded and single stranded), but between 2009 and 2012, DNA viruses belonging to the *Geminiviridae* and *Caulimoviridae* families (Clark et al., 2012) were reported.

2.5.1 Sweet potato DNA viruses

Geminiviruses are plant viruses with a circular ssDNA genome consisting of one (monopartite) or two (bipartite) molecules packed into twin-shaped icosahedral particles

(Fauquet and Stanley, 2003, Hanley-Bowdoin et al., 2013, Vanderschuren et al., 2007) ranging in size from 2.6- 3 kb (Hanley-Bowdoin et al., 1999). They are classified into four genera, the *Begomovirus*, *Curtovirus*, *Mastrevirus*, and *Topocuvirus*. Geminiviruses that infect sweet potato are monopartite, belonging to the genus *Begomovirus* genera, and are transmitted by whiteflies, species *Bemisia tabaci* (Genn.) or distributed by vegetative propagation (Fauquet et al., 2003, Fauquet and Stanley, 2003). Geminiviruses that infect sweet potato are all monopartite and phylogenetically distinct from other begomoviruses and are collectively known as "sweepoviruses" (Clark et al., 2012). The host range of these viruses is narrow, restricted to the *Convolvulaceae* (genus *Ipomoea*) and *Solanacea* families (*N. benthamiana*) (Kreuze and Fuentes, 2008). Infection of sweet potatoes with begomoviruses has been reported in Japan, Israel, Peru, Italy, Spain, China, Taiwan, Korea, Kenya, United States of America; Puerto Rico, Costa Rico, Brazil and South Africa (Esterhuizen et al., 2012, Kreuze and Fuentes, 2008, Paprotka et al., 2010). The symptoms associated with begomovirus infection include upward curling of leaves, vein yellowing or leaf distortion and chlorosis (Kreuze and Fuentes, 2008, Trenado et al., 2011). However, begomoviruses are frequently symptomless in sweet potato but do cause major yield losses and can spread undetected (Ling et al., 2010). Begomoviruses can therefore be a major threat to sweet potato production (Mansoor et al., 2003).

Pararetroviruses - within this group are the sweet potato badnaviruses that usually occur in mixed infections with SPCSV (Clark et al., 2012, Kreuze et al., 2009). These viruses have been detected in sweet potatoes worldwide including Africa (Mbanzibwa et al., 2014). Badnaviruses have a linear double-stranded DNA genome (dsDNA) that ranges between 7.2 and 8.5 kb in size (Bouhida et al., 1993, Hansen and Heslop-Harrison, 2004, Stavolone et al., 2001, James et al., 2011, Lyttle et al., 2011, Geering et al., 2014). Badnaviruses belong to the family *Caulimoviridae* (Borah et al., 2013). Viruses in this family are characterised by bacilliform shaped viral particles 60-900 nm in length and 35-50 nm wide (Borah et al., 2013). Symptoms associated with badnavirus infection vary according to the host and virus strains/isolates. These range from vein-clearing (Zhang et al., 2011a), margin mild leaf distortion, chlorotic lines, necrosis, to tuber flesh discolouration, yellow or bright green patches on the leaf and plant death (Borah et al., 2013). Several insect vectors transmit badnaviruses; these include aphids, mealybugs, leafhoppers and nematodes (Laney et al., 2012). Other dsDNA sweet potato viruses include the sweet potato collusive virus (SPCV) (synonyms, sweet potato caulimo-like virus) in the genus *Cavemovirus*. Sweet potato vein

clearing virus (SPVCV) is in the genus *Solendovirus* and distribution of these viruses is proposed to occur by vertical transmission (Clark et al., 2012).

2.5.2 Sweet potato RNA Viruses

Several potyviruses are known to infect sweet potato. The most studied sweet potato potyvirus is SPFMV (genus *Potyvirus*, family *Potyviridae*), found to occur in most sweet potato growing regions. This virus causes greater damage when involved in a disease complex, with symptoms such as internal root necrosis or external crack lesion on the tubers (Kreuze and Fuentes, 2008) (Figure 2.1). SPFMV has flexuous filamentous particles ranging from 830-850 nm. The genome is a single stranded positive RNA genome, ranging from 10.6-10.8 kb in size (Sakai et al., 1997, Yamasaki et al., 2010). Vectors responsible for transmission of this virus include several aphid species, which transmit this virus in a non-persistent manner (Kreuze and Fuentes, 2008). Previously, two isolates; namely the common strain (C) and the russet crack strain (RC) were characterised based on symptomology, host range studies, molecular and serological methods (Rännäli et al., 2009). Subsequent phylogenetic analysis of the 3'-UTR nucleotide sequences showed that SPFMV is grouped into four phylogenetic groups or strains, namely RC, O, EA (East Africa) and the C strain which is so distantly related to the other 3 strains, it is now classified as a separate virus, sweet potato virus C (SPVC) (Untiveros et al., 2010).

SPCSV, previously known as sweet potato sunken vein virus (SPSVV) (Tesfaye et al., 2011), has a worldwide distribution including the main sweet potato production areas in Africa and the Americas (Cuellar et al., 2011). SPCSV is a single-stranded bipartite positive-sense RNA virus belonging to the *Closteroviridae* family (Cuellar et al., 2011), which acts synergistically with many viruses to cause SPVD (Souto et al., 2003). SPCSV is transmitted by whiteflies, in a semi-persistent, non-circulative manner (Cuellar et al., 2011, Kreuze, 2002, Tairo et al., 2006, Tesfaye et al., 2011). Similar to SPFMV, the host range of SPCSV is limited mainly to *Convolvulaceae* and the genus *Ipomoea* (Kreuze, 2002, Tairo et al., 2006). Symptoms caused by SPCSV are relatively mild in sweet potato and *Ipomoea setosa* (indicator plant) and plants may become mildly stunted, with chlorotic spots or purpling appearing on the leaves (Kreuze, 2002) (Figure 2.1). Infection with SPCSV decreases tuberous root yield by 50% (Gibson et al., 1997, Tesfaye et al., 2011). SPCSV particles are 850-960 nm in length and 12 nm in diameter (Kreuze, 2002) and the size of the coat protein is 33 KDa. The genome of SPCSV consists of two RNA molecules with a total length of 17630 nt. RNA1 (9407 nt) contains five

putative open reading frames (ORFs) and RNA2 (8223 nt) contains seven ORFs (Kreuze, 2002). Based on complete and partial coat protein (CP) and heat shock protein 70h (HSP70h) sequences and characterisation with monoclonal and polyclonal antibodies, SPCSV exists in two distinct strain groups known as east African (EA) and west African (WA) strains (Kokkinos et al., 2006, Kreuze, 2002, Souto et al., 2003, Tairo et al., 2006). Detection of SPCSV is traditionally based on immunohistochemical microscopy (Kreuze, 2002), although this is likely to change with the reduction in sequencing costs. Tanzanian, Ugandan and Peruvian isolates of SPCSV have been characterised (Cuellar et al., 2011). Studies in Africa (Mukasa et al., 2006), Israel, Central America and South America (Kreuze et al., 2009), have confirmed that severe diseases have been associated with mixed viral infections that often include the presence of SPCSV (Cuellar et al., 2011).

Sweet potato virus 2 (SPV2) is a single-stranded, positive-sense RNA virus in the family *Potyviridae*, which also interacts synergistically with SPFMV, to cause SPVD (Tairo et al., 2006). SPV2 was originally discovered in Taiwan in the 1980s from plants that showed symptoms including ring spots and vein yellowing (Ateka et al., 2004a, Tairo et al., 2006). SPV2 has been detected in the United States of America (USA) and Africa (Tairo et al., 2006). The virus has been tentatively named Ipomoea vein mosaic virus (IVMV) and sweet potato virus Y (SPVY) due to incomplete characterisation. However the full genome of SPV2 was made available in 2012. The genome is 10731 nt long excluding the poly (A) tail and contains a single large ORF (Li et al., 2012a).

Sweet potato virus G (SPVG) is a potyvirus that was originally described in China but now has a wider distribution, as it occurs in Oceania, Peru, the United States, and Africa (Kreuze and Fuentes, 2008, Pardina et al., 2012). The SPVG genome sequence of 10798 nt is typical of a potyvirus; and it is the third largest in size after SPFMV and SPVC (Pardina et al., 2012). The genome contains a 5' non-coding region (NCR) of 111 nt, an ORF encoding a 3488 aa polyprotein, a 3'-NCR of 223 nt and a small conserved ORF PIPO (61 aa) (Li et al., 2012a, Pardina et al., 2012).

Other RNA viruses include sweet potato vein mosaic virus (SPVMV), which was first reported in Argentina (Kreuze and Fuentes, 2008). Sweet potato latent virus (SPLV) was first reported in China, but now known to occur in most major growing areas all across Asia (Kreuze and Fuentes, 2008). Sweet potato mild speckling virus (SPMSV) was first reported

in Argentina, and later detected in Peru and Indonesia (Kreuze and Fuentes, 2008). Sweet potato mild mottle virus (SPMMV) originated in East Africa (Tugume et al., 2010) and now has a wide distribution with serological tests showing presence of this virus throughout Africa, Indonesia, China, Philippines, New Guinea, India and New Zealand (Tugume et al., 2010, Valverde et al., 2007). This virus is also transmitted in a non-persistent manner by whiteflies (Kreuze and Fuentes, 2008). Sweet potato yellow dwarf virus is another potyvirus transmitted by whiteflies (Kreuze and Fuentes, 2008).

2.6 Methods for the detection and discovery of plant viruses

2.6.1 Biological assays

One of the oldest techniques of virus detection is based on visual inspection of symptoms on the host plant (López et al., 2003, López et al., 2009, Naidu and Hughes, 2003, Prabha et al., 2013). In the event that plants didn't readily exhibit symptoms, indicator plants were used (graft-inoculation). This method is known as biological indexing (biological assays) (Legrand, 2015). Biological assays utilise indicator plants that are susceptible to viruses, which exhibit symptoms associated with infection upon inoculation. However, indicator plants are not consistent in their expression of symptoms and some viruses do not induce symptoms (López et al., 2009), therefore rendering biological assays inefficient diagnostic tools. Also, plants may exhibit virus-like symptoms in response to environmental stress, and incorrect identification of viruses may occur (Naidu and Hughes, 2003). It is recommended that biological assays should be used in conjunction with serological and molecular methods for more conclusive results (Legrand, 2015).

2.6.2 Electron microscopy

Since the discovery of viruses during 1882 - 1886 (Zaitlin, 1998), several methods have been employed for virus detection. The use of electron microscopy (EM) for virus discovery and characterisation began in the 1930s, shortly after the electron microscope was invented (Liu et al., 2011). This method of detection aims to determine the particle morphology of viruses (Vale et al., 2010). The approach has been useful for the detection of new viruses because specific reagents are not necessary for identification (Liu et al., 2011). However this method is limited in its application because viruses present in the plant at low titres cannot be detected (Naidu and Hughes, 2003, Vale et al., 2010) and viruses can only be characterised up to family or genus level (Liu et al., 2011, Vale et al., 2010).

2.6.3 Serological methods

Serological techniques for the detection and characterisation of viruses were available since the 1960s (Boonham et al., 2014). The most common serological method is the enzyme-linked immunosorbent assay (ELISA), which uses monoclonal or polyclonal antibodies to detect viruses (Clark and Adams, 1977). This method is widely used because of its high throughput, high sensitivity, application simplicity and low cost (Lommel et al., 1982, Naidu and Hughes, 2003, Webster et al., 2004). One major drawback of the ELISA method is the need for prior knowledge of the virus to be detected. In the absence of antibodies, viruses can go undetected, therefore excluding the identification of any new emerging viruses; also the method is limited in distinguishing virus strains (Boonham et al., 2014, Prabha et al., 2013). Some of the ELISA variants used for detection of sweet potato viruses include double antibody sandwich (DAS-ELISA) (Clark and Adams, 1977), triple antibody sandwich (TAS-ELISA) (Gibson et al., 1998) and nitrocellulose membrane-enzyme-linked immunosorbent assay (NCM-ELISA) (Gutierrez et al., 2003) (Table 2.2).

2.6.4 Nucleic acid-acid based methods

Molecular methods used for the direct detection of viruses include polymerase chain reaction (PCR) and DNA arrays (Nicolaisen, 2011, Sankaran et al., 2010) (Table 2.2). In PCR-based methods viral nucleic acids (DNA or RNA) are isolated and amplified using virus specific oligonucleotide primers for the amplification of specific genes (Naidu and Hughes, 2003). Several variants of PCR techniques are available for the detection of plants viruses; they include multiplex PCR, nested PCR, co-operational PCR (co-PCR), quantitative PCR, and reverse-transcription (RT-PCR) (López et al., 2003, López et al., 2009, Webster et al., 2004). These methods are more sensitive and accurate than serological methods because any region of the genome can be targeted (Naidu and Hughes, 2003). However, PCR techniques are primer specific, and in the absence of target genome sequences, many viruses can go undetected and the use of degenerate primers often results in false positives (Prabha et al., 2013). Methods of viral discovery and detection have improved over the years with the aim to overcome the limitations posed by traditional methods. For the detection of microorganisms in an unbiased fashion, metagenomic analysis of different environments has been performed.

2.7 Metagenomics and viral metagenomics

Metagenomics is defined as the study of microorganisms in their natural environment (Rooks et al., 2010) or the genomic analysis of the collective microbial assemblage found in an environmental sample (Kakirde et al., 2010). Metagenomic analysis is not biased towards culturable organisms therefore the total genetic diversity of microorganisms can be studied (Rooks et al., 2010). Therefore, viral metagenomics is the study of virus populations in a specific sample (Edwards and Rohwer, 2005). In earlier viral metagenomic studies, viral particles were purified from samples and then the viral nucleic acids were sequenced directly (Ng et al., 2009, Roossinck, 2012). Viral metagenomics can be used to analyse viral sequences in any sample type, therefore making it a powerful virus discovery tool (Alavandi and Poornima, 2012, Bibby, 2013). It can also be applied to determining disease etiology by investigating viruses that share limited homology with known viruses (Day et al., 2010, Ng et al., 2009). Characterisation of new viruses is often hindered by difficulties to amplify them in cell culture, limited antigenic/serological cross-reactivity or the lack of virus specific primers (Delwart, 2007). The metagenomic approach circumvents the limitations stated above and consequently provides insights into the composition of the viral community (Edwards and Rohwer, 2005). Viral metagenomics began in 2002, when two uncultured marine environments were studied (Edwards and Rohwer, 2005). Since then, many more viral metagenomic studies have been published, with an increase from 5 libraries to over thousands in 2011. Metagenomic approaches to viral characterisation have been applied to seawater, fresh water (Djikeng et al., 2009), faeces (Victoria et al., 2009), serum, plasma and respiratory secretions (Delwart, 2007), grapevine (Coetzee et al., 2010), soil (Kakirde et al., 2010), turkey gut (Day et al., 2010), insects (Liu et al., 2011), crops such as tomato (Li et al., 2012a) and sweet potato (Kreuze et al., 2009) (Table 2.2). Metagenomic studies have also been undertaken in humans to detect viruses that cause human diarrhoea (Finkbeiner et al., 2008) and haemorrhagic fever (Briese et al., 2009), therefore highlighting the wide application of this virus characterisation and detection approach. There is also vector-mediated metagenomics (VEM), where mosquitoes, whiteflies and other vectors are studied to understand the diversity of plant and animal viruses, this method is used because plants and animals are hosts for many bacterial and viral pathogens (Ng et al., 2009). Previously metagenomics employed techniques such as environmental shotgun sequencing (ESS) or random shotgun sequencing (Men et al., 2008, Tammi, 2003). Shotgun sequencing is based on the principle of shredding DNA into smaller fragments, then cloning the fragments into universal vectors also known as plasmids (Tammi, 2003). The fragments are sequenced

individually using the traditional Sanger sequencing method then assembled to create the original contiguous sequence (Men et al., 2008, Tammi, 2003). The advantages of the shotgun sequencing method are easy automation and scalability (Tammi, 2003).

2.7.1 Whole genome sequencing

Whole genome analysis has shifted from Sanger sequencing to newer methods over the years (Metzker, 2010). Prior to this shift, Sanger sequencing had dominated for almost two decades with immense accomplishments like sequencing the human genome in 2004 (International Human Genome Consortium). Sanger sequencing is considered as the "first generation" technology (Metzker, 2010). For many years Sanger sequencing was considered the "golden standard" of sequencing. However, automated Sanger sequencing has several limitations. Firstly the method is based on cloning foreign DNA into vectors, which is time consuming and disadvantageous because some parts of chromosomes such as centromeres cannot be cloned (Men et al., 2008). Secondly, this method has restricted ability to analyse allele frequencies. The third and most significant limitation of Sanger sequencing is the cost. Usually a lot of funds are required to complete large sequencing projects (Men et al., 2008). The development of next generation sequencing (NGS) platforms has led to advances in sequencing technologies that circumvent the limitations of Sanger sequencing.

2.7.2 Next Generation Sequencing

NGS technologies can be distinguished from Sanger sequencing because they do not use chain termination chemistry and electrophoresis (Metzker, 2010). They rely on amplification of single DNA molecules to generate clusters of DNA templates that are attached or immobilized to a solid surface (Metzker, 2010, Shendure and Ji, 2008). This procedure is called solid-phase amplification. The clusters of identical molecules are then sequenced in parallel by cyclic incorporation and measurement of fluorescently labelled nucleotides or short oligonucleotides or the detection of by-products (Metzker, 2010, Shendure and Ji, 2008). Because of the parallel sequencing of the amplified clusters, this technology is also called massive parallel sequencing or high-throughput sequencing (Mardis, 2008, Metzker, 2010). The new technologies are made up of various strategies that rely on a combination of template preparation, sequencing and imaging, genome alignment and assembly methods (Metzker, 2010). The principle behind NGS platforms is to randomly fragment DNA or RNA into smaller pieces and then construct a DNA or cDNA library, respectively. Libraries are sequenced at high coverage and the sequence reads are assembled, either *de novo* (without a

reference), or mapped to a reference genome (de Magalhães et al., 2010). NGS reduces the time and cost of sequencing (Mardis, 2008), and high volumes of data can be generated for a fraction of the price of Sanger sequencing. In addition to parallel sequencing, genetic material is sequenced directly and there is no need for the cloning step (Ansorge, 2009), this increases time and cost savings. Another advantage of NGS is the use of universal adaptors instead of sequence-specific oligonucleotides; this makes it specifically suitable for metagenomics studies (Bentley, 2006, Van Vliet, 2010).

2.7.4 Methods used for virus discovery library preparation

Several methods can be employed to prepare samples for sequencing on NGS platforms. The use of total RNA isolation, followed by ribo-zero prior to direct sequencing is an approach employed to enrich the samples for virus sequences (Massart et al., 2014, Studholme et al., 2011). Isolation of double-stranded RNA has been developed as an approach to enrich samples for RNA viruses, due to that fact that the presence of dsRNA in plants has been associated with the presence of viruses (Tzanetakis and Martin, 2008). Enrichment of DNA viruses is based on rolling circle amplification (RCA) (Haible et al., 2006) because RCA enriches samples for viruses with circular genomes. Another widely utilised indirect strategy for sample preparation is the isolation of small interfering RNAs (siRNAs) (Barba et al., 2014, Kreuze, 2014). Small interfering RNAs are produced in the plant in response to viral infection via the host immune response pathway called RNA interference (RNAi) or RNA silencing (Eamens et al., 2008, Mlotshwa et al., 2008). These siRNAs are complementary to the viral nucleic acids and thus target viral RNA for degradation or transcriptional repression (Baulcombe, 2004). This method has been used to detect and discover viruses in sweet potato (Kashif et al., 2012, Kreuze et al., 2009, Mbanzibwa et al., 2014), tomato (Li et al., 2012a) and various other agriculturally important crops and model plant species as listed in thorough reviews (Barba et al., 2014, Kreuze, 2014). Examples of studies employing these different approaches for virus detection are highlighted in Table 2.2.

2.8 Bioinformatic software utilized for the analysis of NGS data

2.8.1 Alignment/ mapping assembly of reads

Bioinformatic analysis of NGS data begins with quality control (QC) analysis, followed by mapping or aligning reads to the reference or sequence of origin (Horner et al., 2010, Magi et al., 2010). During QC analysis adaptor sequences and indexes that were added during library

preparation are removed (trimmed). Failed or low quality sequence reads, below the Q of 25 (threshold defined), are also removed from the data set prior to mapping. The "clean" data set is then mapped to references of choice. Mapping occurs in two steps: the first step is to search for candidate alignment locations (CAL) by indexing the read sequences or the reference followed by the actual alignment (Fonseca et al., 2012, Lee et al., 2012). The following tools are available for short-read alignment: BLAT, MAQ, Bowtie, SOAPaligner/SOAP2, BWA and BFAST (Bzhalava and Dillner, 2013, Lee et al., 2012). In the absence of a reference genome *de novo* assemblers are used (Magi et al., 2010).

2.8.2 Assembly of short reads generated by NGS platforms

Early genome assemblers (long read assemblers) used a "greedy algorithm" where reads were compared with each other, and the ones that overlapped were merged first (Bzhalava and Dillner, 2013, Miller et al., 2010, Schatz et al., 2010). To allow for sequence errors, assemblers used the Smith-Waterman algorithm (Smith and Waterman, 1981) to compute overlaps with a variant that allowed for a small number of differences in the overlapping sequence (Schatz et al., 2010). Once all the overlaps were computed, the reads with the overlap were merged to form contiguous sequences (also known as contigs). This process was then repeated, each time merging the sequences with the longest overlap until all the overlaps were used (Miller et al., 2010, Schatz et al., 2010). Since NGS platforms produce much shorter reads, assemblers have been developed to address the challenges of assembling short reads. The platforms compensate for short read length by obtaining deeper coverage or sequence depth (Schatz et al., 2010). Coverage is defined as the average number of reads that contain any nucleotide; therefore 40X coverage implies that the genome is sequenced 40 times (Schatz et al., 2010). Popular short read assemblers include Velvet, ALLPATHS, ABySS, SOAPdenovo, Contrail, SSAKE and SaSSY (Magi et al., 2010, Schatz et al., 2010). Instead of using an overlap graph these assemblers use a de Bruijn graph algorithm.

Recently software packages have created user-friendly bioinformatic tools with the ability to map, assemble and interpret NGS data (Horner et al., 2010). These tools have graphical interfaces that are suitable for biologists or scientists without prior bioinformatics training or computational biology experience (Egan et al., 2012). Open-source platforms designed for NGS data analysis include the Galaxy platform (http://main.g2.bx.psu.edu/), iPlant Collaborative (http://www.iplantcollaborative.org) and Bioconductor an R statistical analysis environment (http://www.bioconductor.org/) (Egan et al., 2012). CLC Bio has also developed

a user-friendly graphical interface platform called the CLC Genomics Workbench (http://www.clcbio.com/genomics) that allows one to import, trim, map, assemble and visualise Sanger, 454, Illumina, SOLID and Ion Torrent sequence reads (Egan et al., 2012, Lee et al., 2012).

Table 2.2: Plant virus detection methods, comparing newer strategies such as next generation sequencing and traditional diagnostics methods including electron microscopy, indexing, serology and molecular techniques (PCRs and microarrays)

Host Organism	Method of viral detection/ discovery	Plant virus/viroid detected/characterised	References
Solanum lycopersicum (tomato)	siRNA deep sequencing, dot-ELISA, DTBIA, PCR, multiplex RT-PCR, real-time RT-PCR.	*Potato spindle tuber viroid (PSTVd), Pepino mosaic virus (PepMV), Tomato yellow leaf curl virus (TYLCV), Tomato spotted wilt virus (TSWV), Tomato torrado virus (ToTV).*	(Boonham et al., 2014, Donaire et al., 2009, Kreuze, 2014, Li et al., 2012b, Xie et al., 2013)
Ipomoea batatas (sweet potato)	DAS-ELISA, TAS-ELISA, NCM-ELISA, PCR, multiplex RT-PCR, electron microscopy, siRNA deep sequencing, total RNA-seq, RCA/RFLP.	Sweet potato feathery mottle virus (SPFMV), Sweet potato virus C (SPVC), Sweet potato symptomless virus 1 (SPSMV-1), Sweet potato virus G (SPVG), Sweet potato leaf curl virus (SPLCV), Sweet potato virus 2 (SPV2), Sweet potato chlorotic fleck virus (SPCFV), Sweet potato latent virus (SPLV), Sweet potato mosaic virus (SPMaV), Sweet potato mild mottle virus (SPMMV), Ipomoea vein mosaic virus (IVMV), Sweet potato badnavirus A (SPBVA), Sweet potato badnavirus B (SPBVB), Sweet potato symptomless mastrevirus 1	(Ateka et al., 2004b, Clark et al., 2012, Esterhuizen et al., 2012, Gu et al., 2014, Gutierrez et al., 2003, IsHak and El-Deeb, 2004, Kashif et al., 2012, Kokkinos, 2006, Kreuze et al., 2009, Li et al., 2004, Mbanzibwa et al., 2016, Mbanzibwa et al., 2014, Nyaboga et al., 2008)
Solanum tuberosum (potato)	Multiplex RT-PCR RT-PCR & real-time RT-PCR, microarray technology, siRNA deep sequencing	Potato virus Y (PVY), Potato virus X (PVX), Potato virus A (PVA), Potato virus S (PVS), Potato leaf roll virus (PLRV)	(Agindotan et al., 2007, Boonham et al., 2003, Kutnjak et al., 2014, MacKenzie et al., 2015, Mallik et al., 2012)
Vitis vinifera (Grapevine)	DAS-ELISA, RT-PCR, siRNA deep sequencing, IC-RT-PCR, dsRNA deep sequencing, ELISA, multiplex RT-PCR	Grapevine virus B (GVB), Grapevine vein clearing virus (GVCV), Grapevine fanleaf virus (GFLV), Arabis mosaic virus (ArMV), Grapevine leafroll associated virus 1 (GLRaV-1), Grapevine leafroll associated virus 2 (GLRaV-2), Grapevine leafroll	(Coetzee et al., 2010, Glasa et al., 2014, Hu et al., 2014, Koolivand et al., 2014, Leo et al., 2015, Zhang et al., 2011a)

		associated virus 3 (GLRaV-3), Grapevine virus A (GVA), Grapevine fleck virus (GFkV), Grapevine Pinot gris virus (GPGV),	
Malus domestica (Apple)	siRNA deep sequencing, RT-PCR	Cherry leaf roll virus (CLRV), Apple stem grooving virus (ASGV), Apple chlorotic leaf spot virus (ACLSV); Apple stem pitting virus (ASPV), Apple mosaic virus (ApMV);	(Kundu, 2003, Menzel et al., 2002, Visser et al., 2014, Watpade et al., 2012, Woo et al., 2012)
*Triticum aestivum (*Wheat)	Reverse transcription loop-mediated isothermal amplification (RT-LAMP), RT-PCR, PCR	Wheat yellow mosaic virus (WYMV), Wheat streak mosaic virus (WSMV), Triticum mosaic virus (TriMV), Japanese soil-borne wheat mosaic virus (JSBWMV), Chinese wheat mosaic virus (CWMV), Wheat dwarf India virus (WDIV),	(Fukuta et al., 2013, Kumar et al., 2014, Mar et al., 2013, Price et al., 2010, Zhang et al., 2011b)
Capsicum spp. (Pepper)	PCR (cloning and sequencing), DAS-ELISA, RT-PCR, ELISA	Pepper mild mottle virus (PMMoV), Pepper golden mosaic virus (PepGMV), Pepper veinal mottle virus (PVMV), Cucumber mosaic virus (CMV), Pepper yellow dwarf virus (PeYDV), Pepper yellow leaf curl virus (PYLCV)	(Ala-Poikela et al., 2005, Arogundade et al., 2012, Çağlar et al., 2013, Dombrovsky et al., 2013, Lam et al., 2009, Peng et al., 2015, Rialch et al., 2015)
Zea mays (Maize)	PCR, RT-PCR, RT-PCR/RFLP, siRNA deep sequencing, total RNA & dsRNA sequencing, TAS-ELISA, DIBA, IC-RT-PCR	Maize streak virus (MSV), Maize rough dwarf virus (MRDV), Sugarcane mosaic virus (SCMV), Maize dwarf mosaic virus (MDMV), Maize chlorotic mottle virus (MCMV)	(Achon et al., 2011, Adams et al., 2013b, Dovas et al., 2004, Rybicki and Hughes, 1990, Xia et al., 2014)
*Manihot esculenta (*Cassava)	Total RNA deep sequencing, multiplex RT-PCR, real-time RT-PCR	Cassava brown streak virus (CBCV), African cassava mosaic virus (ACMV), East African cassava mosaic Cameroon virus (EACMCV)	(Abarshi et al., 2012, Adams et al., 2013a, Alabi et al., 2008, Monger et al., 2001)
Citrus	Electron microscopy, PCR, ELISA, Western blotting, ISEM, siRNA deep sequencing, multiplex real-time RT-PCR	Citrus psorosis virus (CPsV), Hop stunt viroid (HSVd), Citrus tristeza virus (CTV), Citrus exocortis viroid (CEVd), Citrus yellow mosaic virus (CYMV), Citrus Leprosis virus (CiLV)	(C. et al., 2015, Gopi et al., 2010, Lin et al., 2015, Loconsole et al., 2006, Roy et al., 2012, Saponari et al., 2008, Su et al., 2015)
Oryza sativa (Rice)	ELISA, RT-PCR, Latex agglutination reaction, RT-LAMP, multiplex-PCR, real-time PCR, microscopy	Rice hoja blanca virus, Rice Tungro virus disease	(Dasgupta et al., 1996, Uehara-Ichiki et al., 2013).

2.10 Disease control strategies

Efforts have been made to develop disease control or management strategies, but viruses accumulate in plants and diseases often become challenging to manage due to vegetative propagation and germplasm exchange of sweet potato (Clark et al., 2012). Accumulation of viruses, resulting in disease complexes also further complicates disease control. Methods for disease control have included chemical and biological control of insect pests or vectors by spraying pesticides, breeding for resistance, increasing knowledge of virus prevalence by developing effective diagnostic tools, and by distribution of virus-free material (Clark et al., 2012, Reddy et al., 2009). Breeding strategies to control virus disease are often hindered because some viruses have the ability to suppress the RNA silencing pathway (Costa et al., 2013, Soosaar et al., 2005). Most viruses, especially potyviruses, express RNA silencing suppressors such as HC-Pro (Pallas and García, 2011, Soosaar et al., 2005), p22 and RNase3 proteins thus blocking or interfering with the host defense mechanism (Costa et al., 2013) ultimately causing disease development on leaves and tubers.

2.11 RNA silencing

RNA silencing is a mechanism of defense plants use to counteract viruses. This mechanism is described as a "sophisticated genetic defense system" (Freeman and Beattie, 2008) that inhibits translation at the post-transcriptional level (post-transcriptional gene silencing - PTGS) (Duan et al., 2012). RNA silencing can also function as a gene expression regulatory mechanism (Guo et al., 2015) and the silencing pathway has the potential to be used as a tool to engineer resistant plants (Duan et al., 2012). Viral RNAs trigger the defense response by producing the opposite sense of viral genome during replication, thus generating dsRNA (Baulcombe, 2004, Mlotshwa et al., 2008, Pallas and García, 2011, Zhang et al., 2013). Plant proteins (Dicer-type dsRNA RNases) cleave viral dsRNA into small interfering RNAs (siRNA), which are incorporated into the Argonaute (AGO) protein to assemble the RNA-induced silencing complex (RISC) (Figure 2.3a). The RISC complex then targets the viral ssRNA or host mRNA for degradation (silencing) or cleavage (Pallas and García, 2011, Vaucheret, 2006). The RISC complex may also target host mRNAs, this would result in the down regulation of host genes, it has been proposed that regulation of host genes can contribute to suppressing antiviral defenses and thus cause symptom development (Pallas and García, 2011, Zhang et al., 2013). A group of small non-coding RNAs called microRNAs (miRNAs) also play a critical role in antiviral defense and regulating gene expression (Guo et al., 2015, Soosaar et al., 2005). These endogenously produced miRNAs are derived from

dsRNA hairpin structures (Guo et al., 2015, Soosaar et al., 2005). The miRNAs are also incorporated into AGO proteins to form the RISC complex. The RISC complex directs the miRNA to target mRNAs for silencing or translational repression (Figure 2.2b) (Guo et al., 2015, Vaucheret, 2006). If RNA silencing is a successful defense mechanism, disease symptoms disappear; this response is known as "recovery" (Freeman and Beattie, 2008). Several studies have shown that some novel miRNAs could be involved in influencing mRNA expression profiles in the defense mechanism process.

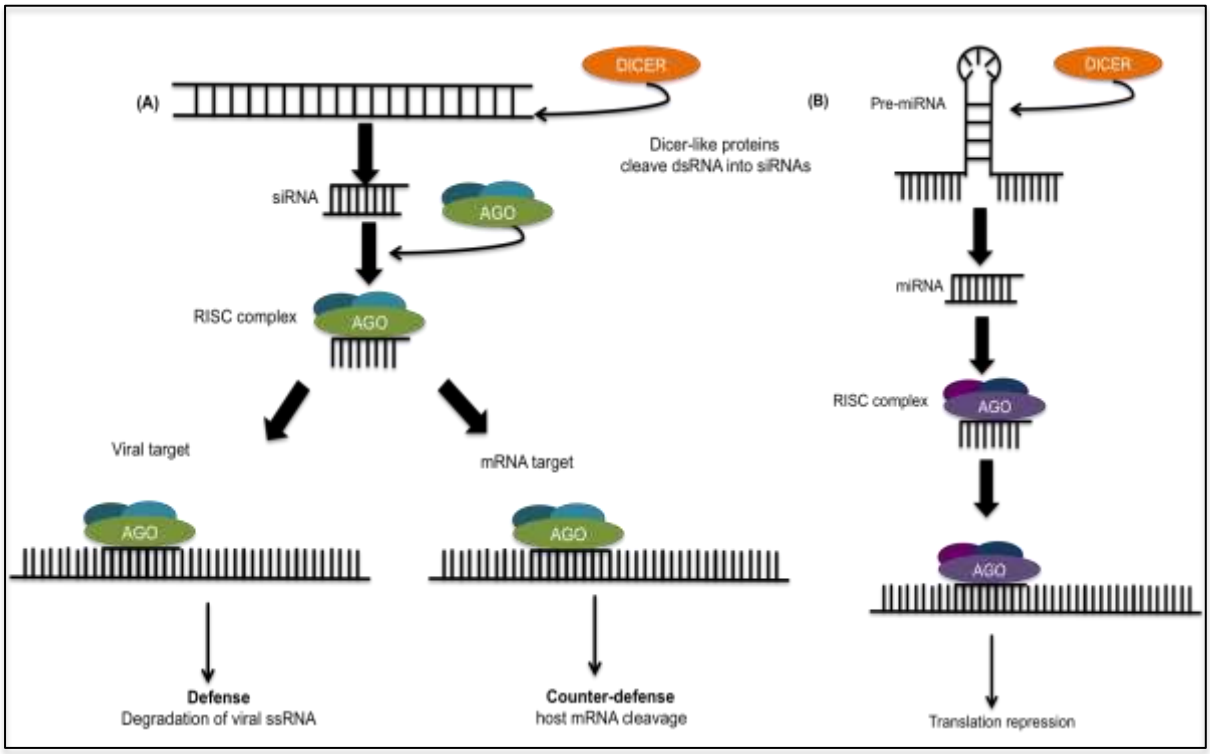

Figure 2.2: The small RNA silencing pathway. **Source**: Redrawn and simplified from (McManus and Sharp, 2002).

Traditional methods coupled with NGS technologies will aid in the best possible classification and characterisation of viruses. As NGS data makes available the genomic information, traditional methods describe other important properties of the virus. Both the traditional and NGS technologies have drawbacks, and it is more feasible to focus on the strengths of each approach in order to obtain informative and conclusive results. Investigating the RNA silencing pathway will further elucidate host pathogen interactions, which will lead to the establishment of disease control strategies either by developing diagnostic tools for the rapid screening of viruses or by breeding disease resistant cultivars suitable for the value market chain.

2.13 References

ABARSHI, M., MOHAMMED, I., JEREMIAH, S., LEGG, J., KUMAR, P. L., HILLOCKS, R. & MARUTHI, M. 2012. Multiplex RT-PCR assays for the simultaneous detection of both RNA and DNA viruses infecting cassava and the common occurrence of mixed infections by two cassava brown streak viruses in East Africa. *Journal of Virological Methods,* 179**,** 176-184.

ACHON, M., ALONSO-DUEÑAS, N. & SERRANO, L. 2011. Maize dwarf mosaic virus diversity in the Johnsongrass native reservoir and in maize: evidence of geographical, host and temporal differentiation. *Plant Pathology,* 60**,** 369-377.

ADAMS, I., ABIDRABO, P., MIANO, D., ALICAI, T., KINYUA, Z., CLARKE, J., MACARTHUR, R., WEEKES, R., LAURENSON, L. & HANY, U. 2013a. High throughput real-time RT-PCR assays for specific detection of cassava brown streak disease causal viruses, and their application to testing of planting material. *Plant Pathology,* 62**,** 233-242.

ADAMS, I., MIANO, D., KINYUA, Z., WANGAI, A., KIMANI, E., PHIRI, N., REEDER, R., HARJU, V., GLOVER, R. & HANY, U. 2013b. Use of next-generation sequencing for the identification and characterization of Maize chlorotic mottle virus and Sugarcane mosaic virus causing maize lethal necrosis in Kenya. *Plant Pathology,* 62**,** 741-749.

AGINDOTAN, B. O., SHIEL, P. J. & BERGER, P. H. 2007. Simultaneous detection of potato viruses, PLRV, PVA, PVX and PVY from dormant potato tubers by TaqMan® real-time RT-PCR. *Journal of Virological Methods,* 142**,** 1-9.

ALA-POIKELA, M., SVENSSON, E., ROJAS, A., HORKO, T., PAULIN, L., VALKONEN, J. & KVARNHEDEN, A. 2005. Genetic diversity and mixed infections of begomoviruses infecting tomato, pepper and cucurbit crops in Nicaragua. *Plant Pathology,* 54**,** 448-459.

ALABI, O. J., KUMAR, P. L. & NAIDU, R. A. 2008. Multiplex PCR for the detection of African cassava mosaic virus and East African cassava mosaic Cameroon virus in cassava. *Journal of Virological Methods,* 154**,** 111-120.

ALAVANDI, S. & POORNIMA, M. 2012. Viral metagenomics: a tool for virus discovery and diversity in aquaculture. *Indian Journal of Virology,* 23**,** 88-98.

ANSORGE, W. J. 2009. Next-generation DNA sequencing techniques. *New Biotechnology,* 25**,** 195-203.

ARITUA, V., BUA, B., BARG, E., VETTEN, H., ADIPALA, E. & GIBSON, R. 2007. Incidence of five viruses infecting sweetpotatoes in Uganda; the first evidence of Sweet potato caulimo-like virus in Africa. *Plant Pathology,* 56**,** 324-331.

AROGUNDADE, O., BALOGUN, O. S. & KAREEM, K. T. 2012. Occurrence and distribution of pepper veinal mottle virus and cucumber mosaic virus in pepper in Ibadan, Nigeria. *Virology Journal,* 9**,** 79.

ATEKA, E., BARG, E., NJERU, R., LESEMANN, D.-E. & VETTEN, H. 2004a. Further characterization of 'sweet potato virus 2': a distinct species of the genus Potyvirus. *Archives of Virology,* 149**,** 225-239.

ATEKA, E., NJERU, R., KIBARU, A., KIMENJU, J., BARG, E., GIBSON, R. & VETTEN, H. 2004b. Identification and distribution of viruses infecting sweet potato in Kenya. *Annals of Applied Biology,* 144**,** 371-379.

BARBA, M., CZOSNEK, H. & HADIDI, A. 2014. Historical perspective, development and applications of next-generation sequencing in plant virology. *Viruses,* 6**,** 106-136.

BAULCOMBE, D. 2004. RNA silencing in plants. *Nature,* 431**,** 356-363.

BENTLEY, D. R. 2006. Whole-genome re-sequencing. *Current Opinion in Genetics & Development,* 16**,** 545-552.

BIBBY, K. 2013. Metagenomic identification of viral pathogens. *Trends in Biotechnology,* 31**,** 275-279.

BOONHAM, N., KREUZE, J., WINTER, S., VAN DER VLUGT, R., BERGERVOET, J., TOMLINSON, J. & MUMFORD, R. 2014. Methods in virus diagnostics: from ELISA to next generation sequencing. *Virus Research,* 186**,** 20-31.

BOONHAM, N., WALSH, K., SMITH, P., MADAGAN, K., GRAHAM, I. & BARKER, I. 2003. Detection of potato viruses using microarray technology: towards a generic method for plant viral disease diagnosis. *Journal of Virological Methods,* 108**,** 181-187.

BORAH, B. K., SHARMA, S., KANT, R., JOHNSON, A., SAIGOPAL, D. V. R. & DASGUPTA, I. 2013. Bacilliform DNA-containing plant viruses in the tropics: commonalities within a genetically diverse group. *Molecular Plant Pathology,* 14**,** 759-771.

BOUHIDA, M., LOCKHART, B. & OLSZEWSKI, N. E. 1993. An analysis of the complete sequence of a sugarcane bacilliform virus genome infectious to banana and rice. *Journal of General Virology,* 74**,** 15-22.

BRIESE, T., PAWESKA, J. T., MCMULLAN, L. K., HUTCHISON, S. K., STREET, C., PALACIOS, G., KHRISTOVA, M. L., WEYER, J., SWANEPOEL, R. & EGHOLM, M. 2009. Genetic detection and characterization of Lujo virus, a new hemorrhagic fever–associated arenavirus from southern Africa. *PLoS Pathogens,* 5**,** e1000455.

BZHALAVA, D. & DILLNER, J. 2013. Bioinformatics for viral metagenomics. *Journal of Data Mining Genomics Proteomics,* 4**,** 134.

C., V., A., O., A., P. J., C., M. & M., C. 2015. Biological and molecular characterization of a distinct Citrus tristeza virus isolate originating from a lemon tree in Greece. *Plant Pathology,* 64**,** 792-798.

ÇAĞLAR, B. K., FIDAN, H. & ELBEAINO, T. 2013. Detection and molecular characterization of Pepper mild mottle virus from Turkey. *Journal of Phytopathology,* 161**,** 434-438.

CLARK, C. A., DAVIS, J. A., ABAD, J. A., CUELLAR, W. J., FUENTES, S., KREUZE, J. F., GIBSON, R. W., MUKASA, S. B., TUGUME, A. K. & TAIRO, F. D. 2012. Sweetpotato viruses: 15 years of progress on understanding and managing complex diseases. *Plant Disease,* 96**,** 168-185.

CLARK, M. F. & ADAMS, A. 1977. Characteristics of the microplate method of enzyme-linked immunosorbent assay for the detection of plant viruses. *Journal of General Virology,* 34**,** 475-483.

COETZEE, B., FREEBOROUGH, M.-J., MAREE, H. J., CELTON, J.-M., REES, D. J. G. & BURGER, J. T. 2010. Deep sequencing analysis of viruses infecting grapevines: virome of a vineyard. *Virology,* 400**,** 157-163.

COSTA, A. T., BRAVO, J. P., MAKIYAMA, R. K., VASCONCELLOS NUNES, A. & MAIA, I. G. 2013. Viral counter defense X antiviral immunity in plants: mechanisms for survival. *Romanowski, V., Current issues in molecular virology-viral genetics and biotechnological applications, InTech*.

CUELLAR, W. J., CRUZADO, R. K., FUENTES, S., UNTIVEROS, M., SOTO, M. & KREUZE, J. F. 2011. Sequence characterization of a Peruvian isolate of Sweet potato chlorotic stunt virus: further variability and a model for p22 acquisition. *Virus Research,* 157**,** 111-115.

CUELLAR, W. J., TAIRO, F., KREUZE, J. F. & VALKONEN, J. P. 2008. Analysis of gene content in sweet potato chlorotic stunt virus RNA1 reveals the presence of the p22 RNA silencing suppressor in only a few isolates: implications for viral evolution and synergism. *Journal of General Virology,* 89**,** 573-582.

DASGUPTA, I., DAS, B. K., NATH, P. S., MUKHOPADHYAY, S., NIAZI, F. R. & VARMA, A. 1996. Detection of rice tungro bacilliform virus in field and glasshouse samples from India using the polymerase chain reaction. *Journal of Virological Methods,* 58**,** 53-8.

DAY, J. M., BALLARD, L. L., DUKE, M. V., SCHEFFLER, B. E. & ZSAK, L. 2010. Metagenomic analysis of the turkey gut RNA virus community. *Virology Journal,* 7**,** 313.

DE MAGALHÃES, J. P., FINCH, C. E. & JANSSENS, G. 2010. Next-generation sequencing in aging research: emerging applications, problems, pitfalls and possible solutions. *Ageing Research Reviews,* 9**,** 315-323.

DELWART, E. L. 2007. Viral metagenomics. *Reviews in Medical Virology,* 17**,** 115-131.

DJIKENG, A., KUZMICKAS, R., ANDERSON, N. G. & SPIRO, D. J. 2009. Metagenomic analysis of RNA viruses in a fresh water lake. *PloS One,* 4**,** e7264.

DOMBROVSKY, A., GLANZ, E., LACHMAN, O., SELA, N., DORON-FAIGENBOIM, A. & ANTIGNUS, Y. 2013. The complete genomic sequence of Pepper yellow leaf curl virus (PYLCV) and its implications for our understanding of evolution dynamics in the genus Polerovirus. *PloS One,* 8**,** e70722.

DOMOLA, M. J. 2006. *Survey and characterisation of sweet potato viruses in South Africa.*

DONAIRE, L., WANG, Y., GONZALEZ-IBEAS, D., MAYER, K. F., ARANDA, M. A. & LLAVE, C. 2009. Deep-sequencing of plant viral small RNAs reveals effective and widespread targeting of viral genomes. *Virology,* 392**,** 203-214.

DOVAS, C., EYTHYMIOU, K. & KATIS, N. 2004. First report of maize rough dwarf virus (MRDV) on maize crops in Greece. *Plant Pathology,* 53**,** 238-238.

DUAN, C.-G., WANG, C.-H. & GUO, H.-S. 2012. Application of RNA silencing to plant disease resistance. *Silence,* 3**,** 5.

EAMENS, A., WANG, M.-B., SMITH, N. A. & WATERHOUSE, P. M. 2008. RNA silencing in plants: yesterday, today, and tomorrow. *Plant physiology,* 147**,** 456-468.

EDWARDS, R. A. & ROHWER, F. 2005. Viral metagenomics. *Nature Reviews Microbiology,* 3**,** 504-510.

EGAN, A. N., SCHLUETER, J. & SPOONER, D. M. 2012. Applications of next-generation sequencing in plant biology. Botanical Soc America.

ESTERHUIZEN, L., VAN HEERDEN, S., REY, M. & VAN HEERDEN, H. 2012. Genetic identification of two sweet-potato-infecting begomoviruses in South Africa. *Archives of Virology,* 157**,** 2241-2245.

FAUQUET, C. M., BISARO, D., BRIDDON, R., BROWN, J., HARRISON, B., RYBICKI, E., STENGER, D. & STANLEY, J. 2003. Virology division news: revision of taxonomic criteria for species demarcation in the family Geminiviridae, and an updated list of begomovirus species. *Archives of Virology,* 148**,** 405-421.

FAUQUET, C. M. & STANLEY, J. 2003. Geminivirus classification and nomenclature: progress and problems. *Annals of Applied Biology,* 142**,** 165-189.

FINKBEINER, S. R., ALLRED, A. F., TARR, P. I., KLEIN, E. J., KIRKWOOD, C. D. & WANG, D. 2008. Metagenomic analysis of human diarrhea: viral detection and discovery. *PLoS Pathogens,* 4**,** e1000011.

FONSECA, N. A., RUNG, J., BRAZMA, A. & MARIONI, J. C. 2012. Tools for mapping high-throughput sequencing data. *Bioinformatics***,** bts605.

FREEMAN, B. C. & BEATTIE, G. A. 2008. An overview of plant defenses against pathogens and herbivores. *The Plant Health Instructor*.

FUKUTA, S., TAMURA, M., MAEJIMA, H., TAKAHASHI, R., KUWAYAMA, S., TSUJI, T., YOSHIDA, T., ITOH, K., HASHIZUME, H. & NAKAJIMA, Y. 2013. Differential detection of Wheat yellow mosaic virus, Japanese soil-borne wheat mosaic virus and Chinese wheat mosaic virus by reverse transcription loop-mediated isothermal amplification reaction. *Journal of Virological Methods,* 189**,** 348-354.

GEERING, A. D., MAUMUS, F., COPETTI, D., CHOISNE, N., ZWICKL, D. J., ZYTNICKI, M., MCTAGGART, A. R., SCALABRIN, S., VEZZULLI, S. & WING,

R. A. 2014. Endogenous florendoviruses are major components of plant genomes and hallmarks of virus evolution. *Nature Communications,* 5.

GIBSON, R., MPEMBE, I., ALICAI, T., CAREY, E., MWANGA, R., SEAL, S. & VETTEN, H. 1998. Symptoms, aetiology and serological analysis of sweet potato virus disease in Uganda. *Plant Pathology,* 47**,** 95-102.

GIBSON, R., MWANGA, R., KASULE, S., MPEMBE, I. & CAREY, E. 1997. Apparent absence of viruses in most symptomless field-grown sweet potato in Uganda. *Annals of Applied Biology,* 130**,** 481-490.

GLASA, M., PREDAJŇA, L., KOMÍNEK, P., NAGYOVÁ, A., CANDRESSE, T. & OLMOS, A. 2014. Molecular characterization of divergent grapevine Pinot gris virus isolates and their detection in Slovak and Czech grapevines. *Archives of Virology,* 159**,** 2103-2107.

GOPI, V., GOPAL, K., SANKAR, T. G. & PALANIVEL, S. 2010. Detection of citrus yellow mosaic virus by PCR and nucleic acid spot hybridisation using non-radioactive probes in commercial citrus species. *Archives of Phytopathology and Plant Protection,* 43**,** 892-899.

GU, Y.-H., TAO, X., LAI, X.-J., WANG, H.-Y. & ZHANG, Y.-Z. 2014. Exploring the polyadenylated RNA virome of sweet potato through high-throughput sequencing. *PloS One,* 9**,** e98884.

GUO, C., LI, L., WANG, X. & LIANG, C. 2015. Alterations in siRNA and miRNA expression profiles detected by deep sequencing of transgenic rice with siRNA-mediated viral resistance. *PLoS One,* 10**,** e0116175.

GUTIERREZ, D., FUENTES, S. & SALAZAR, L. 2003. Sweetpotato virus disease (SPVD): distribution, incidence, and effect on sweetpotato yield in Peru. *Plant Disease,* 87**,** 297-302.

HAIBLE, D., KOBER, S. & JESKE, H. 2006. Rolling circle amplification revolutionizes diagnosis and genomics of geminiviruses. *Journal of Virological Methods,* 135**,** 9-16.

HANLEY-BOWDOIN, L., BEJARANO, E. R., ROBERTSON, D. & MANSOOR, S. 2013. Geminiviruses: masters at redirecting and reprogramming plant processes. *Nature Revies Microbiology,* 11**,** 777-788.

HANLEY-BOWDOIN, L., SETTLAGE, S. B., OROZCO, B. M., NAGAR, S. & ROBERTSON, D. 1999. Geminiviruses: models for plant DNA replication, transcription, and cell cycle regulation. *Critical Reviews in Plant Sciences,* 18**,** 71-106.

HANSEN, C. & HESLOP-HARRISON, J. 2004. Sequences and phylogenies of plant pararetroviruses, viruses, and transposable elements. *Advances in Botanical Research,* 41**,** 165-193.

HORNER, D. S., PAVESI, G., CASTRIGNANÒ, T., DE MEO, P. D. O., LIUNI, S., SAMMETH, M., PICARDI, E. & PESOLE, G. 2010. Bioinformatics approaches for genomics and post genomics applications of next-generation sequencing. *Briefings in Bioinformatics,* 11**,** 181-197.

HU, G., DONG, Y., ZHANG, Z., FAN, X., FANG, R. & ZHU, H. 2014. Detection and sequence analysis of grapevine virus B isolates from China. *Acta Virologica,* 58**,** 180-184.

HULL, R. 2009. *Comparative plant virology,* Academic press.

ISHAK, J. & EL-DEEB, S. 2004. Investigating the effects of Sweetpotato chlorotic stunt virus (SPCSV) infection to sweetpotato plants using light and electron microscopy/Licht-und elektronenmikroskopische Untersuchungen von Süßkartoffelpflanzen nach Infektion mit dem Sweetpotato chlorotic stunt virus. *Zeitschrift für Pflanzenkrankheiten und Pflanzenschutz/Journal of Plant Diseases and Protection***,** 362-370.

JAMES, A., GEIJSKES, R. J., DALE, J. L. & HARDING, R. M. 2011. Molecular characterisation of six badnavirus species associated with leaf streak disease of banana in East Africa. *Annals of Applied Biology,* 158**,** 346-353.

KAKIRDE, K. S., PARSLEY, L. C. & LILES, M. R. 2010. Size does matter: application-driven approaches for soil metagenomics. *Soil Biology and Biochemistry,* 42**,** 1911-1923.

KAPINGA, R. & CAREY, E. 2003. Present status of sweetpotato breeding for eastern and southern Africa. *Sweetpotato post harvest assessment: Experiences from East Africa, NRI, CPHP, DFID, CIP and Ministry of Agriculture Tanzania, Chatman, UK,* 88.

KASHIF, M., PIETILÄ, S., ARTOLA, K., JONES, R., TUGUME, A., MÄKINEN, V. & VALKONEN, J. 2012. Detection of viruses in sweetpotato from Honduras and Guatemala augmented by deep-sequencing of small-RNAs. *Plant Disease,* 96**,** 1430-1437.

KAYS, S. J. Sweetpotato production worldwide: Assessment, trends and the future. I International Symposium on Root and Tuber Crops: Food Down Under 670, 2004. 19-25.

KHAN, J. A. & DIJKSTRA, J. 2001. *Plant viruses as molecular pathogens,* CRC Press.

KOKKINOS, C., CLARK, C., MCGREGOR, C. & LABONTE, D. 2006. The effect of sweet potato virus disease and its viral components on gene expression levels in sweetpotato. *Journal of the American Society for Horticultural Science,* 131**,** 657-666.

