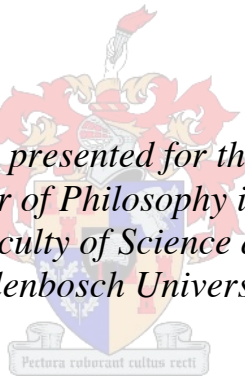
Small RNA profiling of grapevine leafroll-associated virus 3 infected grapevine plants

By
Rachelle Bester



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

Supervisor: Dr. H.J. Maree
Co-supervisor: Prof. J.T. Burger

December 2016

Declaration

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

Rachelle Bester

December 2016

Abstract

One of the most important viral diseases of grapevine worldwide is grapevine leafroll disease (GLD). A number of viruses from the family *Closteroviridae* have been associated with this disease, though *Grapevine leafroll-associated virus 3* is considered the leading causative agent due to its consistent association with GLD. To better understand the disease and develop effective control strategies, it is necessary to characterise the molecular interactions between the virus and the plant. Small RNA (sRNA) molecules have been shown to play an important role in gene regulation of normal development and defence responses to biotic and abiotic stresses in plants. Therefore, the aim of this study was to characterise the sRNA species in healthy and infected grapevine to contribute to the growing database of sRNAs present in *Vitis vinifera*. Microarray analysis and next-generation sequencing was used to identify sRNA species in Chardonnay, Chenin blanc, Cabernet Sauvignon and own-rooted Cabernet Sauvignon plants. Differential expression of sRNAs was evaluated to identify sRNAs associated with GLRaV-3 infection. The modulation of the differentially expressed microRNAs (miRNAs) was validated with stemloop RT-qPCR assays. Transcriptome NGS was also performed to validate the differential expression of the predicted miRNA targets, and to identify metabolic pathways modulated in response to GLRaV-3 independently from sRNA regulation. The transcriptome NGS transcripts that were differentially expressed in all cultivar groups, and transcripts that anti-correlated with miRNA expression, were validated with RT-qPCR assays. These high-throughput approaches identified several differentially expressed sRNAs and (target) genes in infected plants. The anti-correlation of miRNA expression and putative target expression were shown for two miRNAs. Cultivar specificity was identified in the sRNA and gene expression analyses, and both approaches identified Chenin blanc-specific responses. This comparison of symptomatic and asymptomatic GLRaV-3-infected plants provides the first insight into the disease symptom inhibition observed in certain cultivars. The differentially expressed genes identified in all cultivar groups, using the NGS transcriptome data, provides a collection of genes displaying a potentially universal molecular response against GLRaV-3. These genes showed strong associations with cell wall biosynthesis and signalling during pathogen recognition. This study has contributed significantly to the knowledge of sRNAs produced in grapevine and significantly extended the existing sRNA reference database for grapevine. The knowledge generated in this study can be utilised as potential targets for grapevine functional studies, and be translated into potential management strategies to control the disease. A better understanding of both the host defence and viral counter-defence strategies can lead to the prevention of virus replication or the impaired ability of the virus to induce pathogenesis in plants.

Opsomming

Een van die belangrikste virussiektes van wingerd wêreldwyd is wingerd-rolblaarsiekte (GLD). 'n Aantal virusse van die familie *Closteroviridae* hou verband met hierdie siekte, maar *Grapevine leafroll-associated virus 3* (GLRaV-3) word beskou as die hoof-bydraende faktor van GLD as gevolg van die korrelasie met GLD. Om die siekte beter te verstaan en effektiewe beheer toe te pas, is dit nodig om die molekulêre interaksie tussen die virus en die plant te ondersoek. Klein RNA (sRNA) molekules het getoon dat hulle 'n belangrike rol speel in die geenregulering van normale plant ontwikkeling, asook tydens die plant se verdediging teen biotiese en abiotiese stresfaktore. Die doel van hierdie studie was dus om die sRNA spesies in gesonde en geïnfekteerde wingerdstokke te karakteriseer en sodoende by te dra tot die snelgroeiende databasis van *Vitis vinifera* sRNAs. ‘n Mikro-DNA-volgorde-raamwerk analise, asook nuwe-generasie volgordebepaling (NGS) is gebruik om sRNAs spesies te identifiseer in die kultivars Chardonnay, Chenin blanc, Cabernet Sauvignon en eie-gewortelde Cabernet Sauvignon plante. Daarna is die differensiële uitdrukking van sRNAs geëvalueer om sodoende sRNAs te identifiseer wat verbandhou met GLRaV-3 infeksie. Die regulering van die differensiële uitgedrukte mikroRNA’s (miRNAs) is bevestig met stam-lus tru-transkripsie kwantitatiewe polimerase ketting reaksie (stemloop-RT-qPCR) toetse. Transkriptoom-NGS is ook uitgevoer om die differensiële uitdrukking van die miRNAs se voorspelde teikens te valideer en om gemoduleerde metaboliese paaie in reaksie op GLRaV-3, onafhanklik van sRNA regulasie, te identifiseer. Die transkriptoom NGS gene wat differensieël uitgedruk was in al die kultivar groepe, asook die gene wat ’n anti-korrelasie getoon het met miRNA uitdrukking, is bevestig met RT-qPCR toetse. Hierdie hoë-deurset benaderings het verskeie sRNAs en gene geïdentifiseer wat differensieël uitgedruk was in geïnfekteerde plante. Die anti-korrelasie van miRNA uitdrukking en voorspelde teiken uitdrukking is geïdentifiseer vir twee miRNAs. Kultivar-spesifisiteit was geïdentifiseer in beide die sRNA en geenuitdrukking analises en beide benaderings het ook Chenin blanc-spesifieke reaksies geïdentifiseer. Hierdie vergelyking van simptomatiese en asimptomatiese GLRaV-3-geïnfekteerde plante bied die eerste insig in die simtoom-inhibisie wat waargeneem word in sekere kultivars. Die differensieël-uitgedrukte gene wat geïdentifiseer was in alle kultivar groepe, met behulp van die NGS transkriptoom data, bied 'n versameling van gene wat ‘n potensiële universele molekulêre reaksie toon teen GLRaV-3. Hierdie gene het sterk assosiasies met selwand-biosintese en patogeen herkenningseine. Hierdie studie het aansienlik bygedra tot die kennis van wingerd-geassosieerde sRNAs en dra beduidend by tot die uitgebreide sRNA databasis. Die kennis wat gegenereer is in hierdie studie kan gebruik word om teikens te identifiseer vir wingerd funksionele studies en om verwerk te word na potensiële strategieë om die siekte te beheer. 'n Beter begrip van beide die gasheer verdedigings- en die virale teen-verdedigingstrategieë, kan lei tot die voorkoming van virus replikasie of ‘n verlaging in die vermoë van die virus om die siekte in plante te veroorsaak.