KOKKINOS, C. D. 2006. Assessment of interactions amoung viruses infecting sweetpotato.

KOOLIVAND, D., SOKHANDAN-BASHIR, N., BEHJATNIA, S. A. A. & JAFARI JOOZANI, R. A. 2014. Detection of Grapevine fanleaf virus by immunocapture reverse transcription-polymerase chain reaction (IC-RT-PCR) with recombinant antibody. *Archives of Phytopathology and Plant Protection,* 47**,** 2070-2077.

KREUZE, J. 2002. *Molecular studies on the sweet potato virus disease and its two causal agents*.

KREUZE, J. 2014. siRNA deep sequencing and assembly: piecing together viral infections. *Detection and Diagnostics of Plant Pathogens.* Springer.

KREUZE, J. & FUENTES, S. 2008. Sweetpotato Viruses A2 - Mahy, Brian W.J. *In:* REGENMORTEL, M. H. V. V. (ed.) *Encyclopedia of Virology (Third Edition).* Oxford: Academic Press.

KREUZE, J. F., PEREZ, A., UNTIVEROS, M., QUISPE, D., FUENTES, S., BARKER, I. & SIMON, R. 2009. Complete viral genome sequence and discovery of novel viruses by deep sequencing of small RNAs: a generic method for diagnosis, discovery and sequencing of viruses. *Virology,* 388**,** 1-7.

KUMAR, J., KUMAR, J., SINGH, S. P. & TULI, R. 2014. Association of satellites with a mastrevirus in natural infection: complexity of wheat dwarf India virus disease. *Journal of Virology,* 88**,** 7093-7104.

KUNDU, J. 2003. The occurrence of Apple stem pitting virus and Apple stem grooving virus within field-grown apple cultivars evaluated by RT-PCR. *Plant Protection Science,* 39**,** 88-92.

KUTNJAK, D., SILVESTRE, R., CUELLAR, W., PEREZ, W., MÜLLER, G., RAVNIKAR, M. & KREUZE, J. 2014. Complete genome sequences of new divergent potato virus X isolates and discrimination between strains in a mixed infection using small RNAs sequencing approach. *Virus Research,* 191**,** 45-50.

LAM, N., CREAMER, R., RASCON, J. & BELFON, R. 2009. Characterization of a new curtovirus, pepper yellow dwarf virus, from chile pepper and distribution in weed hosts in New Mexico. *Archives of Virology,* 154**,** 429-436.

LANEY, A. G., HASSAN, M. & TZANETAKIS, I. E. 2012. An integrated badnavirus is prevalent in fig germplasm. *Phytopathology,* 102**,** 1182-1189.

LATHAM, J. R. & WILSON, A. K. 2008. Transcomplementation and synergism in plants: implications for viral transgenes? *Molecular Plant Pathology,* 9**,** 85-103.

LEE, H. C., LAI, K., LORENC, M. T., IMELFORT, M., DURAN, C. & EDWARDS, D. 2012. Bioinformatics tools and databases for analysis of next-generation sequence data. *Briefings in Functional Genomics,* 11**,** 12-24.

LEGRAND, P. 2015. Biological assays for plant viruses and other graft-transmissible pathogens diagnoses: a review. *EPPO Bulletin,* 45**,** 240-251.

LEO, G., LUISON, D., GRANDE, S. B., ALBANESE, G. & FAGGIOLI, F. 2015. Grapevine Viruses' Detection and Sanitary Selection in Grapevine Germplasm of Calabria (Southern Italy). *Journal of Phytopathology,* 163**,** 690-693.

LI, F., ZUO, R., ABAD, J., XU, D., BAO, G. & LI, R. 2012a. Simultaneous detection and differentiation of four closely related sweet potato potyviruses by a multiplex one-step RT-PCR. *Journal of Virological Methods,* 186**,** 161-166.

LI, R., GAO, S., HERNANDEZ, A. G., WECHTER, W. P., FEI, Z. & LING, K.-S. 2012b. Deep sequencing of small RNAs in tomato for virus and viroid identification and strain differentiation. *PloS One,* 7**,** e37127.

LI, R., SALIH, S. & HURTT, S. 2004. Detection of geminiviruses in sweetpotato by polymerase chain reaction. *Plant Disease,* 88**,** 1347-1351.

LIN, C. Y., WU, M. L., SHEN, T. L., YEH, H. H. & HUNG, T. H. 2015. Multiplex detection, distribution, and genetic diversity of Hop stunt viroid and Citrus exocortis viroid infecting citrus in Taiwan. *Virology Journal,* 12.

LING, K.-S., JACKSON, D. M., HARRISON, H., SIMMONS, A. M. & PESIC-VANESBROECK, Z. 2010. Field evaluation of yield effects on the USA heirloom sweetpotato cultivars infected by Sweet potato leaf curl virus. *Crop Protection,* 29**,** 757-765.

LIU, S., VIJAYENDRAN, D. & BONNING, B. C. 2011. Next generation sequencing technologies for insect virus discovery. *Viruses,* 3**,** 1849-1869.

LOCONSOLE, G., CASTELLANO, M. A., DELL'ORCO, M., BOSCIA, D. & SAVINO, V. 2006. SEROLOGICAL DETECTION OF CITRUS PSOROSIS VIRUS USING A POLYCLONAL ANTISERUM TO RECOMBINANT VIRUS COAT PROTEIN. *Journal of Plant Pathology,* 88**,** 171-177.

LOEBENSTEIN, G., FUENTES, S., COHEN, J. & SALAZAR, L. 2003. Sweet potato. *Virus and virus-like diseases of major crops in developing countries.* Springer.

LOMMEL, S., MCCAIN, A. & MORRIS, T. 1982. Evaluation of indirect enzyme-linked immunosorbent assay for the detection of plant viruses. *Phytopathology,* 72**,** 1018-1022.

LÓPEZ, M. M., BERTOLINI, E., OLMOS, A., CARUSO, P., GORRIS, M. T., LLOP, P., PENYALVER, R. & CAMBRA, M. 2003. Innovative tools for detection of plant pathogenic viruses and bacteria. *International Microbiology,* 6**,** 233-243.

LÓPEZ, M. M., LLOP, P., OLMOS, A., MARCO-NOALES, E., CAMBRA, M. & BERTOLINI, E. 2009. Are molecular tools solving the challenges posed by detection of plant pathogenic bacteria and viruses? *Current issues in Molecular Biology,* 11**,** 13.

LYTTLE, D. J., ORLOVICH, D. A. & GUY, P. L. 2011. Detection and analysis of endogenous badnaviruses in the New Zealand flora. *AoB Plants,* 2011**,** plr008.

MACKENZIE, T. D., NIE, X. & SINGH, M. 2015. RT-PCR and real-time RT-PCR methods for the detection of potato virus Y in potato leaves and tubers. *Plant Virology Protocols.* Springer.

MADIGAN, M. T., MARTINKO, J. M. & PARKER, J. 1997. *Brock biology of microorganisms*, prentice hall Upper Saddle River, NJ.

MAGI, A., BENELLI, M., GOZZINI, A., GIROLAMI, F., TORRICELLI, F. & BRANDI, M. L. 2010. Bioinformatics for next generation sequencing data. *Genes,* 1**,** 294-307.

MALLIK, I., ANDERSON, N. R. & GUDMESTAD, N. C. 2012. Detection and differentiation of Potato virus Y strains from potato using immunocapture multiplex RT-PCR. *American Journal of Potato Research,* 89**,** 184-191.

MANAMELA, M. T. 2009. Morphological characterisation and cryopreservation of sweet potato, Ipomoea batatas (L.) Lam., accessions at the NPGRC of South Africa.

MANSOOR, S., BRIDDON, R. W., ZAFAR, Y. & STANLEY, J. 2003. Geminivirus disease complexes: an emerging threat. *Trends in Plant Science,* 8**,** 128-134.

MAR, T. B., LAU, D., SCHONS, J., PEREIRA, P. R. V. D. S. & CARMINATTI, A. J. 2013. Identification and characterization of wheat streak mosaic virus isolates in wheat-growing areas in Brazil. *International Journal of Agronomy,* 2013.

MARDIS, E. R. 2008. The impact of next-generation sequencing technology on genetics. *Trends in Genetics,* 24**,** 133-141.

MARTIN, F. 1998. Sweet potato. *Sweet potato*.

MASSART, S., OLMOS, A., JIJAKLI, H. & CANDRESSE, T. 2014. Current impact and future directions of high throughput sequencing in plant virus diagnostics. *Virus Research,* 188**,** 90-96.

MBANZIBWA, D., TAIRO, F., GWANDU, C., KULLAYA, A. & VALKONEN, J. 2016. First report of sweetpotato symptomless virus 1 and sweetpotato virus A in sweetpotatoes in Tanzania. *Plant Disease,* 100**,** 865.

MBANZIBWA, D., TUGUME, A., CHIUNGA, E., MARK, D. & TAIRO, F. 2014. Small RNA deep sequencing-based detection and further evidence of DNA viruses infecting sweetpotato plants in Tanzania. *Annals of Applied Biology,* 165**,** 329-339.

MCMANUS, M. T. & SHARP, P. A. J. N. R. G. 2002. Gene silencing in mammals by small interfering RNAs. 3**,** 737.

MEN, A. E., WILSON, P., SIEMERING, K. & FORREST, S. 2008. Sanger DNA sequencing. *Next-Generation Genome Sequencing: Towards Personalized Medicine***,** 1-11.

MENZEL, W., JELKMANN, W. & MAISS, E. 2002. Detection of four apple viruses by multiplex RT-PCR assays with coamplification of plant mRNA as internal control. *Journal of Virological Methods,* 99**,** 81-92.

METZKER, M. L. 2010. Sequencing technologies—the next generation. *Nature Reviews Genetics,* 11**,** 31-46.

MILLER, J. R., KOREN, S. & SUTTON, G. 2010. Assembly algorithms for next-generation sequencing data. *Genomics,* 95**,** 315-327.

MLOTSHWA, S., PRUSS, G. J. & VANCE, V. 2008. Small RNAs in viral infection and host defense. *Trends in Plant Science,* 13**,** 375-382.

MONGER, W., SEAL, S., COTTON, S. & FOSTER, G. 2001. Identification of different isolates of Cassava brown streak virus and development of a diagnostic test. *Plant Pathology,* 50**,** 768-775.

MUKASA, S. B., RUBAIHAYO, P. R. & VALKONEN, J. 2006. Interactions between a crinivirus, an ipomovirus and a potyvirus in coinfected sweetpotato plants. *Plant Pathology,* 55**,** 458-467.

MUKASA, S. B., RUBAIHAYO, P. R. & VALKONEN, J. P. 2003. Incidence of viruses and virus like diseases of sweetpotato in Uganda. *Plant Disease,* 87**,** 329-335.

NAIDU, R. & HUGHES, J. 2003. Methods for the detection of plant virus diseases. *Plant Virology in Sub Saharan Africa***,** 233-253.

NDERITU, J., SILAI, M., NYAMASYO, G. & KASINA, M. 2009. Insect species associated with sweetpotatoes (Ipomoea batatas (L.) Lam.) in eastern Kenya. *Int J Sustain Crop Prod,* 4**,** 14-18.

NG, T. F. F., MANIRE, C., BORROWMAN, K., LANGER, T., EHRHART, L. & BREITBART, M. 2009. Discovery of a novel single-stranded DNA virus from a sea turtle fibropapilloma by using viral metagenomics. *Journal of Virology,* 83**,** 2500-2509.

NICOLAISEN, M. 2011. An oligonucleotide-based microarray for detection of plant RNA viruses. *Journal of Virological Methods,* 173**,** 137-143.

NYABOGA, E., ATEKA, E. & BULIMO, W. 2008. Serological detection of virus diseases of sweet potato in Kenya. *Journal of Applied Biosciences,* 7**,** 222-229.

O'BRIEN, P. J. 1972. The sweet potato: its origin and dispersal. *American Anthropologist***,** 342-365.

PALLAS, V. & GARCÍA, J. A. 2011. How do plant viruses induce disease? Interactions and interference with host components. *Journal of General Virology,* 92**,** 2691-2705.

PAPROTKA, T., BOITEUX, L., FONSECA, M., RESENDE, R., JESKE, H., FARIA, J. & RIBEIRO, S. 2010. Genomic diversity of sweet potato geminiviruses in a Brazilian germplasm bank. *Virus Research,* 149**,** 224-233.

PARDINA, P. R., BEJERMAN, N., LUQUE, A. & DI FEO, L. 2012. Complete nucleotide sequence of an Argentinean isolate of sweet potato virus G. *Virus Genes,* 45**,** 593-595.

PENG, J., SHI, B., ZHENG, H., LU, Y., LIN, L., JIANG, T., CHEN, J. & YAN, F. 2015. Detection of pepper mild mottle virus in pepper sauce in China. *Archives of virology,* 160**,** 2079-2082.

PRABHA, K., BARANWAL, V. & JAIN, R. 2013. Applications of next generation high throughput sequencing technologies in characterization, discovery and molecular interaction of plant viruses. *Indian Journal of Virology,* 24**,** 157-165.

PRICE, J., SMITH, J., SIMMONS, A., FELLERS, J. & RUSH, C. 2010. Multiplex real-time RT-PCR for detection of Wheat streak mosaic virus and Tritcum mosaic virus. *Journal of Virological Methods,* 165**,** 198-201.

RÄNNÄLI, M., CZEKAJ, V., JONES, R., FLETCHER, J., DAVIS, R., MU, L. & VALKONEN, J. 2009. Molecular characterization of Sweet potato feathery mottle virus (SPFMV) isolates from Easter Island, French Polynesia, New Zealand, and southern Africa. *Plant Disease,* 93**,** 933-939.

REDDY, D., SUDARSHANA, M., FUCHS, M., RAO, N. & THOTTAPPILLY, G. 2009. Genetically engineered virus-resistant plants in developing countries: current status and future prospects. *Advances in Virus Research,* 75**,** 185-220.

RIALCH, N., SHARMA, V., SHARMA, A. & SHARMA, P. N. 2015. Characterization and complete nucleotide sequencing of pepper mild mottle virus infecting bell pepper in India. *Phytoparasitica,* 43**,** 327-337.

ROOKS, D. J., SMITH, D. L., MCDONALD, J. E., WOODWARD, M. J., MCCARTHY, A. J. & ALLISON, H. E. 2010. 454-pyrosequencing: a molecular battiscope for freshwater viral ecology. *Genes,* 1**,** 210-226.

ROOSSINCK, M. J. 2012. Plant virus metagenomics: biodiversity and ecology. *Annual Review of Genetics,* 46**,** 359-369.

ROY, A., CHOUDHARY, N., GUILLERMO, L. M., SHAO, J., GOVINDARAJULU, A., ACHOR, D., WEI, G., PICTON, D. D., LEVY, L., NAKHLA, M. K., HARTUNG, J. S. & BRLANSKY, R. H. 2012. A Novel Virus of the Genus Cilevirus Causing Symptoms Similar to Citrus Leprosis. *Phytopathology,* 103**,** 488-500.

RYBICKI, E. P. & HUGHES, F. L. 1990. Detection and typing of maize streak virus and other distantly related geminiviruses of grasses by polymerase chain reaction amplification of a conserved viral sequence. *Journal of General Virology,* 71**,** 2519-2526.

SAKAI, J., MORI, M., MORISHITA, T., TANAKA, M., HANADA, K., USUGI, T. & NISHIGUCHI, M. 1997. Complete nucleotide sequence and genome organization of sweet potato feathery mottle virus (S strain) genomic RNA: the large coding region of the P1 gene. *Archives of Virology,* 142**,** 1553-1562.

SANKARAN, S., MISHRA, A., EHSANI, R. & DAVIS, C. 2010. A review of advanced techniques for detecting plant diseases. *Computers and Electronics in Agriculture,* 72**,** 1-13.

SAPONARI, M., MANJUNATH, K. & YOKOMI, R. K. 2008. Quantitative detection of Citrus tristeza virus in citrus and aphids by real-time reverse transcription-PCR (TaqMan®). *Journal of Virological Methods,* 147**,** 43-53.

SCHATZ, M. C., DELCHER, A. L. & SALZBERG, S. L. 2010. Assembly of large genomes using second-generation sequencing. *Genome Research,* 20**,** 1165-1173.

SHENDURE, J. & JI, H. 2008. Next-generation DNA sequencing. *Nature Biotechnology,* 26**,** 1135-1145.

SMITH, T. F. & WATERMAN, M. S. 1981. Identification of common molecular subsequences. *Journal of Molecular Biology,* 147**,** 195-197.

SOOSAAR, J. L., BURCH-SMITH, T. M. & DINESH-KUMAR, S. P. 2005. Mechanisms of plant resistance to viruses. *Nature Reviews Microbiology,* 3**,** 789-798.

SOUTO, E., SIM, J., CHEN, J., VALVERDE, R. & CLARK, C. 2003. Properties of strains of Sweet potato feathery mottle virus and two newly recognized potyviruses infecting sweet potato in the United States. *Plant Disease,* 87**,** 1226-1232.

STAVOLONE, L., HERZOG, E., LECLERC, D. & HOHN, T. 2001. Tetramerization is a conserved feature of the virion-associated protein in plant pararetroviruses. *Journal of Virology,* 75**,** 7739-7743.

STUDHOLME, D. J., GLOVER, R. H. & BOONHAM, N. 2011. Application of high-throughput DNA sequencing in phytopathology. *Annual Review of Phytopathology,* 49**,** 87-105.

SU, X., FU, S., QIAN, Y., XU, Y. & ZHOU, X. 2015. Identification of Hop stunt viroid infecting Citrus limon in China using small RNAs deep sequencing approach. *Virology Journal,* 12**,** 103.

TAIRO, F., JONES, R. A. & VALKONEN, J. P. 2006. Potyvirus complexes in sweetpotato: occurrence in Australia, serological and molecular resolution, and analysis of the Sweet potato virus 2 (SPV2) component. *Plant Disease,* 90**,** 1120-1128.

TAMMI, M. T. 2003. The principles of shotgun sequencing and automated fragment assembly. *Center for Genomics and Bioinformatics, Karolinska Institute, Stockholm, Sweden*.

TESFAYE, T., FEYISSA, T. & ABRAHAM, A. 2011. Survey and serological detection of sweet potato (Ipomoea batatas (L.) Lam) viruses in Ethiopia. *Journal of Applied Biosciences,* 41**,** 2746-2756.

TRENADO, H. P., ORÍLIO, A. F., MÁRQUEZ-MARTÍN, B., MORIONES, E. & NAVAS-CASTILLO, J. 2011. Sweepoviruses cause disease in sweet potato and related Ipomoea spp.: fulfilling koch's postulates for a divergent group in the genus Begomovirus. *PloS One,* 6**,** e27329.

TUGUME, A. K., MUKASA, S. B., KALKKINEN, N. & VALKONEN, J. P. 2010. Recombination and selection pressure in the ipomovirus Sweet potato mild mottle virus (Potyviridae) in wild species and cultivated sweetpotato in the centre of evolution in East Africa. *Journal of General Virology,* 91**,** 1092-1108.

TZANETAKIS, I. E. & MARTIN, R. R. 2008. A new method for extraction of double-stranded RNA from plants. *Journal of Virological Methods,* 149**,** 167-170.

UEHARA-ICHIKI, T., SHIBA, T., MATSUKURA, K., UENO, T., HIRAE, M. & SASAYA, T. 2013. Detection and diagnosis of rice-infecting viruses. *Frontiers in Microbiology,* 4.

UNTIVEROS, M., QUISPE, D. & KREUZE, J. 2010. Analysis of complete genomic sequences of isolates of the Sweet potato feathery mottle virus strains C and EA: molecular evidence for two distinct potyvirus species and two P1 protein domains. *Archives of Virology,* 155**,** 2059-2063.

VALE, F., CORREIA, A., MATOS, B., MOURA NUNES, J. & ALVES DE MATOS, A. 2010. Applications of transmission electron microscopy to virus detection and identification. *Microscopy: Science. Technology, Applications and Education; Mendez-Vilas, A., Diaz, J., Eds***,** 128-136.

VALVERDE, R. A., CLARK, C. A. & VALKONEN, J. P. 2007. Viruses and virus disease complexes of sweetpotato. *Plant Viruses,* 1**,** 116-126.

VAN DIJK, E. L., JASZCZYSZYN, Y. & THERMES, C. 2014. Library preparation methods for next-generation sequencing: tone down the bias. *Experimental Cell Research,* 322**,** 12-20.

VAN JAARSVELD, P. J., FABER, M., TANUMIHARDJO, S. A., NESTEL, P., LOMBARD, C. J. & BENADÉ, A. J. S. 2005. β-Carotene–rich orange-fleshed sweet potato improves the vitamin A status of primary school children assessed with the modified-relative-dose-response test. *The American Journal of Clinical Nutrition,* 81**,** 1080-1087.

VAN VLIET, A. H. 2010. Next generation sequencing of microbial transcriptomes: challenges and opportunities. *FEMS Microbiology Letters,* 302**,** 1-7.

VANDERSCHUREN, H., STUPAK, M., FÜTTERER, J., GRUISSEM, W. & ZHANG, P. 2007. Engineering resistance to geminiviruses–review and perspectives. *Plant Biotechnology Journal,* 5**,** 207-220.

VAUCHERET, H. 2006. Post-transcriptional small RNA pathways in plants: mechanisms and regulations. *Genes & Development,* 20**,** 759-771.

VICTORIA, J. G., KAPOOR, A., LI, L., BLINKOVA, O., SLIKAS, B., WANG, C., NAEEM, A., ZAIDI, S. & DELWART, E. 2009. Metagenomic analyses of viruses in stool samples from children with acute flaccid paralysis. *Journal of Virology,* 83**,** 4642-4651.

VISSER, M., MAREE, H. J., REES, D. J. G. & BURGER, J. T. 2014. High-throughput sequencing reveals small RNAs involved in ASGV infection. *BMC Genomics,* 15**,** 568.

WATPADE, S., RAIGOND, B., THAKUR, P., HANDA, A., PRAMANICK, K., SHARMA, Y. & TOMAR, M. 2012. Molecular detection of latent Apple chlorotic leaf spot virus in elite mother plants of apple. *Indian Journal of Virology,* 23**,** 359-363.

WEBSTER, C. G., WYLIE, S. J. & JONES, M. G. 2004. Diagnosis of plant viral pathogens. *Current Science,* 86**,** 1604-1607.

WOO, E., CLOVER, G. & PEARSON, M. 2012. First report of Cherry leaf roll virus (CLRV) in Malus domestica. *Australasian Plant Disease Notes,* 7**,** 151-156.

XIA, Z., PENG, J., LI, Y., CHEN, L., LI, S., ZHOU, T. & FAN, Z. 2014. Characterization of small interfering RNAs derived from Sugarcane mosaic virus in infected maize plants by deep sequencing. *PLoS One,* 9**,** e97013.

XIE, Y., JIAO, X., ZHOU, X., LIU, H., NI, Y. & WU, J. 2013. Highly sensitive serological methods for detecting tomato yellow leaf curl virus in tomato plants and whiteflies. *Virology Journal,* 10**,** 142.

YAMASAKI, S., SAKAI, J., FUJI, S., KAMISOYAMA, S., EMOTO, K., OHSHIMA, K. & HANADA, K. 2010. Comparisons among isolates of Sweet potato feathery mottle virus using complete genomic RNA sequences. *Archives of Virology,* 155**,** 795-800.

ZAITLIN, M. 1998. The discovery of the causal agent of the tobacco mosaic disease. World Scientific Publishing Co Pte Ltd: Singapore.

ZHANG, Y., LUBBERSTEDT, T. & XU, M. 2013. The genetic and molecular basis of plant resistance to pathogens. *Journal of Genetics and Genomics,* 40**,** 23-35.

ZHANG, Y., SINGH, K., KAUR, R. & QIU, W. 2011a. Association of a novel DNA virus with the grapevine vein-clearing and vine decline syndrome. *Phytopathology,* 101**,** 1081-1090.

ZHANG, Z.-Y., LIU, X.-J., LI, D.-W., YU, J.-L. & HAN, C.-G. 2011b. Rapid detection of wheat yellow mosaic virus by reverse transcription loop-mediated isothermal amplification. *Virology Journal,* 8**,** 550.

CHAPTER 3

VIRAL METAGENOMICS REVEALS SWEET POTATO RNA VIRUS DIVERSITY IN SOUTH AFRICA

3.1 Abstract

Limited studies have been undertaken with regard to viruses contributing to the aetiology of sweet potato virus disease (SPVD) in South Africa. In this study, a metagenomic approach was adopted in order to establish the genetic diversity of viruses infecting sweet potato. In order to undertake a comprehensive analysis of viral sequences, total RNA was isolated from 17 asymptomatic and symptomatic sweet potato plants that were collected from the Eastern (EC) and Western Cape (WC) provinces of South Africa. DNase-treated total RNA was depleted of ribosomal RNA (rRNA) and deep-sequenced using the Illumina MiSeq platform. Both *de novo* and reference-guided assemblies were performed resulting in four near full-length RNA virus genomes. Sequences from the transcriptome of two ssDNA geminiviruses were detected in symptomatic plants from the Western Cape province. The sequence data further demonstrated mixed infections of RNA and DNA viruses from 11 of the 17 samples. We recommend the use of deep sequencing of nucleic acids as a reliable diagnostic tool for virus detection as well as for differentiating between diverse viral strains.

3.2 Introduction

Sweet potato (*Ipomoea batatas* (L.) Lam.) is a dicotyledonous perennial plant belonging to the morning glory family *Convolvulaceae*. It produces edible, highly nutritious tubers and is ranked as the third most important root crop and the seventh most important staple crop in the world (Clark et al., 2012, Valverde et al., 2007). Sweet potato is easy to grow, high yielding, is drought and heat tolerant, and crowds out weeds, and therefore an attractive crop to resource-poor farmers (Kays, 2004). In countries such as Zambia, South Africa (SA), Uganda, Kenya, Nigeria, and Tanzania women are the primary growers of sweet potato and they use it to generate income. Sweet potatoes have a high content of carbohydrates and dietary fibre; they are rich in vitamin A, vitamin C, vitamin B6 and because of these nutritional benefits they are used for poverty alleviation (van Jaarsveld et al., 2005). Despite its importance, a range of DNA and RNA viruses, which may accumulate as a result of vegetative propagation (Valverde et al., 2007), pose a serious threat to sweet potato

production (Kreuze et al., 2009, Tesfaye et al., 2011, Valverde et al., 2007). It has been recorded that viral diseases can decrease yield and quality of sweet potato storage roots by 30 – 100 % in countries such as the United States (Valverde et al., 2007), Peru (Cuellar et al., 2008), SA (Domola et al., 2008), Kenya (Ateka et al., 2004) and Ethiopia (Tesfaye et al., 2011).

The most prominent viral disease complex known to affect sweet potato worldwide is sweet potato virus disease (SPVD), which was observed for the first time in Uganda in 1940 (Ateka et al., 2004). Although two viruses, sweet potato chlorotic stunt virus (SPCSV) (family *Closteroviridae*) (Kreuze et al., 2002) and sweet potato feathery mottle virus (SPFMV) (family *Potyviridae*) (Clark et al., 2012, Cuellar et al., 2008) are known to cause SPVD, there are reports in SA suggesting the involvement of several other viruses as role-players in sweet potato disease complexes (Domola et al., 2008, Tesfaye et al., 2011). The last comprehensive survey in SA documenting viral prevalence was carried out in 2003 and employed serology (ELISA assays), reverse transcription PCR (RT-PCR), electron microscopy and indexing methods for virus detection (Domola et al., 2008, Sivparsad and Gubba, 2012). Using these traditional methods, a previous survey detected 9 viruses including SPFMV, SPCSV, sweet potato mild mottle ipomovirus (SPMMV) (family *Potyviridae*), sweet potato chlorotic fleck virus (SPCFV) (family *Flexiviridae*), sweet potato caulimo-like virus (SPCV) (family *Caulimoviridae*), sweet potato virus G (SPVG) (family *Potyviridae*), sweet potato virus 2 (SPV2) (family *Potyviridae*), sweet potato latent (SPLV) (family *Potyviridae)*, and sweet potato mild speckling virus (SPMSV) (family *Potyviridae*) (Domola et al., 2008). More recently, two geminiviruses, sweet potato leaf curl Sao Paulo virus (SPLCSPV) and sweet potato mosaic virus (SPMaV) were detected in SA using rolling circle amplification (RCA), cloning and sequencing (Esterhuizen et al., 2012).

In the last two decades, virus detection methods have shifted from traditional techniques to metagenomic approaches coupled with high throughput sequencing (Boonham et al., 2014). Viral metagenomics coupled with next generation sequencing (NGS) have been used to identify novel viruses in plants (Idris et al., 2014, Kreuze and Fuentes, 2008). This approach is considered an unbiased one for viral detection since no prior knowledge of the virus is necessary, and neither virus-specific primers, nor antibodies are required. Consequently novel viruses, if present, can be detected, identified and quantified in a single experiment (Studholme et al., 2011). However, in the absence of reference sequences, the use of high

throughput sequencing for virus detection would require *de novo* genome assembly of new viruses, which can be a challenge. A metagenomic approach also means that the entire microbial community within a sample can be described, even in mixed viral interactions, thus simplifying diagnostics (Idris et al., 2014). In most cases, the virus sequences generated in a metagenomic study would form a small proportion of the total nucleic acids making the removal of host sequences critical prior to, or after sequencing (Stobbe and Roossinck, 2014). For this reason, enrichment methods such as isolation of dsRNA (Clark et al., 2012) and small interfering RNA (siRNAs) (Kreuze et al., 2009) have been employed to detect DNA and RNA viruses from different hosts (Kashif et al., 2012).

The availability of NGS platforms such as those supplied by Illumina (Illumina Inc., San Diego, CA, USA) has further revolutionised viral metagenomics studies. NGS technologies generate large amounts of data rapidly at reduced costs and many bioinformatic tools have been developed to handle data analysis (Massart et al., 2014). This study was carried out with the objective of establishing the current status of sweet potato viruses in two South African provinces.

3.3 Materials and methods

3.3.1 Sources of plant material

Sweet potato cuttings were collected from four smallholder farms in the EC province of South Africa: Alice (32°47'13.6"S 26°50'56.8"E), Zwide (33°52'12.8"S 25°34'24.0"E), Kwazakhele (33°53'11.0"S 25°36'00.7"E), and Motherwell (33°48'08.3"S 25°35'47.0"E). In the WC province, material was collected from four commercial farms, the locations included Paarl (33°40'12.0"S 18°58'08.0"E), Klawer (31°46'59.0"S 18°37'00.0"E), Franschhoek (33°55'00.1"S 19°07'59.9"E) and Lutzville (31°33'11.0"S 18°12'57.0"E). The cuttings were transplanted to potting soil in 20 cm diameter pots and grown in a greenhouse at the Agricultural Research Council – Vegetable and Ornamental Plant Institute (ARC-VOPI) in Pretoria, South Africa (25°40'51.67"S 28°17'10.25"E). Plants were grown at optimum temperatures of 25°C for 16 h (day cycle) and 15°C for 8 h (night cycle) (Domola et al., 2008). Plants were watered once a day and soluble nutrient fertilization (Multifeed P, Plaaskem, Pty, LTD) was applied on a weekly basis. Insect pests were also monitored and controlled by spraying with recommended insecticides as required. A list of samples, abbreviations and symptoms are depicted in Table 3.2.

3.3.2 Description of symptoms

Sweet potato plants collected from the field were maintained in the glasshouse and observed for symptom development over a period of six months. Plants exhibiting symptoms typical of viral infection, such as upward curling of the leaves, chlorotic spots, vein clearing, and purple ring spots, were selected for analysis by sequencing. Symptom severity was scored using a 1-5 scoring scale (Domola et al., 2008, Mwanga et al., 2001) where, 1 = no virus symptoms, 2 = mild symptoms (chlorotic spots), 3 = moderate symptoms (chlorosis, chlorotic spots and vein clearing), 4 = severe symptoms (chlorotic spots, leaf curl, and leaf puckering/necrosis) and 5 = very severe symptoms (chlorotic spots, leaf curl, mottling, and stunting).

3.3.4 RNA library preparation and sequencing

After a period of six months, ten symptomatic and seven asymptomatic plants were randomly selected for further analysis (Table 3.2). Total RNA was isolated from the 17 samples using the QIAGEN RNeasy Plant Mini Kit (QIAGEN, Valencia, CA, USA) following the manufacturer's instructions. To remove DNA, total RNA underwent DNase treatment using the QIAGEN RNase-free DNase Set (QIAGEN, Valencia, CA, USA). The integrity of the extracted RNA was analysed by agarose gel electrophoresis and quantified using the Qubit™ RNA BR Assay Kit (Invitrogen, Life Technologies). The total RNA was stored at -80°C until further use. Total RNA was treated with the Ribo-Zero™ Magnetic Kit (Plant Leaf) (Epicentre, Madison, WI, USA) to deplete ribosomal RNA (rRNA). RNA paired-end libraries were prepared using the Illumina TruSeq™ Stranded Total RNA Sample Preparation Kit (Illumina Inc., San Diego, CA, USA) following the manufacturer's instructions. The libraries were quantified using the Qubit™ dsDNA HS Assay Kit (Invitrogen, Life Technologies). The libraries were sequenced on the Illumina MiSeq (Illumina Inc., San Diego, CA, USA).

3.3.5 Sequence assembly and alignments

Adaptor removal and quality trimming was performed using Fastq-mcf (Aronesty, 2013) and the quality of the sequence reads was analysed using FastQC. A quality threshold of 30 and a Phred score of 33 were selected for trimming options. Sequence reads below the length of 50 bp and greater than 180 bp were discarded. The trimmed reads were aligned to the sweet potato chloroplast sequence, and sweet potato expressed sequence tags (ESTs) to subtract host sequences. The unmapped sequence reads were assembled into contigs using the CLC Genomics Workbench *de novo* assembly tool, with default parameters. The contigs were subjected to BLASTn and BLASTx searches against viral sequences downloaded from the

NCBI database. To generate consensus sequences for phylogenetic analysis, sequence reads and contigs matching sweet potato viruses were mapped to the full genomes of the closest hits, using alignment settings: length fraction = 0.5 and similarity = 0.9. The newly assembled sequences for SPVG, SPFMV, SPVC, SPCSV RNA1 and SPCSV RNA2 were deposited in GenBank under accession numbers KT069224, KT069222, KT069223, KX932096 and KT069221, respectively.

3.3.6 Phylogenetic analysis

The complete genome sequences of SPFMV, SPVG, SPVC, and SPCSV were retrieved from the NCBI database and used for multiple sequence alignments (MSA) and phylogenetic analysis in MEGA 6.06 (Tamura et al., 2013). The neighbour-joining method was used to generate phylogenetic trees. The bootstrap tests were conducted using 1000 replicates and the evolutionary distances were computed using the Jukes-Cantor method.

3.3.7 RT-PCR confirmation of viruses identified by NGS

Full-length genomes of closely related isolates of SPVG, SPFMV, SPVC, and SPCSV were downloaded and used in multiple sequence alignments (MSA). Alignments were submitted to the IDT PrimerQuest Tool (http://eu.idtdna.com/primerquest/home/index) for primer design. Conserved regions such as the coat protein gene were targeted for primer design. The primer sequences used in the study are listed in Table 3.1. Since RNA viruses were identified in sample KT10, total RNA was isolated from KT10 using the RNeasy Plant Mini Kit (QIAGEN, Valencia, CA, USA). The total RNA was converted to cDNA using the TaKaRa PrimeScipt 1st strand cDNA synthesis kit (TaKaRa, Japan). The polymerase chain reactions (PCRs) were performed using the TaKaRa EmeraldAmp® GT PCR Master Mix (TaKaRa, Japan) following the manufacturers instructions. PCRs consisted of 12.5µl of the TaKaRa Emerald GT PCR Master Mix (TaKaRa, Japan), 0.2 µM of each primer, 2µl of the template DNA and 9µl of nuclease-free water. The PCRs were performed using the recommended thermal cycling conditions: Initial denaturation at 94°C for 5 min, 35 cycles of 94°C for 30 s, annealing temperature (57-63°C) for 45 s, extension at 72°C for 2 min, and the final extension for 10 min at 72°C. The amplicons were visualised on a 2% agarose gel by electrophoresis. The PCR amplicons were sent to Inqaba Biotechnical Industries (Pty) Ltd. service provider for Sanger sequencing. Sequences were analysed using the Sequence Scanner Software v2.0 (Applied Biosystems). The edited sequences were then subjected to BLASTn and BLASTx searches to determine their identities.

Table 3.1: Oligonucleotide sequences used for confirming the identity of RNA viruses

Virus Name	Primer Name	Sequence (5'-3')	Amplicon Size (bp)
Sweet potato feathery mottle virus	SPFMV-F	TCCAGACCCTAAGTCCAAGAT	587
	SPFMV-R	AGTGCGTCATAATCTGCCTAAA	
Sweet potato chlorotic stunt virus	SPCSV-F	CTCTGACTCCGATGTAGGTTTC	590
	SPCSV-R	CGGTTGCAAGATGCCAATAC	
Sweet potato virus G	SPVG-F	GGAAACACAGGAAGAGGAAGAG	689
	SPVG-R	GGGACAGCATGATCCAATAGAG	
Sweet potato virus C	SPVC-F	GGCCATATACAGCACCAGAAA	338
	SPVC-R	TTCCTGAGTTGAGCGTGTATTC	

3.4 Results

3.4.1 Field symptoms associated with viral infection

In the field a variety of symptoms were observed on sweet potato plants. Symptoms ranged from upward curling of leaves, purple-edged vein feathering and purple ring spots in samples KT10, F11, KF1 and L18; to chlorotic spots in KT6, P2, Z24, and KZ17; and vein clearing in Z24, while other samples were asymptomatic (Table 3.2 and Figure 3.1). Symptom severity scoring of field samples can be viewed in Table 3.2. The most severe symptoms were observed on sample KF1, which was collected from the WC province. After one week of harvesting the leaves from sample KF1, the whole plant died. Samples F11 and KT10 displayed moderate symptoms characterised by purple ring spots and leaf curling respectively. These plants were also collected from the WC province. The phenotypic data suggests that sweet potato viruses were more prevalent in the WC, since most of the samples collected from this province were symptomatic. Interestingly, during sample collection in the WC province, whiteflies were observed on the sweet potato leaves. These could be possible vectors of several viruses that are associated with the symptoms recorded (e.g. leaf curling is associated with begomoviruses). Only two of the six samples from the EC province (KZ17 and Z24) displayed noticeable virus symptoms (Table 3.2). Furthermore during field collection sweet potato leaves were either symptomless or displayed mild symptoms in each of the four sampling locations of the EC province.

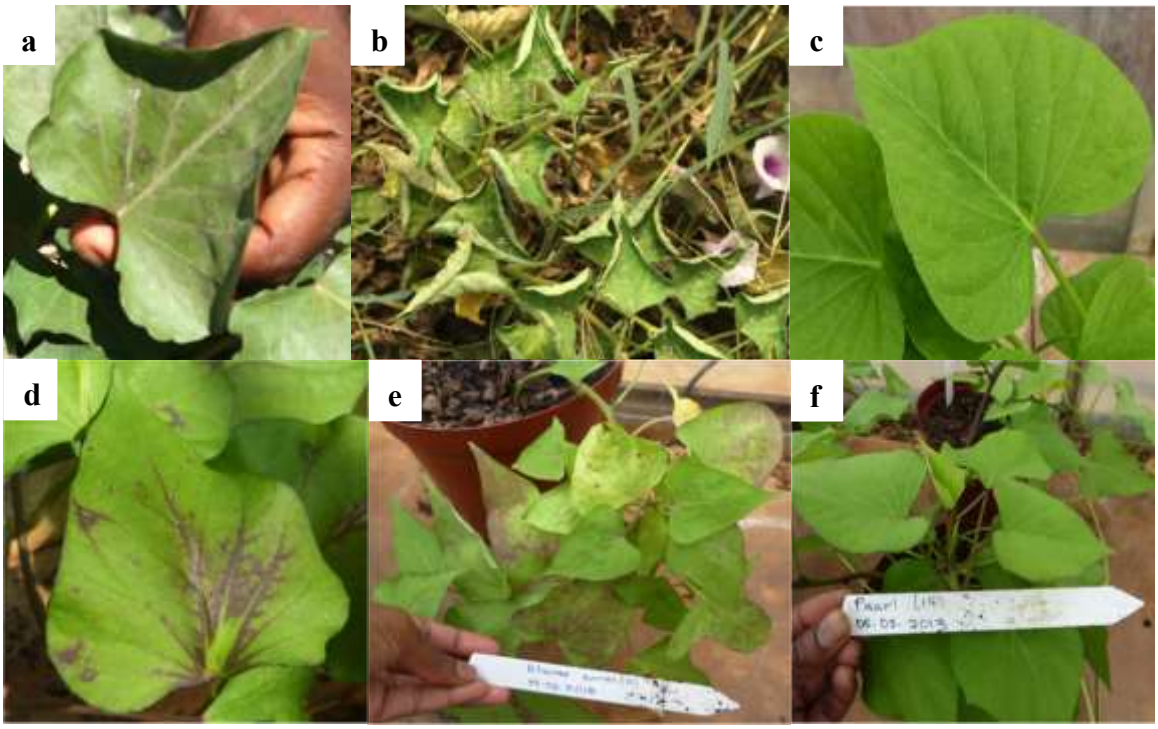

Figure 3.1: Chlorotic spots with purple rings observed on sweet potato plants during field collection in KwaZakhele (Eastern Cape) (a). Sweet potato plants from Klawer (Western Cape) exhibited upward curling of the leaves (b). In the glasshouse KZ17 displayed mild chlorotic spots (c). F11 displayed purple ring spots and purple edged vein feathering (d), KT10 was characterised by leaf curling and purple ring spots (e), and P14 showing no virus symptoms (f).

3.4.3 Sequence data and *de novo* assembly

The 17 individually labelled RNA libraries were sequenced to generate approximately 7 gigabases (Gb) of data. The primary data consisted of over 19 million sequences. After quality control, 56% of the data was retained for further analysis. Sixty eight percent of the retained data aligned to sweet potato chloroplast and EST sequences. For each sample, sequence reads that did not map to host sequences were assembled to generate contigs varying in number and size (Table 3.3). The d*e novo* assemblies generated large numbers of contigs, with the largest contigs being in the range of 5 – 10 kb, while the N50 values were in the range 300 – 486 bp (Table 3.3).

Table 3.2: Symptoms and disease scoring of field plants collected from the Eastern and Western Cape provinces of South Africa

Sample ID	Location	Province	Symptoms	Scoring
KT10[a]	Klawer	Western Cape	Leaf curl, purple ring spots	3
F11	Franschoek	Western Cape	Purple-edged vein feathering, purple ring spots	3
KF1[a]	Klawer	Western Cape	Leaf curl, purple ring spots, stunting	4
L18	Lutzville	Western Cape	Purple ring spots	2
KT6[a]	Klawer	Western Cape	Chlorotic spots	2
P2	Paarl	Western Cape	Chlorotic spots	2
L9	Lutzville	Western Cape	Leaf curl, purple ring spots	3
F4	Franschoek	Western Cape	Necrotic spots	2
K10	Klawer	Western Cape	Asymptomatic	1
P14	Paarl	Western Cape	Asymptomatic	1
L11	Lutzville	Western Cape	Asymptomatic	1
Z24	Zwide	Eastern Cape	Chlorotic spots, vein clearing	3
KZ17	KwaZakhele	Eastern Cape	Chlorotic spots	2
M19[b]	Motherwell	Eastern Cape	Asymptomatic	1
FH22	Fort Hare	Eastern Cape	Asymptomatic	1
M11[b]	Motherwell	Eastern Cape	Asymptomatic	1
FH14	Fort Hare	Eastern Cape	Asymptomatic	1

Footnote: [a] During collection plants were infested with whiteflies. [b] Insect damage was observed on sweet potato leaves.

Table 3.3: Summary of the *de novo* assembly analysis

Sample ID	Sequence reads before filtering	Sequence reads after QC analysis	Sequence reads mapped to host sequences	Unmapped sequence reads	*de novo* assembly		
					Number of contigs	Maximum contig length	N50
KT10	4605084	1474192	929232	544960	3374	10474[a]	336
F11	1400366	1295800	914635	381165	4836	6291[a]	347
KF1	835460	537401	308419	228982	3370	10436[a]	343
L18	541556	521024	341882	179142	4928	10540[a]	430
KT6	931874	551084	399984	151100	2439	5383	337
P2	777294	452674	267809	184865	2390	5646	340
L9	656312	357033	244743	112290	1202	5491	324
F4	724812	334552	220941	113611	1558	3373	319
K10	888036	570397	410811	159586	2118	2826	344
P14	557134	535138	369592	165546	6142	5386	422
L11	584506	566361	447747	118614	3846	5382	389
Z24	1602288	820006	584570	235436	4365	7101	355
KZ17	2015640	1594713	1106418	490295	11207	10442[a]	394
M19	789620	403739	210279	193460	3817	5386	331
FH22	398602	380539	265215	1115324	4767	5386	486
M11	653108	379272	257294	1211978	2708	5403	347
FH14	797028	448469	323391	125078	1814	5519	344

Footnote: *De novo* assembly analysis was performed for each individual sample (library) to identify infecting viruses. [a] Contigs resembling full virus genomes were obtained after host sequence filtration. More detail (virus name, number of sequence reads and coverage) of the identified viruses is given in Table 3.4a & 3.4b.

3.4.4 Detection of RNA viruses

The analysis of the assembled contigs was undertaken by matching all contigs to the viral sequences in the NCBI database using BLASTn and BLASTx. Overall, the majority of contigs showed no similarity to viral sequences, but significant matches to known sweet potato viruses were found in a large number of samples isolated from symptomatic plants (e.g. KT10, KF1, KT6, F11, L18, L9, P2, F4, Z24 and KZ17). Notably, five samples (M19, FH14, M11, K10 and F22) from asymptomatic plants had no detectable viral sequences in this analysis. Low counts of virus sequences were detected in two asymptomatic samples (L11 and P14) collected from the WC province. The viral sequences identified in each sample are shown in Table 3.4a. Viral sequences accounted for 1-2% of the total sequence data that was generated from the RNA libraries. In many cases the RNA virus genomes were

assembled into a single contig (Figure 3.2a and Table 3.4a) with a large number of reads supporting the contig assembly. In other instances there were lower levels of infection and only fragmentary assemblies of the genomes were achieved, however the total contig assemblies represented almost complete genome sequences when aligned with the reference genomes (Figure 3.2 and Table 3.4a).

The BLAST search results revealed the presence of SPCSV RNA1 segment and SPCSV RNA2 segment in 3 samples. The total assembled contigs for SPCSV RNA1 represented a maximum of 72% of the East African strain reference sequence (accession number AJ428554), while for RNA2 the contigs represented a maximum of 92% of the m2-74 RNA2 sequence (accession number HQ291260). The two largest contigs generated for RNA1 were from samples KT10 and L18, at lengths of 7965 bp and 7984 bp respectively. Alignment and pairwise comparisons of these two contigs showed that they were identical, except for the 5' and 3' ends, due to incomplete sequences. The RNA1 sequences shared 76% nucleotide (nt) similarity to the Ugandan reference sequence, while the RNA2 segment shared 97% nt identity to the Peruvian m2-74 isolate. This suggests that the two genome segments are of distinct origin, based on the widely divergent sequences observed. The two contigs matching to RNA1 of SPCSV were merged to generate a reference contig for extension with the PRICE genome assembler software using default parameters (Ruby et al., 2013). The final contig generated using PRICE was 8572 bp.

SPFMV was detected in 8 samples. The largest contig for SPFMV, 10540 bp, was assembled from sample L18. This contig covered 97% of the reference genome and shared 94% sequence identity with the SPFMV 10-O strain (accession number AB439206). Over 80% genome coverage was achieved for the *de novo* assembly of SPFMV in 7 samples (Table 3.4a) and over 50000 sequence reads were assembled into 45 contigs. SPVC was detected in 6 samples and the largest contig of 10442 bp (Table 3.4a), which was assembled from sample KZ17, had a sequence depth of 477-fold. The contig shared 95% nt identity with the SPVC isolate from Argentina (accession number KF386015) and maximum genome coverage of 96% was achieved. BLAST search results also revealed the presence of SPVG in samples KT10 and F11. The largest contig of 6291 bp was assembled from sample F11. A total of 4 contigs were generated for SPVG, and when these were aligned to the reference genome a consensus of 10577 bp, representing 97% of the genome, was generated (Figure 3.2c). The consensus sequence shared 98% nt identity with an isolate from Argentina (accession number

JQ824374). PCR amplification of the coat protein genes for the 4 RNA viruses was successful (Figure 3.4). The PCR amplicons were Sanger sequenced to confirm their identity. The Sanger sequenced amplicons are provided in Figure A1, Figure A2 and Table A5. The optimised PCR assay has the potential to be used in future studies for easy detection of SA virus isolates, especially SPCSV, which could not be detected previously using primers available in the literature.

3.4.5 Identification of DNA viruses

Interestingly gene transcripts of ssDNA of begomoviruses (SPLCSPV and SPMaV) and sweet potato caulimo-like virus (SPCV) were detected in the RNA dataset (Table 3.4a). This could have occurred possibly as a result of purification of DNA viral transcripts. High genome coverage was achieved for SPLCSPV detected in sample KT10. Only fragmentary assemblies were generated for SPMaV, which was detected in samples KT10 and KF1. The begomoviruses (SPLCSPV and SPMaV) were detected previously in the Limpopo province (Esterhuizen et al, 2012), and now detected in this study for the first time the WC province.

3.4.6 *De novo* assembly efficiency

When contigs were aligned to full-length viral genomes, gaps were observed in the consensus sequences. The only virus that was assembled with no gaps was SPVC (Figure 3.2a). From this study we noticed that near full-length virus genome could be *de novo* assembled using datasets of 75 to 300 Mb (e.g. samples L18, F11, KT10, KZ17, Z24 and KF1) (Table 3.4a). This resulted in overall high genome coverage and sequence depth. In cases where large amounts of data were generated and there was low viral sequence count (e.g. P14) or no virus detection (M19, FH14, M11, K10 and F22), it was concluded that viruses were either absent or present at very low concentrations. The sequence data is supported by the phenotypic data. Samples M19, FH14, M11, K10, F22 and P14 showed no visible symptoms prior to sequencing (Table 3.2 and Figure 3.1). The *de novo* assembly approach is effective for virus discovery and for the assembly of near complete viral sequences. This strategy is also efficient in the assembly of distinct viral sequences, where reference-guided assembly could pose a limitation.

3.4.7 Reference sequence-guided assembly

The reference-guided assembly showed that a total of 43224 sequence reads originated from SPFMV (Table 3.4b). SPVC specific sequence reads were 41265, while the crinivirus

(SPCSV) had a total of 12879 sequence reads, and only 3862 sequence reads mapped to SPVG. Four new South African RNA virus genomes were generated from the reference-guided assembly. The SPCSV RNA2 segment sequence was 8210 bp long with a sequence depth of 104-fold (accession number KT069221); the SPFMV genome sequence was 10803 bp long and had a sequence depth186-fold and the new SPVG isolate was 10739 bp long, with a sequence depth of 39-fold (accession number KT069224). The new South African SPVC genome was 1 nucleotide (nt) longer than the reference sequence (10794 bp), and was assembled with a sequence depth of 457-fold (accession number KT069223). Few sequence reads aligned to the Peruvian and Ugandan SPCSV RNA1, possibly as a result of high variability within the South African genome sequence. The longest assembled contig for SPCSV RNA1 (accession number KX932096) was generated from the *de novo* assembly.

3.4.8 Co-infections and mixed virus infections in sweet potato

The sequence data revealed co-infections of potyviruses (SPVC and SPFMV) in samples P2 and Z24 collected from the WC and EC provinces, respectively. Sequence reads matching SPFMV, SPVC and SPVG were detected from F11, which was collected from the WC. Three plant samples from the WC showed evidence of a virus complex including SPLCSPV, SPMaV, SPFMV and SPCSV (KF1, KT6 and L18) (Table 3.4a & 3.4b). A mixed infection of 6 viruses was detected in sample KT10 from the WC (Table 3.4a). The sequence data shows that sweet potato viruses were found mostly as co-infections and mixed infections in plant samples from both provinces.

3.4.9 Phylogenetic analysis of RNA viruses

Near complete sequences of SPVC, SPVG, SPFMV and SPCSV RNA2 generated from the reference-guided assemblies were used for phylogenetic analysis. Phylogenetic trees assigned RNA2 to the East African (EA) strain group (Figure 3.3a). The SPFMV sequence grouped with the ordinary strains (Figure 3.3b). The SPVC from the EC clustered with the SPVC group. The SPVG from the WC clustered with isolates from Taiwan, USA and South Korea (Figure 3.3b).

Table 3.4a Summary statistics of sweet potato viruses detected in symptomatic and asymptomatic plants using *de novo* assembly

Sampled ID	Data after QC (Mb)	Virus detected	Total number of virus contigs[a]	Total number of sequence reads[b]	Maximum contig length	Number of sequence reads[c]	Sequence depth of maximum contig
KT10	303.8	SPFMV	2	27129	10474	26638	225X
		SPCSV-RNA1	4	3649	7675	3365	41X
		SPCSV-RNA2	3	9725	4935	6467	123X
		SPMaV	4	1133	944	693	68X
		SPLCSPV	2	133	1016	133	11X
F11	222	SPVG	4	3941	6291	2094	36X
		SPVC	14	705	3202	210	7X
		SPFMV	3	1727	5649	965	18X
KF1	95.4	SPCSV-RNA1	5	66	696	39	6X
		SPCSV-RNA2	9	287	902	98	12X
		SPFMV	2	4968	10463	4885	52X
		SPLCSPV	3	108	849	81	10X
		SPMaV	1	27	393	27	7X
L18	115.3	SPCSV-RNA1	2	1302	7984	1281	25X
		SPCSV-RNA2	2	2953	7507	2279	46X
		SPFMV	1	14834	10540	14834	222X
		SPLCSPV	4	766	562	61	14X
KT6	99.4	SPFMV	7	955	2423	208	9X
		SPLCSPV	4	127	1089	87	8X
P2	82.1	SPFMV	3	1270	5646	660	13X
		SPVC	6	1176	3095	306	11X
L9	67	SPVC	11	573	2199	154	8X
		SPFMV	8	52	423	10	3X
F4	60.1	SPVC	11	138	955	33	3X
P14	99.6	SPCaLV	1	8	252	8	5X
L11	120.7	SPLCSPV	1	9	356	9	3X
Z24	126.6	SPVC	6	2832	7101	1335	22
		SPFMV	19	564	1214	97	9
KZ17	149.9	SPVC	1	36214	10442	36214	477X
M19	75.1	-	-	-	-	-	-
FH14	162.6	-	-	-	-	-	-
M11	159.1	-	-	-	-	-	-
K10	135.9	-	-	-	-	-	-
F22	89.5	-	-	-	-	-	-

Footnote: Near complete sequences of sweet potato feathery mottle virus (SPFMV), sweet potato virus G (SPVG), sweet potato virus C (SPVC), sweet potato chlorotic stunt virus (SPCSV) and short transcripts of DNA viruses sweet potato leaf curl Sao Paulo virus (SPLCSPV), sweet potato caulimo-like virus (SPCV) and sweet potato mosaic virus (SPMaV) were generated.[a] Contigs identified as viruses during BLASTn and BLASTx searches. [b] Overall number of virus-specific sequence reads. [c] Total number of reads used to assemble maximum contig.

Table 3.4b. Sweet potato viruses identified using reference-guided assembly. New full-length and partial virus genomes were assembled from less than 500 Mb of data

Sample ID	Virus reference genome	Number of sequence reads mapped to reference genome	Size of newly generated sequence (bp)	Genome coverage (%); sequence depth
KT10	SPFMV	22650	10803	(99); 186X
	SPCSV-RNA1	164	1419	(16); 1X
	SPCSV-RNA2	9101	8210	(99); 104X
	SPMaV	785	2544	(91); 26X
	SPLCSPV	510	2690	(97); 16X
	SPVC	171	785	(7); 0.9X
	SPVG	107	439	(4); 0.6X
F11	SPVG	3862	10739	(99); 39X
	SPVC	786	10161	(94); 7X
	SPFMV	1632	10192	(94); 16X
KF1	SPCSV-RNA1	2	230	(2); 0.03X
	SPCSV-RNA2	293	5948	(72); 3X
	SPFMV	4266	10392	(96); 44X
	SPLCSPV	111	1995	(72); 4X
	SPMaV	74	1421	(51); 2X
L18	SPCSV-RNA1	18	842	(9); 0.2X
	SPCSV-RNA2	3301	8193	(99); 62X
	SPFMV	12103	10585	(97); 175X
	SPLCSPV	127	2186	(78); 5X
KT6	SPFMV	839	9859	(91); 8X
	SPLCSPV	53	1612	(57); 2X
	SPMaV	84	1639	(58); 3X
P2	SPFMV	1103	10317	(95); 11X
	SPVC	1150	10534	(97); 12X
L9	SPVC	524	9676	(89); 5X
	SPFMV	99	5411	(50); 1X
F4	SPVC	239	8455	(78); 2X
P14	SPCV	9	344	(4); 0.1
L11	SPLCSPV	22	1172	(42); 1X
	SPMaV	8	603	(7); 0.4X
Z24	SPVC	2756	10486	(97); 30X
	SPFMV	532	9845	(90); 5X
KZ17	SPVC	35939	10794	(100); 457X
M19	-	-	-	-
FH14	-	-	-	-
K10	-	-	-	-
M11	-	-	-	-
F22	-	-	-	-

Footnote:Reference genomes used in reference-guided assembly; SPFMV (Accession number: AB439206); SPCSV-RNA1 (Accession number: HQ291259); SPCSV-RNA2 (Accession number: HQ291260); SPVC (Accession number: KF386015); SPVG (Accession number: JQ824374); SPMaV (Accession number: JQ621843): SPLCSPV (Accession number: JQ621844): SPCV (Accession number: NC_015328).

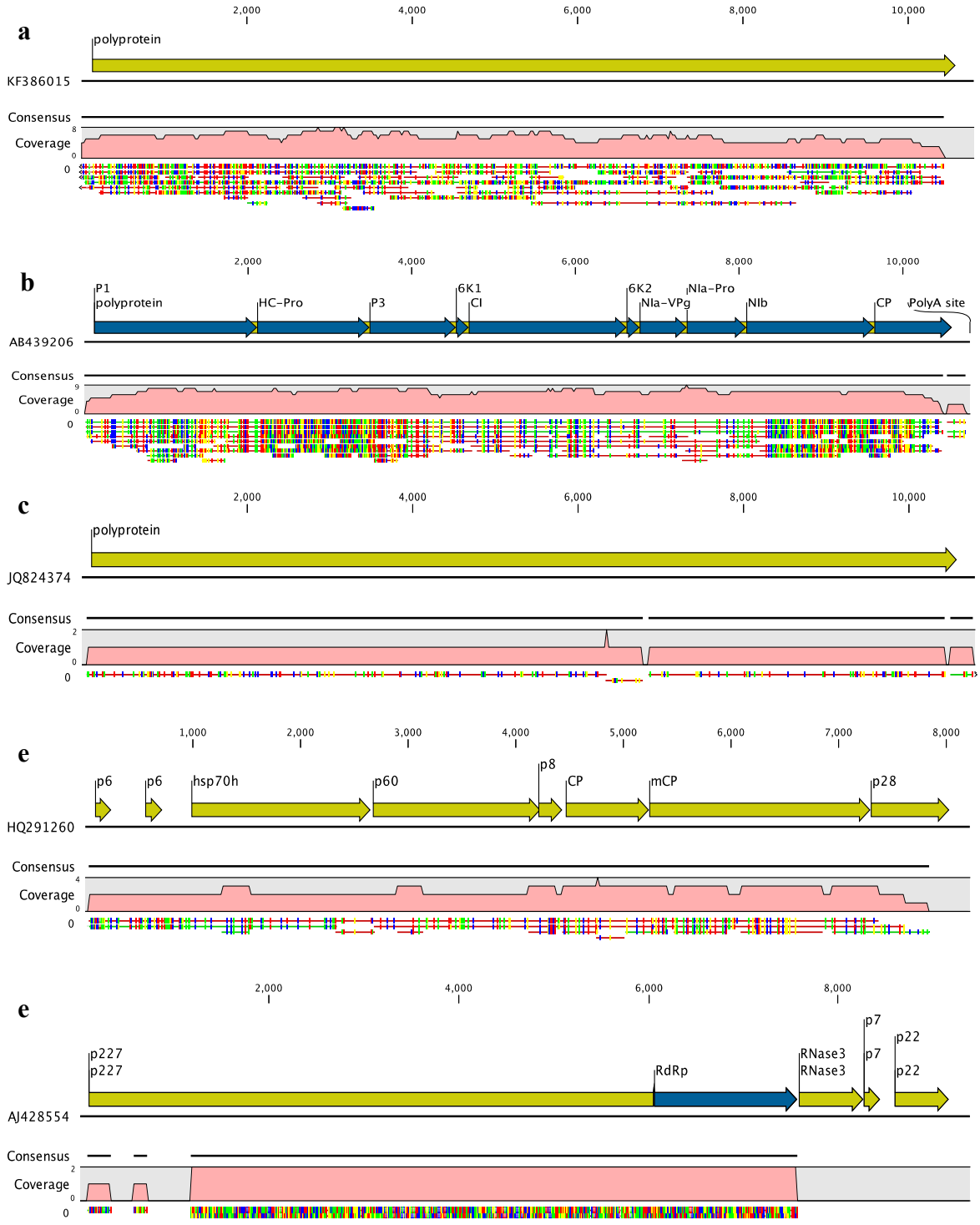

Figure 3.2: Genome coverage achieved by *de novo* assembly. Total contigs assembled for SPVC aligned along the full genome (a); contigs aligned to SPFMV covering the partial polyprotein sequence (b), the complete sequences of the HC-Pro, P3, 6K1, CI, 6K2, Nia-VPg, Nia-Pro, and NIb and partial CP genes were obtained; four contigs aligned to SPVG from sample F11 (c); the RNA2 segment of SPCSV had contigs aligned to the p6, hsp70h, p60, p8, CP, mCP, and partial p28 proteins (d); 4 contigs spanning over the p227 and RdRp proteins for RNA1 segment of SPCSV (e).

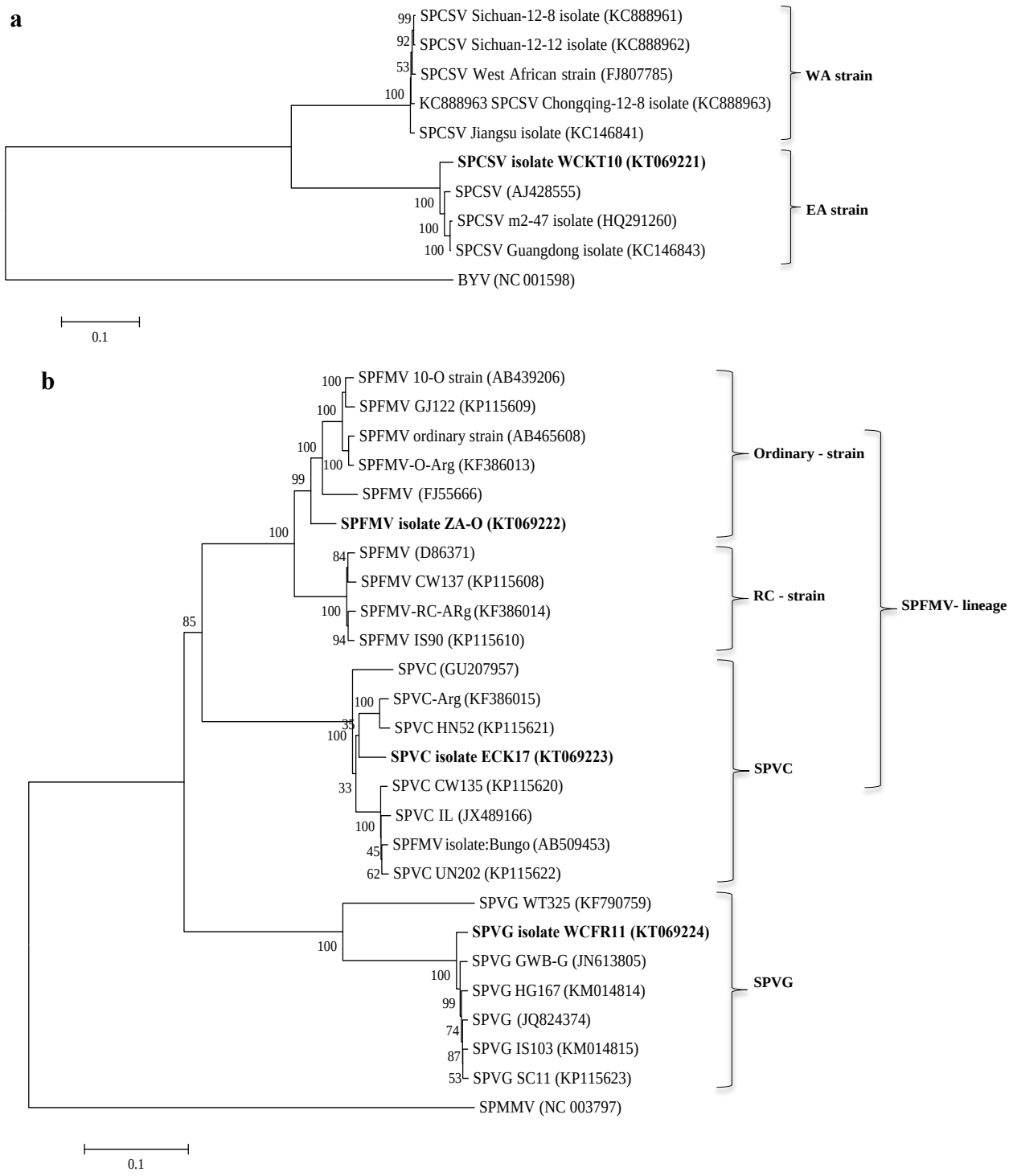

Figure 3.3 Neighbour-joining tree of SPCSV RNA2 segments assigned the SA isolate to the East African (EA) group (a), beet yellows virus (BYV) was used as an outgroup. Phylogenetic analysis of potyviruses (SPVC, SPFMV, and SPVG) based on complete and near-complete sequences (b). Sweet potato mild mottle virus (SPMMV) was included as an outgroup. The trees were generated using the neighbour-joining algorithm and the bootstrap values (1000 replicates) are indicated on the branches.

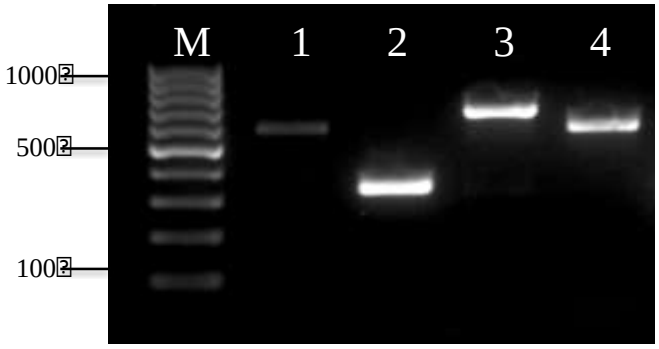

Figure 3.4 Agarose gel electrophoresis confirming the presence of the RNA viruses detected in this study by PCR and RT-PCR. Lane M, 100bp molecular marker; Lane 1, SPFMV; Lane 2, SPVC; Lane 3, SPVG; Lane 4, SPCSV

3.5 Discussion

In this study we detected six different sweet potato viruses in various combinations in the EC and WC provinces of SA. A variety of known symptoms were observed in the infected field samples, which depended on the sweet potato cultivar and virus combination. This has been shown in several other studies (Gibson et al., 2004), where for example a combination of SPCSV and SPFMV caused severe symptom development on susceptible cultivars (Gibson et al., 1998). This study reports for the first time, the detection of two begomoviruses and four RNA viruses in a single plant in SA. Multiple infections of SPFMV, SPCSV, SPLCSPV and SPMaV (found in samples KT10, KF1, KT6 and L18 from the WC province), resulted in severe symptoms including upward curling of leaves, chlorotic spots, mottling and necrosis. The occurrence of multiple viruses in single plants, and correlation between mixed infections and symptom severity, has been reported in sweet potato (Mukasa et al., 2006, Tugume et al., 2016). However the combination of six viruses has not been reported. It is therefore necessary to further investigate how sweet potato cultivars will respond to infection by six viruses because other viruses (such as SPMaV, SPLCSPV, SPVC and SPVG) could be playing a role in causing severe disease symptoms.

Consistent with previous reports, it is evident from the NGS data that SPFMV remains the most common sweet potato virus in SA, occurring wherever sweet potato is grown. Other studies have also reported SPFMV to be the most widely occurring virus in sweet potato to date (Clark et al., 2012, Rännäli et al., 2009, Valverde et al., 2007). Infection with SPFMV often causes no symptoms (Kreuze and Fuentes, 2008), and SPCSV causes mild symptoms in single infection but when the two viruses are in co-infection, they often cause severe

symptoms (Kreuze et al., 2009). The co-infection results in a synergistic interaction ultimately causing SPVD, the most devastating disease affecting sweet potato (Ateka et al., 2004, Cuellar et al., 2011, Kreuze et al., 2009). There are distinct strain groups within SPFMV; these are the O, EA, RC and C strains (Untiveros et al., 2010). In this study, phylogenetic analysis (Fig. 3) clustered SPFMV isolates into two distinct groups; the O strain and the C strain, now classified as a new potyvirus species (SPVC) (Untiveros et al., 2010). As reported previously (Rännäli et al., 2009, Sivparsad and Gubba, 2012), mixed infections of O and C strains have been detected in sweet potato from other regions of South Africa. The SPFMV-O strain detected in the WC province shared 94% nt identity with the isolate from Japan and SPVC shared 95% nt sequence identity with the isolate from Argentina. The full genome sequence of the South African SPVG isolate shared 98% nt identity with the isolate from Argentina. Limited genetic variability observed between SPVG isolates worldwide suggests geographic distribution by infected vegetative material, and that previous diagnostic tests were not sensitive or available at time of material importation/exportation to detect this virus. Co-infection of potyviruses SPVG and SPFMV (both common (C) and ordinary (O) strains) has been previously reported in French Polynesia, New Zealand, Zimbabwe and South Africa (Rännäli et al., 2009). RNA viruses are prone to variation therefore studies investigating the evolution and adaptability of these viruses are necessary in order to develop effective diagnostic assays and disease control strategies (Rubio et al., 2013). Since this is the first study to generate near complete reference sequences of South African isolates, we were also able to successfully design oligonucleotide primers to develop diagnostic assays for all the viruses detected.

The SPCSV RNA2 segment assembled in this study is highly conserved and shares 97% nt identity with the Peruvian isolate. The RNA1 segment shares 76% nt identity with the Ugandan isolate. This result suggested a possible reassortment (Hou and Gilbertson, 1996, Savory et al., 2014) between RNA1 from an "unknown" variant of an East African isolate or RNA2 from a West African isolate. It is also possible that reassortment could have occurred between SPCSV RNA1 and another sweet potato virus species. Since few full-length sequences of RNA1 and RNA2 segments of the SPCSV are available in the GenBank, screening for reassortment becomes a challenge. Reassortment events between RNA segments of two different or closely related viruses have been documented in viruses infecting other crops including tomato (Chen et al., 2009) and banana (Hu et al., 2007). Reassortment between virus species, especially of closteroviruses in the family

Closteroviridae, increases genetic variability and accelerates evolution (Rubio et al., 2013). Genetic diversity observed in SPCSV which belongs to the genus *Crinivirus* within the family *Closteroviridae*, may have arisen from the interaction of mixed viral infections and migration (exchange of sweet potato cuttings) along distant geographical areas (Rubio et al., 2013). Alternatively, reassortment may have occurred in RNA1 as a result of interaction with viruses from natural wild hosts. SPFMV, which co-exists often with SPCSV, was detected in 22 *Ipomoea* spp., *Hewittia sublobata*, and *Lepistemon owariensis* in Uganda (Tugume et al., 2008). SPCSV has been found in complexes with viruses such as SPVG, SPV2, SPLCV, SPMMV and cucumber mosaic virus (CMV), where it enhances replication and ultimately increases virus titers by approximately 1000-fold (Kreuze and Fuentes, 2008, Valverde et al., 2007). The interaction of SPCSV with other viruses also exacerbates viral symptoms (Kreuze and Fuentes, 2008, Mukasa et al., 2006), and it has been documented in many cases that SPCSV plays a major role in the enhancement of disease severity (Mukasa et al., 2006, Valverde et al., 2007). The *Hsp70* gene sequence on RNA2 of SPCSV from KwaZulu-Natal (KZN) province in SA is from the West African (WA) strain group (Sivparsad and Gubba, 2012), while the phylogenetic analysis of the full sequenced segments from this study assigned the SPCSV RNA2 from the WC province to the East African (EA) group. This finding suggests that high genetic diversity of SA SPCSV isolates may exist in different sweet potato growing regions. Our study also demonstrates the need to sequence full-length segments of SPCSV in southern and northern Africa in order to further examine the genetic diversity of SPCSV and to identify potential geographic regions where reassortment could occur, as this could lead to the emergence of new strains and increased disease severity.

Sequence reads identified as sweet potato caulimo-like virus (SPCV) were detected in the RNA dataset. Future studies must explore screening more material from the WC and other provinces in SA for the presence of this virus. Disease control strategies, including the use of virus-free cuttings and vector control, should be implemented especially in the WC province commercial farms, to prevent further spread of the viruses and crop decline. It is necessary to conduct a nationwide survey employing NGS in order to a) screen for viruses, b) assemble full-length genomes and c) gain better understanding on virus diversity and virus complexes.

In conclusion, this study describes a metagenomic approach employing the use of high throughput deep sequencing for the detection of RNA and DNA viruses in sweet potato without *a priori* knowledge. This approach clearly reveals the comprehensive profile of the

entire viral community in a sample. We established that a survey of two provinces detected six viruses in South Africa, including a distinct SPCSV RNA1 sequence. We can also infer that SPCSV, together with SPFMV and begomoviruses, is still a major role player in SPVD.

3.6 Acknowledgements

This work is based on the research support in part by the National Research Foundation of South Africa (Grant reference number UID 79983). The authors would like to acknowledge our colleagues Sidwell Tjale, Thakhani Ramathavhatha, Sunette Laurie, for assistance with sample collection; Jonathan Featherston, Marija Kvas, and Kerry-Anne Pillay for assistance with library preparation.

3.7 Conflicts of Interest

The authors declare no conflict of interest.

3.8 References

ARONESTY, E. 2013. Comparison of sequencing utility programs. *The Open Bioinformatics Journal,* 7.

ATEKA, E., NJERU, R., KIBARU, A., KIMENJU, J., BARG, E., GIBSON, R. & VETTEN, H. 2004. Identification and distribution of viruses infecting sweet potato in Kenya. *Annals of Applied Biology,* 144**,** 371-379.

BOONHAM, N., KREUZE, J., WINTER, S., VAN DER VLUGT, R., BERGERVOET, J., TOMLINSON, J. & MUMFORD, R. 2014. Methods in virus diagnostics: from ELISA to next generation sequencing. *Virus Research,* 186**,** 20-31.

CHEN, L., ROJAS, M., KON, T., GAMBY, K., XOCONOSTLE-CAZARES, B. & GILBERTSON, R. 2009. A severe symptom phenotype in tomato in Mali is caused by a reassortant between a novel recombinant begomovirus (Tomato yellow leaf curl Mali virus) and a betasatellite. *Molecular Plant Pathology,* 10**,** 415.

CLARK, C. A., DAVIS, J. A., ABAD, J. A., CUELLAR, W. J., FUENTES, S., KREUZE, J. F., GIBSON, R. W., MUKASA, S. B., TUGUME, A. K. & TAIRO, F. D. 2012. Sweetpotato viruses: 15 years of progress on understanding and managing complex diseases. *Plant Disease,* 96**,** 168-185.

CUELLAR, W. J., CRUZADO, R. K., FUENTES, S., UNTIVEROS, M., SOTO, M. & KREUZE, J. F. 2011. Sequence characterization of a Peruvian isolate of Sweet potato

chlorotic stunt virus: further variability and a model for p22 acquisition. *Virus Research,* 157**,** 111-115.

CUELLAR, W. J., TAIRO, F., KREUZE, J. F. & VALKONEN, J. P. 2008. Analysis of gene content in sweet potato chlorotic stunt virus RNA1 reveals the presence of the p22 RNA silencing suppressor in only a few isolates: implications for viral evolution and synergism. *Journal of General Virology,* 89**,** 573-582.

DOMOLA, M., THOMPSON, G., AVELING, T., LAURIE, S., STRYDOM, H. & VAN DEN BERG, A. 2008. Sweet potato viruses in South Africa and the effect of viral infection on storage root yield. *African Plant Protection,* 14**,** 15-23.

ESTERHUIZEN, L., VAN HEERDEN, S., REY, M. & VAN HEERDEN, H. 2012. Genetic identification of two sweet-potato-infecting begomoviruses in South Africa. *Archives of Virology,* 157**,** 2241-2245.

GIBSON, R., MPEMBE, I., ALICAI, T., CAREY, E., MWANGA, R., SEAL, S. & VETTEN, H. 1998. Symptoms, aetiology and serological analysis of sweet potato virus disease in Uganda. *Plant Pathology,* 47**,** 95-102.

GIBSON, R. W., ARITUA, V., BYAMUKAMA, E., MPEMBE, I. & KAYONGO, J. 2004. Control strategies for sweet potato virus disease in Africa. *Virus Research,* 100**,** 115-122.

HOU, Y.-M. & GILBERTSON, R. L. 1996. Increased pathogenicity in a pseudorecombinant bipartite geminivirus correlates with intermolecular recombination. *Journal of Virology,* 70**,** 5430-5436.

HU, J.-M., FU, H.-C., LIN, C.-H., SU, H.-J. & YEH, H.-H. 2007. Reassortment and concerted evolution in Banana bunchy top virus genomes. *Journal of Virology,* 81**,** 1746-1761.

IDRIS, A., AL-SALEH, M., PIATEK, M. J., AL-SHAHWAN, I., ALI, S. & BROWN, J. K. 2014. Viral metagenomics: analysis of begomoviruses by illumina high-throughput sequencing. *Viruses,* 6**,** 1219-1236.

KASHIF, M., PIETILÄ, S., ARTOLA, K., JONES, R., TUGUME, A., MÄKINEN, V. & VALKONEN, J. 2012. Detection of viruses in sweetpotato from Honduras and Guatemala augmented by deep-sequencing of small-RNAs. *Plant Disease,* 96**,** 1430-1437.

KAYS, S. J. Sweetpotato production worldwide: Assessment, trends and the future. I International Symposium on Root and Tuber Crops: Food Down Under 670, 2004. 19-25.

KREUZE, J. & FUENTES, S. 2008. Sweetpotato Viruses A2 - Mahy, Brian W.J. *In:* REGENMORTEL, M. H. V. V. (ed.) *Encyclopedia of Virology (Third Edition).* Oxford: Academic Press.

KREUZE, J., SAVENKOV, E. & VALKONEN, J. 2002. Complete genome sequence and analyses of the subgenomic RNAs of Sweet potato chlorotic stunt virus reveal several new features for the genus Crinivirus. *Journal of Virology,* 76**,** 9260-9270.

KREUZE, J. F., PEREZ, A., UNTIVEROS, M., QUISPE, D., FUENTES, S., BARKER, I. & SIMON, R. 2009. Complete viral genome sequence and discovery of novel viruses by deep sequencing of small RNAs: a generic method for diagnosis, discovery and sequencing of viruses. *Virology,* 388**,** 1-7.

MASSART, S., OLMOS, A., JIJAKLI, H. & CANDRESSE, T. 2014. Current impact and future directions of high throughput sequencing in plant virus diagnostics. *Virus Research,* 188**,** 90-96.

MUKASA, S. B., RUBAIHAYO, P. R. & VALKONEN, J. 2006. Interactions between a crinivirus, an ipomovirus and a potyvirus in coinfected sweetpotato plants. *Plant Pathology,* 55**,** 458-467.

MWANGA, R., MOYER, J., ZHANG, D., CAREY, E. & YENCHO, G. Nature of resistance of sweetpotato to sweetpotato virus disease. I International Conference on Sweetpotato. Food and Health for the Future 583, 2001. 113-119.

RÄNNÄLI, M., CZEKAJ, V., JONES, R., FLETCHER, J., DAVIS, R., MU, L. & VALKONEN, J. 2009. Molecular characterization of Sweet potato feathery mottle virus (SPFMV) isolates from Easter Island, French Polynesia, New Zealand, and southern Africa. *Plant Disease,* 93**,** 933-939.

RUBIO, L., GUERRI, J. & MORENO, P. 2013. Genetic variability and evolutionary dynamics of viruses of the family Closteroviridae. *Frontiers in Microbiology,* 4.

RUBY, J. G., BELLARE, P. & DERISI, J. L. 2013. PRICE: software for the targeted assembly of components of (Meta) genomic sequence data. *G3: Genes*| *Genomes*| *Genetics,* 3**,** 865-880.

SAVORY, F. R., VARMA, V. & RAMAKRISHNAN, U. 2014. Identifying geographic hot spots of reassortment in a multipartite plant virus. *Evolutionary Applications,* 7**,** 569-579.

SIVPARSAD, B. & GUBBA, A. 2012. Molecular resolution of the genetic variability of major viruses infecting sweet potato (Ipomoea batatas L.) in the province of KwaZulu-Natal in the Republic of South Africa. *Crop Protection,* 41**,** 49-56.

STOBBE, A. H. & ROOSSINCK, M. J. 2014. Plant virus metagenomics: what we know and why we need to know more. *Frontiers in Plant science,* 5**,** 150.

STUDHOLME, D. J., GLOVER, R. H. & BOONHAM, N. 2011. Application of high-throughput DNA sequencing in phytopathology. *Annual Review of Phytopathology,* 49**,** 87-105.

TAMURA, K., STECHER, G., PETERSON, D., FILIPSKI, A. & KUMAR, S. 2013. MEGA6: molecular evolutionary genetics analysis version 6.0. *Molecular Biology and Evolution,* 30**,** 2725-2729.

TESFAYE, T., FEYISSA, T. & ABRAHAM, A. 2011. Survey and serological detection of sweet potato (Ipomoea batatas (L.) Lam) viruses in Ethiopia. *Journal of Applied Biosciences,* 41**,** 2746-2756.

TUGUME, A., MUKASA, S. & VALKONEN, J. 2008. Natural wild hosts of Sweet potato feathery mottle virus show spatial differences in virus incidence and virus-like diseases in Uganda. *Phytopathology,* 98**,** 640-652.

TUGUME, A. K., MUKASA, S. B. & VALKONEN, J. P. 2016. Mixed infections of four viruses, the incidence and phylogenetic relationships of Sweet potato chlorotic fleck virus (Betaflexiviridae) isolates in wild species and sweetpotatoes in Uganda and evidence of distinct isolates in East Africa. *PloS One,* 11**,** e0167769.

UNTIVEROS, M., QUISPE, D. & KREUZE, J. 2010. Analysis of complete genomic sequences of isolates of the Sweet potato feathery mottle virus strains C and EA: molecular evidence for two distinct potyvirus species and two P1 protein domains. *Archives of Virology,* 155**,** 2059-2063.

VALVERDE, R. A., CLARK, C. A. & VALKONEN, J. P. 2007. Viruses and virus disease complexes of sweetpotato. *Plant Viruses,* 1**,** 116-126.

VAN JAARSVELD, P. J., FABER, M., TANUMIHARDJO, S. A., NESTEL, P., LOMBARD, C. J. & BENADÉ, A. J. S. 2005. β-Carotene–rich orange-fleshed sweet potato improves the vitamin A status of primary school children assessed with the modified-relative-dose-response test. *The American Journal of Clinical Nutrition,* 81**,** 1080-1087.

CHAPTER 4

ROLLING CIRCLE AMPLIFICATION-NEXT GENERATION SEQUENCING DETECTS BEGOMOVIRUSES AND BADNAVIRUSES IN SOUTH AFRICA

4.1 Abstract

The prevalence of sweet potato DNA viruses in South Africa (SA) is not well documented. We explored the current status of DNA viruses in the Western and Eastern provinces of South Africa since the countrywide survey conducted in 2008. We coupled rolling circle amplification (RCA) of genomic DNA with Illumina next generation sequencing (NGS) for the detection of DNA viruses. Sequence reads were analysed with the CLC Bio Genomics Workbench (version 7.5.1). Over two million sequence reads were generated from 17 independently sequenced libraries. BLASTn and BLASTx searches using *de novo* assembled sequences against published virus genomes confirmed the presence of previously detected begomoviruses in SA namely sweet potato mosaic virus (SPMaV) and sweet potato leaf curl Sao Paulo virus (SPLCSPV). Sequences identified as sweet potato badnavirus B (SPBVB) were found in the RCA dataset. Following this finding, seven small RNA (sRNA) libraries were prepared, to further investigate the presence of badnaviruses in sweet potato plants. Small RNAs (sRNA) were isolated from five symptomatic and two asymptomatic plants. The libraries were independently sequenced on the Illumina MiSeq platform. Each of the libraries underwent quality trimming and downstream analysis. Over 6.9 million sequence reads from the seven libraries remained after quality control (QC) analysis. Sequence reads were assembled into contiguous (contigs) sequences using Velvet. BLASTn and BLASTx searches against viral sequences revealed the presence of sweet potato badnavirus A (SPBVA) and SPBVB in all of the libraries sequenced. Both badnaviruses were detected in the symptomatic and asymptomatic samples as co-infections. Conventional Sanger sequencing of amplified PCR products confirmed the identity of the begomoviruses and badnaviruses. This is the first report of sweet potato badnaviruses in South Africa.

4.2 Introduction

Sweet potato is a very important root crop with nutritional and economic value in sub-Saharan countries such as South Africa (SA) (SAHRC 2008), Kenya (Nyaboga et al., 2008) and Uganda (Wasswa et al., 2011). Sweet potato has especially gained much attention in the recent past for its role in poverty alleviation and food security among small-scale farmers and Vitamin A deficient children and women (Kreuze, 2002) respectively. Pests and diseases constitute some of the major production drawbacks in sweet potato and can considerably reduce yield and crop quality. Viral diseases have been reported to reduce sweet potato crop quality and yield by up to 100%, depending on cultivar, type of virus and plant susceptibility (Gibson et al., 1997, Valverde et al., 2007). Several viruses occur in disease complexes, where an individual plant can be infected with multiple RNA and DNA viruses (Hanley-Bowdoin et al., 2013).

While several studies have been performed on sweet potato RNA viruses in SA, relatively little is known about DNA virus diversity. Only three DNA viruses have been reported to date, namely sweet potato caulimo-like virus (SPCaLV) (Aritua and Adipala, 2005, Domola et al., 2008), and the begomoviruses, sweet potato mosaic virus (SPMaV-ZA) and sweet potato leaf curl Sao Paulo virus (SPLCSPV-ZA) (Esterhuizen et al., 2012). They occur either as single, co-infections or multiple infections. Geminiviruses are plant viruses with circular ss-DNA genomes consisting of one (monopartite) or two (bipartite) molecules, ranging in size from 2.6 – 3Kb, packed into twin-shaped icosahedral particles (Albuquerque et al., 2012, Albuquerque et al., 2011, Paprotka et al., 2010). The International Committee on the Taxonomy of viruses classifies geminiviruses into seven genera, namely *Begomovirus*, *Curtovirus*, *Mastrevirus*, *Bercurtovirus*, *Eragrovirus*, *Turnocurtovirus* and *Topocuvirus* (Hanley-Bowdoin et al., 2013). Sweet potato geminiviruses are monopartite, belonging to the genus *Begomovirus*, and are transmitted either by whiteflies of species *Bemisia tabaci* (Genn.), or distributed by vegetative propagation (Albuquerque et al., 2012, Albuquerque et al., 2011, Paprotka et al., 2010). Infection of sweet potato with sweepoviruses has been reported in Japan, Israel, Peru, Italy, Spain, China, Taiwan, Korea, Kenya, United States of America, Puerto Rico, Costa Rico, Brazil and South Africa (Albuquerque et al., 2012, Albuquerque et al., 2011, Clark et al., 2012, Esterhuizen et al., 2012, Kreuze and Fuentes, 2008, Paprotka et al., 2010). The symptoms associated with begomovirus infection include upward curling of leaves, vein yellowing, leaf distortion and chlorosis (Albuquerque et al., 2012, Kreuze and Fuentes, 2008). Although begomoviruses are frequently symptomless in

sweet potato, they do cause yield losses and can spread undetected (Albuquerque et al., 2011, Clark et al., 2012). Other emerging sweet potato mastreviruses, badnaviruses and badnavirus-like particles, have been reported in the USA (Sim and Clark, 2008), Peru (Kreuze et al., 2009) and Tanzania (Mbanzibwa et al., 2014). Badnaviruses are pararetroviruses with a circular double-stranded DNA genome (dsDNA) that ranges between 7.2 – 8.5 kb in size (Bhat et al., 2016, Borah et al., 2013, Bouhida et al., 1993, Chabannes and Iskra-Caruana, 2013). They belong to the family *Caulimoviridae* (Borah et al., 2013).

Traditionally, the detection and characterisation of begomoviruses and badnaviruses has relied on graft inoculation, serological assays (Ndowora and Lockhart, 1996), electron microscopy (Dahal et al., 2000) and molecular methods such as microarrays, restriction fragment length polymorphism (RFLP) analysis (Ndowora et al., 1999) and polymerase chain reaction (PCR) using degenerate primers and virus specific primers (Elbeaino et al., 2013, Geering et al., 2005, Kashif et al., 2012, Li et al., 2004). More recently, rolling circle amplification (RCA) or RCA coupled with RFLP (RCA-RFLP) have become popular diagnostic tools for the detection of both begomoviruses and badnaviruses (James et al., 2011, Laney et al., 2012, Wambulwa et al., 2013). Sequencing of RCA products has aided in the detection of novel DNA viral species and their strains or variants (Albuquerque et al., 2012, Clark et al., 2012, Esterhuizen et al., 2012, Kreuze et al., 2009, Paprotka et al., 2010). The aim of this chapter was to explore the presence of DNA viruses by sequencing RCA products and small RNAs (sRNAs) on Next Generation Sequencing (NGS) platforms. Transcripts of DNA viruses were detected in chapter 3, and partial genome sequences for SPLCSPV and SPMaV were generated. Therefore the objective of this chapter was to generate full-length genome sequences of begomoviruses.

4.3 Materials and Methods

4.3.1 Plant Material

Plant material was sourced from 8 locations in South Africa. The 4 locations in the Eastern Cape (EC) province included Alice, Zwide, Kwazakhele and Motherwell. In the Western Cape (WC) province plant material was collected from Paarl, Klawer, Franschhoek and Lutzville. In the field, cuttings from asymptomatic and symptomatic plants, displaying virus symptoms such as upward curling of the leaves, vein clearing and chlorotic spots, were collected. Plants were grown in potting soil at optimum temperatures of 25°C for 16 h (day cycle) and 15°C for 8 h (night cycle), in summer and winter, respectively (Domola et al.,

2008). The plants were maintained in glasshouses, and watered once a day and given soluble nutrient fertilization (Multifeed P, Plaaskem, Pty, LTD) on a weekly basis. Insect pests were monitored and controlled by spraying with recommended insecticides as required. A total of 17 plants (Table 3.2) were selected for DNA isolation and deep sequencing.

4.3.2 Rolling circle amplification (RCA) and library preparation

Genomic DNA (gDNA) was isolated from plant leaf material using the QIAGEN DNeasy Plant Mini Kit (*QIAGEN* Inc., Valencia, CA, USA) following manufacturer's instructions. The integrity of the extracted DNA was visualized by electrophoresis and quantified using the Qubit™ dsDNA BR Assay Kit (Life Technologies, Thermo Fisher Scientific Inc.). The gDNA was stored at -80°C until further use. Prior to sequencing, genomic DNA underwent rolling circle amplification (RCA) using the Illustra™ TempliPhi™ 100 Amplification Kit (GE Healthcare, Amersham, UK) following the manufacturer's instructions. The RCA products were prepared for sequencing using the Nextera DNA Sample Preparation Kit (Illumina Inc., San Diego, CA, USA). A total of 17 individually labelled RCA libraries were sequenced on the Illumina MiSeq platform (Illumina Inc., San Diego, CA, USA).

4.3.3 Small RNA isolation and sequencing

It is important to note that this experiment was conducted in order to establish and optimise a small RNA sequencing protocol. At the time of this study small RNA library preparation was costly, therefore only 7 plant samples were chosen for sequencing and further analysis. Small RNAs (sRNAs) were isolated from 5 symptomatic and 2 asymptomatic plants using the Ambion mirVana™ miRNA RNA Isolation Kit (Ambion, USA). The sRNAs were quantified using the Qubit™ RNA HS Assay Kit (Life Technologies, Thermo Fisher Scientific Inc.) and then stored at -80°C until further use. The sRNA samples were prepared for sequencing using the Illumina TruSeq Small RNA Sample Preparation Kit (Illumina Inc., San Diego, CA, USA). Size selection was done using the Blue Pippin Prep Instrument (SAGE Science, Beverly, MA, USA). The 7 libraries were independently sequenced on the Illumina MiSeq platform (Illumina Inc., San Diego, CA, USA).

4.3.4 Sequence analysis

Two data sets were generated from the MiSeq platform, the RCA data set (17 libraries) and the sRNA data set (7 libraries). The data was trimmed for adaptor sequences, and quality control (QC) was performed using Fastq-mcf (Aronesty, 2013) before performing further

analysis. For the RCA data set, a quality threshold of 30 and a Phred score of 33 were selected for trimming options. Sequence reads below the length of 50 bp and greater than 150 bp were treated as low quality and discarded. The trimmed sequence reads generated from the RCA libraries were aligned to the partial sweet potato chloroplast genome (Accession: KF242475), the sweet potato mitochondrial DNA (Accession: FN421476) and sweet potato ESTs using the CLC Bio Genomics Workbench (version 7.5.1) (CLC bio, Aarhus, Denmark) in order to filter out host sequences. The unmapped sequence reads were collected and then assembled into contigs using the CLC Bio Genomics Workbench *de novo* assembly tool. The contigs generated from the *de novo* assembly were subjected to BLASTn and BLASTx searches against the viral databases downloaded from the NCBI database. The full-length reference sequences of the viruses detected in the BLASTn and BLASTx searches were retrieved and used in subsequent reference-guided assemblies. The sequence reads and contigs matching sweet potato viruses were mapped to the full genomes of the closest hits, using mapping settings: length fraction = 0.7 and similarity = 0.9 in the CLC Bio Genomics Workbench.

Sequences from the sRNA library were trimmed for adaptor sequences and the trimming options selected for this data set was a quality threshold of 30 and a Phred score of 33. Sequence reads below the length of 15 bp and greater than 30 bp were discarded. The trimmed sequences were mapped to sweet potato reference sequences for host sequence filtration as described above. All the unmapped sequences were collected and utilized to generate contiguous sequences with the Velvet software (Zerbino and Birney, 2008). The contigs were subjected to BLASTn and BLASTx searches against viral sequences obtained from the NCBI database. The full-length genome sequences of the viruses that were detected in the BLAST searches were retrieved from the database and used as references in the reference-guided assemblies. Small RNA data was processed with the SearchSmallRNA tool to generate read mapping graphs (de Andrade and Vaslin, 2014).

4.3.5 PCR confirmation of DNA viruses detected by NGS

Genomic DNA was isolated from sample KT10 using the QIAGEN DNeasy Plant Mini Kit, according to the manufacturer's instructions. The identity of the begomoviruses and badnaviruses was confirmed by polymerase chain reaction (PCR). The PCR products were sent to Inqaba Biotechnological for Sanger sequencing. Oligonucleotide sequences (Table 4.1) were designed using the IDT PrimerQuest Tool (https://eu.idtdna.com/site). The PCRs

were performed using the DreamTaq Green PCR Mix (2X) (Fermentas, Thermo Scientific, Waltham, MA, USA). The 25 µl reactions consisted of 12.5 µl of the DreamTaq Green PCR Master Mix, 0.2 µM of each primer, 2 µl of the template DNA and 9 µl of nuclease-free water. The reactions were amplified at the recommended thermal cycling conditions as follows: Initial denaturation at 95°C for 3 min, 35 cycles of 95°C for 30 s, annealing temperature at 57°C for 30 s, an extension at 72°C for 1 min, the final extension for 10 min at 72°C. The amplicons were visualized on a 2% agarose gel by electrophoresis.

Table 4.1: Oligonucleotides sequences used for confirming the identity of DNA viruses

Virus Name	**Primer Name**	**Sequence (5'-3')[a]**	**Amplicon Size (bp)**
Sweet potato badnavirus A	SPBVA-F	TCCCACCTAAGGCTCAAGAA	698
	SPBVA-R	GCAAACTGTTGCCCCTGTAT	
Sweet potato badnavirus B	SPBVB-F	TGGGTGCAATTTCATCAGAA	700
	SPBVB-R	GTGCATTTACCAGCCCAAAT	
Sweet potato mosaic virus	SPMaV-F	CCGAAGCTATGTCCCGATTT	314
	SPMaV-R	GGTCCTTATTGGGCCTTCTATC	
Sweet potato leaf curl Sao Paulo virus	SPLCSPV-F	TCGAGATAGGAGGCCCAATAA	322
	SPLCSPV-R	GCAACGCAGAGTCTGATACA	

4.4 Results

4.4.1 Symptom description of plants

The samples described and analysed in Chapter three were also used for RCA-deep sequencing in this chapter. In chapter 3 we described the symptoms observed on sample KT10, which included purple ring spots and leaf curling (Table 3.2 and Figure 4.1). We were able to detect sequences with matches to RNA viruses and traces of DNA begomoviruses in sample KT10. During field collected, unidentified whitefly species were observed on the plants collected from Klawer (KT10, KF1, KT6) in the WC province. Leaf curl symptoms were not observed on plants collected from the EC province. When sample KT10 was grown in the glasshouse, after a period of 6 months, the symptoms that were observed in the field (Figure 4.1a), were persistent and observed again the glasshouse (Figure 4.1b). Sample KF1 displayed leaf curl symptoms, purple ring spots and stunting of the plant (Figure 4.1c), the plant died one week after the leaves were harvested. Sample KT6 also displayed curling of

the young and old leaves (Figure 4.1d). Upward curling of the leaves is a symptom that is usually associated with sweet potato leaf curl virus (SPLCV).

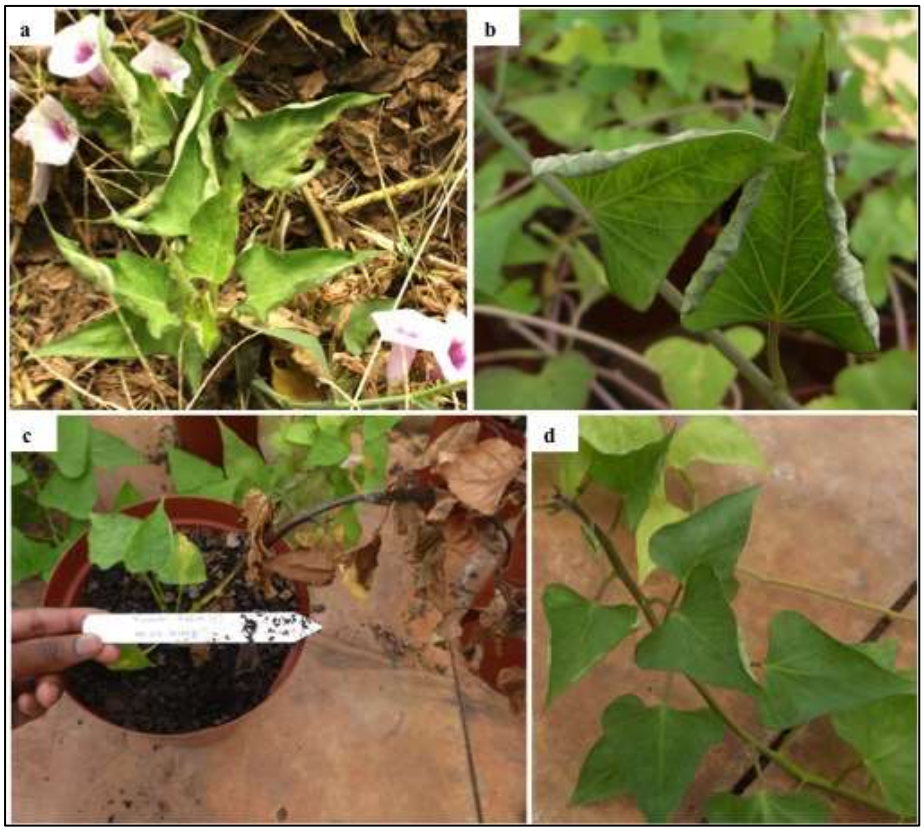

Figure 4.1: a) Sweet potato leaves displaying leaf curl symptoms in the field. Whiteflies were observed on the leaves during collection. b) Sample KT10 showing persistent leaf curl symptoms in the glasshouse. Leaf curling was observed on KF1 (c) and KT6 (d).

4.4.2 RCA data sequence analysis and *de novo* assembly

The 17 individually labelled RCA libraries generated over 2 million sequence reads. Sixty five percent of the paired-end data was retained after adaptor and quality trimming (Table 4.2). The sequence reads mapping to the host sequences accounted for 18% of the data. From this data we were able to assemble the *Ipomoea batatas* mitochondrial DNA complete sequence (Accession: FN421476) (Figure 4.2a). The remaining 82% of the unmapped sequence reads were assembled into contigs (Table 4.3).

Table 4.2: Summary statistics for the 17 RCA libraries. Paired-end sequence reads of 250 bp were generated from the MiSeq Platform. The raw data and cleaned sequences are shown.

Sample ID	Sequence reads before QC	Sequence reads after QC	Sequence reads mapped to host sequences	Unmapped sequence reads
KT10	445010	275122	30367	244755
F11	174216	63580	25049	38531
KF1	535160	381060	17330	363730
L18	389762	256818	44411	212407
KT6	473674	312578	12094	300484
P2	281750	72370	40142	32228
L9	288108	191138	82558	108580
F4	145236	96180	35951	60229
P14	243534	165320	82068	83252
L11	342090	261298	142685	118613
Z24	228190	156188	75446	80742
KZ17	193318	131826	59314	72512
M19	172030	114664	33459	81205
FH14	257518	158962	54630	104332
M11	145748	92980	50625	42355
K10	207852	135508	17340	118168
F22	176194	106074	53468	52606

4.4.3 Identification of begomoviruses from *de novo* assembly

The *de novo* assembled contigs were subjected to BLASTn and BLASTx searches using the virus database downloaded from the NCBI. We detected sweet potato begomoviruses (SPMaV and SPLCSPV); sweet potato badnavirus B (SPBVB) and sequence reads with matches to sweet potato caulimo-like virus (SPCV). Very few contigs (20) matched to sweet potato viruses. SPLCSPV was detected from samples KT10, KT6, KF1, and L18. *De novo* assembly of sequence reads was able to generate the near full-length genome of SPLCSPV. A total of 103646 sequence reads were identified as SPLCSPV in the *de novo* assembly (Table 4.4). The remaining symptomatic samples did not have any sequences or contigs matching to viruses. Co-infections of SPLCSPV and SPMaV were found samples KT10, KT6, KF1 and L18. The *de novo* assembly generated contigs that resembled near-full length viral sequences of SPMaV, which were found in samples L18 (2504 bp) and KT6 (2784 bp) (Table 4.4). The largest SPMaV contig generated from the *de novo* assembly was 2781 bp long (Table 4.4). The *de novo* assembly therefore achieved over 99% genome coverage for SPMaV; this result further highlights, as in the previous chapter 3, the efficiency of the *de novo* assembly approach in virus discovery and complete genome assembly. Sweet potato begomoviruses

were detected in only 4 of the 17 plant samples that were sequenced. The samples were collected in Klawer (KT10, KT6 & KF1) and Lutzville (L18), which are located in the WC province. Contigs from other samples did not have any significant hits to any plant viruses in the NCBI database. Sweet potato caulimovirus (SPCV) was found in sample P14.

4.4.4 Identification of sweet potato badnavirus B (SPBVB) from *de novo* assembly

Interestingly, a 322 bp contig found in sample P14 showed similarity to SPBVB (Table 4.4). This was the only contig identified as SPBVB in the 17 RCA *de novo* assembly libraries. Following this finding we conducted reference-guided assembly and detected 185 and 113 sequence reads with matches to SPBVA and SPBVB, respectively (Table 4.4). The sequence read mapping achieved very poor fragmentary assemblies. Since traces of badnavirus sequences were detected from the RCA data set, we decided to use another enrichment/library preparation method (sRNA isolation), which has been effectively used to identify and assemble full-length genomes of badnaviruses.

Table 4.3. Unmapped sequence reads from the RCA libraries *de novo* assembled into contigs.

Sample ID	Total number of contigs	Maximum contig length (bp)	Minimum contig length (bp)	N50
KT10	264	6256	186	356
F11	256	2288	160	329
KF1	423	6844	174	345
L18	329	6288	140	337
KT6	843	5906	167	335
P2	109	1459	203	475
L9	170	2230	183	350
F4	144	2869	196	338
P14	164	1786	197	337
L11	185	2573	202	334
Z24	208	4470	167	341
KZ17	201	2973	193	355
M19	228	4145	138	349
FH14	342	3156	199	333
M11	122	1124	200	369
K10	209	3397	129	344
F22	182	1667	200	347

4.4.5 Reference-guided assembly of begomoviruses

The full-length sequences of the two begomoviruses were retrieved from the NCBI database and used in reference-guided assemblies. Sequence depth for SPMaV ranged from 75 - 826-fold while it was 140 - 1932-fold for SPLCSPV (Table 4.4). The reference-guided assembly achieved high genome coverage and sequence depth for both viruses (Figure 4.2b & 4.2c). The trimmed sequence reads could be mapped as paired reads and single reads (Figure 4.2b & 4.2c). The high sequence depth observed was expected since samples underwent rolling circle amplification, which enriched the samples for circular genomes. The newly generated consensus sequences were 2781 bp and 2769 bp respectively for SPMaV (SPMaV-[ZA:WP:2011]) (Accession: JQ621843) and SPLCSPV (SPLCSPV-[ZA:WP:2011]) (Accession: JQ621844), sharing 98 and 99% sequence identity respectively. The newly assembled genome sequences for SPLCSPV and SPMaV were deposited in GenBank under accession numbers KX859238 and KX859239, respectively. The positive identification of SPLCSPV and SPMaV using sequence data, correlated with the phenotypic data (leaf curl symptoms observed on three sweet potato samples KF1, KT6 and KT10) (Figure 4.1). Begomovirus sequences were not detected in the remaining samples from the WC (samples F11, P2, L9, F4, K10, P14, L11) or any of the samples collected from the EC province (KZ17, Z24, M19, FH22, M11, FH14). Even though typical viral symptoms had been observed on samples Z24 and K17, RNA viruses were detected from these samples instead (Chapter 3).

Table 4.4 Summary statistics of the DNA viruses (SPLCSPV, SPMaV, SPBVA, SPBVB) detected by rolling circle amplification (RCA) and deep sequencing.