Acknowledgements

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

- Dr. H.J. Maree, for excellent mentorship, intellectual input, support, guidance, encouragement and suggestions throughout this study.
- Prof. J.T. Burger, for giving me the opportunity to be a part of an outstanding research group and for his support and guidance.
- Miss B. Coetzee, for bioinformatic assistance, moral support and all the encouraging conversations.
- Dr. T. Pepler, for all the help with statistical analyses and for being part of Harbin.
- Dr. M. Visser, for bioinformatic assistance and moral support.
- Mr D.J. Aldrich and Miss I. Mostert, for the opportunity to be part of their research and for the enjoyable working environment in Lab 246.
- Dr. A.E.C Jooste, for the opportunity to be part of her research.
- Vititec, for supplying the plant material for this study.
- National Research Foundation (NRF) for both a personal bursary and project funding. Opinions expressed and conclusions arrived at, are those of the author and are not necessarily to be attributed to the NRF.
- Stellenbosch University, for a personal bursary.
- Winetech for a personal bursary.
- NRF, the Vitis laboratory and Whitehead Scientific, for a travel bursary to attend the 18th meeting of the ICVG in Ankara, Turkey.
- Vitis laboratory, to attend Virology Africa, Cape Town, South Africa.
- NRF and the Vitis laboratory for a travel bursary to attend the Advances in Virology conference in Greenwich, London.
- Everyone from the Vitis laboratory, for their support and encouragement.
- My family and friends, especially my mother and sister, for love and moral support.

Table of contents

Chapter 1: Introduction

1.1 General introduction

Grapevine is cultivated worldwide for the production of wine, table grapes, juice and raisins and is a key contributor to the global economy. South Africa is the eighth largest wine producing country contributing 4.1 % to the world production [1]. This industry contributes more than R36 billion to the country's gross domestic product (GDP), including the agriculture, manufacturing, trade and hospitality sectors [1]. The viticulture industry remains one of South Africa's leading agricultural exporters. The country's temperate climate makes it ideal for grapevine cultivation and nine different wine producing regions are recognised including Stellenbosch, Paarl, Robertson, Swartland, Breedekloof Olifants River, Worcester, Northern Cape and the Little Karoo [1].

Diseases caused by the various pathogens that can infect grapevine, threaten the production potential of the grapevine industry and from 2006 a steady decrease in surface area utilised for grapevine cultivation, was observed in South Africa. Grapevine leafroll disease (GLD) is considered the most economically important viral disease threatening the viticulture industry. This disease is associated with viruses from the family *Closteroviridae*, though it is *Grapevine leafroll-associated virus 3* (GLRaV-3) that has the highest correlation with the disease [2]. Substantial economic losses due to GLRaV-3 have been reported in many countries, including New Zealand, USA and South Africa [2–4].

The establishment of genetically resistant grapevine would significantly decrease the financial burden of the disease. A better understanding of the molecular interaction between the plant and the pathogen would aid in the development of better control strategies. The disease symptoms are the consequence of the modulation of genes involved in a wide range of biological functions. These genes can be associated with normal plant development or specific to the plant's defence response, which can both be regulated by small RNAs (sRNAs). Comparative RNA profiling may therefore lead to the identification of differentially expressed sRNAs and possible target genes to gain knowledge of the molecular mechanisms underlying the GLRaV-3 stress response in grapevine.

The establishment of next-generation sequencing (NGS) technology has provided a tool to study sRNA and gene expression on a genome-wide scale. In this study, three of the noble grapevine cultivars, Chardonnay, Chenin blanc and Cabernet Sauvignon, were selected as experimental plants. Chenin blanc and Chardonnay are the most and third most widely planted white cultivars in South Africa. Cabernet Sauvignon is the most planted red cultivar [1]. Furthermore, Chardonnay and Cabernet Sauvignon represent "GLD-sensitive" cultivars, displaying typical symptoms associated with

white and red cultivars, respectively, while Chenin blanc is a non-symptomatic white cultivar. Different cultivars were included in the study in an attempt to identify a universal GLRaV-3-induced molecular response, and a potential explanation for the differential symptom expression observed between cultivars. This study set out to enhance the knowledge of the molecular basis of *Vitis vinifera* susceptibility to this viral pathogen.