		***De novo* assembly**					**Reference-guided assembly**		
Sample ID	**Virus detected**	**Length of reference genome (bp)**	**Maximum contig length (bp)**	**Total number of sequence reads**	**Total number of contigs**	**Total genome coverage (%)**	**Sequence reads mapped to reference**	**Length of new sequence (bp); coverage (%); depth (X)**	**Data after QC (Mb)**
KT10	SPMaV	2783	757	25649	6	80	28993	2781; 99; 826	22,4
	SPLCSPV	2769	1349	28695	3	90	27083	2769; 100; 794	
	SPBVA	8082	-	-	-	-	33	240; 0,02; 0,08	
	SPBVB	7961	-	-	-	-	17	556; 0,06; 0,09	
F11	-	-	-	-	-	-	-	-	14
KF1	SPMaV	2783	1689	9194	3	80	22748	2780; 99; 585	33,6
	SPLCSPV	2769	561	53085	9	85	68800	2769; 100; 1932	
	SPBVA	-	-	-	-	-	47	599; 0,07; 0,12	
	SPBVB	-	-	-	-	-	21	693; 0,08; 0,11	
L18	SPMaV	2783	1295	7097	1	46	18209	2781; 99; 533	30,3
	SPLCSPV	2769	1251	15875	5	45	18328	2769; 100; 553	
	SPBVA		-	-	-	-	24	350; 0,04; 0.09	
	SPBVB						14	423; 0,05; 0.06	
KT6	SPMaV	2783	2781	1900	1	99	2631	2781; 99; 75	24,8
	SPLCSPV	2769	1499	5991	4	80	4808	2769; 100; 140	
	SPBVA	-	-	-	-	-	61	628; 0,07; 0,16	
	SPBVB	-	-	-	-	-	30	808; 0,10; 0,13	
P2	-	-	-	-	-	-	-	-	28,2
L9	-	-	-	-	-	-	-	-	21
F4	-	-	-	-	-	-	-	-	17,1
P14	SPCV	7723	-	-	-	-	39	608; 0.07; 0.07	271,1
	SPBVB	7961	322	5	1	0,04	21	337; 0,07; 0.08	
L11	-	-	-	-	-	-	-	-	43,9
Z24	-	-	-	-	-	-	-	-	27,3
KZ17	-	-	-	-	-	-	-	-	23,1
M19	-	-	-	-	-	-	-	-	20,3
FH14	-	-	-	-	-	-	-	-	30,1
M11	-	-	-	-	-	-	-	-	16,6
K10	SPBVA	-	-	-	-	-	20	170; 0,02; 0,05	21,6
	SPBVB						10	156; 0,01; 0,02	
F22	-	-	-	-	-	-	-	-	22,6

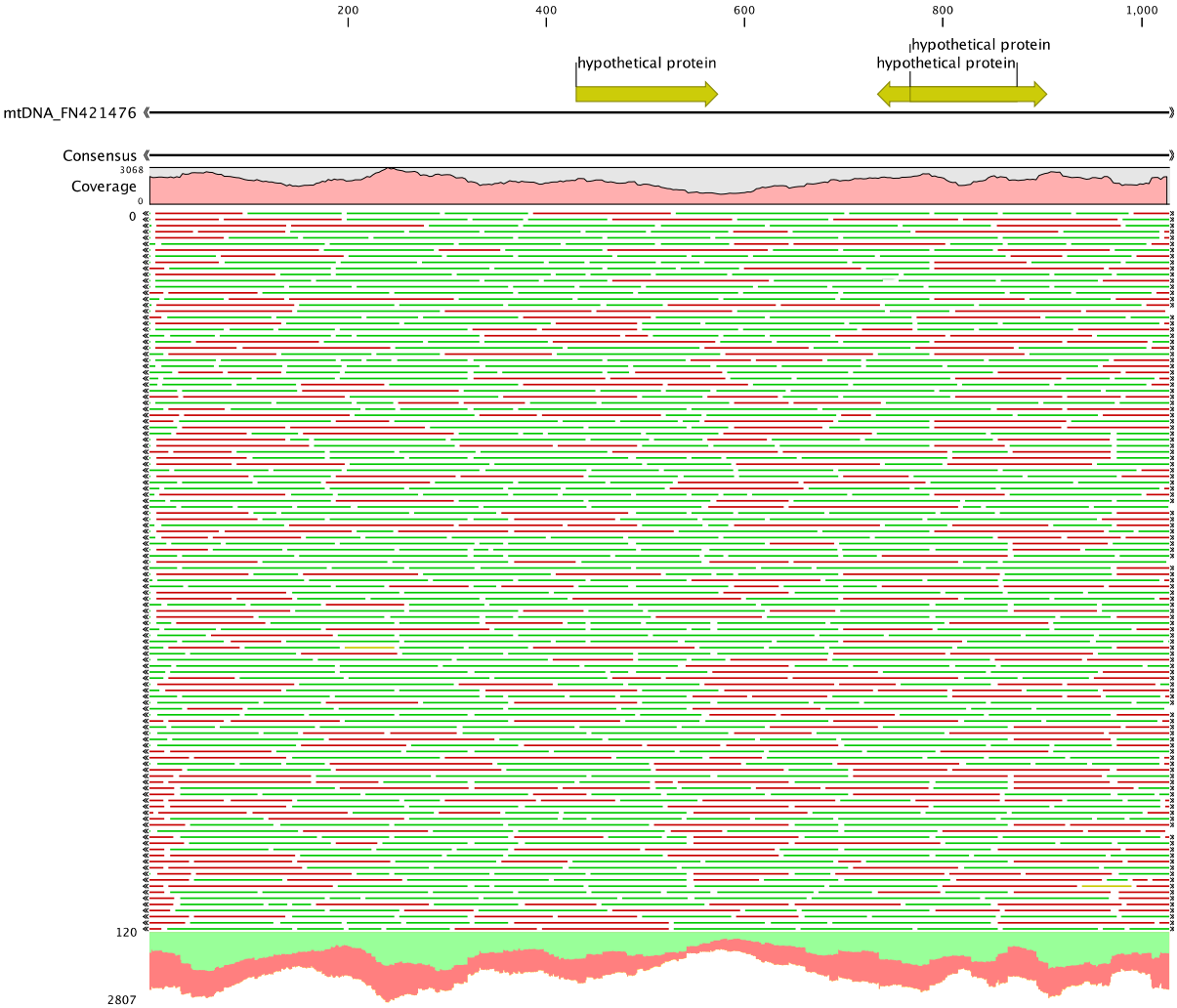

Figure 4.2a: Sequence reads mapped to the sweet potato mitochondrial DNA.

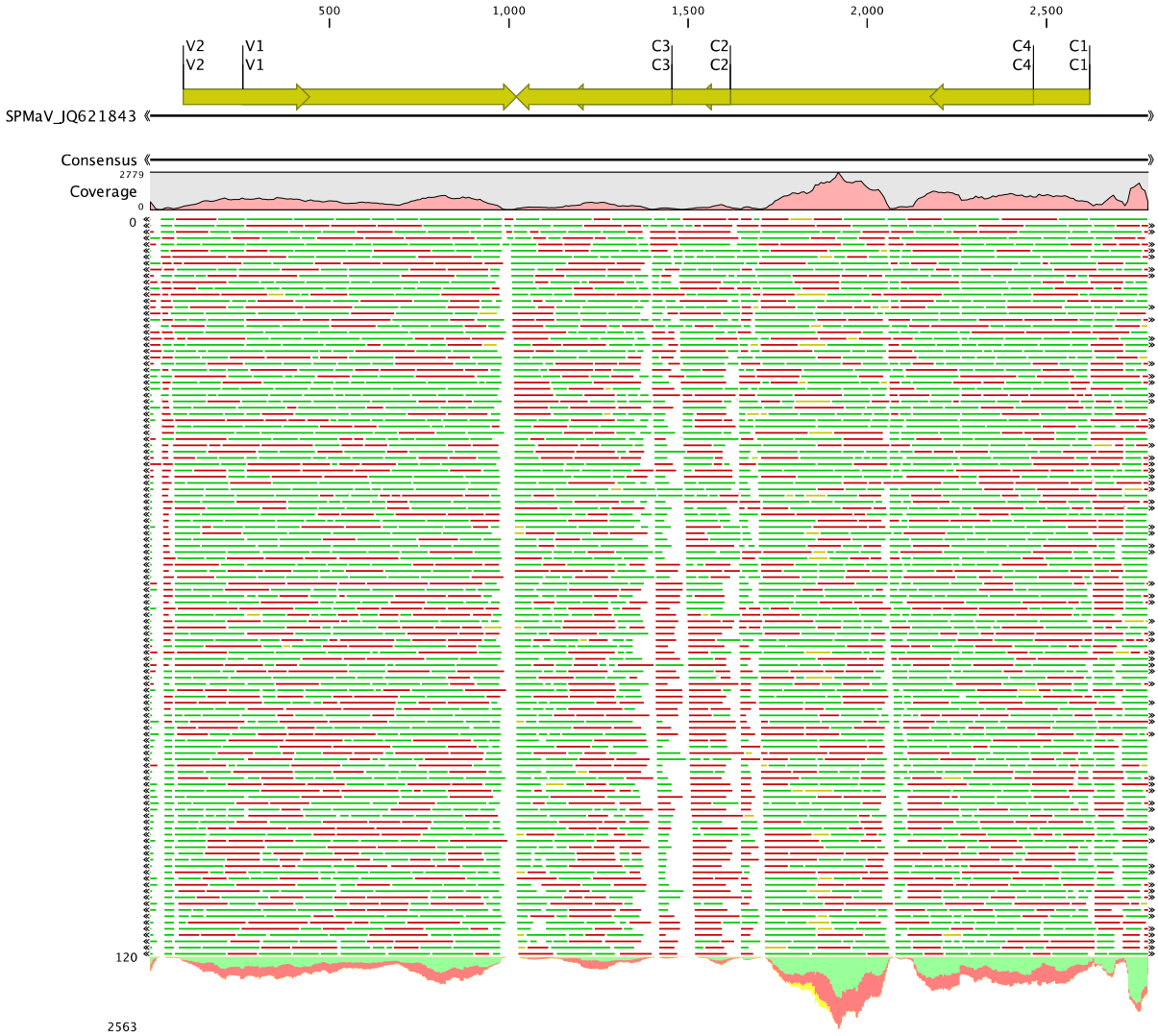

Figure 4.2b: Total number of sequence reads from sample KT10 (RCA library) mapped to SPMaV full-length genome sequence. The single reads that are mapped in the forward direction are green and the red reads are mapped in the reverse direction. The consensus sequence represents the newly generated sequence.

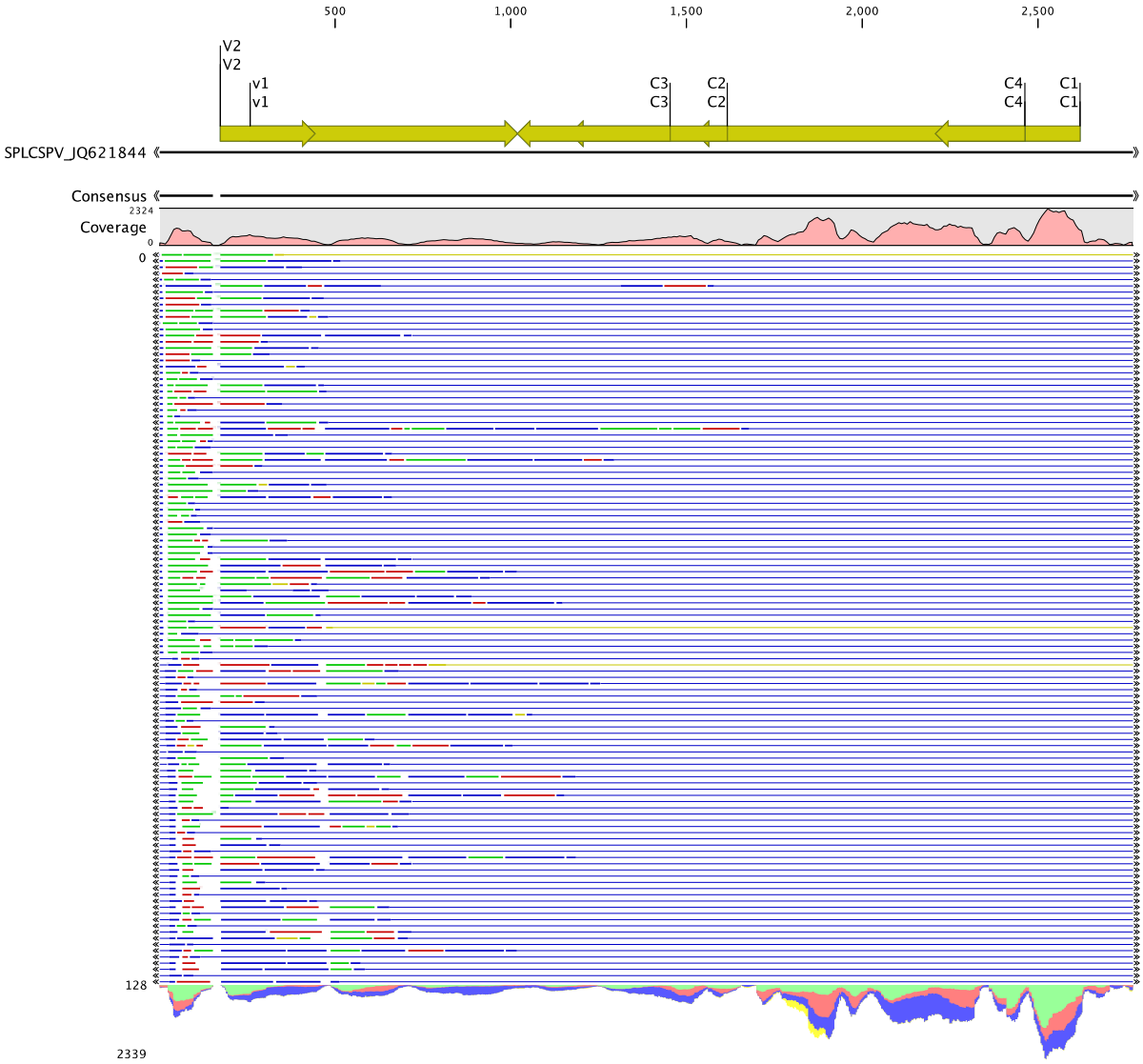

Figure 4.2c: Paired sequence reads (blue), forward reads (green) and reverse reads (red) mapped to SPLCSPV. The coverage (sequence depth) is shown below the consensus sequence.

4.4.6 Phylogenetic analysis of begomoviruses

A phylogenetic tree (Figure 4.3) placed the two begomoviruses from the coastal WC province of SA with the SPLCSPV and SPMaV isolates detected from Waterpoort, Limpopo Province, South Africa in 2012 (Esterhuizen et al., 2012). The results from our study infer that the begomoviruses may be more widespread in the country, necessitating screening of these viruses in all 9 provinces of South Africa. SPLCSPV and SPMaV isolates could be contributing significantly to disease severity as they have been found co-infecting sweet potato with other RNA viruses such as SPCSV and SPFMV, a result from chapter 3 in our study.

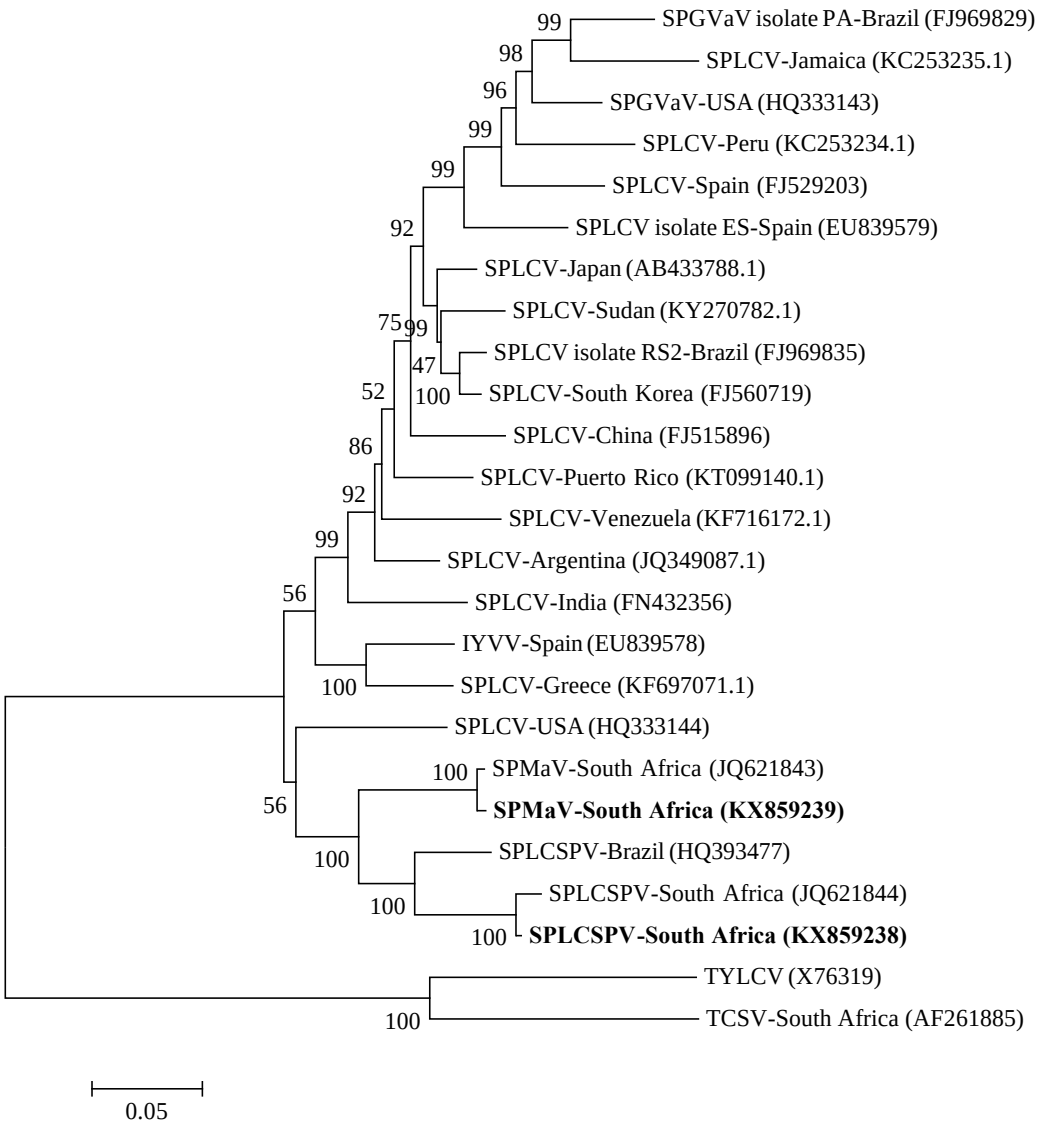

Figure 4.3: Neighbour-joining tree showing the phylogenetic relationship of begomovirus isolates from different geographic locations. Bootstrap values (of 1000 replicates) are indicated on the branches. The acronyms denote the following viruses: SPLCV – sweet potato leaf curl virus; SPLCSPV – sweet potato leaf curl Sao Paulo virus; SPMaV - sweet potato mosaic virus; IYVV – Ipomoea yellow vein virus; TYLCV – tomato yellow leaf curl virus and tomato curly stunt virus (TCSV) were included as outgroups.

4.4.8 Small RNA sequence analysis

Over 141 million sequences were generated from 7 sRNA libraries. After quality control (QC) analysis, sequence reads between 15 – 30 bp were retained for downstream analysis. The lengths of the retained reads ranged between 21 – 25 bp (Figure 4.4), confirming their likely association with virus-derived small interfering RNAs (vsiRNAs) or microRNAs (miRNAs). Over 6.9 million sequences were retained after QC analysis.

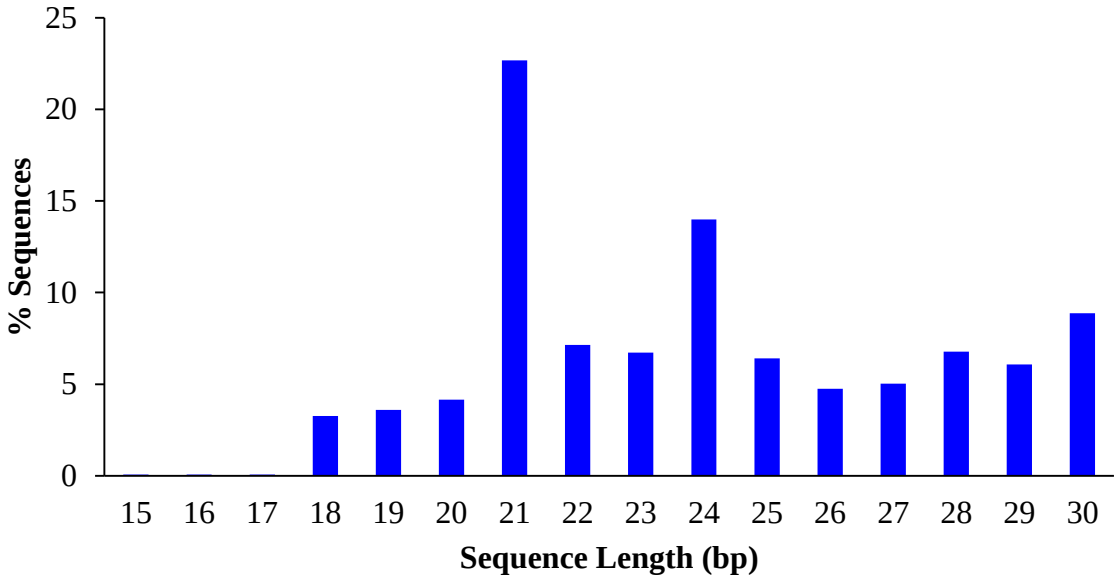

Figure 4.4: Read length distribution of the sequence reads generated from the small RNA (sRNA) libraries.

4.4.9 *De novo* assembly of unmapped sequence reads

The 6.9 million cleaned sequence reads were mapped to host sequences, which lead to the removal of 12% of the sRNA sequence reads (Table 4.5). The remaining 88% of the sequence reads were assembled into 906 contigs using Velvet (Table 4.6). A k-mer length of k-15 was selected and default parameters were used to assemble the contigs. The largest contig of 5358 bp was assembled from sample KT6. While sample K10 had the smallest contig of 193 bp (Table 4.6). The contigs from the 7 sRNA libraries were subjected to BLASTn and BLASTx searches against the virus database. Twenty-five % (225 out of 906) of the contigs generated gave hits after BLAST searches against virus genomes. A total of 103 contigs were identified as sweet potato badnavirus A (SPBVA) and 28 contigs were

identified as sweet potato badnavirus B (SPBVB). Only 3 contigs showed similarity to SPLCSPV and SPMaV, which were detected in the RCA dataset.

Table 4.5: Summary statistics of small RNA libraries. Sequence reads mapped to host sequences (sweet potato chloroplast, mitochondrial DNA and EST sequences).

Sample ID	Raw Data	Sequence reads after QC	Mapped to host sequences	Unmapped sequence reads
KT10	30581766	377766	70385	307381
KZ17	10699940	835368	76738	758630
KT6	42736660	2395850	353000	2042850
L18	34721208	1531129	189418	1341711
K10	8349374	699803	66281	633522
M11	7624924	580973	53350	527623
Z24	6910808	517776	42396	475380
Total	**141624680**	**6938665**	**851568**	**6087097**

Table 4.6: Contigs assembled using Velvet (k-mer length 15)

Sample ID	Total contigs	Maxium contig	N50	Number of sequence reads
KT10	91	4116	1282	307381
KZ17	124	1180	16	758630
KT6	135	5385	30	2042850
L18	259	3958	9	1341711
K10	107	193	12	633522
M11	102	426	21	527623
Z24	88	2428	432	475380

Table 4.7: Number of virus derived small interfering RNAs (vsiRNAs) mapped to sweet potato DNA viruses.

Sample ID	SPBVA	SPBVB	SPLCSPV	SPMaV
KT10	919	4867	86	71
KZ17	1391	1387	-	-
KT6	1387	1412	1867	4307
L18	5531	5648	2684	2554
K10	1208	1218	-	-
M11	684	795	-	-
Z24	930	952	-	-
Total	**12050**	**16279**	**4637**	**6932**

4.4.10 Reference-guided assembly of SPBVA and SPBVB

Since short contigs of SPBVA and SPBVB were generated from the *de novo* assembly, full-length genome sequences of the sweet potato badnaviruses were retrieved from the NCBI database in order to achieve higher genome coverage. When sequence reads from individual samples were mapped to badnaviruses, we detected the highest number of SPBVA specific siRNAs in sample L18, and the highest number of SPBVB specific siRNAs also in L18 (Table 4.7). Sample KT10 had a high number of SPBVB specific siRNAs and a lower number of SPBVA specific siRNAs. The SPBVA and SPBVB siRNA read count in samples KZ17, KT6 and K10 was over 1000. Samples M11 and Z24 had the lowest number of SPBVA and SPBVB specific reads (Table 4.7). The total number of SPBVA and SPBVB specific siRNAs from the 7 sRNA libraries was 12050 and 16279, respectively. Badnaviruses were detected in the symptomatic (KT6, KT10, L9, L11, L18, K17, Z24) and asymptomatic samples (K10, M11) as co-infections. Mixed infections of begomoviruses (SPLCSPV and SPMaV) and badnaviruses (SPBVA and SPBVB) were detected in 3 samples (L18, KT6 and KT10) (Table 4.7).

The siRNAs read mappings from the individual samples could not generate full-length genomes, therefore low genome coverage and sequence depth was observed. All the reads from the 7 sRNA libraries were combined and mapped to the sweet potato badnavirus reference genomes. Read mapping achieved over 80% genome coverage for SPBVA and over 90% genome coverage for SPBVB (Figure 4.5). The partial sequence assembled for SPBVA was 6786 bp long and shared 90% nucleotide similarity with the Peruvian SPBVA genome (Accession: FJ560943) and the SPBVB sequence was 7336 bp long and 96% identical to the Peruvian SPBVB isolate (Accession: FJ560944).

4.4.11 PCR confirmation of DNA viruses identified by NGS

A PCR assay for the two begomoviruses was performed and the expected band sizes of approximately 322 bp and 314 bp for SPLCSPV and SPMaV respectively were obtained (Fig. 5). Sanger sequencing confirmed the identity of the begomoviruses. The PCR products were subjected to BLASTn and BLASTx searches. The SPLCSPV coat protein sequence shared 100% nucleotide (nt) identity with the KT10AII (Accession: KX859238) and 100% similarity with the SPLCSPV coat protein (Accession: AFO66458). The SPMaV shared 100% nt identity with the KT10B11 isolate (Accession: KX859239), and 100% with the SPMaV coat protein (Accession: YP_009338000). Partial amplification of the SPBVA

polyprotein gene and SPBVB ORFb gene was performed to confirm the identity of badnaviruses using the oligonucleotides designed from the sequences generated in this study. The SPBVA sequence shared 100% nt identity with the SPBVA isolate from China (Accession: KT448733), while the SPBVB sequence shared 99% nt identity with the Spanish isolate (Accession: KU511272). The sequences were submitted to the GenBank under the accession numbers KY829453 and KY829454, for SPBVA and SPBVB, respectively

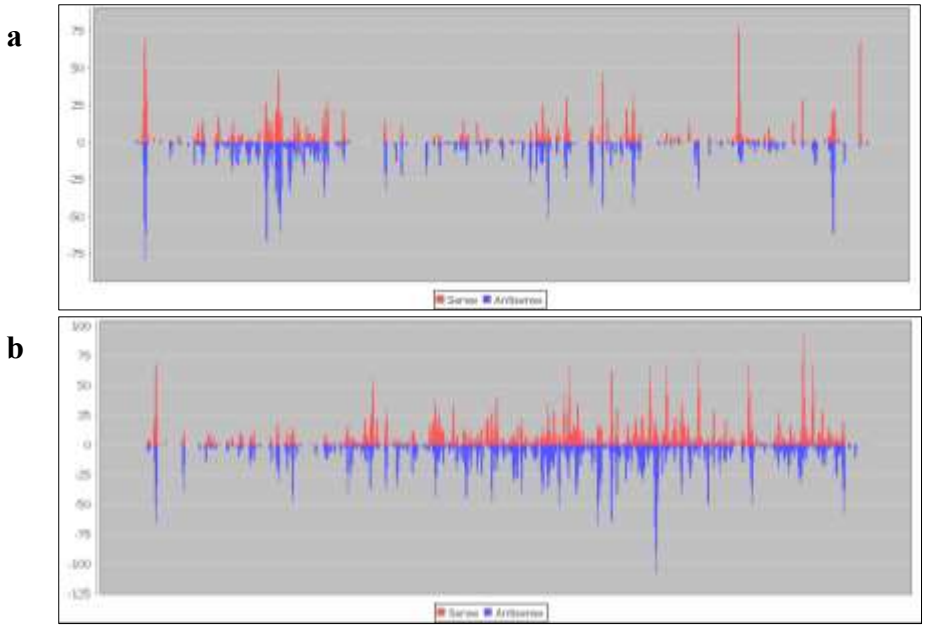

Figure 4.5: The distribution of vsiRNAs mapped along the genomes of SPBVA (a) and SPBVB (b) in the sense and antisense direction, showing the genome coverage.

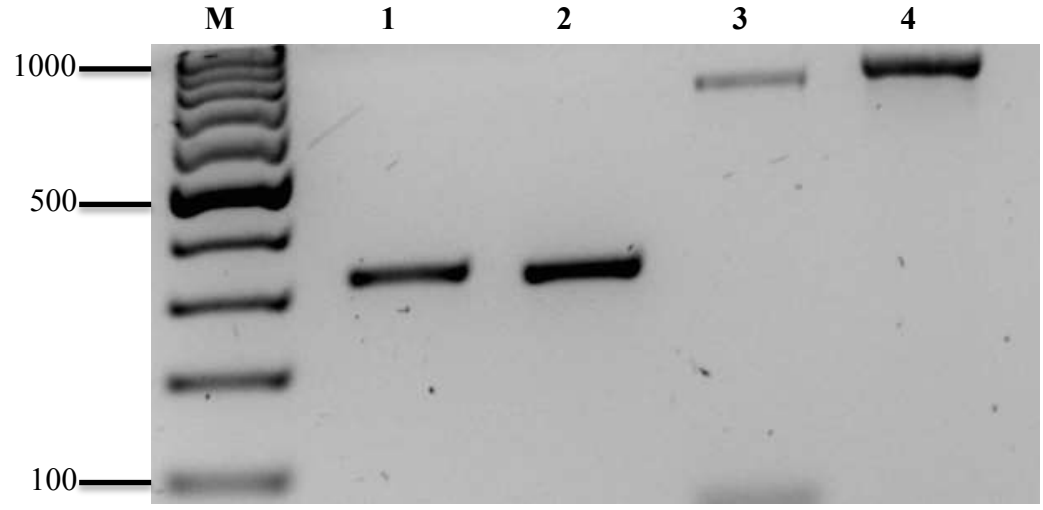

Figure 4.6: Gel electrophoresis showing the PCR amplification of SPLCSPV (lane 1), SPMaV (lane 2), SPBVA (lane 3) and SPBVB (lane 4), done using virus-specific oligonucleotide primers designed from this study. Lane M shows the 100bp DNA ladder.

4.4.12 Biological indexing on the indicator plant

After detecting and confirming the identity of 8 viruses (SPCSV, SPFMV, SPVG, SPVC, SPBVA, SPBVB, SPMaV and SPLCSPV) in sample KT10, we went on to examine the symptoms that mixed viruses would induce in *I. setosa*. We also wanted to determine the symptoms that would be induced when a plant is infected with SPCSV only. A sweet potato plant that was infected with SPCSV was obtained from the ARC-VOPI (GAU15). Cuttings from sample KT10 and sample GAU15 were multiplied and planted in seedling trays. Six scions from sample GAU15 were graft-inoculated onto six indicator plants (treatment 1) and six scions from KT10 were graft-inoculated onto six indicator plants (treatment 2). Inoculation was performed when the first true leaf had expanded (usually 10-20 days post emergence). Six non-inoculated *I. setosa* plants were used as controls (treatment 3). The plants were observed for symptom development for six weeks. All plants underwent disease scoring using a scale of 1-5 (Domola et al., 2008, Mwanga et al., 2001). After a period of three weeks, the indicator plants infected with multiple viruses exhibited mild symptoms (e.g. vein clearing). At the end of the six-week period, the symptoms had intensified to include chlorosis, chlorotic spots, leaf curl, and vein clearing (Figures 4.7e and 4.7f) and leaf necrosis (Table 4.8). Two of the six indicator plants infected with SPCSV displayed moderate to severe symptoms (Table 4.8). Indeed multiple viruses cause more severe symptoms on the indicator plants. At this point, we do not know or cannot conclude which viruses are responsible for the different symptoms. Further investigation is required.

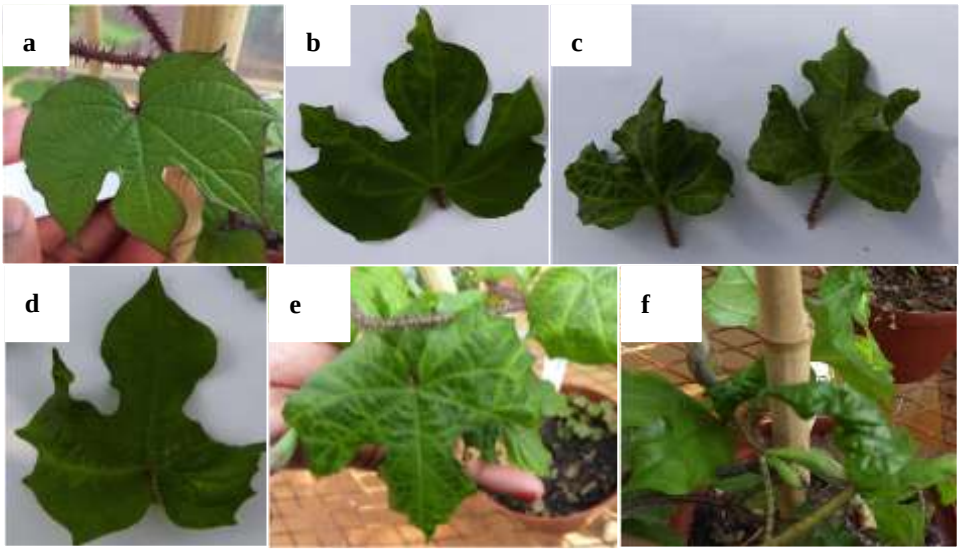

Figure 4.7: Biological assays for SPCSV. Healthy *I. setosa* control (a); Old and young *I. setosa* leaves subjected to single infection of SPCSV respond by displaying vein clearing and chlorosis (b & c). Indicator plants infected with multiple viruses exhibiting leaf curl and vein clearing (d), chlorosis and mottling (e), and puckering (f).

Table 4.8: Disease index scoring of *I. setosa* subjected to single and multiple viral infections

Replicate number	Treatment 1- Single SPCSV[a]	Treatment 2 – Mixed Infection[b]	Treatment 3 – Control[c]
1	No symptom [1]	Chlorosis/leaf curl [3]	No symptom [1]
2	No symptom [1]	Leaf necrosis [4]	No symptom [1]
3	Chlorosis, vein clearing [3]	Chlorosis, chlorotic spots, mottling [4]	No symptom [1]
4	Mild vein clearing/chlorosis on young leaves [3]	Vein clearing/chlorosis/severe leaf curling [4]	No symptom [1]
5	No symptom [1]	Leaf necrosis [5]	No symptom [1]
6	No symptom [1]	Vein clearing/necrosis/chlorotic spots [4]	No symptom [1]

Footnote: [a] Replicates were graft-inoculated with scions that tested positive for sweet potato chlorotic stunt (SPCSV) only. [b] Clean test plants were graft-inoculated with scions infected with SPCSV, sweet potato feathery mottle virus (SPFMV), sweet potato virus C (SPVC), sweet potato virus G (SPVG), sweet potato mosaic virus (SPMaV) and sweet potato leaf curl Sao Paulo virus (SPLCSPV). [b] Control test plants were not grafted.

4.5 Discussion

Studies that focus on the discovery or detection of DNA viruses have relied on the use of molecular methods such as polymerase chain reaction (PCR) (Li et al., 2004) and rolling circle amplification (RCA) coupled with restriction fragment length polymorphism (RFLP) (Haible et al., 2006, Paprotka et al., 2010, Schubert et al., 2007, Wyant et al., 2012). Due to next generation sequencing (NGS) technologies, virus discovery has shifted towards the use of high throughput sensitive and accurate methods. Deep sequencing of small RNAs is a universal and sensitive method for virus detection (Kashif et al., 2012, Kreuze et al., 2009). By using RCA-NGS two begomoviruses (SPMaV and SPLCSPV), previously detected from sweet potato cuttings in the Limpopo province in SA in 2012 (Esterhuizen et al., 2012), were detected for the first time in the coastal WC province in this study. Further investigations of the distribution of begomoviruses in other sweet potato cultivation areas of the country are warranted. A countrywide screening for the presence of begomoviruses is necessary since a small sample size was used in the study. It is possible that gemiviruses are widespread. Oligonucleotides sequences designed from this work can be used in multiplex assays for screening the sweet potato germplasm housed at the ARC-VOPI and for screening plants from other sweet potato growing regions of South Africa.

Based on previous reports, begomoviruses (Clark et al., 2012) have a wide global distribution. Begomoviruses have been detected in *Ipomoea* spp. including *Ipomoea indica* and *Ipomoea batatas* from China (Luan et al., 2006, Zulfiqar et al., 2011), Brazil (Albuquerque et al., 2012, Paprotka et al., 2010), Peru (Kreuze et al., 2009), Uganda (Wasswa et al., 2011) and Taiwan (Li et al., 2004). Typical symptoms associated with begomoviruses include upward curling of leaves, chlorosis and vein clearing. In this study, all of the plants that were infected with SPLCSPV and SPMaV displayed symptoms including but not limited to leaf curling. SPLCSPV and SPMaV have been found co-infecting sweet potato (Paprotka et al., 2010), and in this study the viruses were also found in co-infections in three plant samples. Since there is evidence demonstrating that these two viruses have undergone recombination (Esterhuizen et al., 2012), it is therefore likely that these two viruses occur globally in mixed infections in sweet potato germplasm, and were introduced into SA. The symptom development observed in the plants infected with SPLCSPV and SPMaV could be attributed to the fact that begomoviruses have evolved mechanisms that can evade or suppress the RNA silencing mechanism (Pooggin, 2013, Rajeswaran et al., 2014) through the expression of a viral silencing suppressor, in this case a transcriptional activator (TrAP) protein, encoded by the C2 gene (Pooggin, 2013). Both the begomoviruses detected in this study contain a potential silencing protein encoded by the C2 gene. Evasion of the host silencing mechanisms increases viral replication and infectious viral copies that spread throughout the plant and cause leaf curl symptoms. SPLCSPV and SPMaV were also found in mixed infections with two badnaviruses in this study.

Sweet potato badnaviruses were first discovered in Peru using siRNA deep sequencing (Kreuze et al., 2009). Since then they have been detected in symptomatic and asymptomatic plants from Tanzania (Mbanzibwa et al., 2016, Mbanzibwa et al., 2014), Honduras and Guatemala (Kashif et al., 2012). Consistent with previous studies performed in Peru and Tanzania (Iskra-Caruana et al., 2014, Kreuze et al., 2009, Mbanzibwa et al., 2014), our data shows that badnaviruses (SPBVA and SPBVB) are found in symptomatic and asymptomatic plants. The identity of the badnaviruses was confirmed by PCR amplification and Sanger sequencing. To our knowledge, this is the first report of sweet potato badnaviruses in South Africa. Badnaviruses appear to have a wide global distribution suggesting that they have been integrated into the sweet potato genome a long time ago and are present in germplasm in many countries in the world. There is growing evidence showing that plant virus sequences have integrated into plant genomes during the co-evolution of plant families and viruses

(Harper et al., 2002). The integration of virus sequences into the host genome has been reported for plant viruses in the family *Geminiviridae* and family *Caulimoviridae* (which includes the genus *Badnavirus*) (Harper et al., 2002). Studies have reported integrated badnavirus sequences in yam (Bhat et al., 2016), banana (Ndowora et al., 1999) and members of the *Solanaceae* family (Harper et al., 2002). Although badnaviruses currently do not cause any exogenous infections that result in symptom development, they may play a role in disease severity. Examples in banana indicate that badnaviruses can exist as endogenous sequences, or episomal forms that infect plants causing visible symptoms (Iskra-Caruana et al., 2014, Staginnus and Richert-Pöggeler, 2006). Endogenous sequences can generate circular viral genomes that trigger infections, and this phenomenon can occur when the plants are exposed to stress conditions such as abiotic stress or *in vitro* tissue culture (Bhat et al., 2016, Chabannes and Iskra-Caruana, 2013, Iskra-Caruana et al., 2014). Future studies are necessary to determine whether badnavirus sequences have been integrated into the sweet potato genome. Also, it will be important to also investigate the role badnaviruses play in etiology of sweet potato disease, since they are known to trigger infections and cause emerging diseases.

Begomoviruses and badnaviruses were found in mixed infections in three samples (KT10, KT6, L18). In the previous chapter we also detected multiple virus infection of RNA and DNA viruses (KT10, KT6, L18, KF1). In one of the samples, KT10, we detected sequences originating from 4 RNA viruses (SPFMV, SPVC, SPVG and SPCSV) and 4 DNA viruses (SPLCSPV, SPMaV, SPBVA and SPBVB). The biological assays conducted in this study showed that mixed viruses cause severe symptom development on the indicator plant. When sweet potato plants were inoculated with SPCSV only, mild symptoms were observed (chlorosis and vein clearing). Multiple infections of SPFMV, SPCSV, SPVG, SPVC and SPLCSPV and SPMaV, resulted in severe symptoms including upward curling of leaves, chlorotic spots, mottling and necrosis. The occurrence and correlation between mixed infections and symptom severity has been reported in sweet potato (Mukasa et al., 2006). In this study our data demonstrates that SPCSV acts in synergy with many viruses including SPFMV, SPLCSPV, SPMaV, SPVG, and SPVC to cause severe symptoms.

In conclusion, this study revealed the presence of DNA viruses, begomoviruses (SPLCSPV and SPMaV) and badnaviruses (SPBVA and SPBVB) in sweet potato plants from the WC province. This is the first report of endogenous pararetroviruses belonging to the genus

Badnavirus in sweet potato in South Africa. The badnaviruses were detected by deep sequencing of sRNAs, this method emphasises the advantages of NGS as a diagnostic strategy. Furthermore, SPLCSPV and SPMaV caused symptomatic infection, while SPBVA and SPBVB were detected in some asymptomatic plants. Different sweet potato cultivars and indicator plants should be indexed (biological assays) under normal and stress conditions in order to investigate the role of badnaviruses in single, double or mixed viral infections with other viruses. Since the sweet potato genome sequence is not available, it will be interesting in the future to search for integrated badnavirus or geminivirus-like sequences in the sweet potato genome.

4.6 Acknowledgements

This work is based on the research support in part by the National Research Foundation of South Africa (Grant reference number UID 79983). The authors would like to acknowledge Mr. Thobejane for assistance with maintaining the glasshouse plants.

4.7 Literature Cited

ALBUQUERQUE, L. C., INOUE-NAGATA, A. K., PINHEIRO, B., RESENDE, R. O., MORIONES, E. & NAVAS-CASTILLO, J. 2012. Genetic diversity and recombination analysis of sweepoviruses from Brazil. *Virology Journal,* 9**,** 241.

ALBUQUERQUE, L. C., INOUE-NAGATA, A. K., PINHEIRO, B., RIBEIRO, S. D. G., RESENDE, R. O., MORIONES, E. & NAVAS-CASTILLO, J. 2011. A novel monopartite begomovirus infecting sweet potato in Brazil. *Archives of Virology,* 156**,** 1291-1294.

ARITUA, V. & ADIPALA, E. Characteristics and diversity in sweetpotato-infecting viruses in Africa. II International Symposium on Sweetpotato and Cassava: Innovative Technologies for Commercialization 703, 2005. 175-182.

ARONESTY, E. 2013. Comparison of sequencing utility programs. *The Open Bioinformatics Journal,* 7.

BHAT, A. I., HOHN, T. & SELVARAJAN, R. 2016. Badnaviruses: the current global scenario. *Viruses,* 8**,** 177.

BORAH, B. K., SHARMA, S., KANT, R., JOHNSON, A., SAIGOPAL, D. V. R. & DASGUPTA, I. 2013. Bacilliform DNA ‐ containing plant viruses in the tropics:

commonalities within a genetically diverse group. *Molecular Plant Pathology,* 14**,** 759-771.

BOUHIDA, M., LOCKHART, B. & OLSZEWSKI, N. E. 1993. An analysis of the complete sequence of a sugarcane bacilliform virus genome infectious to banana and rice. *Journal of General Virology,* 74**,** 15-22.

CHABANNES, M. & ISKRA-CARUANA, M.-L. 2013. Endogenous pararetroviruses—a reservoir of virus infection in plants. *Current Opinion in Virology,* 3**,** 615-620.

CLARK, C. A., DAVIS, J. A., ABAD, J. A., CUELLAR, W. J., FUENTES, S., KREUZE, J. F., GIBSON, R. W., MUKASA, S. B., TUGUME, A. K. & TAIRO, F. D. 2012. Sweetpotato viruses: 15 years of progress on understanding and managing complex diseases. *Plant Disease,* 96**,** 168-185.

DAHAL, G., ORTIZ, R., TENKOUANO, A., HUGHES, J. D. A., THOTTAPPILLY, G., VUYLSTEKE, D. & LOCKHART, B. 2000. Relationship between natural occurrence of banana streak badnavirus and symptom expression, relative concentration of viral antigen, and yield characteristics of some micropropagated Musa spp. *Plant Pathology,* 49**,** 68-79.

DE ANDRADE, R. R. & VASLIN, M. F. 2014. SearchSmallRNA: a graphical interface tool for the assemblage of viral genomes using small RNA libraries data. *Virology Journal,* 11**,** 45.

DOMOLA, M., THOMPSON, G., AVELING, T., LAURIE, S., STRYDOM, H. & VAN DEN BERG, A. 2008. Sweet potato viruses in South Africa and the effect of viral infection on storage root yield. *African Plant Protection,* 14**,** 15-23.

ELBEAINO, T., CHIUMENTI, M., DE STRADIS, A., DIGIARO, M., MINAFRA, A. & MARTELLI, G. 2013. Identification of a badnavirus infecting mulberry. *Journal of Plant Pathology***,** 207-210.

ESTERHUIZEN, L., VAN HEERDEN, S., REY, M. & VAN HEERDEN, H. 2012. Genetic identification of two sweet-potato-infecting begomoviruses in South Africa. *Archives of Virology,* 157**,** 2241-2245.

GEERING, A. D., OLSZEWSKI, N. E., HARPER, G., LOCKHART, B. E., HULL, R. & THOMAS, J. E. 2005. Banana contains a diverse array of endogenous badnaviruses. *Journal of General Virology,* 86**,** 511-520.

GIBSON, R., MWANGA, R., KASULE, S., MPEMBE, I. & CAREY, E. 1997. Apparent absence of viruses in most symptomless field‐grown sweet potato in Uganda. *Annals of Applied Biology,* 130**,** 481-490.

HAIBLE, D., KOBER, S. & JESKE, H. 2006. Rolling circle amplification revolutionizes diagnosis and genomics of geminiviruses. *Journal of virological methods,* 135**,** 9-16.

HANLEY-BOWDOIN, L., BEJARANO, E. R., ROBERTSON, D. & MANSOOR, S. 2013. Geminiviruses: masters at redirecting and reprogramming plant processes. *Nature Reviews Microbiology,* 11**,** 777-788.

HARPER, G., HULL, R., LOCKHART, B. & OLSZEWSKI, N. 2002. Viral sequences integrated into plant genomes. *Annual Review of Phytopathology,* 40**,** 119-136.

ISKRA-CARUANA, M.-L., DUROY, P.-O., CHABANNES, M. & MULLER, E. 2014. The common evolutionary history of badnaviruses and banana. *Infection, Genetics and Evolution,* 21**,** 83-89.

JAMES, A., GEIJSKES, R. J., DALE, J. L. & HARDING, R. M. 2011. Molecular characterisation of six badnavirus species associated with leaf streak disease of banana in East Africa. *Annals of Applied Biology,* 158**,** 346-353.

KASHIF, M., PIETILÄ, S., ARTOLA, K., JONES, R., TUGUME, A., MÄKINEN, V. & VALKONEN, J. 2012. Detection of viruses in sweetpotato from Honduras and Guatemala augmented by deep-sequencing of small-RNAs. *Plant Disease,* 96**,** 1430-1437.

KREUZE, J. 2002. *Molecular studies on the sweet potato virus disease and its two causal agents*.

KREUZE, J. & FUENTES, S. 2008. Sweetpotato Viruses A2 - Mahy, Brian W.J. *In:* REGENMORTEL, M. H. V. V. (ed.) *Encyclopedia of Virology (Third Edition).* Oxford: Academic Press.

KREUZE, J. F., PEREZ, A., UNTIVEROS, M., QUISPE, D., FUENTES, S., BARKER, I. & SIMON, R. 2009. Complete viral genome sequence and discovery of novel viruses by deep sequencing of small RNAs: a generic method for diagnosis, discovery and sequencing of viruses. *Virology,* 388**,** 1-7.

LANEY, A. G., HASSAN, M. & TZANETAKIS, I. E. 2012. An integrated badnavirus is prevalent in fig germplasm. *Phytopathology,* 102**,** 1182-1189.

LI, R., SALIH, S. & HURTT, S. 2004. Detection of geminiviruses in sweetpotato by polymerase chain reaction. *Plant Disease,* 88**,** 1347-1351.

LUAN, Y., ZHANG, J. & AN, L. 2006. First report of Sweet potato leaf curl virus in China. *Plant Disease,* 90**,** 1111-1111.

MBANZIBWA, D., TAIRO, F., GWANDU, C., KULLAYA, A. & VALKONEN, J. 2016. First report of sweetpotato symptomless virus 1 and sweetpotato virus A in sweetpotatoes in Tanzania. *Plant Disease,* 100**,** 865.

MBANZIBWA, D., TUGUME, A., CHIUNGA, E., MARK, D. & TAIRO, F. 2014. Small RNA deep sequencing ‐ based detection and further evidence of DNA viruses infecting sweetpotato plants in Tanzania. *Annals of Applied Biology,* 165**,** 329-339.

MUKASA, S. B., RUBAIHAYO, P. R. & VALKONEN, J. 2006. Interactions between a crinivirus, an ipomovirus and a potyvirus in coinfected sweetpotato plants. *Plant Pathology,* 55**,** 458-467.

MWANGA, R., MOYER, J., ZHANG, D., CAREY, E. & YENCHO, G. Nature of resistance of sweetpotato to sweetpotato virus disease. I International Conference on Sweetpotato. Food and Health for the Future 583, 2001. 113-119.

NDOWORA, T., DAHAL, G., LAFLEUR, D., HARPER, G., HULL, R., OLSZEWSKI, N. E. & LOCKHART, B. 1999. Evidence That Badnavirus Infection inMusaCan Originate from Integrated Pararetroviral Sequences. *Virology,* 255**,** 214-220.

NDOWORA, T. C. & LOCKHART, B. E. Development of a serological assay for detecting serologically diverse banana streak virus isolates. I International Symposium on Banana: I International Conference on Banana and Plantain for Africa 540, 1996. 377-388.

NYABOGA, E., ATEKA, E. & BULIMO, W. 2008. Serological detection of virus diseases of sweet potato in Kenya. *Journal of Applied Biosciences,* 7**,** 222-229.

PAPROTKA, T., BOITEUX, L., FONSECA, M., RESENDE, R., JESKE, H., FARIA, J. & RIBEIRO, S. 2010. Genomic diversity of sweet potato geminiviruses in a Brazilian germplasm bank. *Virus Research,* 149**,** 224-233.

POOGGIN, M. M. 2013. How can plant DNA viruses evade siRNA-directed DNA methylation and silencing? *International journal of molecular sciences,* 14**,** 15233-15259.

RAJESWARAN, R., SEGUIN, J., CHABANNES, M., DUROY, P.-O., LABOUREAU, N., FARINELLI, L., ISKRA-CARUANA, M.-L. & POOGGIN, M. M. 2014. Evasion of short interfering RNA-directed antiviral silencing in Musa acuminata persistently

infected with six distinct banana streak pararetroviruses. *Journal of Virology,* 88**,** 11516-11528.

SCHUBERT, J., HABEKUS, A., KAZMAIER, K. & JESKE, H. 2007. Surveying cereal-infecting geminiviruses in Germany—diagnostics and direct sequencing using rolling circle amplification. *Virus Research,* 127**,** 61-70.

SIM, J. G. & CLARK, C. 2008. Virus-like particles and cellular changes in plants infected with sweetpotato viruses. *The Plant Pathology Journal,* 24**,** 36-45.

STAGINNUS, C. & RICHERT-PÖGGELER, K. R. 2006. Endogenous pararetroviruses: two-faced travelers in the plant genome. *Trends in Plant Science,* 11**,** 485-491.

VALVERDE, R. A., CLARK, C. A. & VALKONEN, J. P. 2007. Viruses and virus disease complexes of sweetpotato. *Plant Viruses,* 1**,** 116-126.

WAMBULWA, M., WACHIRA, F., KARANJA, L., KIARIE, S. & MUTURI, S. 2013. The influence of host and pathogen genotypes on symptom severity in banana streak disease. *African Journal of Biotechnology,* 12**,** 27-31.

WASSWA, P., OTTO, B., MARUTHI, M., MUKASA, S., MONGER, W. & GIBSON, R. 2011. First identification of a sweet potato begomovirus (sweepovirus) in Uganda: characterization, detection and distribution. *Plant Pathology,* 60**,** 1030-1039.

WYANT, P. S., STROHMEIER, S., SCHÄFER, B., KRENZ, B., ASSUNÇÃO, I. P., DE ANDRADE LIMA, G. S. & JESKE, H. 2012. Circular DNA genomics (circomics) exemplified for geminiviruses in bean crops and weeds of northeastern Brazil. *Virology,* 427**,** 151-157.

ZERBINO, D. R. & BIRNEY, E. 2008. Velvet: algorithms for de novo short read assembly using de Bruijn graphs. *Genome Research,* 18**,** 821-829.

ZULFIQAR, A., REN, H., ZHANG, J., LI, J., CUI, X. & XIE, Y. 2011. Molecular characterization of two isolates of sweet potato leaf curl virus infecting Ipomoea indica in China. *African Journal of Biotechnology,* 10**,** 9061-9067.