1.2 Aim and objectives

The aim of the study was to use NGS to construct sRNA and gene expression profiles to characterise the plant response to GLRaV-3 infection.

The following objectives were set out to achieve this aim:

- To establish plant material of healthy and GLRaV-3 infected symptomatic and non-symptomatic grapevine cultivars using graft inoculations.
- To design and optimise a GLRaV-3 RT-qPCR assay for relative quantitation of the virus concentration.
- To sequence sRNA and mRNA libraries on an Illumina NGS platform.
- To apply bioinformatic tools for the identification and characterisation of different grapevine sRNA species.
- To identify differentially expressed sRNAs in the samples and to validate these using stemloop RT-qPCRs.
- To apply bioinformatic tools to identify differentially expressed genes in samples.
- To validate differentially expressed genes using RT-qPCR.
- To use bioinformatic tools to predict how the plant's defence pathways are modulated in response to GLRaV-3 infection.

1.3 Chapter layout

The dissertation contains six chapters that are introduced, concluded and referenced individually. A general introduction is followed by a literature overview, three research chapters and a general conclusion.

Chapter 1: Introduction

General introduction, aims and objectives of the study and the chapter layout of the thesis are provided. The scientific outputs generated during the study with the contributions by RB are stated.

Chapter 2: Literature review

An overview of the literature relating to grapevine leafroll disease, GLRaV-3, sRNA species of plants and bioinformatic analysis of next-generation sequencing data is provided.

Chapter 3: Differential expression of miRNAs and associated gene targets in grapevine leafroll-associated virus 3 infected plants.

In this chapter, the use of microarray analysis and NGS to generate miRNA and gene expression profiles to characterise the response of own-rooted Cabernet Sauvignon plants to GLRaV-3 are described. This chapter includes the preliminary study that provided evidence of the anti-correlation in expression between miRNAs and the predicted target genes in response to GLRaV-3 infection. This provided the foundation for the more comprehensive study of sRNA characterisation in GLRaV-3 infected plants.

Chapter 4: The small RNA repertoire of three *Vitis vinifera* cultivars

This chapter describes the use of NGS and bioinformatic analysis to identify and characterise different sRNA species in Chardonnay, Chenin blanc and Cabernet Sauvignon. The grapevine sRNA knowledge base was extended through the identification of novel *Vitis vinifera* miRNAs and novel phased loci, which can lead to the production of phasiRNAs. Furthermore, large numbers of sRNAs originating from repeat sequences in the *Vitis vinifera* genome, and of tRNA-derived sRNAs were identified. Cultivar specificity in the expression levels of the different sRNA species was also observed.

Chapter 5: Characterisation of virus disease-associated plant responses by transcriptome analysis

This chapter describes the differential expression of sRNAs and (target) genes that were identified in GLRaV-3 infected grapevine, in order to identify a universal GLRaV-3-associated stress response. Cultivar-specific sRNA responses and a universal gene level defence response were identified in infected plants.

Chapter 6: Conclusion

This chapter provides general concluding remarks and future prospects.

1.4 Research outputs

The following publications, conference proceedings were generated during the study. The proposed research outputs are also listed.

1.4.1 Publications

- **Bester R**, Pepler PT, Aldrich DJ and Maree HJ (2016) Harbin: A quantitation PCR analysis tool. Biotechnology letters. DOI 10.1007/s10529-016-2221-1 (Appendix A5).
  This paper includes a novel bootstrap test to compare independent RT-qPCR data sets to each other and the development of a Shiny application for RT-qPCR data analysis. RB contributed to the experimental design, data collection and analysis. PTP developed the bootstrap test and wrote the R script associated with the test. DJA generated two of the RT-qPCR data sets. RB re-packaged the R scripts to run as a Shiny web application and added additional known statistical tests to extend the functionality of the application.

- Visser M, **Bester R**, Burger JT, Maree HJ (2016) Next-generation sequencing for virus detection: covering all the bases. Virology Journal 13:85. DOI 10.1186/s12985-016-0539-x (Appendix A6).
  This paper includes the genome coverage obtained at different next-generation sequencing depths for a number of viruses, viroids, hosts and sequencing library types, using both read-mapping and de novo assembly-based approaches. RB was co-first author and contributed to experimental design, data collection, data analysis and manuscript editing.

- Maree HJ, Pirie MD, Oosthuizen K, **Bester R**, Rees DJG, Burger JT (2015) Phylogenomic analysis reveals deep divergence and recombination in an economically important grapevine virus. PLoS ONE 10(5):e0126819. DOI:10.1371/journal.pone.0126819 (Appendix A2).
  This paper includes the sequencing of a new GLRaV-3 variant and the phylogenetic placement of this variant. RB was involved in data collection, constructing of the GLRaV-3 supermatrix, recombination detection and manuscript editing.

- Jooste AEC, Molenaar N, Maree HJ, **Bester R**, De Koker WC, Burger JT (2015) Identification and distribution of multiple virus infections in grapevine leafroll diseased vineyards. European Journal of Plant Pathology 142:363–375. DOI: 10.1007/s10658-015-0620-0 (Appendix A3).
  This paper includes the results of a survey of viruses infecting grapevine in several of the wine regions in South Africa. RB was involved in data collection, data analysis associated with the RT-PCR high-resolution melting curve assay to differentiate GLRaV-3 variant groups and manuscript editing.

- **Bester R**, Pepler T, Burger JT, Maree, HJ (2014) Relative quantitation goes viral: RT-qPCR assay for a grapevine virus. Journal of Virological Methods 210:67–75. DOI: 10.1016/j.jviromet.2014.09.022 (Appendix A4).

This paper includes the quantitation RT-qPCR assay used for all GLRaV-3 diagnostics during the PhD study and is in its entirety the work of RB.

- Maree HJ, Almeida RPP, **Bester R**, Chooi KM, Cohen D, Dolja VV, Fuchs MF, Golino DA, Jooste AEC, Martelli GP, Rayapati N, Rohwani A, Saldarelli P, Burger JT (2013) Review: Grapevine leafroll-associated virus 3. Frontiers in Microbiology - Virology 4:82. DOI: 10.3389/fmicb.2013.00082. (Appendix A1).
  This paper is a review on GLRaV-3. RB wrote the GLRaV-3 diagnostic section.

1.4.2 Manuscripts in preparation

- **Bester R**, Burger JT, Maree HJ (2016) Differential expression of miRNAs and associated gene targets in grapevine leafroll-associated virus 3 infected plants. Submitted to Archives of Virology. Under review.
  This paper forms the basis for Chapter 3 and is almost entirely the work of RB. She had bioinformatic assistance from the non-authors, B. Coetzee and. M. Visser, who contributed an R script for microarray differential expression and a Python script for miRNA comparisons to miRBase, respectively.

- **Bester R**, Burger JT, Maree HJ. Small RNA repertoire of three *Vitis vinifera* cultivars. Submitted to Genes. Under review.
  Chapter 4 forms the basis of this paper and is almost entirely the work of RB. M. Visser provided bioinformatic assistance with the identification of natural antisense transcripts, phased siRNAs and miRNA comparisons to miRBase.

- **Bester R**, Burger JT and Maree HJ. Characterisation of virus disease-associated plant responses by transcriptome analysis. In preparation.
  Chapter 5 will form the basis of this paper and is in its entirety the work of RB.

1.4.3 Conference proceedings (Presenter underlined)

Advances in plant virology 2016, University of Greenwich, Greenwich, UK. 7-9 September 2016.

- **Bester R**, Burger JT, Maree HJ. Differential expression analysis of small RNAs in Grapevine leafroll-associated virus 3 infected plants. Advances in plant virology. (Paper)
  Chapter 4 and 5 form the basis of this proceeding and is almost entirely the work of RB. Bioinformatic assistance to compare miRNAs to miRBase and to identify phasiRNAs and natural antisense transcripts was provided by M. Visser.

- Visser M, **Bester R**, Burger JT, Maree HJ. Next-generation sequencing for virus detection: covering all the bases. (Paper)
  This proceeding includes the genome coverage obtained at different next-generation sequencing depths for a number of viruses, viroids, hosts and sequencing library types, using both read-mapping and de novo assembly-based approaches. RB contributed to experimental design, data collection and data analysis.

- **Bester R**, Pepler PT, Aldrich DJ, Maree HJ. Harbin: An analysis tool for relative quantitation of real-time qPCR data and a quantile-based bootstrap test for data pooling. (Poster)
  This paper includes a novel bootstrap test to compare independent RT-qPCR data sets to each other and the development of a shiny application for RT-qPCR data analysis. RB contributed to the experimental design, data collection and analysis. PTP developed the bootstrap test and wrote the R script associated with the test. DJA generated two of the RT-qPCR data sets. RB re-packaged the R scripts to run as a Shiny web application and added additional known statistical tests to extend the functionality of the application.

Virology Africa, Cape Town, South Africa. 30 November to 3 December 2015.

- Molenaar N, **Bester R**, Pirie MD, Pepler PT, Oosthuizen KO, Burger JT, Maree HJ. Determination of the virus diversity associated with Grapevine leafroll disease. (Paper)
  This conference proceeding included the sequencing of a new GLRaV-3 variant, the phylogenetic placement of this variant and GLRaV-3 diagnostics. RB was involved in data collection, constructing of the GLRaV-3 supermatrix, development of the RT-qPCR assay, recombination detection and RT-qPCR data analysis.

- **Bester R**, Burger JT, Maree HJ. Small RNA analysis of Grapevine leafroll-associated virus 3 infected grapevine. P76 (Poster)
  This poster included the work described in Chapter 4 and 5 and is almost entirely the work of RB. Bioinformatic assistance for miRNA comparisons to miRBase and phased siRNA analysis was provided by M. Visser.

- Aldrich DJ, **Bester R**, Burger JT, Maree HJ. Characterisation of Micro-RNA expression profiles of *Vitis vinifera* in response to Grapevine leafroll-associated virus 3 infection. P79 (Poster)
  This poster included the expression profiles of known microRNAs in GLRaV-3 infected Cabernet Sauvignon plants. RB contributed to sample collection and data analysis.

18th meeting of the International Council for the Study of Virus and Virus-like Diseases of the Grapevine (ICVG), Ankara, Turkey. 7-11 September 2015.

- Maree HJ, **Bester R**, Pirie MD, Pepler PT, Oosthuizen K, Burger JT. GLRaV-3: diversity, detection and quantitation. p.53. (Paper)
  This proceeding included the sequencing of a new GLRaV-3 variant and the phylogenetic placement of this variant. The current status of GLRaV-3 diagnostics is also discussed. RB was involved in data collection, constructing of the GLRaV-3 supermatrix, development of the RT-qPCR assay, recombination detection and RT-qPCR data analysis.