South African Human Rights Commission (SAHRC), 2008 http://www.sahrc.org.za

CHAPTER 5

IDENTIFICATION OF MIRNAS ASSOCIATED WITH MIXED VIRAL INFECTIONS IN SUSCEPTIBLE AND RESISTANT SWEET POTATO CULTIVARS

5.1 Abstract

Micro RNAs (miRNAs) are small non-coding RNAs involved in the regulation of important biological processes such as plant development, biotic and abiotic stress and pathogen defense. Studies in sweet potato have only focused on the characterisation of virus-derived small interfering RNAs (vsiRNAs). This study therefore aimed not only to identify the vsiRNAs associated with mixed virus infections but also to identify and compare the differential expression of miRNAs between sweet potato susceptible (Blesbok) and resistant (NASPOT 1) cultivars after infection with multiple viruses. The cultivars were infected with a combination of 8 viruses that were found infecting sample KT10 from Klawer in the Western Cape province of South Africa. Small RNA was isolated from leaf tissue at 60 days post infection (dpi) and 12 libraries were prepared for sequencing. Sequence analysis was performed with the CLC Bio Genomics Workbench and psRNATarget platforms. A total of 60% of the sequence reads matched to rRNA, nc-RNA, and tRNA sequences in the Rfam database. VsiRNAs accounted for 1% of the data. The vsiRNA count was significantly higher in both the Blesbok and NASPOT 1 after infection with multiple viruses. Further, the vsiRNA sequence reads in the susceptible cultivar were 3 times higher than the vsiRNA sequence reads in the resistant cultivar. The majority of vsiRNAs were associated with SPVC, SPFMV, and SPVG in the NASPOT 1 cultivar, while SPFMV, SPVC and SPBVB vsiRNA reads were significantly higher in the Blesbok cultivar. Twenty one percent (21%) of the sequence reads matched to known miRNAs in the miRBase (Release 21) database, and 60 miRNA families were identified from the NASPOT 1 and Blesbok cultivars collectively. The remaining 18% of the reads were unannotated. Conserved miRNAs including miR398, miR319, miR168, miR166 were present in both Blesbok and NASPOT 1 cultivars. Non-conserved miRNAs such as miR6300 found in *Glycine max* were identified in both *Ipomoea batatas* cultivars. Amongst the miRNAs identified from our data set, we detected upregulation of miR393, miR398, miR168, miR162, miR167 and miR397 in both cultivars.

We also observed upregulation of a defense-related miRNA, miR482, in both cultivars. Two miRNAs (miR6300 and miR160) were downregulated in both cultivars after virus infection. Interestingly, miR399, miR319, miR530 and miR5077 were upregulated in the resistant cultivar and downregulated in the susceptible cultivar. MiR403 was downregulated in the resistant cultivar and upregulated in the susceptible cultivar. The expression of miR393 and miR397 populations was significantly higher in the Blesbok cultivar than the NASPOT 1 post infection. The changes in the expression patterns of these miRNAs infer that they could play a role in pathogen response. The differentially expressed miRNAs identified in this study, miR168, miR403, miR162 and miR482, are known to regulate Argonaute 1 (AGO1), AGO 2 & 3, Dicer-like 1 (DCL1) and NB-LRR resistance genes respectively. These proteins and genes are known to play key roles in the silencing pathway. Other differentially expressed miRNAs namely, miR167, miR399, miR393 and miR397, target auxin response factors (AFR), PHO2, (miR393 target) and laccases respectively. These targets are known to be associated with stress response in plants. NGS expression profiles of miR482, miR393, miR397 and miR168 were validated by quantitative real-time PCR (RT-qPCR).

5.2 Introduction

Plants have developed defense mechanisms in response to viral infection (Costa et al., 2013). The first line of defense known as basal resistance or innate immunity is induced when plant cells recognise pathogen/microbe-associated molecular patterns (PAMPs or MAMPs) via pathogen recognition receptors (PRRs) in the plasma membrane (Freeman and Beattie, 2008, Sharma et al., 2014). Other defense mechanisms include the hypersensitive response (HR), systemic acquired resistance (SAR) and a recently described system (ubiquitin/26S proteasome system (UPS)) known to be involved in the degradation of ubiquitinated viral proteins (Casassola et al., 2013, Pallas and García, 2011, Sekine et al., 2008, Soosaar et al., 2005). Susceptible and resistant host plants (cultivars) respond differently to virus infection. Different sets of genes are expressed in susceptible and resistant cultivars upon viral entry and replication. Studies have shown that during HR, resistant cultivars express defense-related genes such as resistance genes (R genes) (McDowell and Woffenden, 2003, Sharma et al., 2014, Soosaar et al., 2005) or pathogenesis-related (PR) proteins (Thakur and Sohal, 2013) that localise virus infection, and prevent the infection from spreading further (Costa et al., 2013, Goyer et al., 2015, Ishihara et al., 2004, Ruiz-Ferrer and Voinnet, 2009, Thakur and Sohal, 2013). Systematic acquired resistance (SAR), the second defense mechanism, then

protects the plant against future attacks by the same pathogens or other new pathogens (Costa et al., 2013, Huang et al., 2005, Ishihara et al., 2004). On the contrary, susceptible hosts succumb to virus invasion and cells become systemically infected which often results in symptom development such as chlorosis and stunting (Agudelo-Romero et al., 2008, Goyer et al., 2015, Havelda et al., 2008, Ishihara et al., 2004, Thakur and Sohal, 2013, Whitham et al., 2006).

RNA silencing (virus induced gene silencing - VIGS) is another mechanism of defense plants use to counteract viruses (Freeman and Beattie, 2008). RNA silencing can also function as a gene expression regulatory mechanism (Guo et al., 2015) and the silencing pathway has the potential to be used as a tool to engineer resistant plants (Duan et al., 2012). The silencing pathway is comprised of different proteins, enzymes and short molecules (small RNAs) that interact to regulate gene expression. Over the years much interest has gone into investigating and characterising small RNAs and their role in gene silencing. Small RNAs are non-coding RNAs (sncRNAs) found in animals and plants. These sncRNAs are important protein-coding gene expression regulators that function by causing either transcriptional gene silencing (TGS) or post-transcriptional gene silencing (PTGS) (Khraiwesh et al., 2012, Ku et al., 2015). Small RNAs are 18 – 30 nt molecules categorised into different classes based on biogenesis and function (Jones and Dangl, 2006, López et al., 2012, Wang et al., 2012)**.** Reviews extensively covering the characterisation of sncRNAs in plants, animals and humans are available (Axtell, 2013, Borges and Martienssen, 2015, Finnegan and Matzke, 2003, Guleria et al., 2011, He and Hannon, 2004, Mallory and Vaucheret, 2006).

One of the well-studied classes of sncRNAs are microRNAs (miRNAs), which are defined as endogenous molecules that play a role in regulating gene expression at the post-transcriptional level (Pantaleo et al., 2010, Pérez-Quintero et al., 2010). Most of these miRNAs are conserved across plant species and are important regulators of development, hormone responses, feedback mechanisms, abiotic and biotic stress responses and pathogen responses (Jagtap and Shivaprasad, 2014, Pantaleo et al., 2010). Almost 50% of the conserved plant miRNAs target transcription factors and the remaining percentage play a role in regulating the expression of protein coding genes involved in metabolic processes, RNA silencing and disease resistance (Jagtap and Shivaprasad, 2014). The miRNA biogenesis pathway involves the production of short (21 bp) duplexes generated from long non-coding genome-encoded transcripts. Primary mRNA transcripts form secondary transcripts (pre-

miRNA stem-loop structure), can be recognized and cleaved by Dicer-like (DCL1) enzymes, to further generate a duplex (Abreu et al., 2014, Guleria et al., 2011). The duplex is made up of the mature miRNA strand and the complementary miRNA star strand (miRNA*) (miRNA: miRNA*). The mature miRNA is then incorporated into an Argonaute protein (AGO1) to form the RNA-induced silencing complex (RISC), and the miRNA* strand is degraded (Abreu et al., 2014, Guleria et al., 2011). The RISC is guided to target protein-coding RNAs, which are cleaved by AGO1 between position 10 and 11 (Pantaleo et al., 2010). The plant miRNAs then target mRNAs for degradation or translational repression (Axtell, 2013, Jagtap and Shivaprasad, 2014, Mallory and Vaucheret, 2006).

In plants, another class of sncRNAs are small interfering RNAs (siRNAs), which are produced from long double-stranded RNA (dsRNAs) that originate from viruses, transgenes and transposons (López et al., 2012). There are several different types of siRNAs that are produced from different genes, these include 21 nt ta-siRNAs (*trans*-acting siRNAs), which are processed by DCL4 and originate from a nuclear *TAS* gene; the heterochromatin siRNAs are 24 nt long and are generated by DCL3. The third type of siRNA, known as the natural antisense siRNAs (nat-siRNAs), is made up of primary and secondary nat-siRNAs (Pantaleo et al., 2010). The 21 nt secondary nat-siRNAs are generated by DCL2 (Pantaleo et al., 2010). Virus-derived small RNAs (vsiRNAs) are generated from viral dsRNA by DCL2 and DCL4 during virus infection (Várallyay et al., 2010). Similarly to the miRNA pathway, vsiRNAs are incorporated into the RISC, which guides cleavage of target RNAs (Mallory and Vaucheret, 2006). Viral silencing suppressors (VSRs) can inhibit this mechanism at different stages by binding to vsiRNAs, dsRNAs or interacting with Argonaute proteins (AGO1) (Mlotshwa et al., 2008, Pumplin and Voinnet, 2013). VsiRNAs have been studied and identified in sweet potato (Kashif et al., 2012, Kreuze et al., 2009, Mbanzibwa et al., 2014), however very little is known about miRNAs and the target mRNAs they regulate. The aim of our study was to identify and compare miRNAs expressed in resistant vs. susceptible cultivars in response to mixed virus infections, which included SPFMV and SPCSV, the two causal agents of sweet potato virus disease (SPVD). Yield losses of up to 100% caused by SPVD have significant negative impact on food security and income for rural poor farmers (Ngailo et al., 2013). Control measures set in place to mitigate the effects of SPVD have been limited, as chemical and biological methods are not efficient (Ngailo et al., 2013). To date the most cost-effective control strategy is the use of resistant varieties. Breeding for resistance against SPVD and other viral diseases has become an important long-term approach

employed to improve crop quality and yield (Ngailo et al., 2013). In sweet potato breeding, techniques such as marker-assisted selection and genetic engineering have been developed (Mwanga et al., 2011, Ngailo et al., 2013). At the start of the new millennium, breeding for SPVD resistance in Africa (Uganda) and Asia (Japan and China) progressed considerably. The breeding efforts resulted in the development of cultivars such as the New Kawogo, NASPOT 1, 5, 6, 7, 8, 9, 10 and 11, which are resistant to SPVD (Mwanga et al., 2011, Ngailo et al., 2013). The resistant cultivars were produced via the traditional virus–induced gene silencing (VIGS) approach. This approach utilises viral vectors that carry genes of interest, usually coat proteins, which generate long double-stranded RNAs within the plant thus triggering the production of vsiRNAs resulting in VIGS (Costa et al., 2013, Duan et al., 2012, Lindbo et al., 2001, Tiwari et al., 2014). The use of artificial miRNAs to genetically engineer virus-resistant cultivars has not been explored in sweet potato. Breeding efforts can employ this aspect of the silencing pathway as another strategy for controlling major sweet potato diseases and other viruses that affect crop quality. Since sweet potato miRNA sequences are absent in small RNA repositories such as miRBase and the plant microRNA database (PMRD), the identification of miRNAs in this study will contribute to building a reference database of sweet potato miRNA sequences. The availability of this information will elucidate the role of miRNAs in pathogen defense and researchers can then explore their use in genetic engineering. To our knowledge, this is the first study that identifies and characterises pathogen defense related miRNAs in sweet potato, and it will serve as a baseline study for future research.

5.3 Materials and methods

5.3.1 Plant multiplication and graft inoculation

The sample (KT10), which tested positive for a combination of RNA and DNA viruses, was chosen as a source of inoculum for this study. In previous chapters (chapters three and four), we performed RT-PCR and PCR to confirm the presence and identity of SPFMV, SPCSV, SPLCSPV, SPMaV, SPVG, SPVC, SPBVA and SPBVB in sample KT10 from Klawer in the Western Cape province. It was for this reason that KT10 was selected and used as a source of inoculum for the investigation of small RNAs (vsiRNAs and miRNAs) associated with mixed virus infections in the Blesbok (susceptible) and NASPOT 1 (resistant) cultivars. The experiment comprised of four treatments, Blesbok infected vs. Blesbok uninfected and NASPOT 1 infected vs. NASPOT 1 uninfected. The experimental design is depicted in

Figure 5.1. Three biological replicates per treatment and four plants per biological replicate were set up for the experiment, totalling 48 plants. Cuttings of the Blesbok and NASPOT 1 cultivars were sourced from the ARC-VOPI and used for multiplication and graft inoculation. Plants were vegetatively propagated in seedling trays and then transplanted to potting soil in 20 cm diameter pots and grown in a greenhouse for 8 weeks. After 8 weeks 12 Blesbok (Blesbok infected) and 12 NASPOT 1 (NASPOT 1 infected) plants were graft inoculated with KT10 scions, while the 12 Blesbok (Blesbok uninfected) and 12 NASPOT 1 plants (NASPOT 1 uninfected) were not graft inoculated and used as controls. The plants were grown in a greenhouse and observed for symptom development for 60 days. At 60 days post infection (dpi), young and old leaves from the four plants within each biological replicate and treatment were pooled to generate 12 individual samples that were used for small RNA isolation, library preparation and deep sequencing (Figure 5.1).

5.3.2 Genomic DNA isolation, total RNA isolation and confirmation of virus presence by PCR & RT-PCR

Prior to deep sequencing, PCR and RT-PCR were conducted for the 12 individual samples to confirm infection. Total RNA and genomic DNA (gDNA) were isolated from leaf tissue using the QIAGEN RNeasy Plant Mini and QIAGEN DNeasy Plant Mini Kits (QIAGEN, Valencia, CA, USA) respectively. The total RNA underwent DNase treatment using the QIAGEN RNase-free DNase Set (QIAGEN, Valencia, CA, USA) to remove genomic DNA. The integrity of the RNA and DNA was analysed by agarose gel electrophoresis. Quantitation was performed on the Qubit fluorometer using the Qubit^{TM} RNA BR Assay Kit and Qubit^{TM} dsDNA BR Assay Kit (Invitrogen, Life Technologies), for the total RNA and gDNA respectively. The total RNA was reverse transcribed into cDNA using the TaKaRa PrimeScipt 1st strand cDNA synthesis kit (TaKaRa, Japan). PCR and RT-PCR reactions were performed according to previously described instructions (chapters three and four).

5.3.3 Small RNA isolation, library preparation and sequencing

Small RNA was isolated from leaf tissue collected from the 12 samples (Figure 5.1) using the Ambion mirVana^{TM} miRNA RNA Isolation Kit (Ambion, USA), following the manufacturer's instructions. The small RNA was quantified using the Qubit™ RNA HS Assay Kit (Life Technologies, Thermo Fisher Scientific Inc.) and then stored at -80°C until further use. The samples underwent sample preparation using the Illumina TruSeq Small

RNA Sample Preparation Kit (Illumina Inc., San Diego, CA, USA) according to the manufacturer's instructions. A total of 12 individual libraries were sequenced on the Illumina MiSeq platform.

5.3.4 Sequence analysis, siRNA and miRNA identification

The raw sequence reads generated from the 12 individual libraries were trimmed for adapter sequences and indexes utilising the CLC Bio Genomics Workbench v9. All the trimmed files underwent quality control analysis using FastQC. Sequence reads less than 18 nt and greater than 30 nt were discarded. After quality trimming the sequence reads were aligned to the Rfam database to filter out other non-coding small RNAs sncRNAs (e.g. rRNAs, tRNAs) (Figure 5.2). To detect virus-derived small RNAs (vsiRNAs), the sequences were aligned to virus reference genomes (SPCSV, SPFMV, SPVG, SPVC, SPMaV, SPLCSPV, SPBVA, SPBVB). The remaining reads underwent analysis using the transcriptomics and small RNA analysis tools within the CLC Bio Genomics Workbench. Small RNA reads were aligned to the miRBase (Release 21) database to annotate known miRNAs. For the identification of sweet potato miRNAs, we allowed 1-2 mismatches during the sequence similarity search against the miRBase database (Release 21).

5.3.5 Differential miRNA expression analysis and target prediction

Differential expression analysis was performed using the RNASeq tool on the CLC Bio Workbench v9. The putative targets of the differentially expressed (DE) miRNAs were predicted using the psRNATarget webserver (Dai and Zhao, 2011).

5.3.6 Real-Time PCR validation of DE miRNAs

Real-Time PCR (RT-qPCR) was performed to validate the differentially expressed miRNAs. Four differentially expressed miRNAs (miR482, miR168, miR393 and miR397) were selected for validation. The miRNA sequences generated from this study were first subjected to BLASTn searches on the microRNA database (miRBase) (http://www.mirbase.org/) to confirm the identity of the miRNAs and to compare the experimentally/bioinformatically generated miRNAs from this study with the miRNA sequences available miRBase. We then designed four Custom TaqMan MicroRNA Assays (containing small RNA-specific RT primer, small RNA-specific forward PCR primer, specific reverse PCR primer and small RNA-specific TaqMan ® MGB probe with a FAM dye), using the Custom TaqMan Small

RNA Assay Design Tool (https://www.thermofisher.com/order/custom-genomic-products/tools/small-rna/). Total RNA was extracted from the 12 samples using the Ambion mirVana™ miRNA RNA Isolation Kit (Ambion, USA), according to the manufacturer's instructions. The total RNA from each sample was then diluted to 10 ng. Then reverse transcription was performed using the TaqMan MicroRNA Reverse Transcription Kit, following the manufacturer's instructions. The real-time PCR reactions were performed using the custom designed TaqMan® assays and the Applied Biosystems TaqMan 2X Universal PCR Master Mix. The qPCR reactions were run in triplicates on the Roche LightCycler® 96 Instrument. Sweet potato ADP-ribosylation factor (ARF) and ubiquitin extension protein (UBI) genes were used as endogenous controls (reference genes) (Park et al., 2012), in order to normalise the amount of sample RNA in the qPCR reactions. Applied Biosystems Custom TaqMan Gene Expression Assays (containing primer pairs and a TaqMan probe with an Applied Biosystems FAM dye) for the ARF and UBI genes were designed using the Custom TaqMan® Assay Design Tool (https://www.thermofisher.com/order/custom-genomic-products/tools/gene-expression/). Prior to performing the PCR reactions, total RNA from the 12 samples was reverse transcribed into cDNA using the Thermo Scientific™ First Strand cDNA Synthesis Kit, according to the manufacturer's instructions. Amplification of the ARF and UBI genes was performed using the Custom TaqMan® Gene Expression Assays and the TaqMan® Fast Advanced Master Mix (2X) following the manufacturer's instructions. The qPCR reactions were run in triplicates on the Roche LightCycler® 96 Instrument. The changes in miRNA expression levels were estimated using the $2^{-\Delta\Delta Ct}$ relative quantification method (Livak and Schmittgen, 2001), and the stability of each reference gene was checked using BestKeeper (Pfaffl et al., 2004).

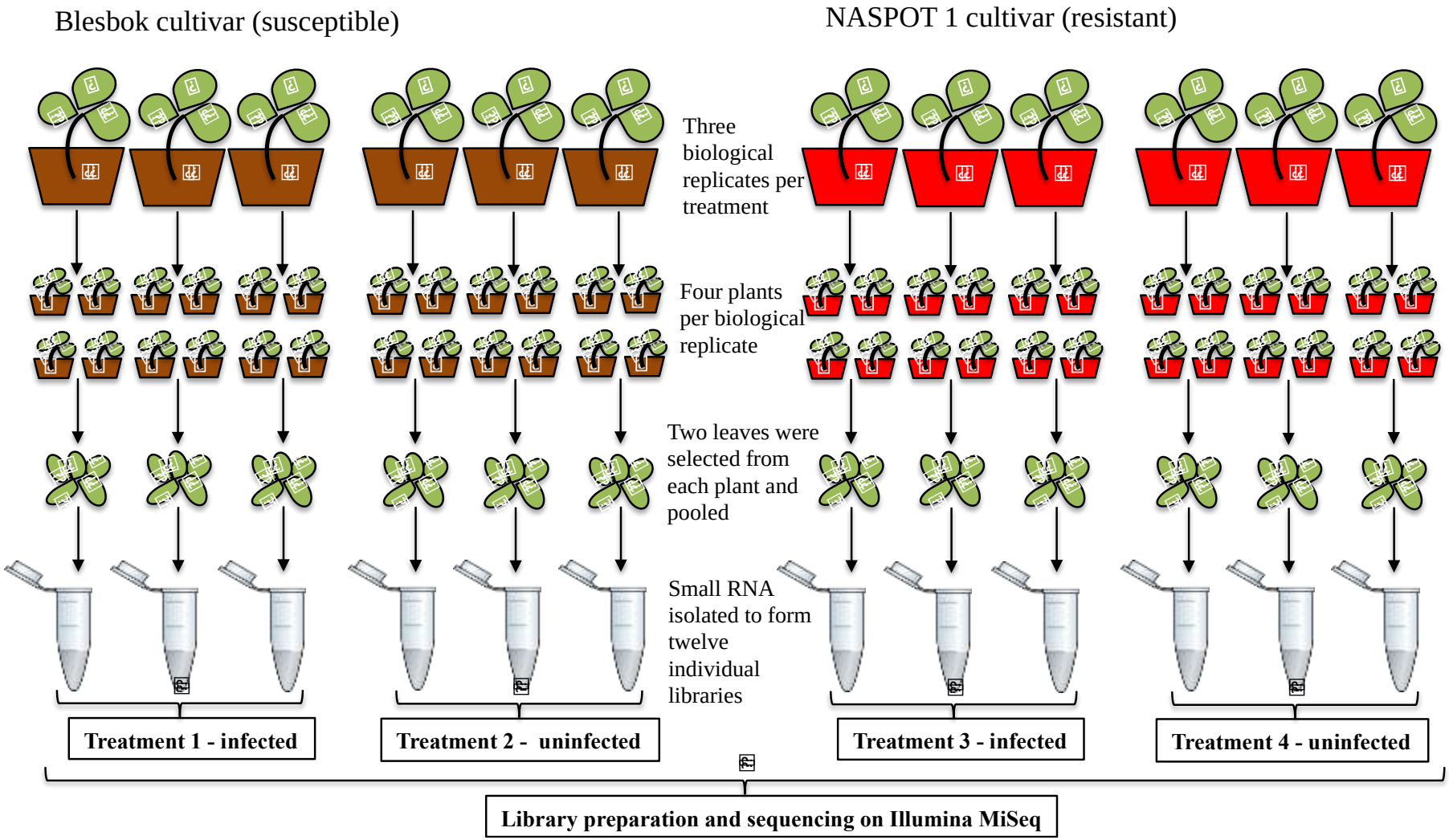

Figure 5.1: Experimental design for comparing miRNA expression between virus infected and non-infected plants of the Blesbok (susceptible) and NASPOT 1 (resistant) cultivars.

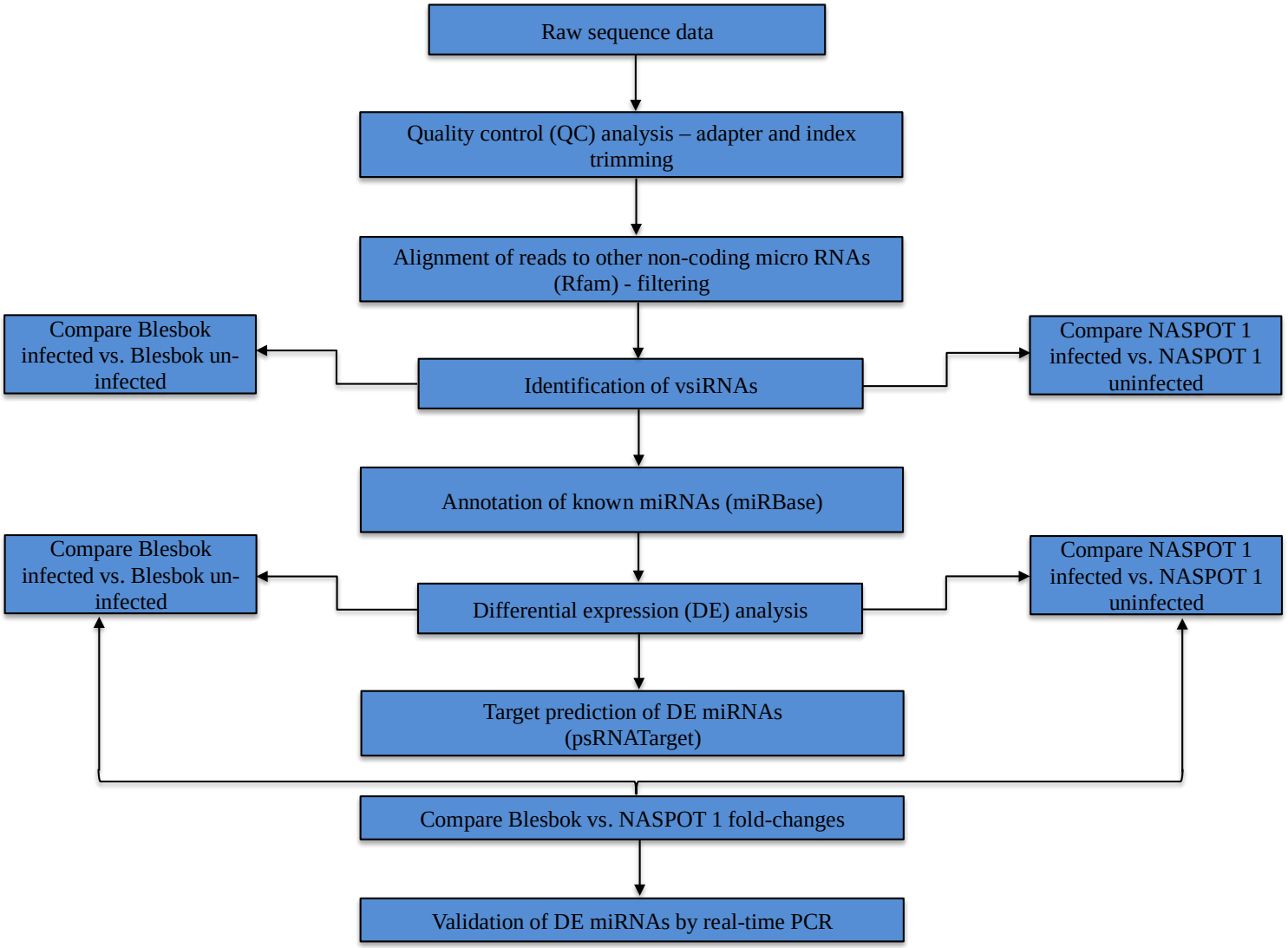

Figure 5.2: Bioinformatics workflow for the analysis of small RNA data.

5.4 Results and Discussion

5.4.1 Symptom development observed on the resistant and susceptible varieties post infection

At the end of the 60-day period symptom development was recorded. The NASPOT 1 uninfected plants displayed no viral symptoms while the infected NASPOT 1 (resistant) replicates displayed chlorotic spots. The NASPOT 1 infected replicates displayed milder symptom compared to Blesbok (susceptible) replicates (Figure 5.3). The infected Blesbok replicates displayed severe symptoms including upward leaf curling, chlorosis, vein clearing, leaf malformation and stunting (Figure 5.3). The chlorosis and leaf curl symptoms displayed on the Blesbok replicates are usually associated with potyvirus SPFMV, and begomovirus SPLCV, respectively. Single infection of sweet potato plants with SPFMV or begomovirus is often symptomless or mild, but co-infection of begomoviruses and potyviruses with SPCSV (crinivirus) causes severe symptoms often characterized by stunting (Carey et al., 1997, Cuellar et al., 2015, Gibson and Kreuze, 2014, Gibson et al., 1998). The synergistic interaction between SPFMV and SPCSV causes sweet potato virus disease (SPVD) (Kreuze, 2002) and the interaction between SPLCV and SPCSV causes leaf curl disease (Cuellar et al., 2015). It was interesting to note that all the NASPOT 1 replicates did not show typical leaf curl symptoms, even though they tested positive for both begomoviruses via PCR and vsiRNA deep sequencing (Table 5.3). Our findings suggest that NASPOT 1 is not only resistant to sweet potato virus disease caused by SPFMV and SPCSV but could also be resistant to sweet potato leaf curl disease. The result was expected, since the NASPOT 1 cultivar was reported to be highly resistant to major disease causing viruses, SPFMV and SPCSV, in field trials (Gibson et al., 2011, Mwanga et al., 2003). Studies have shown that disease symptoms in NASPOT 1 are usually mild, characterized by chlorotic spots and mild vein clearing (McGregor et al., 2009, Miano et al., 2008). Moreover virus resistant cultivars, such as the NASPOT and New Kawogo varieties, generally display mild disease symptoms compared to susceptible cultivars (Aritua et al., 1999). Therefore, it can be deduced from our results that the NASPOT 1 cultivar effectively counteracts infection by mixed sweet potato viruses, thus resulting in less severe symptoms.

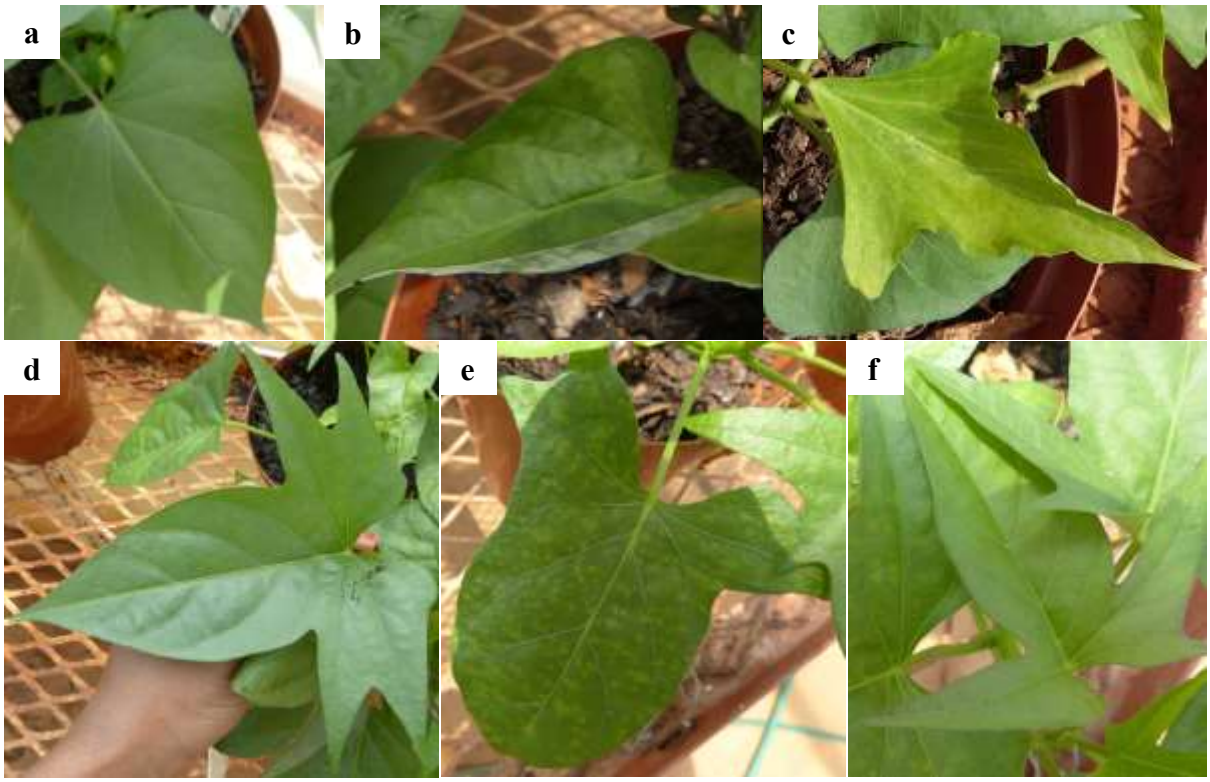

Figure 5.3: Viral symptoms observed on sweet potato varieties infected with mixed viruses. Asymptomatic leaves of the uninfected Blesbok cultivar (a). Severe symptoms on infected Blesbok leaves characterized by leaf curling (b), chlorosis and leaf malformation (c). NASPOT 1 uninfected leaves displaying no symptoms (d), chlorotic spots observed on infected NASPOT 1 leaves (e). Some NASPOT 1 infected leaves showed no symptoms at 60 dpi (f).

5.4.2 Confirmation of virus infection in the inoculated and non-inoculated plant samples

Prior to small RNA deep sequencing, the 12 plant samples were screened for the 8 viruses. The gel electrophoreses shows that the test plants that were inoculated with the KT10 scions tested positive for all 8 viruses, while the control plants were negative (Figure 5.4a & b). This statement was true only for six viruses (SPCSV, SPFMV, SPVC, SPVG, SPLCSPV, SPMaV). SPBVA was detected in the inoculated Blesbok plants (BT1, BT2 & BT3) and two control plants (BC1 & BC2) by PCR. SPBVB was detected in the Blesbok infected (BT1, BT2, BT3) and uninfected plants (BC1, BC2, BC3); in all the NASPOT 1 infected samples (NT1, NT2, NT3) and two NASPOT 1 uninfected plant samples (NC1 & NC3) (Figure 5.4b).

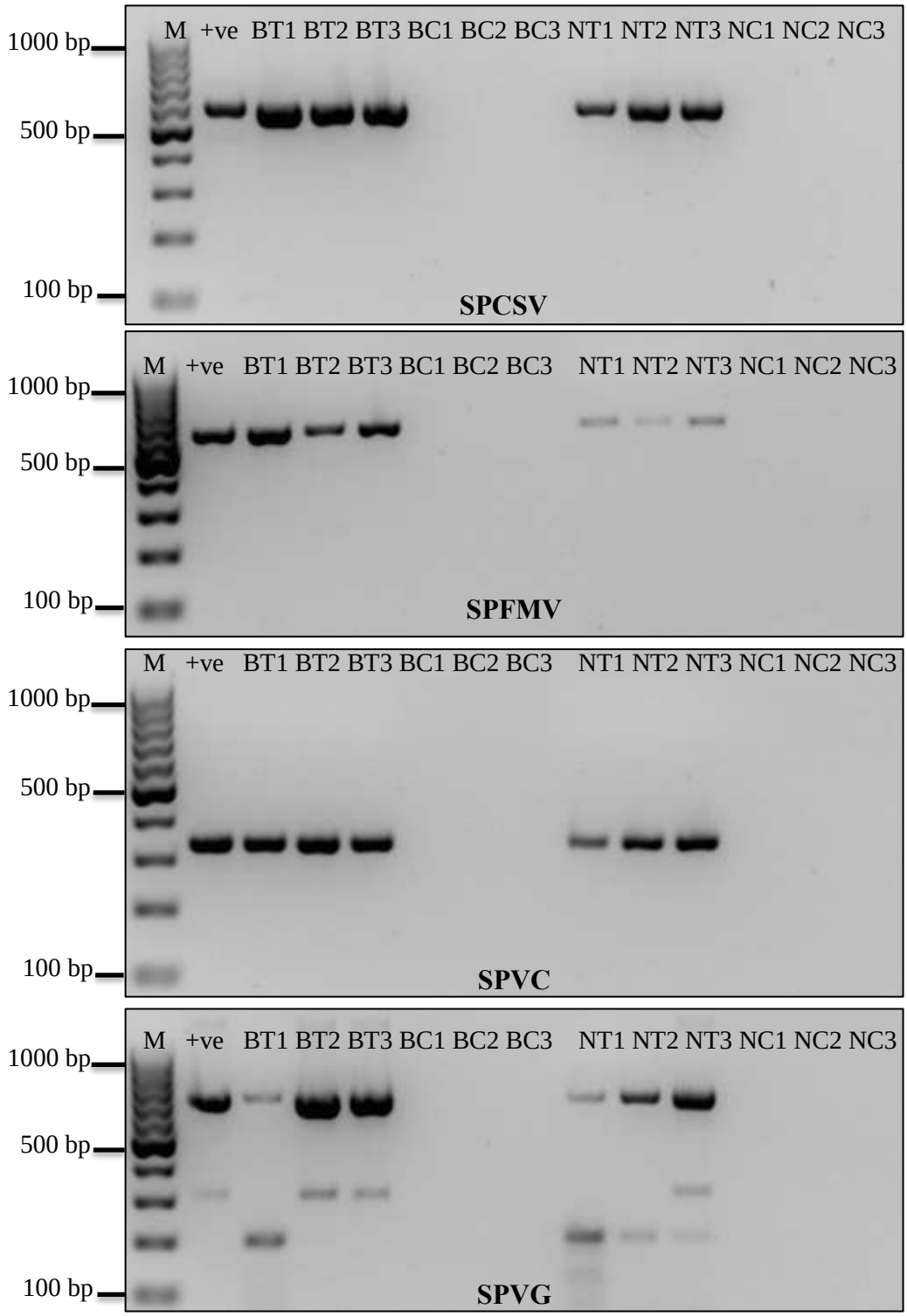

Figure 5.3a: RT-PCR of RNA viruses in susceptible (Blesbok) and resistant (NASPOT 1) cultivars. Plants were inoculated with a scion infected with multiple viruses (BT1, BT2, BT3, NT1, NT2, NT3) while 6 plant samples were used as non-inoculated controls (BC1, BC2, BC3, NC1, NC2, NC3).

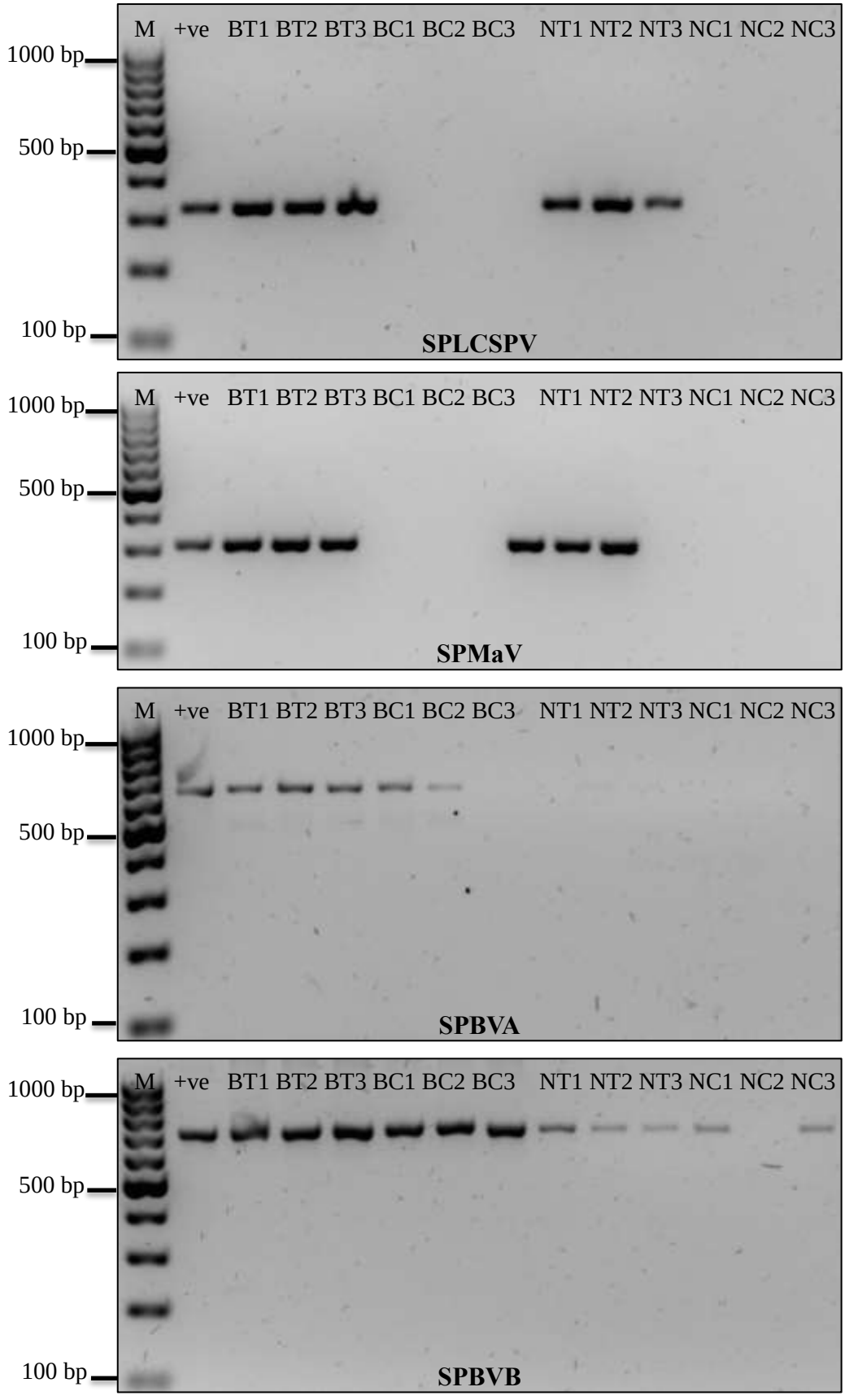

Figure 5.3b: PCR amplification of sweet potato DNA viruses.

5.4.2 Small RNA data analysis

Over 68 million raw reads were generated for the 12 small RNA sequenced libraries. After quality trimming over 7 million reads were retained (Table 5.1). Sequence reads ranging from 18 – 30 nt were retained for downstream analysis. Size class distribution indicated that the most abundant reads were in the 21, 24 and 30 nt size classes (Figure 5.4a & Figure 5.4b). Many studies show that plant regulatory microRNAs are within this size class distribution (Chen, 2005, Rogers and Chen, 2013, Sun, 2012). Sixty percent of the data sRNAs aligned to sequences in the Rfam database. This includes non-coding RNA genes (ncRNAs) such as transfer RNA (tRNAs), ribosomal RNA (rRNAs), small nuclear RNAs (snRNAs) and small nucleolar RNAs (snoRNAs) (Griffiths-Jones et al., 2003, Nawrocki et al., 2015). Only 1% of the data aligned to viruses, and these will be described in more detail below. Over 20% of the reads were annotated using the miRBase database. The plant species that were selected as references for miRNA annotation and identification included *Arabidopsis thaliana*, *Citrus sinensis*, *Glycine max*, *Solanum lycopersicum*, *Solanum tuberosum*, *Oryza sativa*, *Populus trichocarpa*, and *Medicago truncatula*, all which have been studied extensively (Bonnet et al., 2004, Fahlgren et al., 2007, Lu et al., 2005, Moxon et al., 2008, Rajagopalan et al., 2006, Song et al., 2009, Szittya et al., 2008, Xie et al., 2011, Zhang et al., 2008, Zhou et al., 2010). The remaining 18% of the data could not be annotated using the reference sequences; the reason for this could be that these sequences are unique to the *Ipomoea batatas* species. Literature shows that there is a considerable amount of non-conserved miRNAs that are family or species-specific (Amiteye et al., 2011, Campo et al., 2013, Cuperus et al., 2011, Islam et al., 2015, Lenz et al., 2011, Qin et al., 2014). This category of non-conserved miRNAs is usually identified as novel (Lindow and Krogh, 2005, Lu et al., 2005, Sunkar and Zhu, 2004). The most recent strategies used in the discovery or identification of novel miRNAs, are mainly deep sequencing and bioinformatic approaches (Taylor et al., 2014, Thiebaut et al., 2012).

Table 5.1: Data generated for the 12 small RNA (sRNA) libraries. The number of sequence reads before and after quality control (QC) analysis is presented. The raw read counts that aligned to vsiRNAs, miRNAs and other non-coding small RNAs are shown

Sample ID	Reads after QC	Virus derived siRNAs	Reads mapped to miRBase	Reads mapped to Rfam	Unmapped reads
NASPOT 1 infected replicate 1	520931	4858	128111	316419	4197371
NASPOT 1 infected replicate 2	616350	4328	129187	389277	4117400
NASPOT 1 infected replicate 3	592373	3718	117099	373599	5112549
NASPOT 1 uninfected replicate 1	464662	362	104766	278243	3569273
NASPOT 1 uninfected replicate 2	696838	258	137477	427211	5933532
NASPOT 1 uninfected replicate 3	712313	442	109807	493366	6136816
Blesbok infected replicate 1	626574	13513	170137	308809	4500612
Blesbok infected replicate 2	689569	13333	166108	353712	5457781
Blesbok infected replicate 2	816636	18778	191056	436536	6979824
Blesbok uninfected replicate 1	560794	3126	92610	371285	4297931
Blesbok uninfected replicate 2	796783	4210	129248	497695	5484886
Blesbok uninfected replicate 3	545726	3609	99135	341686	5199173

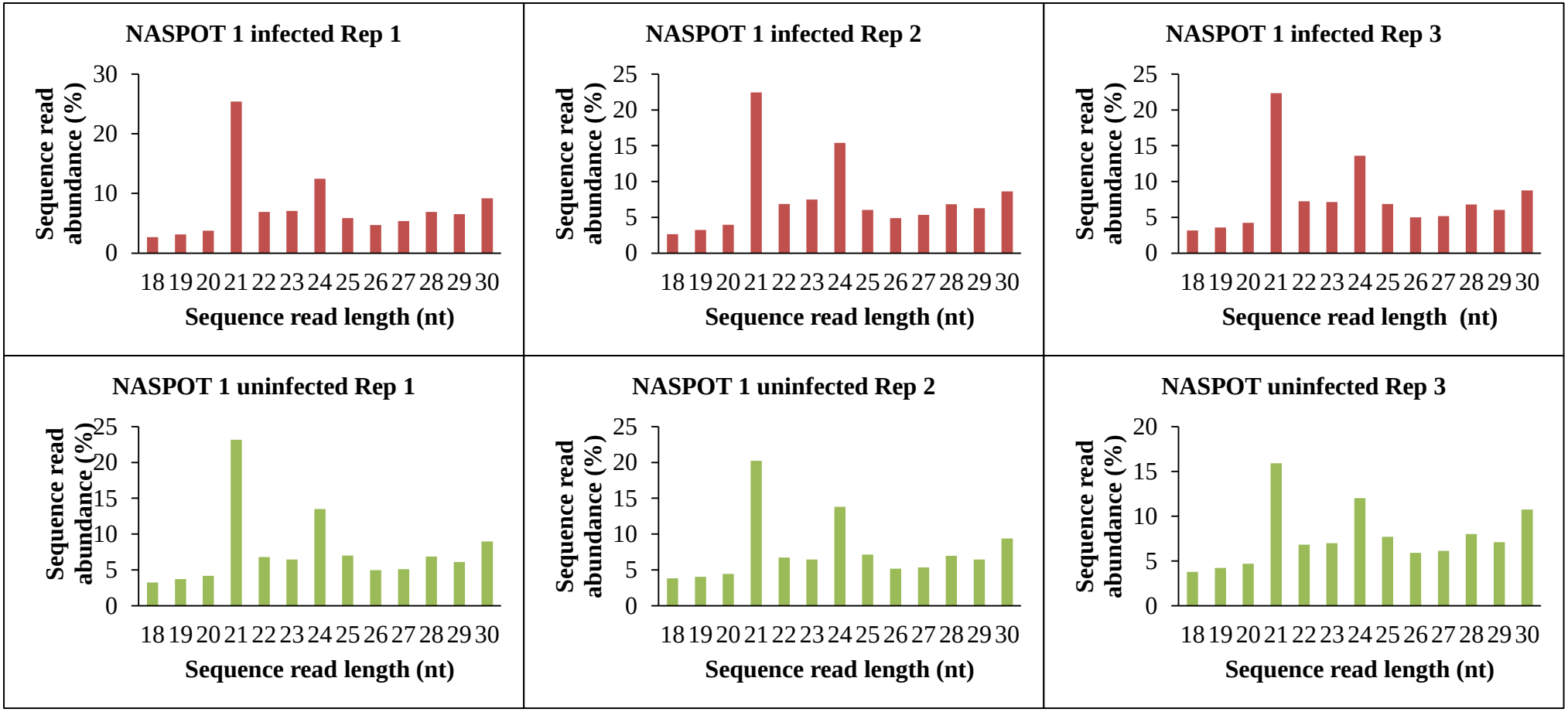

Figure 5.4a: Sequence read size class distribution of NASPOT 1 infected and uninfected libraries. The percentage read abundance was calculated after quality trimming.

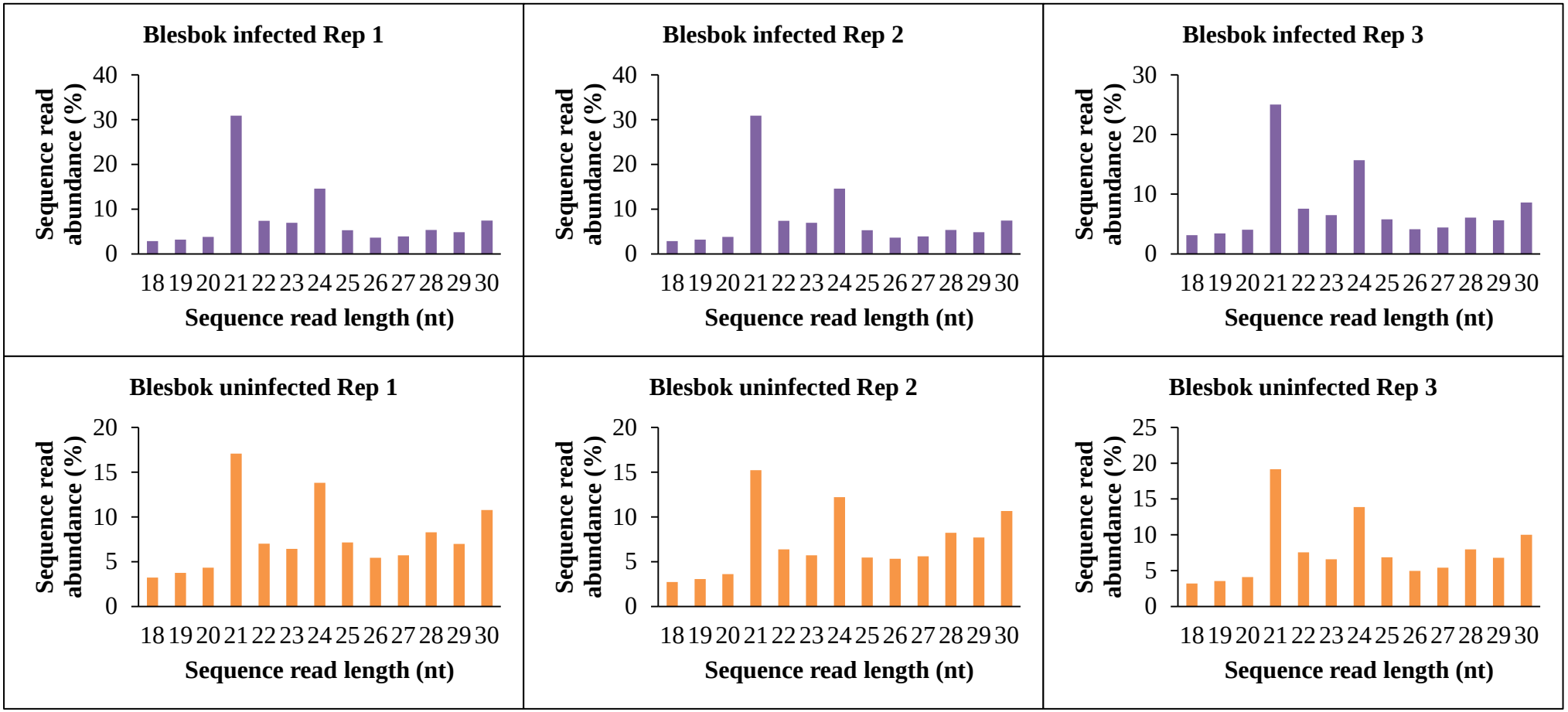

Figure 5.4b: Sequence read size class distribution for Blesbok infected and uninfected libraries. The percentage read abundance was calculated after quality analysis.

5.4.3 Comparison of virus-derived siRNAs (vsiRNAs) populations in resistant and susceptible cultivars

The raw vsiRNA counts for each replicate were normalised to reads per million (RPM), prior to the comparison between infected and uninfected replicates and cultivars (Table 5.2). The overall read length distribution of vsiRNAs for all the sequenced libraries was between 21 – 25 nt long (Figure A1). From our data set we found that the majority of the reads that aligned to viruses were in the 22 and 24 nt size classes (Figure 5.5). A similar result was reported for vsiRNAs associated with SPFMV and SPCSV (Kreuze et al., 2009). Sequence reads in this range are usually associated with virus-derived siRNAs or regulatory miRNAs targeting homologous host mRNA for silencing (Kamthan et al., 2015, Vance and Vaucheret, 2001).

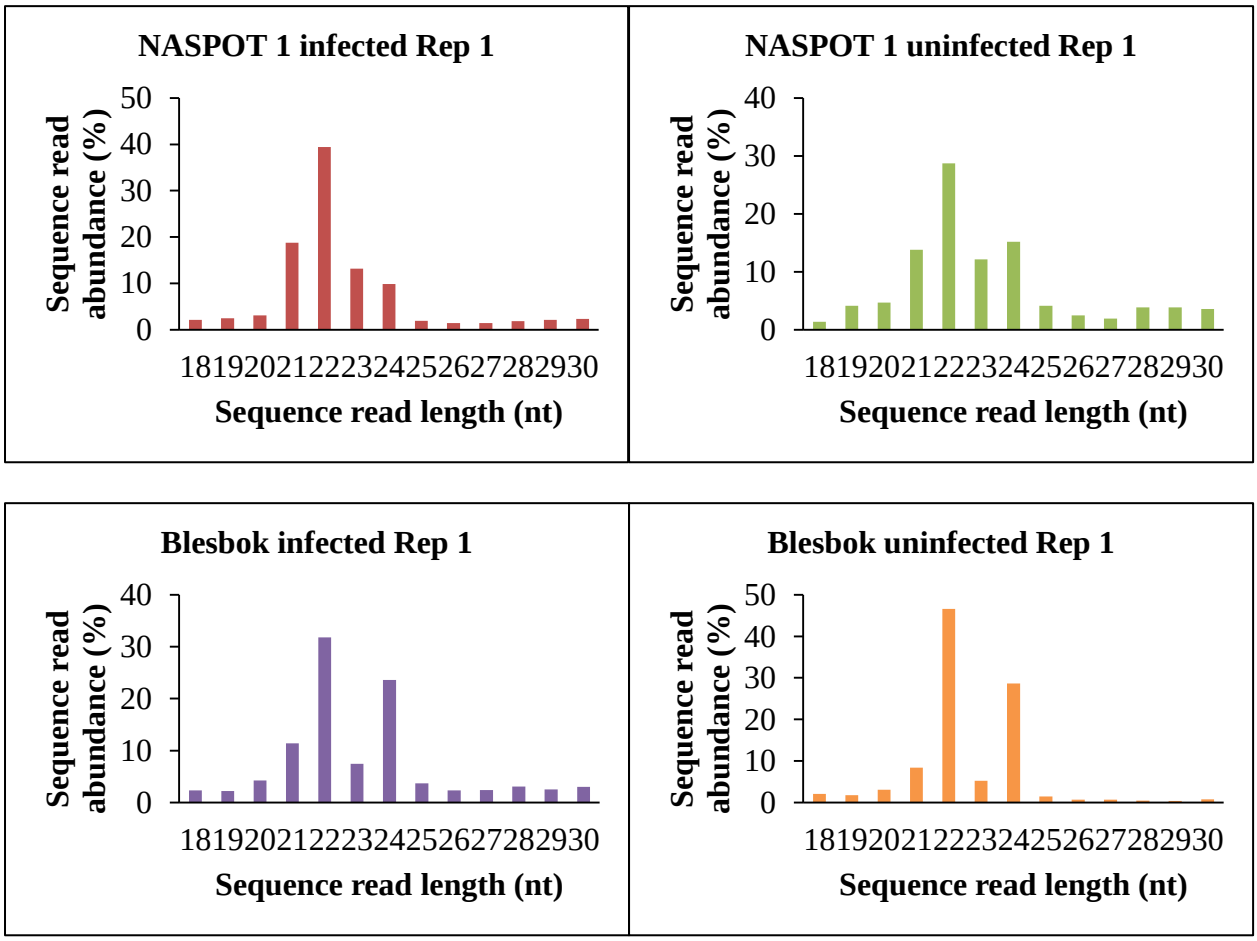

Figure 5.5: Read length distribution of vsiRNAs that mapped to sweet potato virus genomes.