- **Bester R**, Burger JT, Maree HJ. Searching for the needle in a haystack: Small RNA analysis of grapevine leafroll disease in symptomatic and asymptomatic cultivars. p.102. (Paper)
  This proceeding included the work described in Chapter 4 and 5 and is almost entirely the work of RB. Guidance with bioinformatic analysis was provided by M. Visser with regards to miRNA comparisons to miRBase and phased siRNA analysis.

35th Conference of the South African Society for Enology and Viticulture, Somerset West, South Africa. 13-15 November 2013.

- **Bester R**, Maree HJ, Burger JT. Grapevine virus diagnostics: Beyond ELISA. (Paper)
  This proceeding included an overview of GLRaV-3 diagnostics and included the development of a RT-qPCR assay able to detect all variants. This was in its entirety the work of RB.

- Jooste AEC, Maree HJ, Molenaar N, **Bester R**, De Koker WC, Burger JT. Survey of white and red cultivar vineyards affected by Grapevine leafroll disease for genetic variation in Grapevine leafroll-associated virus 3. (Paper)
  This paper includes the results of a survey of viruses infecting grapevine in several of the wine regions in South Africa. RB was involved in data collection and data analysis associated with the RT-PCR high-resolution melting curve assay to differentiate GLRaV-3 variant groups.

1.5 References

1. SAWIS (2016) South African Wine Industry Statistics report nr. 40. SAWIS, Paarl (http://www.sawis.co.za/)
2. Maree HJ, Almeida RPP, Bester R, et al (2013) Grapevine leafroll-associated virus 3. Front Microbiol 4:82. doi: 10.3389/fmicb.2013.00082

3. Atallah SS, Gomez MI, Fuchs MF, Martinson TE (2012) Economic Impact of Grapevine Leafroll Disease on Vitis vinifera cv. Cabernet franc in Finger Lakes Vineyards of New York. Am J Enol Vitic 63:73–79. doi: 10.5344/ajev.2011.11055

4. Freeborough M, Burger JT (2008) Leafroll: Economic implications Michael-John Freeborough & Johan Burger. Wynboer, A technical Guide for Wine Producers

Chapter 2: Literature review

2.1 Introduction

Grapevine (genus *Vitis*, family *Vitaceae*) is the most widely cultivated woody deciduous fruit crop in the world. This species has commercial significance for wine and table grape production and are also valued in the cosmetic and pharmaceutical industries for the high antioxidant level of berries and the subsequent health benefits. There are many grapevine species within the genus *Vitis*, among which *Vitis vinifera* is the most widely grown in the world. The different cultivars of *Vitis vinifera* are usually planted as grafted vines, consisting of a specific cultivar scion grafted onto a rootstock. This can increase vigor, provide protection to soil pests such as phylloxera (*Daktulosphaira vitifoliae*) and promote early ripening [1].

This highly valuable agricultural commodity is vegetatively propagated and therefore exposed to a variety of pests and pathogens. More than 70 infectious agents among viruses, viroids, phytoplasmas, bacteria and fungi are known to infect grapevine, making it the highest number of intracellular pathogens found in a single crop [2]. The detrimental effects of the diseases associated with these pathogens and the importance of the grapevine industry have encouraged intensive research into controlling or preventing disease spread. To date, research has enhanced the establishment and implementation of clean stock certification schemes by focussing on early and accurate detection of pathogens. The development of next-generation sequencing (NGS) has contributed significantly to the detection of novel viruses previously unrecorded to infect grapevine like, *Grapevine Syrah virus 1* (GSyV-1) [3], *Grapevine vein clearing virus* (GVCV) [4], *Grapevine Pinot gris-associated virus* (GPGaV) [5], *Grapevine virus F* (GVF) [6, 7] and *Grapevine red blotch-associated virus* (GRBaV) [8].

There are five major viral diseases of grapevine, these are grapevine leafroll disease; grapevine degeneration and decline; rugose wood complex; graft incompatibility and the fleck disease complex. Grapevine leafroll disease (GLD) is one of the most important and widespread virus diseases of grapevine affecting wine and table grape cultivars, as well as rootstocks [9–11]. A number of viruses from the family *Closteroviridae* have been associated with GLD, however *Grapevine leafroll-associated virus 3* (GLRaV-3) is considered the main causative agent (Appendix A1)[i] [9]. Even though GLD is a threat to wine and grape production, the nature of the disease is not fully understood.

---

[i] Appendix A1. Maree HJ, Almeida RPP, Bester R, Chooi KM, Cohen D, Dolja VV, Fuchs MF, Golino DA, Jooste AEC, Martelli GP, Rayapati N, Rohwani A, Saldarelli P and Burger JT (2013) Review: Grapevine leafroll-associated virus 3. Frontiers in Microbiology. Virology 4:82.

The current state of knowledge on GLRaV-3 is reviewed by focusing on most aspects of GLRaV-3 research, including molecular characterisation, genome organisation, virus replication, genetic variability between GLRaV-3 isolates, detection assays employed to detect the virus and host-pathogen interactions.

The association of several viruses and virus variants, the differential expression of symptoms in red- and white-berried cultivars, and the complete absence of natural resistance impart on this disease a level of complexity that has intrigued researchers for decades, and probably will for several years to come.

This chapter provides an overview of the literature related to this study. The review focuses on four aspects including GLD, GLRaV-3, the different small RNA (sRNA) species analysed in the study and the use of next-generation sequencing and *in silico* analysis to identify sRNA species and investigate differential gene expression.

2.2 Grapevine leafroll disease (GLD)

2.2.1 Disease symptoms and impact

Grapevine leafroll disease is one of the most widespread and economically detrimental viral diseases of grapevines in all grape-producing regions of the world [9–12]. Depending on the disease incidence, GLD can impact on fruit quality and cause significant yield reductions [13, 14]. This in turn will cause substantial economic losses and negatively impact on the sustainability of the industry. Grapevine leafroll disease was shown to reduce leaf photosynthesis and cause degeneration of phloem cells in leaves, stems and petioles [15–17]. Changes in the berry ripening process, in particular the up-regulation of genes involved in anthocyanin biosynthesis, sugar metabolism, sugar transport, flavonoid biosynthesis and senescence were detected [15, 18, 19]. The reduction of leaf metabolism induced by GLD also produced a significant reduction in CO_2 assimilation, yield, canopy size and cane lignification [20].

The abovementioned physiological and metabolic changes associated with the disease are detectable in the well-known visual symptom expression of the disease. Leaf symptoms are clearly visible in red cultivars, expressed as red and reddish-purple discolorations in the interveinal areas while the midrib and main veins remain green (Figure 1A). Leaves often curl downwards and become brittle. In white cultivars symptoms are less obvious and if symptoms are visible, yellowing and downwards rolling of the leaf margins can be seen (Figure 1B). The degree of symptom expression varies considerably amongst cultivars and in some cultivars downward rolling of leaf margins may not be evident. Additionally, the identification of disease symptoms in red cultivars is complicated by differences observed between scion-rootstock combinations and environmental factors such as nutritional disorders (potassium deficiency), mechanical damage and other diseases (red blotch disease) that can cause similar discolorations as GLD [10].

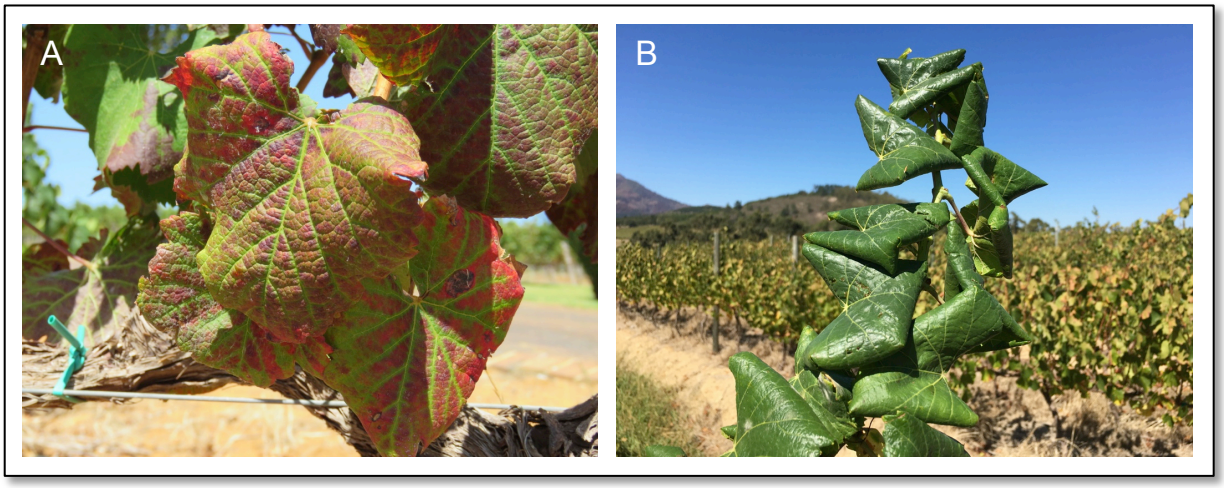


Figure 1. Grapevine leafroll diseased grapevine. Typical symptom expression observed in the red cultivar Cabernet Sauvignon (A) and in the white cultivar Chardonnay (B).

2.2.2 Transmission and disease management

Grapevine leafroll disease is predominantly transmitted by grafting or the propagation of infected plant material, however, insect vectors can mediate natural GLD transmission from plant to plant [21–23]. The transmission of grapevine leafroll-associated viruses (GLRaVs) has been shown for a number of species of mealybugs (*Pseudococcidae*) and a few species of soft scale insects (*Coccidae*) [24, 25]. These small, sap-feeding insects, transmit viruses in a semi-persistent manner [23]. The exchange and trade of infected plant material is the primary cause of disease spread, and for this reason many countries developed certification programmes to provide virus-free plant material to growers. Unfortunately, no natural resistance to GLRaV-3 in *Vitis vinifera* has been identified and the disease is best managed by prevention through accurate detection assays. Although disease spread can be prevented by implementing an effective vector control program, the most effective disease management strategy involves the removal of infected vines (rouging) and extending the period before re-planting [11, 26].

2.2.3 Grapevine leafroll-associated viruses

Different viruses in the family *Closteroviridae* have been reported to be associated with GLD; these include GLRaV-1 to -9, GLRaV-Pr, GLRaV-De and GLRaV-Car [9]. Although the existence of GLRaV-8 has been reported, GLRaV-8 is not recognised as a virus species anymore [27]. All these viruses belong to the genus *Ampelovirus* except for GLRaV-2, which is in the genus *Closterovirus*, and GLRaV-7, which is in the genus *Velarivirus* [27, 28]. Subsequently it was shown that GLRaV-4, -5, -6, -9, -Pr, -De, and -Car are closely related and could be considered as different strains of GLRaV-4 [27]. Among the currently known GLRaVs, GLRaV-3 is the most widespread, found in both single and mixed GLRaV infections associated with GLD, and is therefore considered as the main causative

agent correlated with the disease [9–11].