Table 5.2: Comparison of vsiRNAs within and between virus-infected and uninfected cultivars of NASPOT 1 (resistant) and Blesbok (susceptible)

Sample ID	Normalised vsiRNA counts (reads per million - RPM)	P-value <0.05
NASPOT 1 infected replicate 1	9326	0,016
NASPOT 1 infected replicate 2	7022	
NASPOT 1 infected replicate 3	6276	
NASPOT 1 uninfected replicate 1	779	
NASPOT 1 uninfected replicate 2	370	
NASPOT 1 uninfected replicate 3	621	
Blesbok infected replicate 1	21566	0,002
Blesbok infected replicate 2	19335	
Blesbok infected replicate 2	22994	
Blesbok uninfected replicate 1	5574	
Blesbok uninfected replicate 2	5284	
Blesbok uninfected replicate 3	6613	
Between cultivar comparison		
NASPOT 1 infected vs. Blesbok infected		0,001
NASPOT 1 uninfected vs. Blesbok uninfected		0,003

It was found that the number of vsiRNAs was significantly different between the NASPOT 1 infected and NASPOT 1 uninfected libraries (Table 5.2). Similarly, the number of vsiRNAs in the infected Blesbok libraries was significantly higher than the Blesbok uninfected libraries. Our data shows that both cultivars respond to mixed viral infection by producing vsiRNAs, which potentially target the virus genomes for degradation. The normalised vsiRNAs reads from the NASPOT 1 virus infected libraries were significantly lower compared to the Blesbok infected libraries (Table 5.2). Unexpectedly, there were vsiRNAs detected in the NASPOT 1 uninfected replicates. The number of reads observed in the uninfected libraries could be background noise or contamination from the sequencing run, since all the twelve libraries were sequenced on one flow cell lane. The RT-PCR and PCR results determined that uninfected replicates of NASPOT 1 and Blesbok were all virus-free. With that said, it is not unknown that sweet potatoes can harbour undetectable virus sequences (Clark et al., 2012, Kreuze et al., 2009) Interestingly the uninfected Blesbok cultivar showed a significantly higher number of vsiRNAs originating from SPBVA and SPBVB. Previous studies suggest that in resistant cultivars post-transcriptional silencing of viral mRNA is usually very effective, therefore the production or accumulation of vsiRNAs in the resistant cultivar is significantly reduced (Ogwok et al., 2016). A similar finding in cassava infected with *South African cassava mosaic virus* was reported by (Rogans et al.,

2016) who suggest that in the virus-tolerant cassava TME3 landrace, low vsiRNA counts may represent efficient PTGS of viral mRNA, leading to a depletion/sequestration of vsiRNA populations, supporting a role for PTGS in tolerance/recovery in TME3. Further, because viral replication in resistant cultivars is usually localised to a specific part of the plant, disease symptoms are also localised or even absent (Kang et al., 2005). Inversely, the higher levels of vsiRNAs observed in the susceptible cultivar may represent accumulation of vsiRNAs that did not target viral mRNA for degradation, resulting in virus replication and symptom development (Peláez and Sanchez, 2013). Therefore in an attempt for the host to counter-attack viruses, the susceptible host continues to produce vsiRNAs which fail to target homologous virus sequences, possibly due to interference by virus suppressor proteins (VSRs) (Ogwok et al., 2016). It was found that higher vsiRNAs counts were associated with potyviruses SPFMV, SPVC and SPVG in the NASPOT 1 infected replicates (Table 5.3). While the majority of vsiRNAs in the Blesbok infected replicates were aligned to SPVC, SPFMV, badnaviruses (SPBVA & SPBVB), and begomoviruses (SPLCSPV & SPMaV). It was observed that the vsiRNA mapping patterns along each of the viral genomes differed (Figure 5.6). The mapping distribution of the vsiRNAs was not uniform as peaks were observed in certain regions (high numbers of vsiRNAs) and low amounts of vsiRNAs were observed in other regions of the genome (Figure 5.6). It is possible that low genome coverage was achieved by the sequencing approach used in this study or the peaks could represent specific vsiRNAs target regions. A region where higher vsiRNA abundance is observed is called a hotspot. Hotspots have been reported for plant viruses such as cucumber mosaic virus (CMV) (Wang et al., 2011), potato virus Y (PVY) (Kutnjak et al., 2015) and brassica yellow virus (BrYV) (Zhou et al., 2017). Previous studies suggest that high vsiRNAs production (hotspots) is observed in sub-genomic RNAs (sgRNAs) that are translated into proteins such as the coat protein (CP) region (Bronkhorst et al., 2013, Li et al., 2016). In this study we found that hotspots were observed in various ORFs including the minor coat protein (mCP) in SPCSV and V1, V2 and C2 in SPLCSPV (Table 5.4). The ORFs play an important role during the viral replication cycle. Identification of hotspot regions (high vsiRNA production) on virus genomes is important and can have future amplifications for targeting viral silencing suppressors (VSRs) of RNA and DNA viruses, which can result in the development of transgenic sweet potato cultivars with resistance against multiple viruses. Similar models have been used for other crops such as tomato (Sharma et al., 2015).

Table 5.3: The average number of vsiRNA reads aligned to sweet potato viruses

Treatment	Virus name	Raw read count	Normalised read count (RPM)
NASPOT 1 infected	SPFMV	2950	1706
	SPBVA	106	61
	SPBVB	69	40
	SPCSV RNA 2	1164	673
	SPVG	2087	1207
	SPVC	2910	1682
	SPCSV RNA 1	1734	1003
	SPMaV	1000	578
	SPLCSPV	884	511
NASPOT 1 uninfected	SPFMV	81	43
	SPBVA	71	38
	SPBVB	66	35
	SPCSV RNA 2	284	152
	SPVG	382	204
	SPVC	104	56
	SPCSV RNA 1	61	33
	SPMaV	9	5
	SPLCSPV	4	2
Blesbok infected	SPFMV	6567	3076
	SPBVA	3801	1781
	SPBVB	11468	5372
	SPCSV RNA 2	1164	545
	SPVG	117	55
	SPVC	12513	5861
	SPCSV RNA 1	1673	784
	SPMaV	3443	1613
	SPLCSPV	5058	2369
Blesbok uninfected	SPFMV	105	55
	SPBVA	2771	1456
	SPBVB	7522	3952
	SPCSV RNA 2	364	191
	SPVG	42	22
	SPVC	86	45
	SPCSV RNA 1	34	18
	SPMaV	12	6
	SPLCSPV	9	5

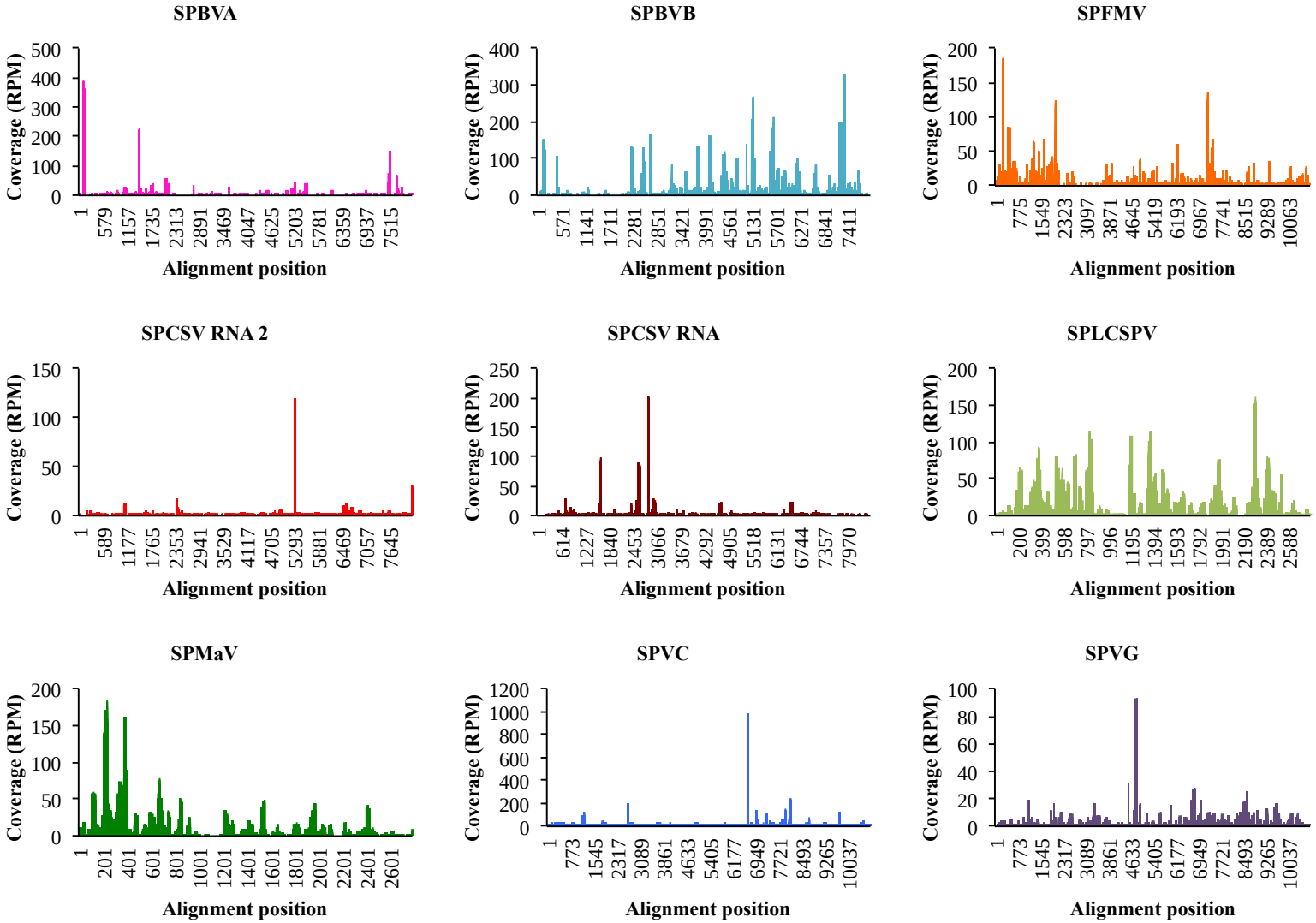

Figure 5.6: Graphical representation of hotspot regions in the genomes of SPBVA, SPBVB, SPFMV, SPCSV, SPLCSPV, SPMaV, and SPVC from the Blesbok infected treatment. The graph for SPVG was generated using reads from NASPOT 1 infected treatment. Coverage for each virus was calculated using the average read count (RPM) generated from the three replicates in each treatment.

Table 5.4: Open reading frames (ORFs) where hotspots were observed on virus genomes

Virus name	Hotspot position on genome (nt)	Name of ORF
SPCSV RNA 1	1621	p227
	2831 - 2845	p227
SPCSV RNA 2	5317	mCP
SPBVA	1450 - 1452	ORF 2
	7509	ORF 4
SPBVB	2252 - 7332	ORF 3a, ORF 3b, ORF 4
SPFMV	250 - 2066	P1
	7293	NIa-VPg
SPLCSPV	175 - 841	V2 & V1
	1184 - 1306	C3 & C2
SPMaV	209 - 381	V2
SPVC	6690	-
SPVG	4807	-

5.4.4 The detection of conserved and non-conserved microRNA families

Using the reference species for annotation (*Arabidopsis thaliana*, *Citrus sinensis*, *Glycine max*, *Medicago truncatula*, *Oryza sativa*, *Populus trichocarpa*, *Solanum lycopersicum*, and *Solanum tuberosum*), we were able to identify 60 miRNA families and 252 family members collectively from both Blesbok and NASPOT 1 cultivars (Figure 5.7 & Table A1.2). Some of the most well characterised miRNAs identified in both cultivars included miR166, miR398 miR399 and miR393 (Table A2). MiRNA166 has been extensively studied in model organism *Arabidopsis thaliana* and together with miR165, targets *HOMEODIMAIN-LEUCINE ZIPPER* (*HD-ZIP*) transcription factor genes, which function to regulate plant development, specifically floral and shoot apical stem (SAM) development (Jung and Park, 2007, Zhu et al., 2011b). MiR166 has also been identified in over 40 plant species including soybean (*Glycine* max) (Li et al., 2017, Song et al., 2011), maize (*Zea Mays*) (Ding et al., 2009) and tomato (*Solanum lycopersicum*) (Valiollahi et al., 2014). Reports show that miR166 has been associated with abiotic stress response such as drought (Ferdous et al., 2015) and biotic stress responses have been reported in virus-infected grapevines (Pantaleo et al., 2016). Other conserved miRNAs for which a high number of reads were observed include miR398, miR482, mIR408 and miR168 (Table A2). These miRNAs have been studied at length and reported in cocoa *(Theobroma cacao*) cucumber (*Cucumis sativus* L.) (Liu et al., 2015a), *Arabidopsis thaliana* (Stief et al., 2014) and tomato (*Solanum lycopersicum*) (Huang et al., 2017). Most of the miRNAs identified in this study regulate genes that play important roles in plant development (Liu et al., 2015b, Manavella et al., 2013, Morea et al., 2016),

abiotic stress responses (Bilichak et al., 2015, Ma et al., 2015, Shriram et al., 2016, Trindade et al., 2010, Zhang et al., 2014, Zhu et al., 2011a) and pathogen defense (Abreu et al., 2014, Baldrich and San Segundo, 2016, Wu et al., 2015, Xin et al., 2010). Significant read counts where observed for miR6300 in all four treatments (NASPOT infected, NASPOT 1 uninfected, Blesbok infected, Blesbok uninfected). MiR6300 is a non-conserved miRNA and has been detected in a few species including soybean (*Glycine max*) (Turner et al., 2012), tomato (Jin and Wu, 2015), peanut (Gao et al., 2017) and wheat (Liu et al., 2015b). The specific role of miR6300 is not clearly elucidated, but there's evidence to show that miR6300 sequences are found among legumes including chickpea (Srivastava et al., 2015). Other non-conserved microRNAs identified from our data were miR827, miR2111, miR1511 and miR6024. Non-conserved miRNAs are species-specific or found only in certain plant families (Jagadeeswaran et al., 2009, Zhang et al., 2006). It is assumed that species-specific miRNAs also play roles in stress responses or some biological processes (Campo et al., 2013, Lenz et al., 2011, Xin et al., 2010).

5.4.5 Identification of pathogen defense related microRNAs

Sequence data revealed the presence of miR482, a miRNA often associated with pathogen-defense, in both resistant and susceptible cultivars. The miR482 family members detected in this study include miR482a, miR482b, miR482c, miR482d, and miR482f. The detected miR482 family members were compared to miR482 sequences from a range of plant species in order to show sequence diversity (Figure 5.8). The miR482 sequences detected in our study were most similar to miR482 found in *Prunus persica* (peach) (ppe-miR482), *Pinus densata* (pine) (pde-miR482) and *Solanum lycopersicum* (tomato) (sly-miR482) (Figure 5.8).

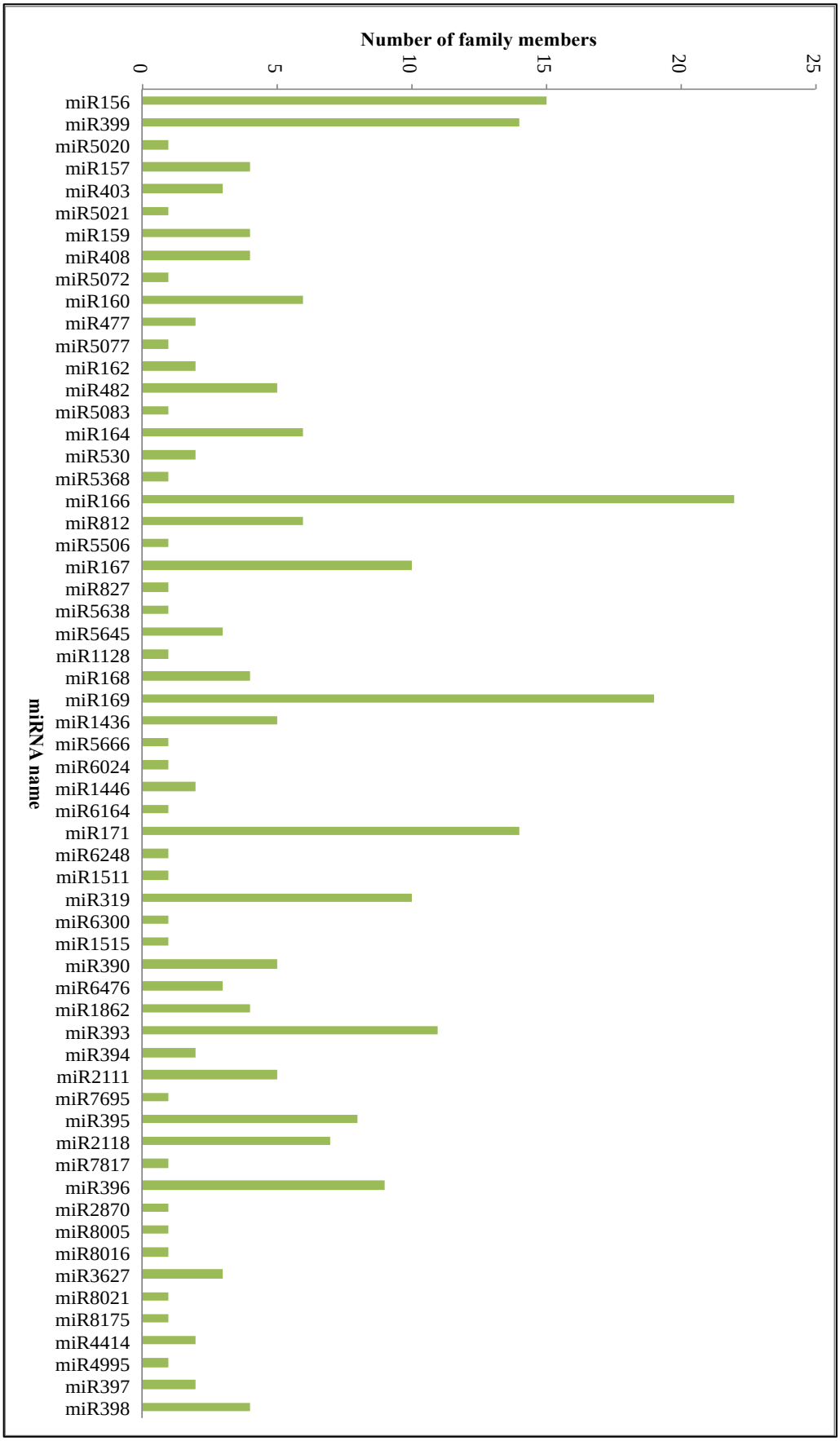

Figure 5.7: MicroRNA (miRNA) families identified using known sequences in miRBase as references.

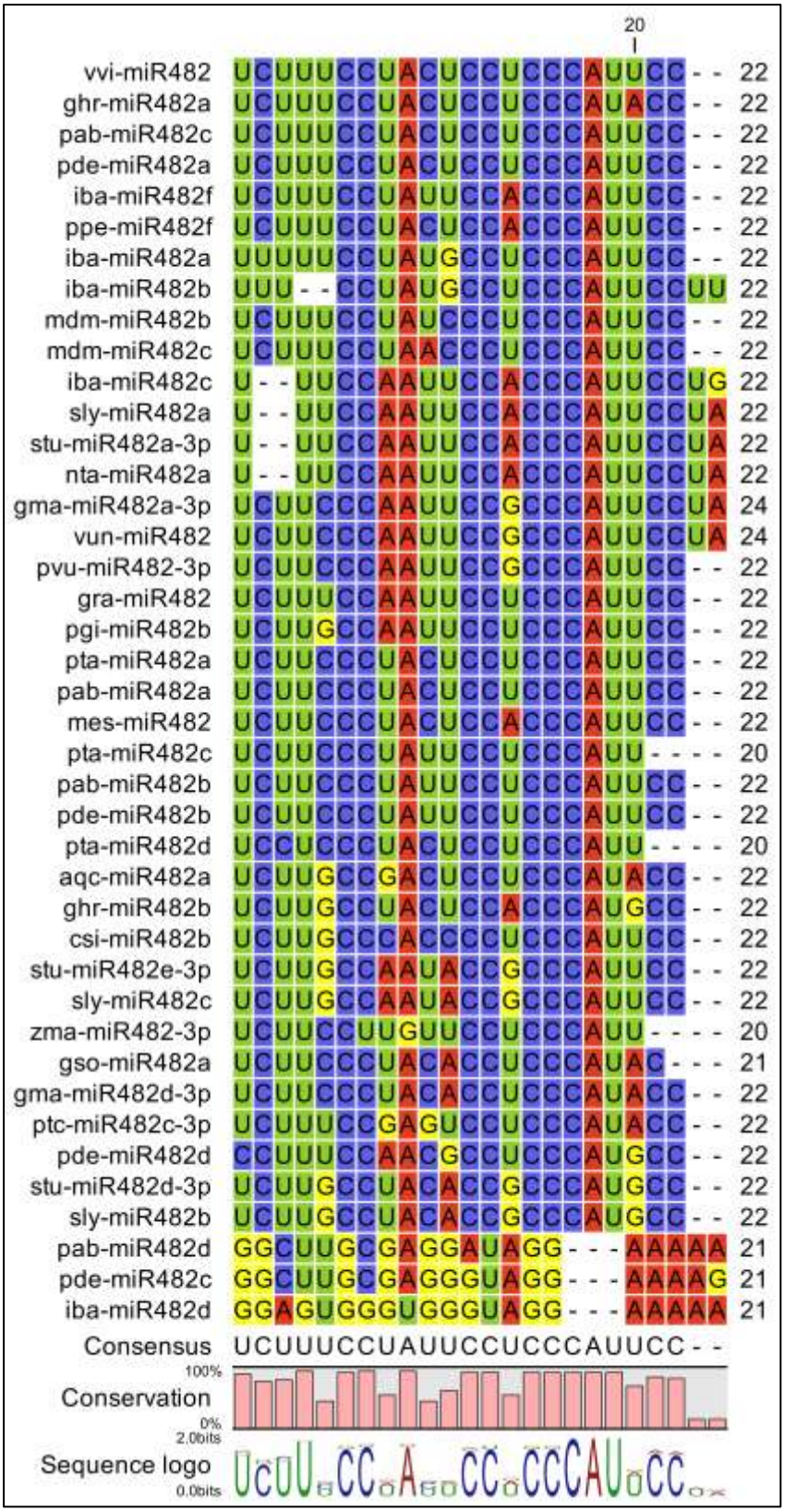

Figure 5.8: Multiple sequence alignment of miR482 family members from various plant species including *Citrus sinesis*, *Solanum lycopersicum*, *Solanum tuberosum*, *Vitis vinifera* and the newly sequenced miR482 sequences from *Ipomoea batatas* (proposed names, iba-miR482a, iba-miR482b, iba-miR482c, iba-miR482d and iba-miR482f).

5.4.6 Differential expression analysis of miRNAs in response to mixed viral infections

Differential expression (DE) analysis was conducted to firstly compare expression levels of miRNAs within each cultivar (infected vs. uninfected plants), and then to compare expression levels between the two cultivars (resistant vs. susceptible) in response to mixed infections. The analysis was performed utilising the small RNA analysis tool in the CLC Bio Workbench v9. The expression data (raw reads) was normalised to reads per million (RPM) before the differential expression analysis was conducted. The mean fold-change and the log2 fold-change were calculated on the normalised reads. A total of 17 miRNAs were differentially expressed in the NASPOT 1 and Blesbok cultivars after infection with the 8 viruses (Table 5.5). The analysis shows that miR393, miR398, miR168, miR162, miR167 and miR397 were upregulated in both cultivars after virus infection (Table 5.5). MiRNAs of particular interest are miR168 and miR162, which regulate RNA silencing pathway by guiding the cleavage of Argonaute 1 (AGO1) and Dicer-like 1 (DCL1) mRNAs, respectively (Vaucheret et al., 2006). Plants induce AGO1 mRNA expression during viral infection, as a defense mechanism against viruses (Várallyay and Havelda, 2013, Várallyay et al., 2010). Viruses have also developed strategies to counteract the host defense mechanism by directly or indirectly suppressing RNA silencing (Pumplin and Voinnet, 2013). In this study we observed the upregulation of miR162 and miR168, which could be an indication that the VSRs from multiple viruses (e.g. H-Pro from SPFMV and RNAse3 from SPCSV) are interfering with the RNA silencing pathway by causing the suppression of AGO proteins (Várallyay and Havelda, 2013, Várallyay et al., 2010). There is evidence suggesting that the high expression of miR168 induced by viruses could play a role in symptom development (Várallyay and Havelda, 2013, Várallyay et al., 2010).

In this study we observed the upregulation of miR482 post viral infection in both NASPOT 1 and Blesbok cultivars (Table 5.5). Studies show that upon viral infection, expression of miR482 is downregulated so that the levels of NB-LRR proteins in the host plant are increased (Balmer and Mauch-Mani, 2013, Li et al., 2012, Shivaprasad et al., 2012). Our results are not in agreement with the findings from previous studies. It is possible that miR482 was downregulated during early infection stages and because our data was collected from plants during late viral infection, the downregulation of miR482 could not be detected. It is also probable that viruses encoding VSR have already suppressed or counteracted the host pathogen defense mechanism during late infection. Further analysis of miR482 expression levels during early infection should therefore be conducted in future studies. We

found that miR482 targets disease resistance (R) genes that encode nucleotide-binding site leucine-rich (NBS-LRR) proteins found in sweet potato (Table 5.6) (Chen et al., 2007, Wang et al., 2010). Studies show that the miR482 super family target and regulate disease resistance (R) genes encoding proteins belonging to the NBS-LRR protein super family (Li et al., 2012, Shivaprasad et al., 2012). The regulation of NBS-LRR genes by mir482 has been reported for many plant species including tobacco, tomato, potato (Li et al., 2012, Shivaprasad et al., 2012, Yang et al., 2015), soybean (Yin et al., 2013), cotton (Zhu et al., 2013) and peach (González et al., 2015). Over time plants have evolved multi-layered defense mechanisms in order to counteract pathogens. Resistance (R) genes induce the effector-triggered immunity (ETI), which results in hypersensitive response and programmed cell dell (PCD), thus conferring resistance to pathogens including insects, bacteria, fungi and viruses (Jones and Dangl, 2006, Zvereva and Pooggin, 2012). In this study we illustrate the possible involvement of small RNAs (miRNAs and vsiRNAs) in defense response within sweet potato. Upon viral infection, the first innate immune response identified as RNA silencing was induced and resulted in the production of virus-derived siRNAs (vsiRNAs). This was followed by expression of various endogenous miRNAs (Balmer and Mauch-Mani, 2013), which regulate R genes that confer resistance to viruses. Studies show that cleavage of resistance gene mRNAs can trigger the production of secondary siRNAs that regulate NB-LRR genes (Li et al., 2012, Shivaprasad et al., 2012), further contributing to the innate immune response. Therefore the second layer of defense (R gene response) is more effective than RNA silencing because viruses are localised or contained in less than 5 days post infection (early infection) and additional siRNA generation makes this mechanism more robust (de Ronde et al., 2014).

Two miRNAs (miR160 and miR6300) were downregulated in both cultivars after infection. A target gene (auxin response factor 18-like - ARF18) was predicted for miR160 (Table 5.7) and none for miR6300. MiR160 is associated with root and flower development, abiotic stress and biotic responses (Feng et al., 2013). Downregulation of AFRs (upregulation of miR160) often results in developmental defects (Liu et al., 2014) and positive regulation (downregulation of miR160) results in the development of roots. Our data demonstrates the involved of these miRNAs in viral infection (Feng et al., 2013). We predicted laccase-12-like as the target for miR397 (Table 5.7). Laccases are involved in stress responses, lignin synthesis and cell wall structure among many other functions (Abdel-Ghany and Pilon, 2008, Wang et al., 2010). MiR397 was upregulated in both the resistant and susceptible cultivars

(Table 5.5). Studies have shown the downregulation of miR397 in cotton infected with fungi, this response then triggers laccase accumulation (Thiebaut et al., 2015). An increase in laccase levels ensures cell wall lignification when pathogen infection occurs (Bellincampi et al., 2014, Miedes et al., 2014). Our study demonstrates the potential role of miR397 in viral infection in sweet potato cultivars. Significant upregulation of miR397 may result in a decrease in laccase accumulation which could be attributed to the degradation of the cell wall upon viral infection (Bellincampi et al., 2014, Miedes et al., 2014). Other miRNAs that are significantly regulated between the two cultivars (miR399, miR319, miR530 and miR5077) are possibly involved in the regulation of abiotic and biotic stress responses (Hicks and Liu, 2013).

Table 5.5 Differentially expressed miRNAs between the infected and uninfected NASPOT 1 and Blesbok cultivars. Mean fold changes and log2 fold changes were calculated on normalised read counts. Reads were normalised and expressed as reads per million (RPM). The significant changes in miRNA expression between the two cultivars are highlighted in bold (p value<0.05)

	NASPOT 1		**Blesbok**		
miRNA name	**Fold change**	**Log2 fold change**	**Fold change**	**Log2 fold change**	**p-value < 0,05**
miR166	1,150	0,201	1,497	0,582	0,105
miR399	1,084	0,117	0,768	-0,380	0,394
miR393	**15,443**	**3,949**	**70,996**	**6,150**	**0,010**
miR398	3,195	1,676	4,829	2,272	0,329
miR482	1,286	0,363	1,410	0,496	0,988
miR408	1,086	0,119	1,347	0,430	0,692
miR168	14,513	3,859	18,181	4,184	0,912
miR6300	0,813	-0,299	0,790	-0,339	0,967
miR160	0,697	-0,520	0,596	-0,746	0,354
miR162	2,688	1,427	4,677	2,226	0,147
miR164	1,328	0,409	1,741	0,800	0,426
miR167	1,575	0,655	3,412	1,770	0,129
miR319	**1,301**	**0,380**	**0,529**	**-0,919**	**0,024**
miR397	**1,139**	**0,188**	**4,921**	**2,299**	**0,042**
miR403	0,470	-1,090	1,031	0,044	0,265
miR530	**1,033**	**0,047**	**0,182**	**-2,460**	**0,035**
miR5077	**1,601**	**0,679**	**0,390**	**-1,359**	**0,016**

Table 5.6: Targets genes of miR482. The resistance genes and their different homologues are targeted by miR482. A perfect miRNA-mRNA match causes cleavage and an imperfect match causes translation repression

miRNA	**Description of target sequence**	**Inhibition**
miR482a	IBRGA-13 NBS-LRR protein gene	Cleavage
	IBRGA-10 NBS-LRR protein gene	Cleavage
miR482d	IBRGA-5 NBS-LRR protein gene	Cleavage
	SPRGA-2 NBS-LRR protein rsp-2 gene	Cleavage
	IBRGA-3 NBS-LRR protein gene	Translation
	IBRGA-6 NBS-LRR protein gene	Translation
	SP2_E05 NBS-LRR type disease resistance protein gene	Cleavage
	IBRGA-1 NBS-LRR protein gene	Translation
	IBRGA-10 NBS-LRR protein gene	Translation

Table 5.7: Putative target genes of significantly expressed miRNAs in sweet potato

miRNA name	**Target gene in sweet potato**	**Target gene in other species**
iba-miR164	NAC domain-containing 92-like	NAC-domain
iba-miR168	N/A	Argonaute 1 (AGO1)
iba-miR166	N/A	*HD-ZIP* transcription factors & disease resistance protein RPM1
iba-miR397	Laccase-12-like	Laccase and beta-6 tubulin
iba-miR167	N/A	Auxin response factors (ARF transcription factors)
iba-miR160	Auxin response factor 18-like	Auxin response factor proteins
iba-miR398	N/A	Copper superoxide dismutases (CSD1& CSD2) and cytochrome C oxidase subunit V
iba-miR403	N/A	Argonaute 2 & 3 (AGO2 & AGO3)
iba-miR162	N/A	Dicer-like 1 (DCL1)
iba-miR399	Inorganic phosphate transporter 1-4-like	Phosphate 2 (PHO2)
iba-miR319	N/A	Transcription factors of TCP family
iba-miR530	N/A	Argonaute 1 (AGO1)
iba-miR5077	N/A	Amine oxidase and ATP-dependent RNA helicase
iba-miR482	NBS-LRR rsp- partial	NB-LRR resistance genes
iba-miR6300	N/A	N/A

5.4.6 RT-qPCR of miR482, miR168, miR393 and miR397

Four differentially expressed miRNAs were selected and successfully validated by RT-qPCR (Figure 5.9). Relative expression analysis was conducted (Table A4 – A7), and the results confirmed the expression patterns shown using miRNA-Seq (Figure 5.10). RT-qPCR and miRNA-Seq show upregulation of all four miRNAs after infection with multiple viruses. Based on our analysis using BestKeeper, the UBI gene was selected as the endogenous control (reference gene/ housekeeping gene) since it was more stable than the ARF gene. According to BestKeeper if a gene has a standard deviation (SD) higher than 1, it can be considered inconsistent or unstable. Our analysis shows that UBI is more stable as it has a SD value of 0.67, while the ARF gene has a SD value of 1.30 (Table 5.8).

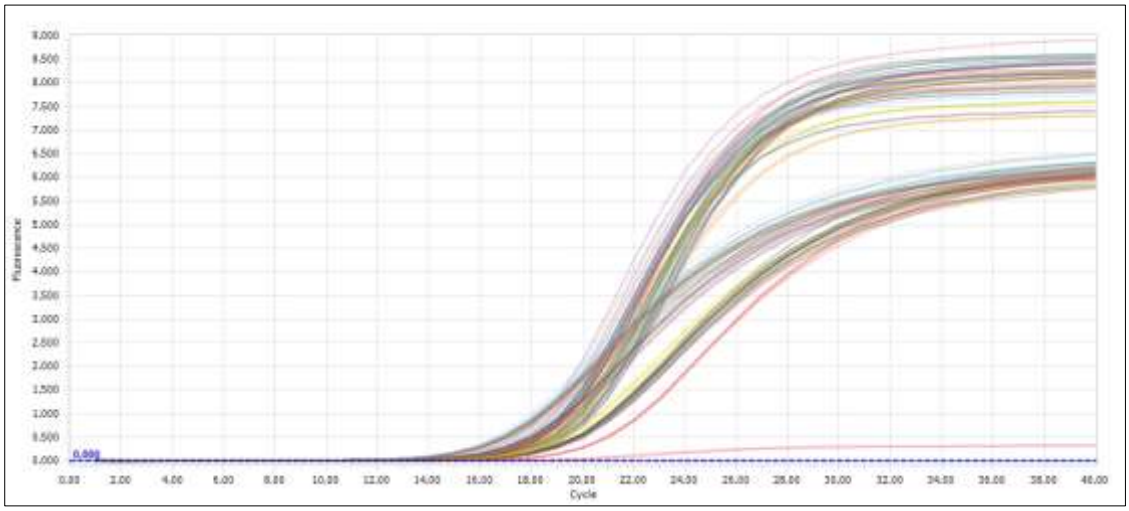

Figure 5.9a: RT-qPCR amplification curves of sweet potato defense related miRNAs (miR482 and miR168).

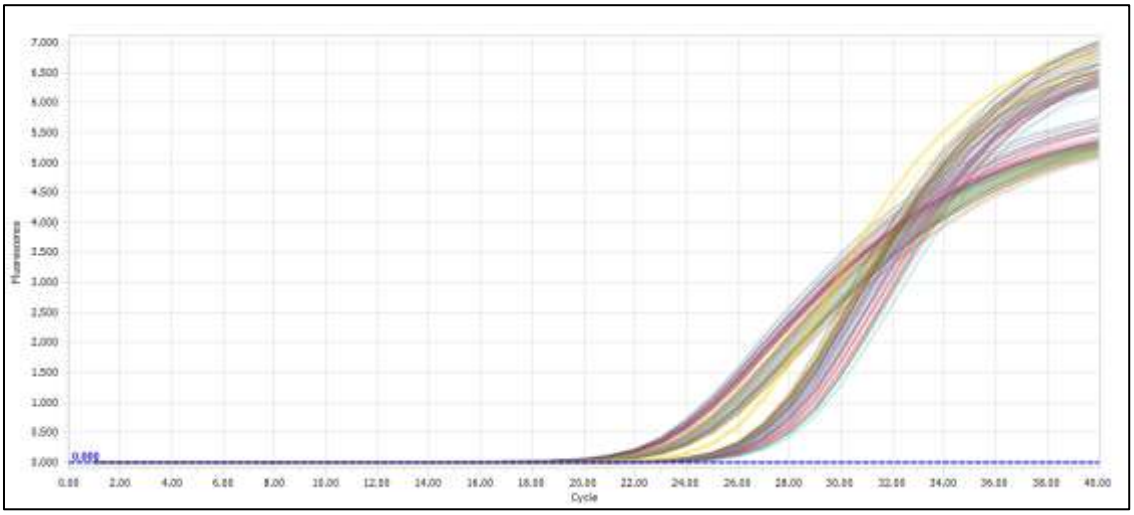

Figure 5.9b: RT-qPCR amplification curves of sweet potato miR393 and miR397.

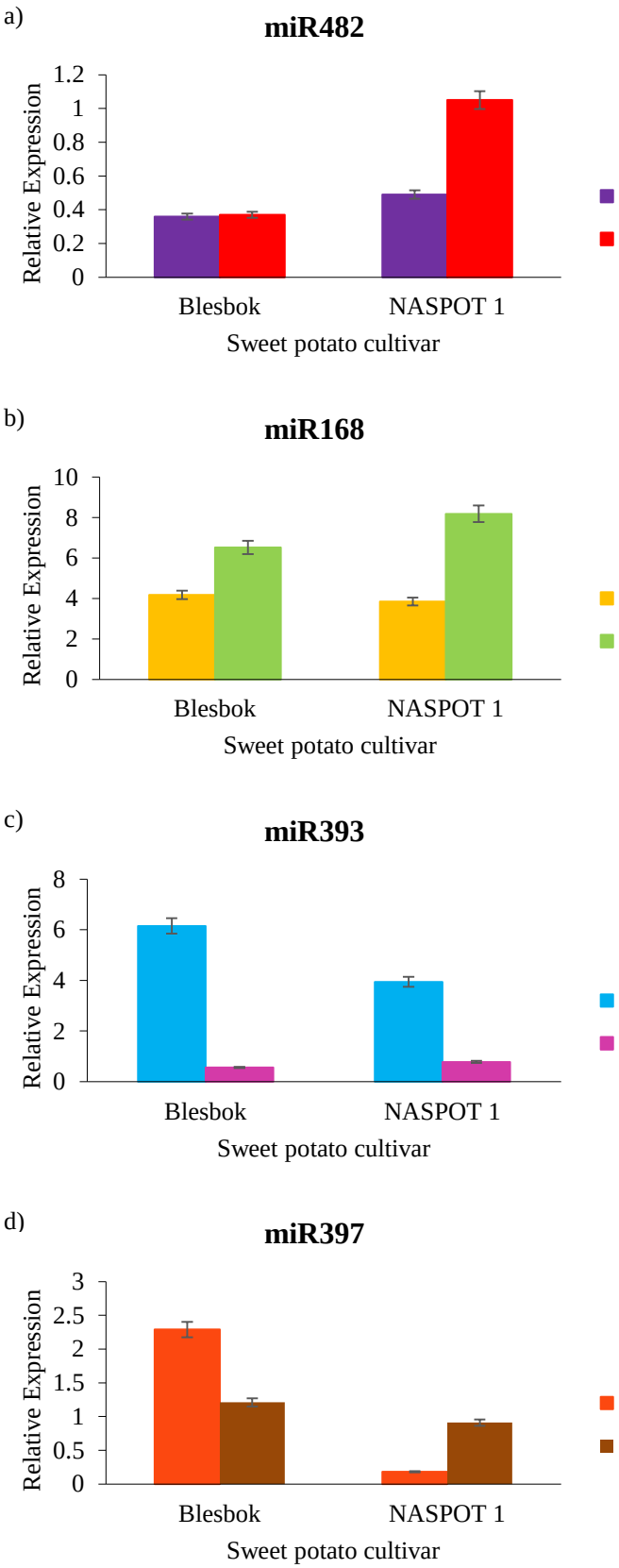

Figure 5.10. The relative expression levels of four miRNAs (miR482, mR168, miR393, miR397) validated by RT-qPCR.

Table 5.8: The statistics of two sweet potato housekeeping genes (HKG) or endogenous controls based on the crossing point (CP) values calculated using BestKeeper

	HKG 1 (UBI)	HKG 2 (ARF)
n	36	36
geo Mean [CP]	**24.95**	**29.55**
ar Mean [CP]	**24.96**	**29.60**
min [CP]	23.79	26.65
max [CP]	26.78	33.81
std dev [± CP]	**0.67**	**1.30**
CV [% CP]	**2.68**	**4.39**

5.5 Conclusion

This is the first study to report miRNA sequences involved in biotic stress responses in sweet potato *(Ipomoea batatas*). We show that the pathogen related miRNAs found in this study target previously detected sweet potato R-genes. We also illustrate the intricate multi-layered defense mechanism (RNA silencing & effector triggered immunity) in sweet potato cultivars, facilitated by siRNAs and miRNAs, which target viral sequences and regulate host genes, respectively. Future studies should focus on conducting RNA-Seq analysis to determine the expression levels of target genes post-viral infection. Moreover, quantification of miRNA levels at different time points (early and late infection) will further elucidate the host-pathogen interactions in susceptible and resistant sweet potato cultivars. Understanding host-pathogen interactions will assist researchers (breeders) to devise effective strategies that will significantly contribute to crop improvement in sweet potato.

5.6 References

ABDEL-GHANY, S. E. & PILON, M. 2008. MicroRNA-mediated systemic down-regulation of copper protein expression in response to low copper availability in Arabidopsis. *Journal of Biological Chemistry,* 283**,** 15932-15945.

ABREU, P. M., GASPAR, C. G., BUSS, D. S., VENTURA, J. A., FERREIRA, P. C. & FERNANDES, P. M. 2014. Carica papaya microRNAs are responsive to Papaya meleira virus infection. *PloS One,* 9**,** e103401.

AGUDELO-ROMERO, P., CARBONELL, P., DE LA IGLESIA, F., CARRERA, J., RODRIGO, G., JARAMILLO, A., PÉREZ-AMADOR, M. A. & ELENA, S. F. 2008. Changes in the gene expression profile of Arabidopsis thaliana after infection with Tobacco etch virus. *Virology Journal,* 5**,** 92.

AMITEYE, S., CORRAL, J. M., VOGEL, H. & SHARBEL, T. F. 2011. Analysis of conserved microRNAs in floral tissues of sexual and apomictic Boechera species. *BMC Genomics,* 12**,** 500.

ARITUA, V., LEGG, J., SMIT, N. & GIBSON, R. 1999. Effect of local inoculum on the spread of sweet potato virus disease: limited infection of susceptible cultivars following widespread cultivation of a resistant sweet potato cultivar. *Plant Pathology,* 48**,** 655-661.

AXTELL, M. J. 2013. Classification and comparison of small RNAs from plants. *Annual Review of Plant Biology,* 64**,** 137-159.

BALDRICH, P. & SAN SEGUNDO, B. 2016. MicroRNAs in rice innate immunity. *Rice,* 9**,** 6.

BALMER, D. & MAUCH-MANI, B. 2013. Small yet mighty–microRNAs in plant-microbe interactions. *MicroRNA,* 2**,** 73-80.

BELLINCAMPI, D., CERVONE, F. & LIONETTI, V. 2014. Plant cell wall dynamics and wall-related susceptibility in plant–pathogen interactions. *Frontiers in Plant Science,* 5.

BILICHAK, A., ILNYTSKYY, Y., WÓYCICKI, R., KEPESHCHUK, N., FOGEN, D. & KOVALCHUK, I. 2015. The elucidation of stress memory inheritance in Brassica rapa plants. *Frontiers in Plant Science,* 6**,** 5.

BONNET, E., WUYTS, J., ROUZÉ, P. & VAN DE PEER, Y. 2004. Detection of 91 potential conserved plant microRNAs in Arabidopsis thaliana and Oryza sativa

identifies important target genes. *Proceedings of the National Academy of Sciences of the United States of America*, 101**,** 11511-11516.

BORGES, F. & MARTIENSSEN, R. A. 2015. The expanding world of small RNAs in plants. *Nature Reviews Molecular Cell Biology*.

BRONKHORST, A. W., MIESEN, P. & VAN RIJ, R. P. 2013. Small RNAs tackle large viruses: RNA interference-based antiviral defense against DNA viruses in insects. *Fly*, 7**,** 216-223.

CAMPO, S., PERIS- PERIS, C., SIRÉ, C., MORENO, A. B., DONAIRE, L., ZYTNICKI, M., NOTREDAME, C., LLAVE, C. & SAN SEGUNDO, B. 2013. Identification of a novel microRNA (miRNA) from rice that targets an alternatively spliced transcript of the Nramp6 (Natural resistance- associated macrophage protein 6) gene involved in pathogen resistance. *New Phytologist*, 199**,** 212-227.

CAREY, E., GIBSON, R., FUENTES, S., MACHMUD, M., MWANGA, R., TURYAMUREEBA, G., ZHANG, L., MA, D., ABO EL-ABBAS, F. & EL-BEDEWY, R. 1997. The causes and control of virus diseases of sweetpotato in developing countries: is sweetpotato virus disease the main problem. *Impact on a changing world. International Potato Center Program Report,* 1998**,** 241-248.

CASASSOLA, A., BRAMMER, S. P., CHAVES, M. S., ANT, J. & GRANDO, M. F. 2013. Gene expression: A review on methods for the study of defense-related gene differential expression in plants. *American Journal of Plant Sciences,* 2013.

CHEN, G., PAN, D., ZHOU, Y., LIN, S. & KE, X. 2007. Diversity and evolutionary relationship of nucleotide binding site-encoding disease-resistance gene analogues in sweet potato (Ipomoea batatas Lam.). *Journal of Biosciences,* 32**,** 713-721.

CHEN, X. 2005. MicroRNA biogenesis and function in plants. *FEBS letters,* 579**,** 5923-5931.

CLARK, C. A., DAVIS, J. A., ABAD, J. A., CUELLAR, W. J., FUENTES, S., KREUZE, J. F., GIBSON, R. W., MUKASA, S. B., TUGUME, A. K. & TAIRO, F. D. 2012. Sweetpotato viruses: 15 years of progress on understanding and managing complex diseases. *Plant Disease*, 96**,** 168-185.

COSTA, A. T., BRAVO, J. P., MAKIYAMA, R. K., VASCONCELLOS NUNES, A. & MAIA, I. G. 2013. Viral counter defense X antiviral immunity in plants: mechanisms for survival. *Romanowski, V., Current issues in molecular virology-viral genetics and biotechnological applications, InTech*.

CUELLAR, W. J., GALVEZ, M., FUENTES, S., TUGUME, J. & KREUZE, J. 2015. Synergistic interactions of begomoviruses with Sweet potato chlorotic stunt virus

(genus Crinivirus) in sweet potato (Ipomoea batatas L.). *Molecular Plant Pathology,* 16**,** 459-471.

CUPERUS, J. T., FAHLGREN, N. & CARRINGTON, J. C. 2011. Evolution and functional diversification of MIRNA genes. *The Plant Cell,* 23**,** 431-442.

DAI, X. & ZHAO, P. X. 2011. psRNATarget: a plant small RNA target analysis server. *Nucleic Acids Research,* 39**,** W155-W159.

DE RONDE, D., BUTTERBACH, P. & KORMELINK, R. 2014. Dominant resistance against plant viruses. *Frontiers in Plant Science,* 5**,** 307.

DING, D., ZHANG, L., WANG, H., LIU, Z., ZHANG, Z. & ZHENG, Y. 2009. Differential expression of miRNAs in response to salt stress in maize roots. *Annals of Botany,* 103**,** 29-38.

DUAN, C.-G., WANG, C.-H. & GUO, H.-S. 2012. Application of RNA silencing to plant disease resistance. *Silence,* 3**,** 5.

FAHLGREN, N., HOWELL, M. D., KASSCHAU, K. D., CHAPMAN, E. J., SULLIVAN, C. M., CUMBIE, J. S., GIVAN, S. A., LAW, T. F., GRANT, S. R. & DANGL, J. L. 2007. High-throughput sequencing of Arabidopsis microRNAs: evidence for frequent birth and death of MIRNA genes. *PloS One,* 2**,** e219.

FENG, J., WANG, Y., LIN, R. & CHEN, J. 2013. Altered expression of microRNAs and target mRNAs in tomato root and stem tissues upon different viral infection. *Journal of Phytopathology,* 161**,** 107-119.

FERDOUS, J., HUSSAIN, S. S. & SHI, B. J. 2015. Role of microRNAs in plant drought tolerance. *Plant Biotechnology Journal,* 13**,** 293-305.

FINNEGAN, E. J. & MATZKE, M. A. 2003. The small RNA world. *Journal of Cell Science,* 116**,** 4689-4693.

FREEMAN, B. C. & BEATTIE, G. A. 2008. An overview of plant defenses against pathogens and herbivores. *The Plant Health Instructor*.

GAO, C., WANG, P., ZHAO, S., ZHAO, C., XIA, H., HOU, L., JU, Z., ZHANG, Y., LI, C. & WANG, X. 2017. Small RNA profiling and degradome analysis reveal regulation of microRNA in peanut embryogenesis and early pod development. *BMC Genomics,* 18**,** 220.

GIBSON, R. & KREUZE, J. 2014. Degeneration in sweetpotato due to viruses, virus-cleaned planting material and reversion: a review. *Plant Pathology,* 64**,** 1-15.

GIBSON, R., MPEMBE, I., ALICAI, T., CAREY, E., MWANGA, R., SEAL, S. & VETTEN, H. 1998. Symptoms, aetiology and serological analysis of sweet potato virus disease in Uganda. *Plant Pathology,* 47**,** 95-102.

GIBSON, R., MPEMBE, I. & MWANGA, R. 2011. Benefits of participatory plant breeding (PPB) as exemplified by the first-ever officially released PPB-bred sweet potato cultivar. *The Journal of Agricultural Science,* 149**,** 625-632.

GONZÁLEZ, V. M., MÜLLER, S., BAULCOMBE, D. & PUIGDOMÈNECH, P. 2015. Evolution of NBS-LRR gene copies among dicot plants and its regulation by members of the miR482/2118 superfamily of miRNAs. *Molecular Plant,* 8**,** 329-331.

GOYER, A., HAMLIN, L., CROSSLIN, J. M., BUCHANAN, A. & CHANG, J. H. 2015. RNA-Seq analysis of resistant and susceptible potato varieties during the early stages of potato virus Y infection. *BMC Genomics,* 16**,** 472.

GRIFFITHS-JONES, S., BATEMAN, A., MARSHALL, M., KHANNA, A. & EDDY, S. R. 2003. Rfam: an RNA family database. *Nucleic acids research,* 31**,** 439-441.

GULERIA, P., MAHAJAN, M., BHARDWAJ, J. & YADAV, S. K. 2011. Plant small RNAs: biogenesis, mode of action and their roles in abiotic stresses. *Genomics, Proteomics & Bioinformatics,* 9**,** 183-199.

GUO, C., LI, L., WANG, X. & LIANG, C. 2015. Alterations in siRNA and miRNA expression profiles detected by deep sequencing of transgenic rice with siRNA-mediated viral resistance. *PLoS One,* 10**,** e0116175.

HAVELDA, Z., VÁRALLYAY, É., VÁLÓCZI, A. & BURGYÁN, J. 2008. Plant virus infection- induced persistent host gene downregulation in systemically infected leaves. *The Plant Journal,* 55**,** 278-288.

HE, L. & HANNON, G. J. 2004. MicroRNAs: small RNAs with a big role in gene regulation. *Nature Reviews Genetics,* 5**,** 522-531.

HICKS, J. & LIU, H.-C. 2013. Involvement of eukaryotic small RNA pathways in host defense and viral pathogenesis. *Viruses,* 5**,** 2659-2678.

HUANG, W., PENG, S., XIAN, Z., LIN, D., HU, G., YANG, L., REN, M. & LI, Z. 2017. Overexpression of a tomato miR171 target gene SlGRAS24 impacts multiple agronomical traits via regulating gibberellin and auxin homeostasis. *Plant Biotechnology Journal,* 15**,** 472-488.

HUANG, Z., YEAKLEY, J. M., GARCIA, E. W., HOLDRIDGE, J. D., FAN, J.-B. & WHITHAM, S. A. 2005. Salicylic acid-dependent expression of host genes in compatible Arabidopsis-virus interactions. *Plant Physiology,* 137**,** 1147-1159.

ISHIHARA, T., SAKURAI, N., SEKINE, K.-T., HASE, S., IKEGAMI, M., SHIBATA, D. & TAKAHASHI, H. 2004. Comparative analysis of expressed sequence tags in resistant and susceptible ecotypes of Arabidopsis thaliana infected with Cucumber mosaic virus. *Plant and Cell Physiology,* 45**,** 470-480.

ISLAM, M. T., FERDOUS, A. S., NAJNIN, R. A., SARKER, S. K. & KHAN, H. 2015. High-throughput sequencing reveals diverse sets of conserved, nonconserved, and species-specific miRNAs in jute. *International Journal of Genomics,* 2015.

JAGADEESWARAN, G., ZHENG, Y., LI, Y. F., SHUKLA, L. I., MATTS, J., HOYT, P., MACMIL, S. L., WILEY, G. B., ROE, B. A. & ZHANG, W. 2009. Cloning and characterization of small RNAs from Medicago truncatula reveals four novel legume- specific microRNA families. *New Phytologist,* 184**,** 85-98.

JAGTAP, S. & SHIVAPRASAD, P. V. 2014. Diversity, expression and mRNA targeting abilities of Argonaute-targeting miRNAs among selected vascular plants. *BMC Genomics,* 15**,** 1049.

JIN, W. & WU, F. 2015. Characterization of miRNAs associated with Botrytis cinerea infection of tomato leaves. *BMC Plant Biology,* 15**,** 1.

JONES, J. D. & DANGL, J. L. 2006. The plant immune system. *Nature,* 444**,** 323-329.

JUNG, J.-H. & PARK, C.-M. 2007. MIR166/165 genes exhibit dynamic expression patterns in regulating shoot apical meristem and floral development in Arabidopsis. *Planta,* 225**,** 1327-1338.

KAMTHAN, A., CHAUDHURI, A., KAMTHAN, M. & DATTA, A. 2015. Small RNAs in plants: recent development and application for crop improvement. *Frontiers in Plant Science,* 6**,** 208.

KANG, B.-C., YEAM, I. & JAHN, M. M. 2005. Genetics of plant virus resistance. *Annu. Rev. Phytopathol.,* 43**,** 581-621.

KASHIF, M., PIETILÄ, S., ARTOLA, K., JONES, R., TUGUME, A., MÄKINEN, V. & VALKONEN, J. 2012. Detection of viruses in sweetpotato from Honduras and Guatemala augmented by deep-sequencing of small-RNAs. *Plant Disease,* 96**,** 1430-1437.

KHRAIWESH, B., ZHU, J.-K. & ZHU, J. 2012. Role of miRNAs and siRNAs in biotic and abiotic stress responses of plants. *Biochimica et Biophysica Acta (BBA)-Gene Regulatory Mechanisms,* 1819**,** 137-148.

KREUZE, J. 2002. *Molecular studies on the sweet potato virus disease and its two causal agents*.

KREUZE, J. F., PEREZ, A., UNTIVEROS, M., QUISPE, D., FUENTES, S., BARKER, I. & SIMON, R. 2009. Complete viral genome sequence and discovery of novel viruses by deep sequencing of small RNAs: a generic method for diagnosis, discovery and sequencing of viruses. *Virology,* 388**,** 1-7.

KU, Y.-S., WONG, J. W.-H., MUI, Z., LIU, X., HUI, J. H.-L., CHAN, T.-F. & LAM, H.-M. 2015. Small RNAs in plant responses to abiotic stresses: regulatory roles and study methods. *International Journal of Molecular Sciences,* 16**,** 24532-24554.

KUTNJAK, D., RUPAR, M., GUTIERREZ-AGUIRRE, I., CURK, T., KREUZE, J. F. & RAVNIKAR, M. 2015. Deep sequencing of virus-derived small interfering RNAs and RNA from viral particles shows highly similar mutational landscapes of a plant virus population. *Journal of Virology,* 89**,** 4760-4769.

LENZ, D., MAY, P. & WALTHER, D. 2011. Comparative analysis of miRNAs and their targets across four plant species. *BMC Research Notes,* 4**,** 483.

LI, F., PIGNATTA, D., BENDIX, C., BRUNKARD, J. O., COHN, M. M., TUNG, J., SUN, H., KUMAR, P. & BAKER, B. 2012. MicroRNA regulation of plant innate immune receptors. *Proceedings of the National Academy of Sciences,* 109**,** 1790-1795.

LI, J., ZHENG, H., ZHANG, C., HAN, K., WANG, S., PENG, J., LU, Y., ZHAO, J., XU, P. & WU, X. 2016. Different Virus-Derived siRNAs Profiles between Leaves and Fruits in Cucumber Green Mottle Mosaic Virus-Infected Lagenaria siceraria Plants. *Frontiers in Microbiology,* 7.

LI, X., XIE, X., LI, J., CUI, Y., HOU, Y., ZHAI, L., WANG, X., FU, Y., LIU, R. & BIAN, S. 2017. Conservation and diversification of the miR166 family in soybean and potential roles of newly identified miR166s. *BMC Plant Biology,* 17**,** 32.

LINDBO, J. A., FITZMAURICE, W. P. & DELLA-CIOPPA, G. 2001. Virus-mediated reprogramming of gene expression in plants. *Current Opinion in Plant Biology,* 4**,** 181-185.

LINDOW, M. & KROGH, A. 2005. Computational evidence for hundreds of non-conserved plant microRNAs. *BMC Genomics,* 6**,** 119.

LIU, H., LUO, L., LIANG, C., JIANG, N., LIU, P. & LI, J. 2015a. High-throughput sequencing identifies novel and conserved cucumber (Cucumis sativus L.) microRNAs in response to cucumber green mottle mosaic virus infection. *PloS One,* 10**,** e0129002.

LIU, H., SEARLE, I. R., WATSON-HAIGH, N. S., BAUMANN, U., MATHER, D. E., ABLE, A. J. & ABLE, J. A. 2015b. Genome-wide identification of microRNAs in

leaves and the developing head of four durum genotypes during water deficit stress. *PloS One,* 10**,** e0142799.

LIU, N., WU, S., VAN HOUTEN, J., WANG, Y., DING, B., FEI, Z., CLARKE, T. H., REED, J. W. & VAN DER KNAAP, E. 2014. Down-regulation of AUXIN RESPONSE FACTORS 6 and 8 by microRNA 167 leads to floral development defects and female sterility in tomato. *Journal of Experimental Botany,* 65**,** 2507-2520.

LIVAK, K. J. & SCHMITTGEN, T. D. J. M. 2001. Analysis of relative gene expression data using real-time quantitative PCR and the 2− ΔΔCT method. 25**,** 402-408.

LÓPEZ, C., SZUREK, B. & PEREZ-QUINTERO, A. L. L. 2012. Small Non-Coding RNAs in Plant Immunity. *Plant Science.* InTech.

LU, S., SUN, Y.-H., SHI, R., CLARK, C., LI, L. & CHIANG, V. L. 2005. Novel and mechanical stress–responsive microRNAs in Populus trichocarpa that are absent from Arabidopsis. *The Plant Cell,* 17**,** 2186-2203.

MA, C., BURD, S. & LERS, A. 2015. miR408 is involved in abiotic stress responses in Arabidopsis. *The Plant Journal,* 84**,** 169-187.

MALLORY, A. C. & VAUCHERET, H. 2006. Functions of microRNAs and related small RNAs in plants. *Nature Genetics,* 38**,** S31-S36.

MANAVELLA, P. A., KOENIG, D., RUBIO-SOMOZA, I., BURBANO, H. A., BECKER, C. & WEIGEL, D. 2013. Tissue-specific silencing of Arabidopsis SU (VAR) 3-9 HOMOLOG8 by miR171a. *Plant Physiology,* 161**,** 805-812.

MBANZIBWA, D., TUGUME, A., CHIUNGA, E., MARK, D. & TAIRO, F. 2014. Small RNA deep sequencing- based detection and further evidence of DNA viruses infecting sweetpotato plants in Tanzania. *Annals of Applied Biology,* 165**,** 329-339.

MCDOWELL, J. M. & WOFFENDEN, B. J. 2003. Plant disease resistance genes: recent insights and potential applications. *TRENDS in Biotechnology,* 21**,** 178-183.

MCGREGOR, C. E., MIANO, D. W., LABONTE, D. R., HOY, M., CLARK, C. A. & ROSA, G. J. 2009. Differential gene expression of resistant and susceptible sweetpotato plants after infection with the causal agents of sweet potato virus disease. *Journal of the American Society for Horticultural Science,* 134**,** 658-666.

MIANO, D. W., LABONTE, D. R. & CLARK, C. A. 2008. Identification of molecular markers associated with sweet potato resistance to sweet potato virus disease in Kenya. *Euphytica,* 160**,** 15-24.

MIEDES, E., VANHOLME, R., BOERJAN, W. & MOLINA, A. 2014. The role of the secondary cell wall in plant resistance to pathogens. *Frontiers in Plant Science,* 5.

MLOTSHWA, S., PRUSS, G. J. & VANCE, V. 2008. Small RNAs in viral infection and host defense. *Trends in Plant Science,* 13**,** 375-382.

MOREA, E. G. O., DA SILVA, E. M., E SILVA, G. F. F., VALENTE, G. T., ROJAS, C. H. B., VINCENTZ, M. & NOGUEIRA, F. T. S. 2016. Functional and evolutionary analyses of the miR156 and miR529 families in land plants. *BMC Plant Biology,* 16**,** 40.

MOXON, S., JING, R., SZITTYA, G., SCHWACH, F., PILCHER, R. L. R., MOULTON, V. & DALMAY, T. 2008. Deep sequencing of tomato short RNAs identifies microRNAs targeting genes involved in fruit ripening. *Genome Research,* 18**,** 1602-1609.

MWANGA, R., ODONGO, B., TURYAMUREEBA, G., ALAJO, A., YENCHO, G., GIBSON, R., SMIT, N. & CAREY, E. 2003. Release of six sweetpotato cultivars ('NASPOT 1'to'NASPOT 6') in Uganda. *HortScience,* 38**,** 475-476.

MWANGA, S., GHISLAIN, M., KREUZE, J., SSEMAKULA, G. & YENCHO, G. C. 2011. Exploiting the use of biotechnology in sweetpotato for improved nutrition and food security: Progress and future outlook.

NAWROCKI, E. P., BURGE, S. W., BATEMAN, A., DAUB, J., EBERHARDT, R. Y., EDDY, S. R., FLODEN, E. W., GARDNER, P. P., JONES, T. A. & TATE, J. 2015. Rfam 12.0: updates to the RNA families database. *Nucleic Acids Research,* 43**,** D130-D137.

NGAILO, S., SHIMELIS, H., SIBIYA, J. & MTUNDA, K. 2013. Sweet potato breeding for resistance to sweet potato virus disease and improved yield: progress and challenges. *African Journal of Agricultural Research,* 8**,** 3202-3215.

OGWOK, E., ILYAS, M., ALICAI, T., REY, M. E. & TAYLOR, N. J. 2016. Comparative analysis of virus-derived small RNAs within cassava (Manihot esculenta Crantz) infected with cassava brown streak viruses. *Virus Research,* 215**,** 1-11.

PALLAS, V. & GARCÍA, J. A. 2011. How do plant viruses induce disease? Interactions and interference with host components. *Journal of General Virology,* 92**,** 2691-2705.

PANTALEO, V., SZITTYA, G., MOXON, S., MIOZZI, L., MOULTON, V., DALMAY, T. & BURGYAN, J. 2010. Identification of grapevine microRNAs and their targets using high- throughput sequencing and degradome analysis. *The Plant Journal,* 62**,** 960-976.

PANTALEO, V., VITALI, M., BOCCACCI, P., MIOZZI, L., CUOZZO, D., CHITARRA, W., MANNINI, F., LOVISOLO, C. & GAMBINO, G. 2016. Novel functional microRNAs from virus-free and infected Vitis vinifera plants under water stress. *Scientific Reports,* 6.

PARK, S.-C., KIM, Y.-H., JI, C. Y., PARK, S., CHEOL JEONG, J., LEE, H.-S. & KWAK, S.-S. 2012. Stable internal reference genes for the normalization of real-time PCR in different sweetpotato cultivars subjected to abiotic stress conditions. *PloS One,* 7**,** e51502.

PELÁEZ, P. & SANCHEZ, F. 2013. Small RNAs in plant defense responses during viral and bacterial interactions: similarities and differences. *Frontiers in Plant Science,* 4**,** 343.

PÉREZ-QUINTERO, Á. L., NEME, R., ZAPATA, A. & LÓPEZ, C. 2010. Plant microRNAs and their role in defense against viruses: a bioinformatics approach. *BMC Plant Biology,* 10**,** 138.

PFAFFL, M. W., TICHOPAD, A., PRGOMET, C. & NEUVIANS, T. P. J. B. L. 2004. Determination of stable housekeeping genes, differentially regulated target genes and sample integrity: BestKeeper–Excel-based tool using pair-wise correlations. 26**,** 509-515.

PUMPLIN, N. & VOINNET, O. 2013. RNA silencing suppression by plant pathogens: defence, counter-defence and counter-counter-defence. *Nature Reviews Microbiology,* 11**,** 745-760.

QIN, Z., LI, C., MAO, L. & WU, L. 2014. Novel insights from non-conserved microRNAs in plants. *Frontiers in Plant Science,* 5**,** 586.

RAJAGOPALAN, R., VAUCHERET, H., TREJO, J. & BARTEL, D. P. 2006. A diverse and evolutionarily fluid set of microRNAs in Arabidopsis thaliana. *Genes & Development,* 20**,** 3407-3425.

ROGANS, S. J., ALLIE, F., TIRANT, J. E. & REY, M. E. C. 2016. Small RNA and methylation responses in susceptible and tolerant landraces of cassava infected with South African cassava mosaic virus. *Virus Research,* 225**,** 10-22.

ROGERS, K. & CHEN, X. 2013. Biogenesis, turnover, and mode of action of plant microRNAs. *The Plant Cell,* 25**,** 2383-2399.

RUIZ-FERRER, V. & VOINNET, O. 2009. Roles of plant small RNAs in biotic stress responses. *Annual Review of Plant Biology,* 60**,** 485-510.

SEKINE, K.-T., KAWAKAMI, S., HASE, S., KUBOTA, M., ICHINOSE, Y., SHAH, J., KANG, H.-G., KLESSIG, D. F. & TAKAHASHI, H. 2008. High level expression of

a virus resistance gene, RCY1, confers extreme resistance to Cucumber mosaic virus in Arabidopsis thaliana. *Molecular Plant-microbe Interactions,* 21**,** 1398-1407.

SHARMA, T., DAS, A., THAKUR, S. & JALALI, S. Recent understanding on structure, function and evolution of plant disease resistance genes. Proc. Indian Nat. Sci. Acad, 2014. 83-93.

SHARMA, V. K., KUSHWAHA, N., BASU, S., SINGH, A. K. & CHAKRABORTY, S. 2015. Identification of siRNA generating hot spots in multiple viral suppressors to generate broad-spectrum antiviral resistance in plants. *Physiology and Molecular Biology of Plants,* 21**,** 9-18.

SHIVAPRASAD, P. V., CHEN, H.-M., PATEL, K., BOND, D. M., SANTOS, B. A. & BAULCOMBE, D. C. 2012. A microRNA superfamily regulates nucleotide binding site–leucine-rich repeats and other mRNAs. *The Plant Cell,* 24**,** 859-874.

SHRIRAM, V., KUMAR, V., DEVARUMATH, R. M., KHARE, T. S. & WANI, S. H. 2016. MicroRNAs as potential targets for abiotic stress tolerance in plants. *Frontiers in Plant Science,* 7.

SONG, C., FANG, J., LI, X., LIU, H. & CHAO, C. T. 2009. Identification and characterization of 27 conserved microRNAs in citrus. *Planta,* 230**,** 671-685.

SONG, Q.-X., LIU, Y.-F., HU, X.-Y., ZHANG, W.-K., MA, B., CHEN, S.-Y. & ZHANG, J.-S. 2011. Identification of miRNAs and their target genes in developing soybean seeds by deep sequencing. *BMC Plant Biology,* 11**,** 5.

SOOSAAR, J. L., BURCH-SMITH, T. M. & DINESH-KUMAR, S. P. 2005. Mechanisms of plant resistance to viruses. *Nature Reviews Microbiology,* 3**,** 789-798.

SRIVASTAVA, S., ZHENG, Y., KUDAPA, H., JAGADEESWARAN, G., HIVRALE, V., VARSHNEY, R. K. & SUNKAR, R. 2015. High throughput sequencing of small RNA component of leaves and inflorescence revealed conserved and novel miRNAs as well as phasiRNA loci in chickpea. *Plant Science,* 235**,** 46-57.

STIEF, A., ALTMANN, S., HOFFMANN, K., PANT, B. D., SCHEIBLE, W.-R. & BÄURLE, I. 2014. Arabidopsis miR156 regulates tolerance to recurring environmental stress through SPL transcription factors. *The Plant Cell,* 26**,** 1792-1807.

SUN, G. 2012. MicroRNAs and their diverse functions in plants. *Plant Molecular Biology,* 80**,** 17-36.

SUNKAR, R. & ZHU, J.-K. 2004. Novel and stress-regulated microRNAs and other small RNAs from Arabidopsis. *The Plant Cell,* 16**,** 2001-2019.

SZITTYA, G., MOXON, S., SANTOS, D. M., JING, R., FEVEREIRO, M. P., MOULTON, V. & DALMAY, T. 2008. High-throughput sequencing of Medicago truncatula short RNAs identifies eight new miRNA families. *BMC Genomics,* 9**,** 593.

TAYLOR, R. S., TARVER, J. E., HISCOCK, S. J. & DONOGHUE, P. C. 2014. Evolutionary history of plant microRNAs. *Trends in Plant Science,* 19**,** 175-182.

THAKUR, M. & SOHAL, B. S. 2013. Role of elicitors in inducing resistance in plants against pathogen infection: a review. *ISRN Biochemistry,* 2013.

THIEBAUT, F., GRATIVOL, C., CARNAVALE-BOTTINO, M., ROJAS, C. A., TANURDZIC, M., FARINELLI, L., MARTIENSSEN, R. A., HEMERLY, A. S. & FERREIRA, P. C. G. 2012. Computational identification and analysis of novel sugarcane microRNAs. *BMC Genomics,* 13**,** 290.

THIEBAUT, F., GRATIVOL, C., HEMERLY, A. S. & FERREIRA, P. C. G. 2015. MicroRNA networks in plant-microorganism interactions. *Tropical Plant Biology,* 8**,** 40-50.

TIWARI, M., SHARMA, D. & TRIVEDI, P. K. 2014. Artificial microRNA mediated gene silencing in plants: progress and perspectives. *Plant Molecular Biology,* 86**,** 1-18.

TRINDADE, I., CAPITÃO, C., DALMAY, T., FEVEREIRO, M. P. & DOS SANTOS, D. M. 2010. miR398 and miR408 are up-regulated in response to water deficit in Medicago truncatula. *Planta,* 231**,** 705-716.

TURNER, M., YU, O. & SUBRAMANIAN, S. 2012. Genome organization and characteristics of soybean microRNAs. *BMC Genomics,* 13**,** 169.

VALIOLLAHI, E., FARSI, M. & KAKHKI, A. M. 2014. Sly-miR166 and Sly-miR319 are components of the cold stress response in Solanum lycopersicum. *Plant Biotechnology Reports,* 8**,** 349-356.

VANCE, V. & VAUCHERET, H. 2001. RNA silencing in plants--defense and counterdefense. *Science,* 292**,** 2277-2280.

VÁRALLYAY, É. & HAVELDA, Z. 2013. Unrelated viral suppressors of RNA silencing mediate the control of ARGONAUTE1 level. *Molecular Plant Pathology,* 14**,** 567-575.

VÁRALLYAY, É., VÁLÓCZI, A., ÁGYI, Á., BURGYÁN, J. & HAVELDA, Z. 2010. Plant virus- mediated induction of miR168 is associated with repression of ARGONAUTE1 accumulation. *The EMBO Journal,* 29**,** 3507-3519.

VAUCHERET, H., MALLORY, A. C. & BARTEL, D. P. 2006. AGO1 homeostasis entails coexpression of MIR168 and AGO1 and preferential stabilization of miR168 by AGO1. *Molecular Cell,* 22**,** 129-136.

WANG, M.-B., MASUTA, C., SMITH, N. A. & SHIMURA, H. 2012. RNA silencing and plant viral diseases. *Molecular Plant-Microbe Interactions,* 25**,** 1275-1285.

WANG, X.-B., JOVEL, J., UDOMPORN, P., WANG, Y., WU, Q., LI, W.-X., GASCIOLLI, V., VAUCHERET, H. & DING, S.-W. 2011. The 21-nucleotide, but not 22-nucleotide, viral secondary small interfering RNAs direct potent antiviral defense by two cooperative argonautes in Arabidopsis thaliana. *The Plant Cell,* 23**,** 1625-1638.

WANG, Y., ROSEN, B., SCOFFIELD, J., EGNIN, M., MORTLEY, D., STEINER, S., COOK, D. & HE, G. 2010. Isolation and analysis of resistance gene homologues in sweetpotato. *Plant Breeding,* 129**,** 519-525.

WHITHAM, S. A., YANG, C. & GOODIN, M. M. 2006. Global impact: elucidating plant responses to viral infection. *Molecular Plant-microbe Interactions,* 19**,** 1207-1215.

WU, J., YANG, Z., WANG, Y., ZHENG, L., YE, R., JI, Y., ZHAO, S., JI, S., LIU, R. & XU, L. 2015. Viral-inducible Argonaute18 confers broad-spectrum virus resistance in rice by sequestering a host microRNA. *Elife,* 4**,** e05733.

XIE, F., FRAZIER, T. P. & ZHANG, B. 2011. Identification, characterization and expression analysis of MicroRNAs and their targets in the potato (Solanum tuberosum). *Gene,* 473**,** 8-22.

XIN, M., WANG, Y., YAO, Y., XIE, C., PENG, H., NI, Z. & SUN, Q. 2010. Diverse set of microRNAs are responsive to powdery mildew infection and heat stress in wheat (Triticum aestivum L.). *BMC Plant Biology,* 10**,** 123.

YANG, L., MU, X., LIU, C., CAI, J., SHI, K., ZHU, W. & YANG, Q. 2015. Overexpression of potato miR482e enhanced plant sensitivity to Verticillium dahliae infection. *Journal of Integrative Plant Biology,* 57**,** 1078-1088.

YIN, X., WANG, J., CHENG, H., WANG, X. & YU, D. 2013. Detection and evolutionary analysis of soybean miRNAs responsive to soybean mosaic virus. *Planta,* 237**,** 1213-1225.

ZHANG, B., PAN, X., CANNON, C. H., COBB, G. P. & ANDERSON, T. A. 2006. Conservation and divergence of plant microRNA genes. *The Plant Journal,* 46**,** 243-259.

ZHANG, B., PAN, X. & STELLWAG, E. J. 2008. Identification of soybean microRNAs and their targets. *Planta,* 229**,** 161-182.

ZHANG, S., WANG, Y., LI, K., ZOU, Y., CHEN, L. & LI, X. 2014. Identification of cold-responsive miRNAs and their target genes in nitrogen-fixing nodules of soybean. *International Journal of Molecular Sciences,* 15**,** 13596-13614.

ZHOU, C.-J., ZHANG, X.-Y., LIU, S.-Y., WANG, Y., LI, D.-W., YU, J.-L. & HAN, C.-G. 2017. Synergistic infection of BrYV and PEMV 2 increases the accumulations of both BrYV and BrYV-derived siRNAs in Nicotiana benthamiana. *Scientific Reports,* 7**,** 45132.

ZHOU, L., LIU, Y., LIU, Z., KONG, D., DUAN, M. & LUO, L. 2010. Genome-wide identification and analysis of drought-responsive microRNAs in Oryza sativa. *Journal of Experimental Botany,* 61**,** 4157-4168.

ZHU, C., DING, Y. & LIU, H. 2011a. MiR398 and plant stress responses. *Physiologia Plantarum,* 143**,** 1-9.

ZHU, H., HU, F., WANG, R., ZHOU, X., SZE, S.-H., LIOU, L. W., BAREFOOT, A., DICKMAN, M. & ZHANG, X. 2011b. Arabidopsis Argonaute10 specifically sequesters miR166/165 to regulate shoot apical meristem development. *Cell,* 145**,** 242-256.

ZHU, Q.-H., FAN, L., LIU, Y., XU, H., LLEWELLYN, D. & WILSON, I. 2013. miR482 regulation of NBS-LRR defense genes during fungal pathogen infection in cotton. *PLoS One,* 8**,** e84390.

ZVEREVA, A. S. & POOGGIN, M. M. 2012. Silencing and innate immunity in plant defense against viral and non-viral pathogens. *Viruses,* 4**,** 2578-2597.

CHAPTER 6

GENERAL DISCUSSION AND CONCLUSION

Sweet potato has become an important crop in South Africa due to its high vitamin A and overall nutritional content (Laurie et al., 2012, Laurie and Faber, 2008). This crop is attractive to resource poor farmers because it is drought tolerant, high yielding and is a potential source of income (Faber et al., 2013, Laurie et al., 2015, Motsa et al., 2015). In 2016 the Department of Agriculture, Forestry and Fisheries reported that from 2011 to 2015 sweet potato local and international export markets were making a significant contribution to the national agricultural production gross value (DAFF, 2016). This was attributed to high production output and higher producer prices (DAFF, 2016). To maintain high production yields, quality and a thriving market, any limiting factors such as diseases or pests should be controlled or eradicated. The Agricultural Research Council's Vegetable and Ornamental Plant Institute (ARC-VOPI) maintains the only index collection in the whole country. Researchers at this institute test stock plants for viruses on a yearly. The virus-free material is then multiplied and distributed to vine growers and farmers. Ensuring virus-free limits major crop decline and economic losses. Identification of viruses facilitates accurate indexing, thus guaranteeing virus-free and healthy plant material. Virus detection therefore becomes the principal step in ensuring higher productivity. The ARC-VOPI produces over 1000 nursery bags and approximately 20000 seedling plantlets every year for distribution (http://www.arc.agric.za/arc-vopi/Pages/Plant%20Breeding/Sweet-Potatoes.aspx).

Sweet potato is categorised as an orphan crop (Varshney et al., 2012) and until recently, in the last two decades, genomic resources for sweet potato were scarce (Hirakawa et al., 2015, Tao et al., 2012, Wang et al., 2010, Wang et al., 2011). Researchers at the Michigan State University have developed a website which acts as a portal for sweet potato genomic resources (http://sweetpotato.plantbiology.msu.edu/index.shtml). These researchers have sequenced sweet potato relatives (*Ipomoea trifida*) with the aim to improve genetic and genomic resources of the *Ipomoea* species. Genome sequencing is also essential where virus identification and/or detection is concerned. In this study we found viruses that are known to sometimes integrate into the host genome (badnaviruses), it becomes increasingly critical to sequence the full sweet potato genome in order to screen it for any integrated sequences, so

that we understand how the viruses and host evolved over time. The near full length viral genomes and miRNA sequences obtained from deep sequencing experiments (RCA, RNASeq and small RNA libraries) can now also be added to the existing resources (ESTs, sweet potato viruses and cDNA sequences). The availability of genomic resources can significantly contribute towards the enhancement of nutrition and productivity of orphan crops such as sweet potato, through breeding programs for beneficial agronomic traits and disease or pest resistance (Laurie et al., 2015, Laurie et al., 2004, Laurie et al., 2012, Varshney et al., 2010, Varshney et al., 2009, Varshney et al., 2012). As the world population increases and food security is threatened, genomic resources provided by next generation sequencing (NGS) technologies, now play an key role in crop improvement (Edwards and Batley, 2010, M Perez-de-Castro et al., 2012). Effective and early detection of plant viruses is important to assist in the development of efficient disease control strategies (Jones, 2006). It is for this reason that the methods of plant virus detection and identification need to be accurate/sensitive and rapid as they could have major implications for disease control and management (Naidu and Hughes, 2003). Appropriate control of plant diseases is largely dependent on the correct identification of the causal agent (diagnosis). In this study we show that we were able to rapidly detect viruses infecting sweet potato using a viral metagenomics approach.

In the first research chapter we employed a total RNA sequencing strategy, which was coupled with ribosomal RNA depletion, to determine the diversity sweet potato viruses in two South African provinces. This approach allowed us to detect RNA viruses which represented only 1% of the sequence data. A total of 17 plant samples were individually sequenced and 7 Gb of data was generated. From this sequence data we successfully detected and sequenced 4 partial RNA genomes. The sequence data from this study managed to generate over 70% genome coverage for each virus sequenced. The identified viruses include potyviruses (sweet potato virus C – SPVC; sweet potato virus G – SPVG; sweet potato feathery mottle virus – SPFMV) and a crinivirus (sweet potato chlorotic stunt virus – SPCSV). These viruses were found co-infecting most plants. The two causal agents of SPVD (SPCSV and SPFMV) were also found in mixed infections with other viruses. Plants infected with SPFMV and SPCSV showed severe symptom development, while infections of SPVC and SPVG showed mild symptoms. SPCSV and SPFMV commonly occur in mixed infections and may cause synergistic interactions in the host, which can be observed phenotypically. Interestingly we also observed high sequence similarity/homology between

the South African SPCSV RNA2 segment and the Peruvian isolate (EA-strain) while the RNA1 segment was highly divergent, sharing 76% homology with the Ugandan isolate (WA-strain). These findings lead us to conclude that; 1) both EA and WA-strains are present in the country and 2) there could be a reassortment event that might have occurred between the strain that we found and another unknown parent sequence. We attempted to look for reassortment but due to the fact that there are few SPCSV RNA1 and RNA2 sequences available in the database, we could not detect any reassortment in our analysis. Our study was only limited to screening two provinces of SA by NGS, it becomes increasingly important for future studies to survey all 9 provinces or the entire continent for SPCSV. Rapid detection of SPCSV can be done using the primers we developed in this study to confirm the SPCSV sequence. Once the SPCSV is detected by RT-PCR, the plant samples from which the virus has been detected can then undergo whole genome sequencing using the approach used in this study. The full length sequences will then have to be analysed for reassortment. The results in this chapter have added new information to the existing body of knowledge on viruses infecting sweet potato in South Africa. The study also demonstrates NGS technology as a powerful tool in generating a comprehensive profile of the entire viral community in a given sample.

The second research chapter unveiled the presence of previously reported DNA viruses, namely sweet potato mosaic virus (SPMaV) and sweet potato leaf Sao Paulo virus (SPLCSPV). Additionally, two novel badnaviruses, sweet potato badnavirus A (SPBVA) and sweet potato badnavirus B (SPBVB), were reported for the first time in sweet potato plants in South Africa. Plant samples underwent gDNA isolation, enrichment for circular virus genomes using rolling circle amplification (RCA) and deep sequencing. The sequence data revealed that the begomoviruses (SPMaV and SPLCSPV) were present in mixtures with SPFMV and SPCSV. The two badnaviruses were also found in mixtures in all of the seven plant samples that were analysed. The findings of this study highlight the advantages of RNASeq and DNASeq in viral diagnosis. This approach can find wide application in other plant-viral interactions.

In the final experimental chapter the study aimed to identify and compare miRNAs expressed in resistant vs susceptible sweet potato cultivars in response to multiple virus infection. In this study we found virus-derived small interfering RNAs (vsiRNAs) and defense-related microRNAs (miRNAs) from both resistant and susceptible sweet potato cultivars. The

majority of vsiRNAs were associated with SPVC, SPFMV, and SPVG in the NASPOT 1 cultivar, while SPFMV, SPVC and SPBVB vsiRNA reads were significantly higher in the Blesbok cultivar. Twenty one percent (21%) of the sequence reads matched to known miRNAs in the miRBase (Release 21) database, and 60 miRNA families were identified from the NASPOT 1 and Blesbok cultivars collectively. Amongst the miRNAs identified from our data set, we detected upregulation of miR393, miR398, miR168, miR162, miR167 and miR397 in both cultivars. We also observed upregulation of a defense-related miRNA, miR482, in both cultivars. Two miRNAs (miR6300 and miR160) were downregulated in both cultivars after virus infection. The changes in the expression patterns of these miRNAs infer that they could play a role in pathogen response. The differentially expressed miRNAs identified in this study, miR168, miR403, miR162 and miR482, are known to regulate Argonaute 1 (AGO1), AGO 2 & 3, Dicer-like 1 (DCL1) and NB-LRR resistance genes respectively. These proteins and genes are known to play key roles in the silencing pathway. NGS expression profiles of miR482, miR393, miR397 and miR168 were validated by quantitative real-time PCR (RT-qPCR). The findings from this study add information to the existing miRNA database and will elucidate the role of miRNAs in pathogen defense. This study focused primarily on identifying and analysing defense related miRNAs but information gathered from this study can be used to study miRNAs that are associated with stress responses (drought) or agronomic traits (crop yield, quality or β-carotene content). We recommend undertaking RNASeq analysis in conjunction with the miRNA identification, at specific time points (early and late infection) in order to understand the complete gene expression patterns, which will further elucidate host-pathogen interactions. Functional gene analysis will also assist in the identification of more target genes. This study lays a foundation for understanding host-pathogen interactions in sweet potato. This is significant as it will contribute to devising effective strategies to manage viral diseases in sweet potato and exploring the use of miRNAs for sweet potato improvement.

Studies and reviews have indicated that for breeding strategies to be successful, the availability of genomic information is important (Berkman et al., 2012, Brenton et al., 2016, Cardi et al., 2017, M Perez-de-Castro et al., 2012, Thottathil et al., 2016, Varshney et al., 2010, Varshney et al., 2009). The availability of genomic information on the host and the pathogens is necessary for breeding programmes. Genomic knowledge has enabled the use of virus induced gene silencing (VIGS) and marker-assisted selection methods to engineer virus resistant sweet potato cultivars (Ngailo et al., 2013). MiRNAs have not been utilised to

genetically engineer virus-resistant sweet potato cultivars. Breeding efforts can employ the silencing pathway as a strategy for controlling sweet potato diseases and viruses that affect the production and quality of sweet potato. Many studies have shown that artificial microRNA (amiRNA) technology has contributed greatly to increasing plant resistance to viruses (Gupta, 2015, Kamthan et al., 2015). Therefore this technology can also be used to design a potential multi-layered defense mechanism against multiple viruses in sweet potato. To our knowledge, this is the first study that identifies and characterises miRNAs in sweet potato.

The results found in this study provide new, valuable, metagenomics-level insights into the diversity of RNA and DNA viruses infecting sweet potatoes and inform resistance/susceptible responses to virus infection. This work makes an original contribution to this research field. This study highlights a number of firsts: 1) the first study in South Africa to sequence 8 full-length and near full-length sweet potato virus genomes; 2) the first study to report sweet potato badnaviruses in South Africa; 3) the first study to report a divergent SPCSV RNA1 segment in South Africa; and 4) the first study to report miRNA sequences involved in biotic stress response in sweet potato.

References

AMES, T. 1997. *Sweetpotato: major pests, diseases, and nutritional disorders*, International Potato Center.

BERKMAN, P. J., LAI, K., LORENC, M. T. & EDWARDS, D. 2012. Next-generation sequencing applications for wheat crop improvement. *American Journal of Botany,* 99**,** 365-371.

BRENTON, Z. W., COOPER, E. A., MYERS, M. T., BOYLES, R. E., SHAKOOR, N., ZIELINSKI, K. J., RAUH, B. L., BRIDGES, W. C., MORRIS, G. P. & KRESOVICH, S. 2016. A genomic resource for the development, improvement, and exploitation of sorghum for bioenergy. *Genetics,* 204**,** 21-33.

CARDI, T., D'AGOSTINO, N. & TRIPODI, P. 2017. Genetic transformation and genomic resources for next-generation precise genome engineering in vegetable crops. *Frontiers in Plant Science,* 8.

CUELLAR, W. J., CRUZADO, R. K., FUENTES, S., UNTIVEROS, M., SOTO, M. & KREUZE, J. F. 2011. Sequence characterization of a Peruvian isolate of Sweet potato

chlorotic stunt virus: further variability and a model for p22 acquisition. *Virus Research,* 157**,** 111-115.

CUELLAR, W. J., TAIRO, F., KREUZE, J. F. & VALKONEN, J. P. 2008. Analysis of gene content in sweet potato chlorotic stunt virus RNA1 reveals the presence of the p22 RNA silencing suppressor in only a few isolates: implications for viral evolution and synergism. *Journal of General Virology,* 89**,** 573-582.

EDWARDS, D. & BATLEY, J. 2010. Plant genome sequencing: applications for crop improvement. *Plant Biotechnology Journal,* 8**,** 2-9.

FABER, M., LAURIE, S. M. & VAN JAARSVELD, P. J. 2013. Total β-carotene content of orange sweetpotato cultivated under optimal conditions and at a rural village. *African Journal of Biotechnology,* 12.

GUPTA, P. K. 2015. MicroRNAs and target mimics for crop improvement. *Current Science India,* 108**,** 1624-33.

HIRAKAWA, H., OKADA, Y., TABUCHI, H., SHIRASAWA, K., WATANABE, A., TSURUOKA, H., MINAMI, C., NAKAYAMA, S., SASAMOTO, S. & KOHARA, M. 2015. Survey of genome sequences in a wild sweet potato, Ipomoea trifida (HBK) G. Don. *DNA Research***,** dsv002.

JONES, D. R. 2003. Plant viruses transmitted by whiteflies. *European Journal of Plant Pathology,* 109**,** 195-219.

JONES, R. A. 2006. Control of plant virus diseases. *Advances in Virus Research,* 67**,** 205-244.

KAMTHAN, A., CHAUDHURI, A., KAMTHAN, M. & DATTA, A. 2015. Small RNAs in plants: recent development and application for crop improvement. *Frontiers in Plant Science,* 6**,** 208.

KIM, J., KIL, E. J., KIM, S., SEO, H., BYUN, H. S., PARK, J., CHUNG, M. N., KWAK, H. R., KIM, M. K. & KIM, C. S. 2015. Seed transmission of Sweet potato leaf curl virus in sweet potato (Ipomoea batatas). *Plant Pathology,* 64**,** 1284-1291.

KREUZE, J. 2002. *Molecular studies on the sweet potato virus disease and its two causal agents*.

KREUZE, J. & FUENTES, S. 2008. Sweetpotato Viruses A2 - Mahy, Brian W.J. *In:* REGENMORTEL, M. H. V. V. (ed.) *Encyclopedia of Virology (Third Edition).* Oxford: Academic Press.

LAURIE, S., FABER, M., ADEBOLA, P. & BELETE, A. 2015. Biofortification of sweet potato for food and nutrition security in South Africa. *Food Research International,* 76**,** 962-970.

LAURIE, S., VAN DEN BERG, A., MAGORO, M. & KGONYANE, M. 2004. Breeding of sweetpotato and evaluation of imported cultivars in South Africa. *African Crop Science Journal,* 12**,** 189-196.

LAURIE, S., VAN JAARSVELD, P., FABER, M., PHILPOTT, M. & LABUSCHAGNE, M. 2012. Trans-β-carotene, selected mineral content and potential nutritional contribution of 12 sweetpotato varieties. *Journal of Food Composition and Analysis,* 27**,** 151-159.

LAURIE, S. M. & FABER, M. 2008. Integrated community‐based growth monitoring and vegetable gardens focusing on crops rich in β‐carotene: Project evaluation in a rural community in the Eastern Cape, South Africa. *Journal of the Science of Food and Agriculture,* 88**,** 2093-2101.

LOTRAKUL, P., VALVERDE, R., CLARK, C., SIM, J. & DE LA TORRE, R. 1998. Detection of a geminivirus infecting sweet potato in the United States. *Plant Disease,* 82**,** 1253-1257.

M PEREZ-DE-CASTRO, A., VILANOVA, S., CAÑIZARES, J., PASCUAL, L., M BLANCA, J., J DIEZ, M., PROHENS, J. & PICÓ, B. 2012. Application of genomic tools in plant breeding. *Current Genomics,* 13**,** 179-195.

MOTSA, N. M., MODI, A. T. & MABHAUDHI, T. 2015. Sweet potato (Ipomoea batatas L.) as a drought tolerant and food security crop. *South African Journal of Science,* 111**,** 1-8.

NAIDU, R. & HUGHES, J. 2003. Methods for the detection of plant virus diseases. *Plant Virology in Sub Saharan Africa***,** 233-253.

NGAILO, S., SHIMELIS, H., SIBIYA, J. & MTUNDA, K. 2013. Sweet potato breeding for resistance to sweet potato virus disease and improved yield: progress and challenges. *African Journal of Agricultural Research,* 8**,** 3202-3215.

PERRING, T. M., GRUENHAGEN, N. M. & FARRAR, C. A. 1999. Management of plant viral diseases through chemical control of insect vectors. *Annual Review of Entomology,* 44**,** 457-481.

TAO, X., GU, Y.-H., WANG, H.-Y., ZHENG, W., LI, X., ZHAO, C.-W. & ZHANG, Y.-Z. 2012. Digital gene expression analysis based on integrated de novo transcriptome assembly of sweet potato [Ipomoea batatas (L.) Lam.]. *PloS One,* 7**,** e36234.

THOTTATHIL, G. P., JAYASEKARAN, K. & OTHMAN, A. S. 2016. Sequencing Crop Genomes: A Gateway to Improve Tropical Agriculture. *Tropical Life Sciences Research,* 27**,** 93.

VALVERDE, R. A., SIM, J. & LOTRAKUL, P. 2004. Whitefly transmission of sweet potato viruses. *Virus Research,* 100**,** 123-128.

VARSHNEY, R. K., GLASZMANN, J.-C., LEUNG, H. & RIBAUT, J.-M. 2010. More genomic resources for less-studied crops. *Trends in Biotechnology,* 28**,** 452-460.

VARSHNEY, R. K., NAYAK, S. N., MAY, G. D. & JACKSON, S. A. 2009. Next-generation sequencing technologies and their implications for crop genetics and breeding. *Trends in Biotechnology,* 27**,** 522-530.

VARSHNEY, R. K., RIBAUT, J.-M., BUCKLER, E. S., TUBEROSA, R., RAFALSKI, J. A. & LANGRIDGE, P. 2012. Can genomics boost productivity of orphan crops? *Nature Biotechnology,* 30**,** 1172.

WANG, Z., FANG, B., CHEN, J., ZHANG, X., LUO, Z., HUANG, L., CHEN, X. & LI, Y. 2010. De novo assembly and characterization of root transcriptome using Illumina paired-end sequencing and development of cSSR markers in sweetpotato (Ipomoea batatas). *BMC Genomics,* 11**,** 726.

WANG, Z., LI, J., LUO, Z., HUANG, L., CHEN, X., FANG, B., LI, Y., CHEN, J. & ZHANG, X. 2011. Characterization and development of EST-derived SSR markers in cultivated sweetpotato (Ipomoea batatas). *BMC Plant Biology,* 11**,** 139.

APPENDIX

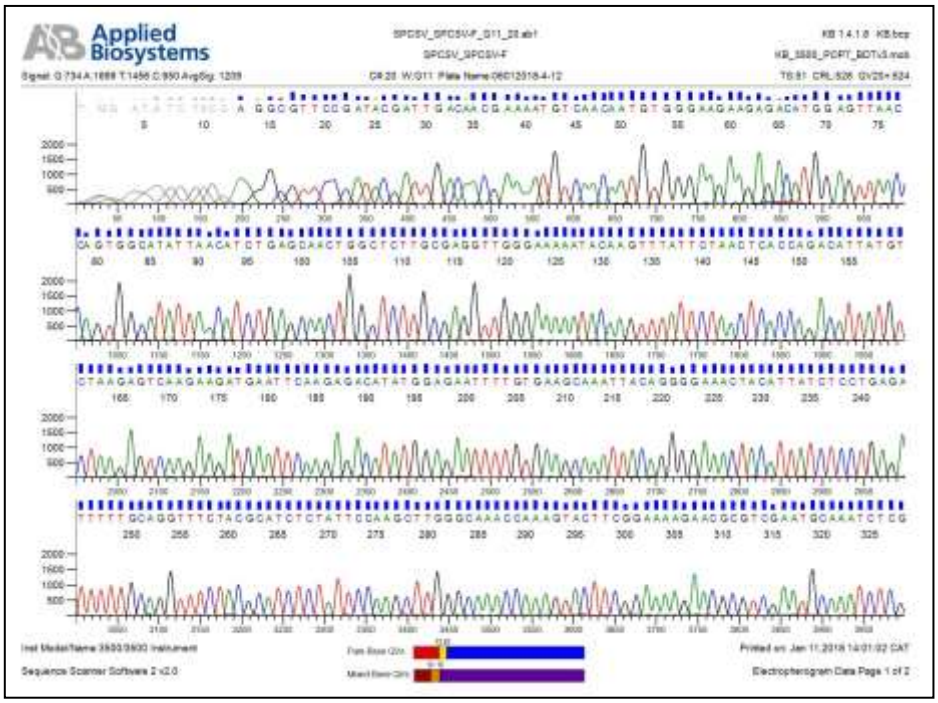

Figure A1a: Sanger sequenced PCR products viewed on the Sequence Scanner Software v2.0. This image shows the raw forward read of SPCSV coat protein before sequence editing. This image is representative of the 8 PCR products that were sequenced for the 8 viruses detected by NGS. Forward and reverse reads underwent Sanger sequencing.

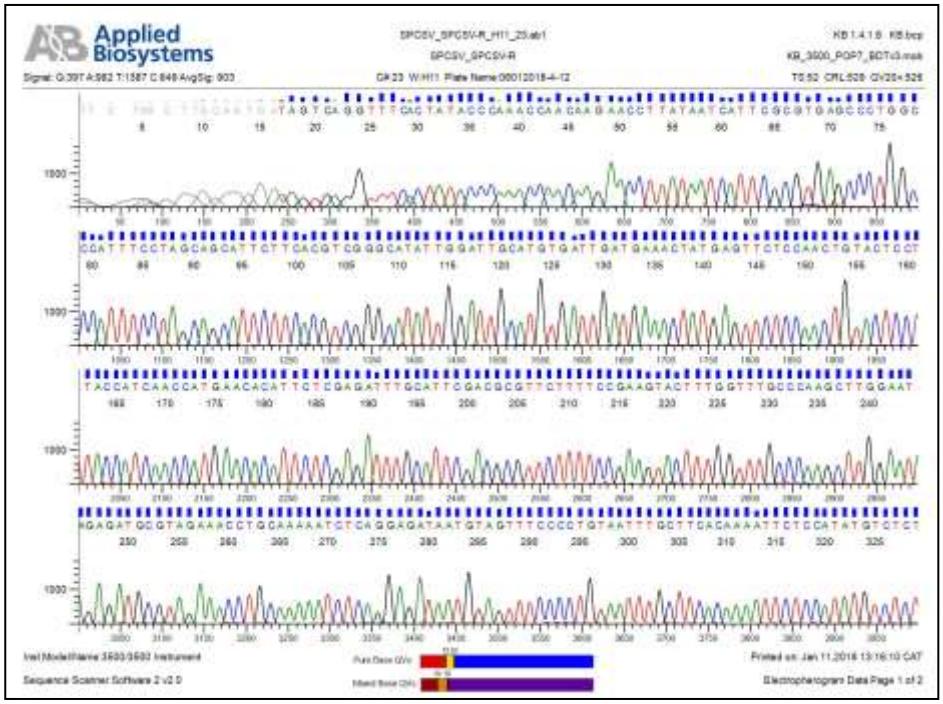

Figure A1b: Sanger sequenced SPCSV reverse read.

Table A1.1. Sanger sequenced PCR products for the 8 viruses detected by NGS.