2.3 Grapevine leafroll-associated virus 3

2.3.1 Genome organisation

Grapevine leafroll-associated virus 3 is the type species of the genus *Ampelovirus* in the family *Closteroviridae* [27]. The genus *Ampelovirus* can be divided into two subgroups based on genome structure and size, of which GLRaV-3 is classified into subgroup I [27]. Ampelovirus virions are flexuous and filamentous particles with GLRaV-3 particles being between 1800 and 2000 nm in length (Figure 2) [29].

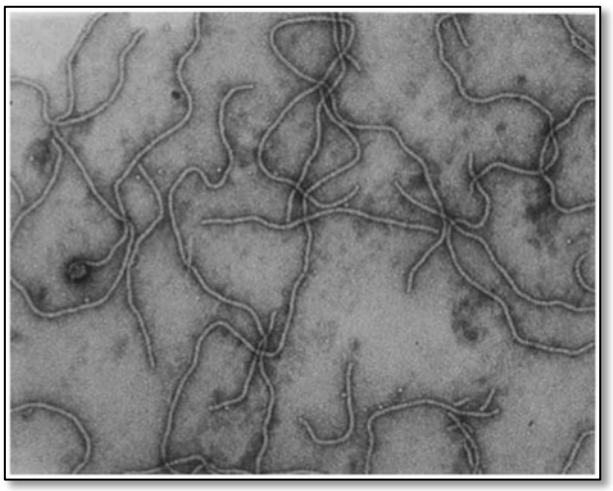


Figure 2. Transmission electron micrograph of purified GLRaV-3 particles. Picture by G. G. F. Kasdorf.

Grapevine leafroll-associated virus 3 has a monopartite, linear, positive sense, single-stranded RNA genome of approximately 18500 nucleotides (nts) [25]. A near complete genome sequence of 17919 nts for GLRaV-3 was reported in 2004 [30], however the first complete genome sequence was presented in 2008 and found to be 18498 nts in length, including a 5′ untranslated region (UTR) of 737 nts (Figure 3) [31]. *Grapevine leafroll-associated virus 3* has 12 open reading frames (ORFs) and follows the convention for closteroviruses, with ORFs designated as ORF1a, 1b and 2-12 [32–34].

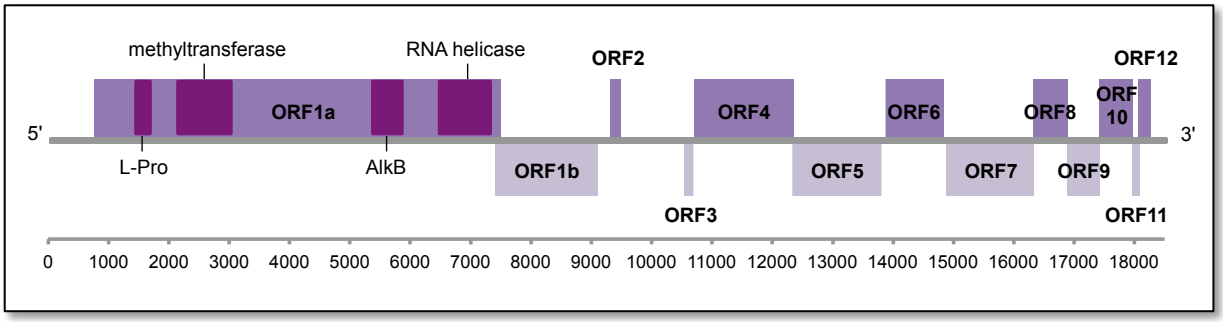


Figure 3. Schematic representation of the genome organisation of GLRaV-3 isolate GP18. The 12 ORFs encode the replication proteins (ORF1a and ORF1b (RdRp)), p6 (ORF2), p5 (ORF3), the HSP70h protein (ORF4), p55 (ORF5), the coat protein (ORF6), the minor capsid protein (ORF7), p21 (ORF8), p20A (ORF9), p20B (ORF10), p4 (ORF11) and p7 (ORF12) [9, 34].

Homologous ORFs in the genomes of other positive-strand RNA viruses were used to infer the putative functions of GLRaV-3's ORFs. Open reading frame 1a and 1b are essential for RNA replication and contains a methyltransferase, RNA helicase, papain-like leader protease (L-Pro), AlkB and RNA-dependent RNA polymerase domain [31, 35–39]. The L-Pro domain of *Beet yellows virus* (BYV) and GLRaV-2 was shown to be involved in the systemic spread of the virus, accumulation and virus invasiveness [40–42]. RNA demethylation has been linked with the AlkB domain and this domain is proposed to be involved in viral RNA repair [43]. Open reading frame 2 potentially encodes a small peptide, except no equivalent ORF has been found in other members of the family *Closteroviridae* and is not present in all isolates of GLRaV-3. Open reading frames 3-7 (quintuple gene block) are conserved in the family *Closteroviridae*, coding for a transmembrane protein (ORF3) [44], heat shock protein 70 (HSP70) molecular chaperone (ORF4), HSP70-like protein (ORF5) [45, 46], coat protein (CP) (ORF6) [47] and the minor capsid protein (ORF7) [48, 49]. Open reading frames 3-7 are considered to be involved in the systemic spread and cell-to-cell movement of the virus [36].

The putative functions of GLRaV-3 ORFs 8-10 were inferred by identifying ORFs in other members of the family *Closteroviridae* at similar locations in the genome. These ORFs are presumed to be involved in RNA interference and long-distance transport [50–53]. Open reading frame 10 was shown to have RNA interference suppressor activity in *Nicotiana benthamiana* [54]. Open reading frames 11 and 12 are unique to GLRaV-3 and not present in other members of the family *Closteroviridae*. Therefore, the functions of these small ORFs are still unknown. The development of a full-length infectious clone will contribute significantly to studying the functions of the GLRaV-3 proteins.