>**SPCSV_PCR_product 570 bp**
CGATGTAGGTTTCGGGGACAGGAATGTGAACGATAATCTACGAAGGCGTTCCGA
TACGATTGACAACGAAAATGTCAACAATGTGGGAAGAAGAGACATGGAGTTAAC
CAGTGGCATATTAACATCTGAGCAACTGGCTCTTGCGAGGTTGGGAAAAATAC
AGTTTATTCTAACTCACCAGACATTATGTCTAAGAGTCAAGAAGATGAATTCAAG
AGACATATGGAGAATTTTGTGAAGCAAATTACAGGGGAAACTACATTATCTCCT
GAGATTTTTGCAGGTTTCTACGCATCTCTATTCCAAGCTTGGGCAAACCAAAGTA
CTTCGGAAAAGAACGCGTCGAATGCAAATCTCGAGAATGTGTTCATGGTTGATG
GTAAGGAGTACAGTTGGAGAACTCATAGTTTCATCAATCACATGCAATCCAATAT
GCCCGACGTGAAGAATGCTGCTAGGAAATGGGCCAGGGCTCACGCGAATGATTA
TAAGGTTCTTGTTGGTTTGGGTATAGTGAAACCTGACTATCATTTGCAAGCTAAA
CAAGGTGTATTACCGGAGTATTGGCA

>SPFMV_PCR_product 557 bp
GATATCCCTCCACCACCCACAATAACTGAGGTTACTGATCCAGAAGACCCAAAG
CAGGCAGCTTTGAGAGCTGCACGAGCTAAGCAACCCGCAACCATTCCAGAATCA
TATGGACGAGACACTAGCAAGGAGAAGGAATCAATAGTGGGAGCATCATCAAA
GGGTGTGAAGGATAAAGATGTAAACGTTGGTACAGTTGGTACATTTGTCGTGCC
ACGTGTTAAGATGAATGCAAACAAGAAAAGGCAACCAATGGTAAATGGAAGGG
CCATTATAAATTTCCAACACTTGTCAACATATGAGCCAGAACAGTTTGAGGTTGC
AAACACCCGGTCGACTCAAGAGCAGTTTCAAGCATGGTATGAAGGAGTGAAAGG
GGACTATGGTGTTGACGATGCAGGAATGGGGATCTTATTGAATGGATTAATGGTT
TGGTGCATTGAAAATGGCACATCCCCAAATATAAATGGTGTGTGGACAATGATG
GATGGTGATGAGCAAGTGACATATCCAATTAAACCATTGTTGGACCATGCAGTG
CCTACTTTTAGGCAGAT

>SPVC_PCR_product 316 bp
GCACCAGAAATCACAGAAATTACTGAGCCCGAGGATCCCAAACAAGCTGCGCTT
CGCGAAGCTAGACAGAAACAACCTGCTGTCACACCCGAATCATACGGTAGAGAT
ACAGGTGAGAAACCTATGCGCTCTGTTTCACCACAAAGGGTAAAAGACAAGGAT
GTCAATGTTGGTACGACAGGTACATTTTTAGTACCACGAGTTAAGCTTCATACCA
GTAAAATGCGCCAACCGAGAGTCAATGGAATCTCTGTAGTAAACTTACAACACC
TTGCAACCTACGAACCTGAGCAACATAACATTGGGAATACACGCT

>SPVG_PCR_product 666 bp
AAGGGGAAGAGGACGAGGTACTGTGCCTCCGCCGCCGCCACCCCCTGGAGCACC
AAGAACAGGTGACCTGCCTCCAGCAGTGCAGACAGGACCATTACCACCAGGTGC
AGCCTCAAAACCACCTATCATCGAGGAAATTCTGCAACCAGAGTCACCGAGATC
GAAGGCATTGCGGGAAGCGAGAGGGAAAGCTCCAGCAACAATTCCAGATAGTA
GAGGGGTTGATACATCACAAATACCGAGTTTTACATCAGGTGGAGACCAAACAA
TGACACCAACCCCTCAAAGAACAAGCACTAGAGTGAGAGATAGAGATGTCAATG
CTGGTACGGTTGGAACTTTCACAGTGCCACGACTCCAGATAACACATAGTAAGA
AAAGAGCACCAATGGCAAATGGAAGAATAGTAGTCAATCTTGACCACTTGACAG
TCTATGACCCTGAACAAACAAGTCTTTCAAATACTCGAGCAACACAGGAACAAT
TTAATGCTTGGTACGAGGGTGTAAGGGAAGATTATGGAGTAAATGATGAGCAAA
TGGGGATATTGCTCAATGGGTTAATGGTTTGGTGCATCGAGAATGGAACATCCCC
GAATATTAATGGAATGTGGGTCATGATGGATGGTGATGAACAAGTTACATATCC
AATAAAACCTCTATTGGA

>SPLCSPV_PCR_product 324 bp
TTCGAGATAGGAGACCCAATAAGGACCCTCTGACGTTTGTACAGGCCTTTACCAT
GTATGATAACGAACCCACTACTGCTAAGATCCGAATGGATCTGAGAGATAGAAT
GCAAGTGTTGAAGAAGTTTTCTGTTACTGTATCTGGTGGCCCTTACAATCACAAG
GAGCAGGCTTTGGTTAGGAAGTTTTTTAATGGATTGTACAATCATGTTACCTATA
ATCACAAGGAAGAAGCTAAGTACGAAAACCATTTGGAAAATGCTTTGATGTTGT
ATTCAGCTAGCAGTCATGCTAGTAATCCTGTGTATCAGACTCTGCGTTGCA

>SPMaV_PCR_product 302 bp
CCCGATTTCAAGAAGCGTTCAGATGAAGAGAAAGAAGGGGGATCGAATCCCTAA
GGGATGTGTGGGTCCTTGTAAGGTCCAGGATTACGAGTTTAAGATGGATGTCCCA
CATAGTGGGACCTTTGTGTGTGTGTCTGATTTTACTAGAGGTACTGGGCTTACTC
ATCGTCTGGGTAAGCGTGTGTGTATTAAGTCCATGGGTATTGACGGTAAGGTCTG
GATGGATGACAACGTCGCTAAGAGGGATCACACTAACATAATTACTTACTGGTT
GATCCGAGATAGAAGGCCCAATAAGGACC

>SPBVA_PCR_product 593 bp
CCTGCTGAGGTGCTATATTCGTCCACAGAAGGGACAGAAAATCAGAGGGTGTAT
ATACACCGAAGTGAGGAAGAGATCACTTGCCTGGACAATCAGCAAGTGGATCTA
CCACTTATCACCCCTCAAAGTCATGCGCAGCTCCTACGACAAAACTACAGGTTTA
TTCATATCGGAGCAATCCAGGTTAGGGTGCAAGCCCTACACAGAACGCATGCAG
GGACCATGGTGTTGGTCCTAAATACTGATCGAAGGTGGAATGGGGACTTATCCCT
GTTCGGAGGAATTGAAGGTGACCTTACGGAAGGTGCCTTCATGACATATATCATC
CCCAACGTAACAATGACAGTTGAGGATTTCTGCCAGAATATTATGGTAGAATTCC
AGACTAGGGGATACGCTGAGTGGGTTCATGGGTCGAACCTACTAATCACTCGAG
GGATGGTAGGAAGGTTATCCAACACCCCAAATGTTGGATTTAACTATAACATCTC
AGCCGTTACTGATTATTTGGCTAGCAGAGGTGTACGAACCCTACCAGGAAGGCG
GTACAACACGGCTGATCTCCAAGGCCTAAGGTGGAACATACGGAGACC

>SPBVB_PCR_product 679 bp
TCATCAGGAGAATGAAAGGGGGTCTCCGAATAGAAGGACCAACAGTTACCTTCT
ATCGCAACGTGTCTACTATAGAGACACAGGAGAAATCAACTGTAGCGGCTGCCA
TTGGAAACATCAATGAAGAAAGGACCATGGTTTTCCCAAGGTTCAGAAAGGAGG
TAGCCAGACTAATTCAAGAGGGGTATATAGGAGACAACCCTCTTAGACATTGGT
CAAAAAATAAGGTGGAATGCACGCTCAGGATTAAAAACCCTGATCTAGTCATAC
AGGACCCACCTCTGAAGCATGTAACGCCAGCTGCAAGAGAGTTCTTCCAGAACC
AAGTAAGCAGCTTGATCAAGGCTCAGCTCATAAGGCCTTCAAGGAGCAGGCACA
GGACCACTGCCTTTATGGTTGAATCTGGAACCAGTGTGGATCCAAAAACAGGAA
AAGAGGTCAGAGGCAAGCAACGGATGGTGCTCAATTATAAACGCCTCAATGACA
ACACTGAAAAGGATCAATATTCCCTGCCAGGCATCAACACTATCATCAGCCGGG
TAGCAGGGAAGAAGGTCTTTTCTAAATTTGACCTCAAATCAGGCTTCCACCAGAT
CAGAATGAGTAGAGAGTCCATACCCTGGACTGCTTTCTGGACTCCTGATGGGCTT
TACGAATTCCTCGTAATGCCATTTGGGCT

Table A1.2: A list identified mature miRNA sequences present collectively in the NASPOT 1 and Blesbok cultivars sweet potato libraries

Name of miRNA	Sequence	Length (nt)	Reference miRNA
iba-miR1436	AAUGUGUUACUCCCUCCGUCCCAU	24	osa-miR1436
iba-miR1128	ACAAAUUUAAUACUACCUCCGUCC	24	tae-miR1128
iba-miR1446	UGUAUGAACUCUCUCCCUCAA	21	ptc-miR1446
iba-miR482	UUUCCUAUGCCUCCCAUUCCUAA	23	csi-miR482
iba-miR1511	AACCUGGCUCUGAUACCA	18	gma-miR1511
iba-miR1515	UCAUUUUUGCAUGCAGUGAUCC	22	csi-miR1515
iba-miR156	CUCUCUCUCUCUCUCUCU	18	ath-miR156
iba-miR157	UCUCUAUGCUUCUGUCAUCACC	22	aly-miR157
iba-miR159	UUGGAUUGAAGGGAGCUCUGCACC	24	osa-miR159
iba-miR160	UGCCUAGCUCCCUGUAUGCCA	21	ath-miR160
iba-miR162	CGAUAAACCUCUGCAUCCA	19	ath-miR162
iba-miR164	UGGAGAAGCAGGGCACGUGCA	21	ath-miR164
iba-miR166	UGAGGGGAAUGUUGUCUGGCU	21	gra-miR166
iba-miR167	UGAAGCUGCCAGCAUGAUCUGU	22	ath-miR167
iba-miR168	UUCACUUGGUGCAGGUCGGGA	21	ath-miR168
iba-miR169	UGCAGCCAAGGAUGACUUGC	20	ath-miR169
iba-miR171	UUGAGCCGCGUCAAUAUCUCU	21	stu-miR171
iba-miR1862	UGGGACGGAGGGAGUAUUAU	20	osa-miR1862
iba-miR2111	UAAUCUGCAUCCUGAGGUCUA	21	vvi-miR2111
iba-miR2118	UUUUUCCUAUGCCUCCCAUUCCU	23	bdi-miR2118
iba-miR2870	AACUUCGGAGUUCUGAUU	18	osa-miR2870
iba-miR319	AGUGAAUGAUGCGGGAGAUAG	21	sly-miR319
iba-miR3627	UUGUCGCAGGAGAGAAGGCAC	21	vvi-miR3627
iba-miR390	AAGCUCAGGAGGGAUAGCGCC	21	ath-miR390
iba-miR393	AAGGGAUCGCAUUGAUCC	18	ath-miR393
iba-miR394	UUGGCAUUCUGUCCACCUCC	20	ath-miR394
iba-miR395	CUGAAGUGUUUGGGGGAACUC	21	ath-miR395
iba-miR396	UUCUACAGCUUUCUUGAACUG	21	ath-miR396
iba-miR397	UUGAGUGCAGCGUUGAUGAGA	21	mdm-miR397
iba-miR398	UGUGUUCACAGGUCGCCCCUG	21	osa-miR398
iba-miR399	CGCCAAAGGAGAGUUGCCCUG	21	vvi-miR399
iba-miR403	UUUCGGGUUUGUGCGUGAAUC	21	ath-miR403

Table A1.2: A list identified mature miRNA sequences present collectively in the NASPOT 1 and Blesbok cultivars sweet potato libraries

Name of miRNA	Sequence	Length (nt)	Reference miRNA
iba-miR408	CUGCACUGCCUCUUCCCUGGCUU	23	osa-miR408
iba-miR4414	AUCCAACGAUGCAGGAGCUGU	21	mtr-miR4414
iba-miR477	ACUCUCCCUCAAGGGCUUCUAU	22	aqc-miR477
iba-miR4995	AUAGGCAGUGGCUUGGUUAAGGGAAC	26	gma-miR4995
iba-miR5020	AUGGCAUGAGAGAAGGGGAGAA	22	ath-miR5020
iba-miR5021	AAGAAGAAGAAGAAAAGGGAGA	22	ath-miR5021
iba-miR5072	UCCCCAGCAGAGUCGCCA	18	osa-miR5072
iba-miR5077	UUCUUCACGUCGGGUUCACCA	21	osa-miR5077
iba-miR5083	CAAUGGAUCCUUCUGAGCCUA	21	osa-miR5083
iba-miR530	UGCAUUUGCACCUGCACCUCC	21	tcc-miR530
iba-miR5368	CGGGCCGAGGGACAGUUUCAGGUAGAC	27	gma-miR5368
iba-miR5506	AUGGAUCGCUUCGUCUUCUGG	21	osa-miR5506
iba-miR5638	CAAUACUCUCCCACUUUUAAAAG	23	ath-miR5638
iba-miR5645	AAAGAAAAAAAGAAAAAAGAAA	22	ath-miR5645
iba-miR5666	CCAUGGGACAUCGAGCAGUUAA	22	ath-miR5666
iba-miR6024	CUACUCAUCUCUUUUCUUUUGG	22	sly-miR6024
iba-miR6164	ACUCCCUCCGUCUCAUUUUAUGUG	24	nta-miR6164
iba-miR6248	AUAAAAUGCUACUACCUCCGUCCC	24	osa-miR6248
iba-miR6300	UCGUUGUAGUAUAGUGGUGAGUAUUCC	27	gma-miR6300
iba-miR6476	GAGUUGCAGAUCAGUGGAGAUGAAAC	26	ptc-miR6476
iba-miR7695	GGAGAAAAAAGAAAGAAGAAA	21	osa-miR7695
iba-miR7817	CUCGGGAUAGAAAAAACAAAGG	22	ptc-miR7817
iba-miR8005	GUUUAGGGUUUAGGGUUUAGGGU	23	stu-miR8005
iba-miR8016	AUUUUUGAAUGGAAGACUCAUGUG	24	stu-miR8016
iba-miR8021	AGGCUCAAACUCAAGACCUUUGGU	24	stu-miR8021
iba-miR812	AUAUACUCCCUCCGUCCCAUUU	22	osa-miR812
iba-miR8175	GGUUCGUUCCCCGGCAACGGCGCCA	25	ath-miR8175
iba-miR827	UUAGAUGAUCAUCAACAAACU	21	ath-miR827

Table A2.1: Identification of conserved and non-conserved miRNA sequences from the NASPOT 1 infected plants. The raw and normalised counts of some of the identified miRNAs are represented

NASPOT 1 infected											
Sample ID	**miRNA name**	**Raw sequence count**	**Normalised sequence count**	**Sample ID**	**miRNA name**	**Raw sequence count**	**Normalised sequence count**	**Sample ID**	**miRNA name**	**Raw sequence count**	**Normalised sequence count**
T1 (520931)	**miR166**	**83721**	**160714**	**T2 (616350)**	**miR166**	**92376**	**149876**	**T3 (592373)**	**miR166**	**77953**	**131594**
	miR399	29	56		miR399	27	44		miR399	38	64
	miR393	213	409		miR393	534	866		miR393	449	758
	miR398	**17465**	**33527**		**miR398**	**10352**	**16796**		**miR398**	**14218**	**24002**
	miR482	**2593**	**4978**		**miR482**	**2978**	**4832**		**miR482**	**2635**	**4448**
	miR408	**870**	**1670**		**miR408**	**1250**	**2028**		**miR408**	**962**	**1624**
	miR168	**1606**	**3083**		**miR168**	**1306**	**2119**		**miR168**	**1655**	**2794**
	miR6300	**2784**	**5344**		**miR6300**	**2648**	**4296**		**miR6300**	**2242**	**3785**
	miR160	39	75		miR160	39	63		miR160	40	68
	miR162	50	96		miR162	72	117		miR162	74	125
	miR164	17	33		miR164	13	21		miR164	10	17
	miR167	25	48		miR167	49	80		miR167	45	76
	miR319	20	38		miR319	14	23		miR319	11	19
	miR397	31	60		miR397	84	136		miR397	52	88
	miR403	19	36		miR403	59	96		miR403	78	132
	miR530	15	29		miR530	14	23		miR530	29	49
	miR5077	34	65		miR5077	41	67		miR5077	21	35

Table A2.2: Identification of conserved and non-conserved miRNA sequences from the NASPOT 1 uninfected plants. The raw and normalised counts of some of the identified miRNAs are represented

NASPOT 1 uinfected											
Sample ID	**miRNA name**	**Raw sequence count**	**Normalised sequence count**	**Sample ID**	**miRNA name**	**Raw sequence count**	**Normalised sequence count**	**Sample ID**	**miRNA name**	**Raw sequence count**	**Normalised sequence count**
T4 (464662)	**miR166**	**77302**	**166362**	**T5 (696838)**	**miR166**	**96525**	**138519**	**T6 (712313)**	**miR166**	**66914**	**93939**
	miR399	40	86		miR399	20	29		miR399	42	59
	miR393	23	49		miR393	26	37		miR393	37	52
	miR398	**3648**	**7851**		**miR398**	**5433**	**7797**		**miR398**	**5410**	**7595**
	miR482	**1890**	**4067**		**miR482**	**2753**	**3951**		**miR482**	**2244**	**3150**
	miR408	**933**	**2008**		**miR408**	**943**	**1353**		**miR408**	**1247**	**1751**
	miR168	75	161		miR168	131	188		miR168	152	213
	miR6300	**2214**	**4765**		**miR6300**	**3384**	**4856**		**miR6300**	**6238**	**8757**
	miR160	60	129		miR160	50	72		miR160	76	107
	miR162	36	77		miR162	29	42		miR162	22	31
	miR164	6	13		miR164	16	23		miR164	23	32
	miR167	38	82		miR167	29	42		miR167	24	34
	miR319	17	37		miR319	14	20		miR319	8	11
	miR397	57	123		miR397	45	65		miR397	75	105
	miR403	161	346		miR403	169	243		miR403	103	145
	miR530	12	26		miR530	40	57		miR530	22	31
	miR5077	16	34		miR5077	24	34		miR5077	27	38

Table A2.3: Identification of conserved and non-conserved miRNA sequences from the Blesbok infected plants. The raw and normalised counts of some of the identified miRNAs are represented

Blesbok infected											
Sample ID	**miRNA name**	**Raw sequence count**	**Normalised sequence count**	**Sample ID**	**miRNA name**	**Raw sequence count**	**Normalised sequence count**	**Sample ID**	**miRNA name**	**Raw sequence count**	**Normalised sequence count**
T7 (626574)	**miR166**	**96010**	**153230**	**T8 (689569)**	**miR166**	**108461**	**157288**	**T9 (816636)**	**miR166**	**129789**	**158931**
	miR399	26	41		miR399	38	55		miR399	42	51
	miR393	**1448**	**2311**		**miR393**	**1734**	**2515**		**miR393**	**1629**	**1995**
	miR398	**50685**	**80892**		**miR398**	**37847**	**54885**		**miR398**	**22821**	**27945**
	miR482	**4604**	**7348**		**miR482**	**3075**	**4459**		**miR482**	**3195**	**3912**
	miR408	**1101**	**1757**		**miR408**	**2003**	**2905**		**miR408**	**1395**	**1708**
	miR168	**3367**	**5374**		**miR168**	**2690**	**3901**		**miR168**	**2556**	**3130**
	miR6300	**5103**	**8144**		**miR6300**	**2178**	**3158**		**miR6300**	**6617**	**8103**
	miR160	71	113		miR160	61	88		miR160	113	138
	miR162	158	252		miR162	145	210		miR162	134	164
	miR164	75	120		miR164	29	42		miR164	49	60
	miR167	110	176		miR167	142	206		miR167	151	185
	miR319	21	34		miR319	15	22		miR319	22	27
	miR397	214	342		miR397	224	325		miR397	174	213
	miR403	84	134		miR403	166	241		miR403	177	217
	miR530	11	18		miR530	5	7		miR530	13	16
	miR5077	17	27		miR5077	17	25		miR5077	35	43

Table A2.4: Identification of conserved and non-conserved miRNA sequences from the Blesbok uninfected plants. The raw and normalised counts of some of the identified miRNAs are represented

Blesbok uinfected											
Sample ID	**miRNA name**	**Raw sequence count**	**Normalised sequence count**	**Sample ID**	**miRNA name**	**Raw sequence count**	**Normalised sequence count**	**Sample ID**	**miRNA name**	**Raw sequence count**	**Normalised sequence count**
T10 (560794)	**miR166**	**55449**	**98876**	**T11 (796783)**	**miR166**	**76428**	**95921**	**T12 (545726)**	**miR166**	**66601**	**122041**
	miR399	70	125		miR399	31	39		miR399	49	90
	miR393	22	39		miR393	21	26		miR393	19	35
	miR398	**7260**	**12946**		**miR398**	**7723**	**9693**		**miR398**	**5920**	**10848**
	miR482	**1843**	**3286**		**miR482**	**3005**	**3771**		**miR482**	**2631**	**4821**
	miR408	**1190**	**2122**		**miR408**	**1040**	**1305**		**miR408**	**944**	**1730**
	miR168	93	166		miR168	827	1038		miR168	93	170
	miR6300	**5289**	**9431**		**miR6300**	**4718**	**5921**		**miR6300**	**4539**	**8317**
	miR160	103	184		miR160	132	166		miR160	117	214
	miR162	37	66		miR162	30	38		miR162	19	35
	miR164	30	53		miR164	18	23		miR164	29	53
	miR167	35	62		miR167	34	43		miR167	39	71
	miR319	27	48		miR319	50	63		miR319	28	51
	miR397	47	84		miR397	33	41		miR397	42	77
	miR403	172	307		miR403	109	137		miR403	132	242
	miR530	34	61		miR530	65	82		miR530	53	97
	miR5077	37	66		miR5077	87	109		miR5077	44	81

Table A3.1: Differentially expressed miRNAs between NASPOT 1 infected and NASPOT 1 uninfected cultivar. Fold change and log2 fold change values calculated on the normalised miRNA counts

miRNA name	Fold change (3reps)	Log2 fold change
miR166	0,966049939	-0,049830325
	1,081988752	0,113685502
	1,400845229	0,48629757
miR399	0,651162791	-0,618909833
	1,517241379	0,601450624
	1,084745763	0,117356951
miR393	8,346938776	3,061247189
	23,40540541	4,548769849
	14,57692308	3,86561432
miR398	4,270411413	2,094375066
	2,154161857	1,107126654
	3,160236998	1,660032756
miR482	1,223998033	0,29160124
	1,222981524	0,290402608
	1,412063492	0,497804959
miR408	0,831673307	-0,265911167
	1,498891353	0,583895813
	0,927470017	-0,108627451
miR168	19,14906832	4,259202296
	11,2712766	3,494579021
	13,11737089	3,713406685
miR6300	1,121511018	0,165443794
	0,884678748	-0,176774428
	0,432225648	-1,210143411
miR160	0,581395349	-0,782408565
	0,875	-0,192645078
	0,635514019	-0,654004145
miR162	1,246753247	0,31817596
	2,785714286	1,478047297
	4,032258065	2,011587974
miR164	2,538461538	1,343954401
	0,913043478	-0,131244533
	0,53125	-0,912537159
miR167	0,585365854	-0,772589504
	1,904761905	0,929610672
	2,235294118	1,160464672
miR319	1,027027027	0,038474148
	1,15	0,201633861
	1,727272727	0,788495895
miR397	0,487804878	-1,03562391
	2,092307692	1,065095028
	0,838095238	-0,254813899
miR403	0,104046243	-3,264703226
	0,395061728	-1,339850003
	0,910344828	-0,135514971
miR530	1,115384615	0,157541277
	0,403508772	-1,309328058
	1,580645161	0,660513534
miR5077	1,911764706	0,934904972
	1,970588235	0,978626349
	0,921052632	-0,118644496

Table A3.2: Differentially expressed miRNAs between Blesbok infected and Blesbok uninfected cultivar. Fold change and log2 fold change values calculated on the normalised miRNA counts

miRNA name	Fold change (3reps)	Log2 fold change
miR166	1,54971884	0,632006496
	1,639766057	0,713490003
	1,302275465	0,381034648
miR399	0,328	-1,60823228
	1,41025641	0,495957495
	0,566666667	-0,819427754
miR393	59,25641026	5,888899326
	96,73076923	6,595902967
	57	5,832890014
miR398	6,248416499	2,643490622
	5,662333643	2,50139676
	2,576050885	1,365161091
miR482	2,236153378	1,161019146
	1,182444975	0,241773049
	0,811449907	-0,301426059
miR408	0,82799246	-0,272310465
	2,22605364	1,154488357
	0,987283237	-0,018464063
miR168	32,37349398	5,016741175
	3,758188825	1,910037555
	18,41176471	4,202556006
miR6300	0,86353515	-0,211673191
	0,533355852	-0,906829683
	0,974269568	-0,037607091
miR160	0,614130435	-0,703382994
	0,530120482	-0,915607813
	0,644859813	-0,63294253
miR162	3,818181818	1,932885804
	5,526315789	2,466318004
	4,685714286	2,228268988
miR164	2,264150943	1,178970141
	1,826086957	0,868755467
	1,132075472	0,178970141
miR167	2,838709677	1,505235308
	4,790697674	2,260235772
	2,605633803	1,381634341
miR319	0,708333333	-0,497499659
	0,349206349	-1,517848305
	0,529411765	-0,91753784
miR397	4,071428571	2,025535092
	7,926829268	2,986743903
	2,766233766	1,46792308
miR403	0,436482085	-1,196005655
	1,759124088	0,814857253
	0,896694215	-0,157312005
miR530	0,295081967	-1,760812336
	0,085365854	-3,550197083
	0,164948454	-2,599912842
miR5077	0,409090909	-1,289506617
	0,229357798	-2,124328135
	0,530864198	-0,913585248

Table A4: Delta Ct calculations of miR482

miR482														
BT1 vs. BC1	**Treated Sample Rep1**	**Treated Sample Rep2**	**Treated Sample Rep3**	**Mock Sample Rep1**	**Mock Sample Rep2**	**Mock Sample Rep3**	**Average Treated Ct Value**	**Average Treated Ct Value**	**Average Control Ct Value**	**Average Control Ct Value**	**ΔCt Value (Treated)**	**ΔCt Value (Mock)**	**Delta Delta Ct Value**	**Expression Fold Change**
	Raw Ct Value	**Raw Ct Value**	**Raw Ct Value**	**Raw Ct Value**	**Raw Ct Value**	**Raw Ct Value**	**GOI Treated**	**REF Treated**	**GOI Mock**	**REF Mock**	**ΔCTT**	**ΔCTM**	**ΔΔCt**	**2^-ΔΔCt**
Reference Gene (REF)	24.90	24.71	25.22	26.48	26.78	26.61	-	24.94	-	26.62	-8.76	-11.12	2.36	0.194791
miRNA (Gene Of Interest - GOI)	16.15	16.03	16.36	15.39	15.25	15.86	16.18	-	15.5	-				

miR482 Average Expression Fold Change - Blesbok	**miR482 Average Expression Fold Change - NASPOT 1**
0.37	1.05

miR482														
BT2 vs. BC2	**Treated Sample Rep1**	**Treated Sample Rep2**	**Treated Sample Rep3**	**Mock Sample Rep1**	**Mock Sample Rep2**	**Mock Sample Rep3**	**Average Treated Ct Value**	**Average Treated Ct Value**	**Average Control Ct Value**	**Average Control Ct Value**	**ΔCt Value (Treated)**	**ΔCt Value (Mock)**	**Delta Delta Ct Value**	**Expression Fold Change**
	Raw Ct Value	**Raw Ct Value**	**Raw Ct Value**	**Raw Ct Value**	**Raw Ct Value**	**Raw Ct Value**	**GOI Treated**	**REF Treated**	**GOI Mock**	**REF Mock**	**ΔCTT**	**ΔCTM**	**ΔΔCt**	**2^-ΔΔCt**
Reference Gene (REF)	25.81	26.03	26.04	25.21	25.44	25.46	-	25.96	-	25.37	-9.68	-10.33	0.64	0.64
miRNA (Gene Of Interest - GOI)	16.39	16.15	16.29	15.20	14.78	15.15	16.28	-	15.04	-				

miR482														
BT3 vs. BC3	**Treated Sample Rep1**	**Treated Sample Rep2**	**Treated Sample Rep3**	**Mock Sample Rep1**	**Mock Sample Rep2**	**Mock Sample Rep3**	**Average Treated Ct Value**	**Average Treated Ct Value**	**Average Control Ct Value**	**Average Control Ct Value**	**ΔCt Value (Treated)**	**ΔCt Value (Mock)**	**Delta Delta Ct Value**	**Expression Fold Change**
	Raw Ct Value	**Raw Ct Value**	**Raw Ct Value**	**Raw Ct Value**	**Raw Ct Value**	**Raw Ct Value**	**GOI Treated**	**REF Treated**	**GOI Mock**	**REF Mock**	**ΔCTT**	**ΔCTM**	**ΔΔCt**	**2^-ΔΔCt**
Reference Gene (REF)	23.79	23.87	-	24.87	24.91	25.09	-	23.83	-	24.96	-7.73	-9.64	1.91	0.27
miRNA (Gene Of Interest - GOI)	15.94	16.26	-	15.28	15.32	15.36	16.10	-	15.32	-				

miR482														
NT1 vs. NC1	**Treated Sample Rep1**	**Treated Sample Rep2**	**Treated Sample Rep3**	**Mock Sample Rep1**	**Mock Sample Rep2**	**Mock Sample Rep3**	**Average Treated Ct Value**	**Average Treated Ct Value**	**Average Control Ct Value**	**Average Control Ct Value**	**ΔCt Value (Treated)**	**ΔCt Value (Mock)**	**Delta Delta Ct Value**	**Expression Fold Change**
	Raw Ct Value	**Raw Ct Value**	**Raw Ct Value**	**Raw Ct Value**	**Raw Ct Value**	**Raw Ct Value**	**GOI Treated**	**REF Treated**	**GOI Mock**	**REF Mock**	**ΔCTT**	**ΔCTM**	**ΔΔCt**	**2^-ΔΔCt**
Reference Gene (REF)	24.18	24.64	24.33	25.24	24.77	24.88	-	24.38	-	24.96	-8.40	-10.36	1.96	0.26
miRNA (Gene Of Interest - GOI)	15.97	15.77	16.20	14.83	14.22	14.76	15.98	-	14.60	-				

miR482														
NT2 vs. NC2	**Treated Sample Rep1**	**Treated Sample Rep2**	**Treated Sample Rep3**	**Mock Sample Rep1**	**Mock Sample Rep2**	**Mock Sample Rep3**	**Average Treated Ct Value**	**Average Treated Ct Value**	**Average Control Ct Value**	**Average Control Ct Value**	**ΔCt Value (Treated)**	**ΔCt Value (Mock)**	**Delta Delta Ct Value**	**Expression Fold Change**
	Raw Ct Value	**Raw Ct Value**	**Raw Ct Value**	**Raw Ct Value**	**Raw Ct Value**	**Raw Ct Value**	**GOI Treated**	**REF Treated**	**GOI Mock**	**REF Mock**	**ΔCTT**	**ΔCTM**	**ΔΔCt**	**2^-ΔΔCt**
Reference Gene (REF)	24.08	24.12	24.30	24.45	24.11	24.34	-	24.17	-	24.30	-8.97	-9.31	0.33	0.79
miRNA (Gene Of Interest - GOI)	14.91	15.18	15.49	14.75	15.12	15.11	15.19	-	14.99	-				

miR482														
NT3 vs. NC3	**Treated Sample Rep1**	**Treated Sample Rep2**	**Treated Sample Rep3**	**Mock Sample Rep1**	**Mock Sample Rep2**	**Mock Sample Rep3**	**Average Treated Ct Value**	**Average Treated Ct Value**	**Average Control Ct Value**	**Average Control Ct Value**	**ΔCt Value (Treated)**	**ΔCt Value (Mock)**	**Delta Delta Ct Value**	**Expression Fold Change**
	Raw Ct Value	**Raw Ct Value**	**Raw Ct Value**	**Raw Ct Value**	**Raw Ct Value**	**Raw Ct Value**	**GOI Treated**	**REF Treated**	**GOI Mock**	**REF Mock**	**ΔCTT**	**ΔCTM**	**ΔΔCt**	**2^-ΔΔCt**
Reference Gene (REF)	25.54	25.55	25.94	24.18	24.48	24.29	-	25.68	-	24.32	-9.82	-8.75	-1.07	2.10
miRNA (Gene Of Interest - GOI)	15.65	15.88	16.03	15.50	15.57	15.62	15.85	-	15.56	-				

Table A5: Delta Ct calculations of miR393

miR393														
BT1 vs. BC1	**Treated Sample Rep1**	**Treated Sample Rep2**	**Treated Sample Rep3**	**Mock Sample Rep1**	**Mock Sample Rep2**	**Mock Sample Rep3**	**Average Treated Ct Value**	**Average Treated Ct Value**	**Average Control Ct Value**	**Average Control Ct Value**	**ΔCt Value (Treated)**	**ΔCt Value (Mock)**	**Delta Delta Ct Value**	**Expression Fold Change**
	Raw Ct Value	**Raw Ct Value**	**Raw Ct Value**	**Raw Ct Value**	**Raw Ct Value**	**Raw Ct Value**	**GOI Treated**	**REF Treated**	**GOI Mock**	**REF Mock**	**ΔCTT**	**ΔCTM**	**ΔΔCt**	**2^-ΔΔCt**
Reference Gene (REF)	24.90	24.71	25.22	26.48	26.78	26.61	-	24.94	-	26.62	-1.16	-4.03	2.87	0.14
miRNA (Gene Of Interest - GOI)	23.89	23.68	23.79	22.71	22.50	22.57	23.79	-	22.59	-				

miR393 Average Expression Fold Change - Blesbok	**miR393 Average Expression Fold Change - NASPOT 1**
0.56	0.78

miR393														
BT2 vs. BC2	**Treated Sample Rep1**	**Treated Sample Rep2**	**Treated Sample Rep3**	**Mock Sample Rep1**	**Mock Sample Rep2**	**Mock Sample Rep3**	**Average Treated Ct Value**	**Average Treated Ct Value**	**Average Control Ct Value**	**Average Control Ct Value**	**ΔCt Value (Treated)**	**ΔCt Value (Mock)**	**Delta Delta Ct Value**	**Expression Fold Change**
	Raw Ct Value	**Raw Ct Value**	**Raw Ct Value**	**Raw Ct Value**	**Raw Ct Value**	**Raw Ct Value**	**GOI Treated**	**REF Treated**	**GOI Mock**	**REF Mock**	**ΔCTT**	**ΔCTM**	**ΔΔCt**	**2^-ΔΔCt**
Reference Gene (REF)	25.81	26.03	26.04	25.21	25.44	25.46	-	25.96	-	25.37	-1.79	-1.76	-0.02	1.02
miRNA (Gene Of Interest - GOI)	24.33	24.10	24.09	23.70	23.54	23.58	24.17	-	23.61	-				

miR393														
BT3 vs. BC3	**Treated Sample Rep1**	**Treated Sample Rep2**	**Treated Sample Rep3**	**Mock Sample Rep1**	**Mock Sample Rep2**	**Mock Sample Rep3**	**Average Treated Ct Value**	**Average Treated Ct Value**	**Average Control Ct Value**	**Average Control Ct Value**	**ΔCt Value (Treated)**	**ΔCt Value (Mock)**	**Delta Delta Ct Value**	**Expression Fold Change**
	Raw Ct Value	**Raw Ct Value**	**Raw Ct Value**	**Raw Ct Value**	**Raw Ct Value**	**Raw Ct Value**	**GOI Treated**	**REF Treated**	**GOI Mock**	**REF Mock**	**ΔCTT**	**ΔCTM**	**ΔΔCt**	**2^-ΔΔCt**
Reference Gene (REF)	23.79	23.87	23.93	24.87	24.91	25.09	-	23.86	-	24.96	-0.24	-1.17	0.93	0.52
miRNA (Gene Of Interest - GOI)	23.69	23.53	23.65	23.83	23.62	23.91	23.62	-	23.79	-				

miR393														
NT1 vs. NC1	**Treated Sample Rep1**	**Treated Sample Rep2**	**Treated Sample Rep3**	**Mock Sample Rep1**	**Mock Sample Rep2**	**Mock Sample Rep3**	**Average Treated Ct Value**	**Average Treated Ct Value**	**Average Control Ct Value**	**Average Control Ct Value**	**ΔCt Value (Treated)**	**ΔCt Value (Mock)**	**Delta Delta Ct Value**	**Expression Fold Change**
	Raw Ct Value	**Raw Ct Value**	**Raw Ct Value**	**Raw Ct Value**	**Raw Ct Value**	**Raw Ct Value**	**GOI Treated**	**REF Treated**	**GOI Mock**	**REF Mock**	**ΔCTT**	**ΔCTM**	**ΔΔCt**	**2^-ΔΔCt**
Reference Gene (REF)	24.18	24.64	24.33	25.24	24.77	24.88	-	24.38	-	24.96	-0.14	-1.15	1.01	0.50
miRNA (Gene Of Interest - GOI)	24.19	24.27	24.28	23.81	23.81	23.83	24.25	-	23.82	-				

miR393														
NT2 vs. NC2	**Treated Sample Rep1**	**Treated Sample Rep2**	**Treated Sample Rep3**	**Mock Sample Rep1**	**Mock Sample Rep2**	**Mock Sample Rep3**	**Average Treated Ct Value**	**Average Treated Ct Value**	**Average Control Ct Value**	**Average Control Ct Value**	**ΔCt Value (Treated)**	**ΔCt Value (Mock)**	**Delta Delta Ct Value**	**Expression Fold Change**
	Raw Ct Value	**Raw Ct Value**	**Raw Ct Value**	**Raw Ct Value**	**Raw Ct Value**	**Raw Ct Value**	**GOI Treated**	**REF Treated**	**GOI Mock**	**REF Mock**	**ΔCTT**	**ΔCTM**	**ΔΔCt**	**2^-ΔΔCt**
Reference Gene (REF)	24.08	24.12	24.30	24.45	24.11	24.34	-	24.17	-	24.30	0.41	0.19	0.22	0.86
miRNA (Gene Of Interest - GOI)	24.49	24.54	24.69	24.50	24.25	24.72	24.57	-	24.49	-				

miR393														
NT3 vs. NC3	**Treated Sample Rep1**	**Treated Sample Rep2**	**Treated Sample Rep3**	**Mock Sample Rep1**	**Mock Sample Rep2**	**Mock Sample Rep3**	**Average Treated Ct Value**	**Average Treated Ct Value**	**Average Control Ct Value**	**Average Control Ct Value**	**ΔCt Value (Treated)**	**ΔCt Value (Mock)**	**Delta Delta Ct Value**	**Expression Fold Change**
	Raw Ct Value	**Raw Ct Value**	**Raw Ct Value**	**Raw Ct Value**	**Raw Ct Value**	**Raw Ct Value**	**GOI Treated**	**REF Treated**	**GOI Mock**	**REF Mock**	**ΔCTT**	**ΔCTM**	**ΔΔCt**	**2^-ΔΔCt**
Reference Gene (REF)	25.54	25.55	25.94	24.18	24.48	24.29	-	25.68	-	24.32	-0.80	-0.81	0.00	1.00
miRNA (Gene Of Interest - GOI)	24.74	24.94	24.94	23.57	23.41	23.55	24.87	-	23.51	-				

Table A6: Delta Ct calculations of miR397

miR397														
BT1 vs. BC1	**Treated Sample Rep1**	**Treated Sample Rep2**	**Treated Sample Rep3**	**Mock Sample Rep1**	**Mock Sample Rep2**	**Mock Sample Rep3**	**Average Treated Ct Value**	**Average Treated Ct Value**	**Average Control Ct Value**	**Average Control Ct Value**	**ΔCt Value (Treated)**	**ΔCt Value (Mock)**	**Delta Delta Ct Value**	**Expression Fold Change**
	Raw Ct Value	**Raw Ct Value**	**Raw Ct Value**	**Raw Ct Value**	**Raw Ct Value**	**Raw Ct Value**	**GOI Treated**	**REF Treated**	**GOI Mock**	**REF Mock**	**ΔCTT**	**ΔCTM**	**ΔΔCt**	**2^-ΔΔCt**
Reference Gene (REF)	24.90	24.71	25.22	26.48	26.78	26.61	-	24.94	-	26.62	-4.47	-6.66	2.19	0.22
miRNA (Gene Of Interest - GOI)	20.60	20.73	20.09	19.87	19.92	20.09	20.47	-	19.96	-				

miR397 Average Expression Fold Change - Blesbok	**miR397 Average Expression Fold Change - NASPOT 1**
1.21	0.91

miR397														
BT2 vs. BC2	**Treated Sample Rep1**	**Treated Sample Rep2**	**Treated Sample Rep3**	**Mock Sample Rep1**	**Mock Sample Rep2**	**Mock Sample Rep3**	**Average Treated Ct Value**	**Average Treated Ct Value**	**Average Control Ct Value**	**Average Control Ct Value**	**ΔCt Value (Treated)**	**ΔCt Value (Mock)**	**Delta Delta Ct Value**	**Expression Fold Change**
	Raw Ct Value	**Raw Ct Value**	**Raw Ct Value**	**Raw Ct Value**	**Raw Ct Value**	**Raw Ct Value**	**GOI Treated**	**REF Treated**	**GOI Mock**	**REF Mock**	**ΔCTT**	**ΔCTM**	**ΔΔCt**	**2^-ΔΔCt**
Reference Gene (REF)	25.81	26.03	26.04	25.21	25.44	25.46	-	25.96	-	25.37	-6.01	-4.53	-1.49	2.80
miRNA (Gene Of Interest - GOI)	19.84	19.92	20.08	20.84	20.90	20.79	19.95	-	20.84	-				

miR397														
BT3 vs. BC3	**Treated Sample Rep1**	**Treated Sample Rep2**	**Treated Sample Rep3**	**Mock Sample Rep1**	**Mock Sample Rep2**	**Mock Sample Rep3**	**Average Treated Ct Value**	**Average Treated Ct Value**	**Average Control Ct Value**	**Average Control Ct Value**	**ΔCt Value (Treated)**	**ΔCt Value (Mock)**	**Delta Delta Ct Value**	**Expression Fold Change**
	Raw Ct Value	**Raw Ct Value**	**Raw Ct Value**	**Raw Ct Value**	**Raw Ct Value**	**Raw Ct Value**	**GOI Treated**	**REF Treated**	**GOI Mock**	**REF Mock**	**ΔCTT**	**ΔCTM**	**ΔΔCt**	**2^-ΔΔCt**
Reference Gene (REF)	23.79	23.87	23.93	24.87	24.91	25.09	-	23.86	-	24.96	-2.96	-3.66	0.69	0.62
miRNA (Gene Of Interest - GOI)	20.74	20.94	21.02	21.39	21.22	21.29	20.90	-	21.30	-				

miR397														
NT1 vs. NC1	**Treated Sample Rep1**	**Treated Sample Rep2**	**Treated Sample Rep3**	**Mock Sample Rep1**	**Mock Sample Rep2**	**Mock Sample Rep3**	**Average Treated Ct Value**	**Average Treated Ct Value**	**Average Control Ct Value**	**Average Control Ct Value**	**ΔCt Value (Treated)**	**ΔCt Value (Mock)**	**Delta Delta Ct Value**	**Expression Fold Change**
	Raw Ct Value	**Raw Ct Value**	**Raw Ct Value**	**Raw Ct Value**	**Raw Ct Value**	**Raw Ct Value**	**GOI Treated**	**REF Treated**	**GOI Mock**	**REF Mock**	**ΔCTT**	**ΔCTM**	**ΔΔCt**	**2^-ΔΔCt**
Reference Gene (REF)	24.18	24.64	24.33	25.24	24.77	24.88	-	24.38333333	-	24.96333333	-3.113333333	-4.876666667	1.763333333	0.294566785
miRNA (Gene Of Interest - GOI)	21.59	21.02	21.2	20.09	20.09	20.08	21.27	-	20.08666667					

miR397														
NT2 vs. NC2	**Treated Sample Rep1**	**Treated Sample Rep2**	**Treated Sample Rep3**	**Mock Sample Rep1**	**Mock Sample Rep2**	**Mock Sample Rep3**	**Average Treated Ct Value**	**Average Treated Ct Value**	**Average Control Ct Value**	**Average Control Ct Value**	**ΔCt Value (Treated)**	**ΔCt Value (Mock)**	**Delta Delta Ct Value**	**Expression Fold Change**
	Raw Ct Value	**Raw Ct Value**	**Raw Ct Value**	**Raw Ct Value**	**Raw Ct Value**	**Raw Ct Value**	**GOI Treated**	**REF Treated**	**GOI Mock**	**REF Mock**	**ΔCTT**	**ΔCTM**	**ΔΔCt**	**2^-ΔΔCt**
Reference Gene (REF)	24.08	24.12	24.30	24.45	24.11	24.34	-	24.17	-	24.30	-3.60	-3.26	-0.34	1.27
miRNA (Gene Of Interest - GOI)	20.58	20.61	20.51	20.86	21.09	21.17	20.57	-	21.04	-				

miR397														
NT3 vs. NC3	**Treated Sample Rep1**	**Treated Sample Rep2**	**Treated Sample Rep3**	**Mock Sample Rep1**	**Mock Sample Rep2**	**Mock Sample Rep3**	**Average Treated Ct Value**	**Average Treated Ct Value**	**Average Control Ct Value**	**Average Control Ct Value**	**ΔCt Value (Treated)**	**ΔCt Value (Mock)**	**Delta Delta Ct Value**	**Expression Fold Change**
	Raw Ct Value	**Raw Ct Value**	**Raw Ct Value**	**Raw Ct Value**	**Raw Ct Value**	**Raw Ct Value**	**GOI Treated**	**REF Treated**	**GOI Mock**	**REF Mock**	**ΔCTT**	**ΔCTM**	**ΔΔCt**	**2^-ΔΔCt**
Reference Gene (REF)	25.54	25.55	25.94	24.18	24.48	24.29	-	25.68	-	24.32	-4.24	-4.01	-0.23	1.17
miRNA (Gene Of Interest - GOI)	21.56	21.33	21.43	20.31	20.27	20.34	21.44	-	20.31	-				

Table A7: Delta Ct calculations of miR168

miR168														
BT1 vs. BC1	**Treated Sample Rep1**	**Treated Sample Rep2**	**Treated Sample Rep3**	**Mock Sample Rep1**	**Mock Sample Rep2**	**Mock Sample Rep3**	**Average Treated Ct Value**	**Average Treated Ct Value**	**Average Control Ct Value**	**Average Control Ct Value**	**ΔCt Value (Treated)**	**ΔCt Value (Mock)**	**Delta Delta Ct Value**	**Expression Fold Change**
	Raw Ct Value	**Raw Ct Value**	**Raw Ct Value**	**Raw Ct Value**	**Raw Ct Value**	**Raw Ct Value**	**GOI Treated**	**REF Treated**	**GOI Mock**	**REF Mock**	**ΔCTT**	**ΔCTM**	**ΔΔCt**	**2^-ΔΔCt**
Reference Gene (REF)	24.90	24.71	25.22	26.48	26.78	26.61	-	24.94	-	26.62	-12.00	-10.89	-1.11	2.16
miRNA (Gene Of Interest - GOI)	12.96	12.81	13.05	15.89	15.63	15.67	12.94	-	15.73	-				

miR168 Average Expression Fold Change - Blesbok	**miR168 Average Expression Fold Change - NASPOT 1**
6.52	8.14

miR168														
BT2 vs. BC2	**Treated Sample Rep1**	**Treated Sample Rep2**	**Treated Sample Rep3**	**Mock Sample Rep1**	**Mock Sample Rep2**	**Mock Sample Rep3**	**Average Treated Ct Value**	**Average Treated Ct Value**	**Average Control Ct Value**	**Average Control Ct Value**	**ΔCt Value (Treated)**	**ΔCt Value (Mock)**	**Delta Delta Ct Value**	**Expression Fold Change**
	Raw Ct Value	**Raw Ct Value**	**Raw Ct Value**	**Raw Ct Value**	**Raw Ct Value**	**Raw Ct Value**	**GOI Treated**	**REF Treated**	**GOI Mock**	**REF Mock**	**ΔCTT**	**ΔCTM**	**ΔΔCt**	**2^-ΔΔCt**
Reference Gene (REF)	25.81	26.03	26.04	25.21	25.44	25.46	-	25.96	-	25.37	-12.89	-9.13	-3.76	13.55
miRNA (Gene Of Interest - GOI)	13.14	13.02	13.05	16.45	16.17	16.10	13.07	-	16.24	-				

miR168														
BT3 vs. BC3	**Treated Sample Rep1**	**Treated Sample Rep2**	**Treated Sample Rep3**	**Mock Sample Rep1**	**Mock Sample Rep2**	**Mock Sample Rep3**	**Average Treated Ct Value**	**Average Treated Ct Value**	**Average Control Ct Value**	**Average Control Ct Value**	**ΔCt Value (Treated)**	**ΔCt Value (Mock)**	**Delta Delta Ct Value**	**Expression Fold Change**
	Raw Ct Value	**Raw Ct Value**	**Raw Ct Value**	**Raw Ct Value**	**Raw Ct Value**	**Raw Ct Value**	**GOI Treated**	**REF Treated**	**GOI Mock**	**REF Mock**	**ΔCTT**	**ΔCTM**	**ΔΔCt**	**2^-ΔΔCt**
Reference Gene (REF)	23.79	23.87	23.93	24.87	24.91	25.09	-	23.86	-	24.96	-10.59	-8.64	-1.95	3.85
miRNA (Gene Of Interest - GOI)	13.31	13.27	13.24	16.43	16.20	16.31	13.27	-	16.31	-				

miR168														
NT1 vs. NC1	**Treated Sample Rep1**	**Treated Sample Rep2**	**Treated Sample Rep3**	**Mock Sample Rep1**	**Mock Sample Rep2**	**Mock Sample Rep3**	**Average Treated Ct Value**	**Average Treated Ct Value**	**Average Control Ct Value**	**Average Control Ct Value**	**ΔCt Value (Treated)**	**ΔCt Value (Mock)**	**Delta Delta Ct Value**	**Expression Fold Change**
	Raw Ct Value	**Raw Ct Value**	**Raw Ct Value**	**Raw Ct Value**	**Raw Ct Value**	**Raw Ct Value**	**GOI Treated**	**REF Treated**	**GOI Mock**	**REF Mock**	**ΔCTT**	**ΔCTM**	**ΔΔCt**	**2^-ΔΔCt**
Reference Gene (REF)	24.18	24.64	24.33	25.24	24.77	24.88	-	24.38	-	24.96	-11.12	-8.77	-2.36	5.12
miRNA (Gene Of Interest - GOI)	13.35	13.33	13.10	16.19	16.16	16.24	13.26	-	16.20	-				

miR168														
NT2 vs. NC2	**Treated Sample Rep1**	**Treated Sample Rep2**	**Treated Sample Rep3**	**Mock Sample Rep1**	**Mock Sample Rep2**	**Mock Sample Rep3**	**Average Treated Ct Value**	**Average Treated Ct Value**	**Average Control Ct Value**	**Average Control Ct Value**	**ΔCt Value (Treated)**	**ΔCt Value (Mock)**	**Delta Delta Ct Value**	**Expression Fold Change**
	Raw Ct Value	**Raw Ct Value**	**Raw Ct Value**	**Raw Ct Value**	**Raw Ct Value**	**Raw Ct Value**	**GOI Treated**	**REF Treated**	**GOI Mock**	**REF Mock**	**ΔCTT**	**ΔCTM**	**ΔΔCt**	**2^-ΔΔCt**
Reference Gene (REF)	24.08	24.12	24.30	24.45	24.11	24.34	-	24.17	-	24.30	-9.94	-6.87	-3.07	8.38
miRNA (Gene Of Interest - GOI)	14.26	14.18	14.24	17.48	17.40	17.40	14.23	-	17.43	-				

miR168														
NT1 vs. NC3	**Treated Sample Rep1**	**Treated Sample Rep2**	**Treated Sample Rep3**	**Mock Sample Rep1**	**Mock Sample Rep2**	**Mock Sample Rep3**	**Average Treated Ct Value**	**Average Treated Ct Value**	**Average Control Ct Value**	**Average Control Ct Value**	**ΔCt Value (Treated)**	**ΔCt Value (Mock)**	**Delta Delta Ct Value**	**Expression Fold Change**
	Raw Ct Value	**Raw Ct Value**	**Raw Ct Value**	**Raw Ct Value**	**Raw Ct Value**	**Raw Ct Value**	**GOI Treated**	**REF Treated**	**GOI Mock**	**REF Mock**	**ΔCTT**	**ΔCTM**	**ΔΔCt**	**2^-ΔΔCt**
Reference Gene (REF)	25.54	25.55	25.94	24.18	24.48	24.29	-	25.68	-	24.32	-11.50	-8.05	-3.45	10.93
miRNA (Gene Of Interest - GOI)	14.28	14.30	13.96	16.34	16.16	16.31	14.18	-	16.27	-				

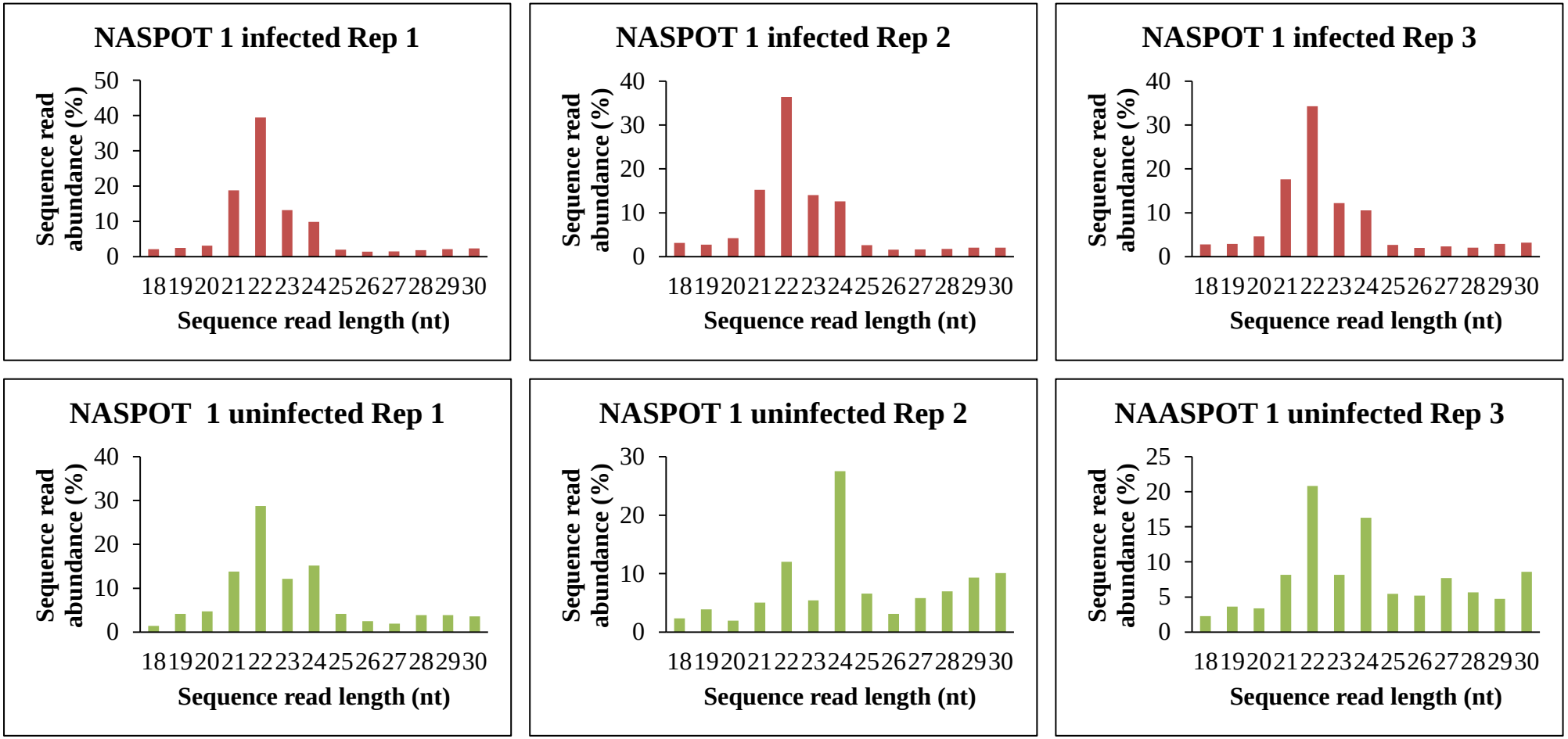

Figure A1.1: Read length distribution of vsiRNAs from the NASPOT 1 infected and uninfected treatments.

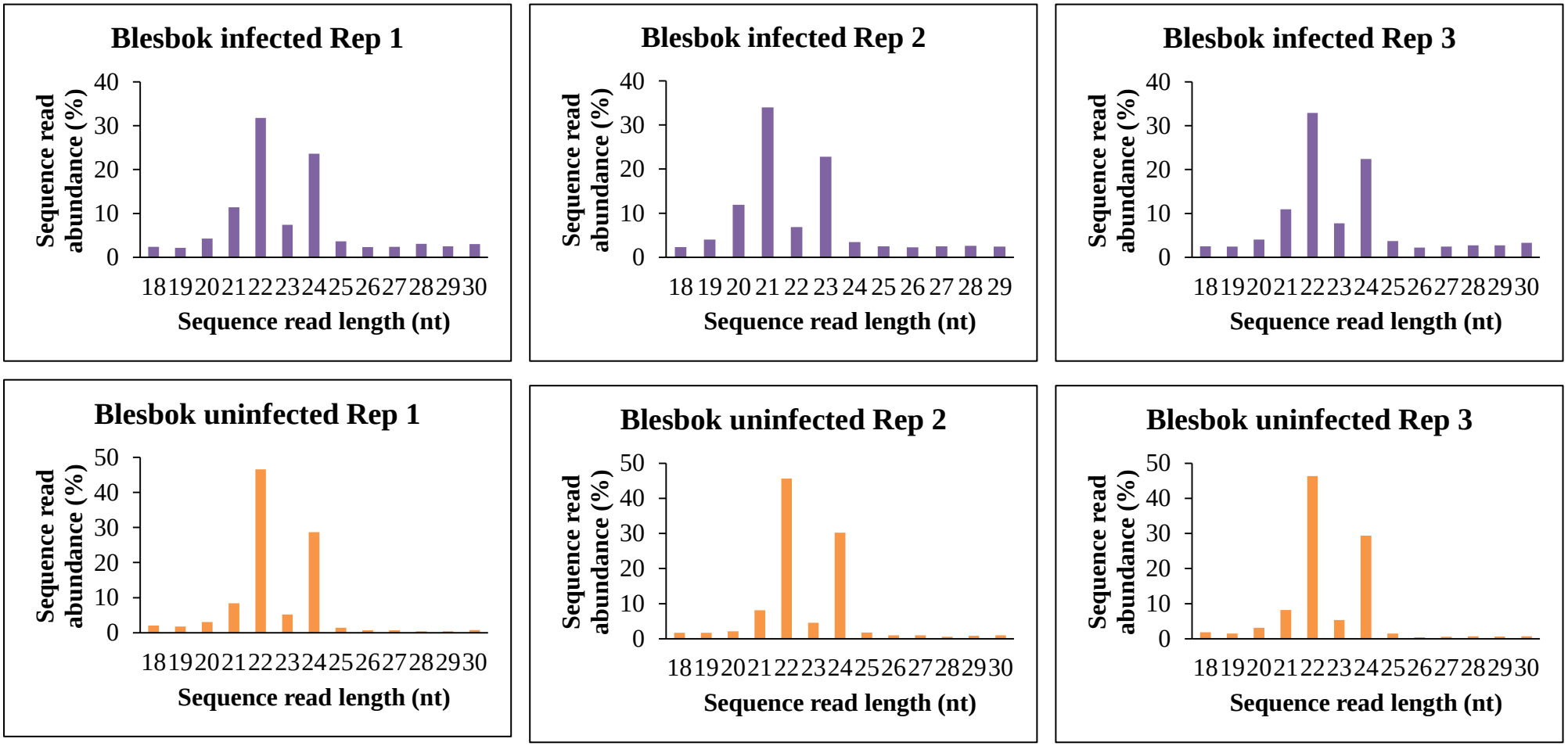

Figure A1.2: Read length distribution of vsiRNAs from the Blesbok infected and uninfected treatments.