2.3.2 GLRaV-3 replication

The replication mechanism of GLRaV-3 is assumed to follow an equivalent strategy to other

closteroviruses like *Citrus tristeza virus* (CTV) and BYV [36]. The replication of positive stranded RNA viruses involves first the uncoating of the virus to expose the nucleic acid for translation to produce structural and non-structural proteins, then replication of the genome, and finally the encapsidation of the progeny genomic strands for the virus to spread [36, 55]. The replication-associated proteins encoded by ORFs 1a and 1b are translated from the genomic RNA [56] and as with the other members of the family *Closteroviridae*, the GLRaV-3 ORFs downstream of ORF1b are expressed via the formation of subgenomic RNAs (sgRNAs) [57, 58]. These sgRNAs are believed to be transcribed from the viral replicase that recognises the internal sgRNA promoters [59].

2.3.3 Genetic variants

Research worldwide showed the existence of several genetic variants of GLRaV-3. To date, the complete genomes of 13 GLRaV-3 isolates [31, 58, 60–65] representing five major genetic variant groups are available and also partial sequences of three additional variant groups [65, 66]. Higher order groups to classify the different variant groups were proposed as supergroups A to D (Appendix A2)[ii] [65]. Supergroup A includes variant groups I-V and supergroup B includes group VI and its related unclassified isolates. Supergroups C and D includes the newly identified variant groups VII (represented by the complete genome sequence isolate GH24) and VIII, respectively. Variant groups within supergroup A are more than 85 % identical in sequence compared to variant group VII that is only 63-65 % similar to variant groups I-III and VI [65]. Both variant groups VI and VII lack ORF2 and very low levels of sequence similarity for ORF11 and ORF12 was identified between variants groups. This high level of diversity observed in these ORFs among GLRaV-3 variant groups, suggest that these ORFs are unlikely to code for proteins with conserved functions.

These different genetic variant groups occur as single or mix variant infections and in combinations with other grapevine viruses. Based on data collected from a survey of South African vineyards, variant groups II and VI infections were the most prevalent (Appendix A3)[iii] [67]. Little is known about the biological properties of the different GLRaV-3 genetic variants and whether there is significant variation in their pathogenicity remains to be determined.

[ii] Appendix A2. Maree HJ, Pirie MD, Oosthuizen K, Bester R, Rees DJG, Burger JT (2015) Phylogenomic Analysis Reveals Deep Divergence and Recombination in an Economically Important Grapevine Virus. PLOS ONE 10:e0126819.

In this study we trace the evolutionary history of GLRaV-3, focussing on isolate GH24, a newly discovered variant. GH24 was discovered through the use of next-generation sequencing (NGS) and the whole genome sequence determined and validated with Sanger sequencing. We assembled an alignment of all 13 available whole genomes of GLRaV-3 isolates and all other publicly available GLRaV-3 sequence data. Using multiple recombination detection methods we identified a clear signal for recombination in one whole genome sequence and further evidence for recombination in two more, including GH24. We inferred phylogenetic trees and networks and estimated the ages of common ancestors of GLRaV-3 clades by means of relaxed clock models calibrated with asynchronous sampling dates.

[iii] Appendix A3. Jooste AEC, Molenaar N, Maree HJ, Bester R, Morey L, De Koker WC, Burger JT (2015) Identification and distribution of multiple virus infections in Grapevine leafroll diseased vineyards. Eur J Plant Pathol 142:363–375.

A survey of viruses infecting grapevine in the wine regions of the Western Cape Province in South Africa was conducted. The survey determined the relative abundance of five different grapevine leafroll-associated virus 3 (GLRaV-3) variants. Virus profiles were also determined for individual vines. A total of 315 plants were sampled and analysed over two growing seasons. The complexity of virus populations detected in this study, highlights the need for detection methods able to identify all viruses and their variants in vineyards.

2.3.4 GLRaV-3 detection

Several established techniques are available to detect GLRaV-3 in plants. These include biological indexing, serology and nucleic acid-based methods. Although biological indexing is a successful technique to detect the disease rather than the associated viruses, skilled virologists and a timeframe of at least two seasons are required. The fact that plants can be infected with more than one virus causing similar disease symptoms can complicate biological indexing. Serological assays, like enzyme-linked immunosorbent assays (ELISA), [68, 69] and nucleic acid-based methods including single-strand conformation polymorphism (SSCP) analysis, RT-PCRs and microarrays [70–77] have been very successful in detecting GLRaV-3, however false negative results can occur due to low virus titre, diverse variants and the uneven distribution in infected plants.

Next-generation sequencing avoids the need for prior sequence information and allows for the detection of known and unknown viruses simultaneously in the same sample. This technique has contributed to the construction of plant viromes [78, 79] as well as the discovery of new viruses and virus variants [3–5, 7, 62, 65, 80, 81], which enabled the development of more accurate nucleic acid-based detection assays [71, 77].

The high genome sequence variation between genetic variants of GLRaV-3 illustrates the importance of having sensitive and rapid detection methods. A pathogen-independent detection assay that focuses on the plant's response to the virus and not the viral genome, can potentially be more sensitive and circumvent the problem with unidentified diverse variants. Therefore, a better understanding of the molecular interaction between the plant and the virus is essential.

2.3.5 Host-pathogen interaction

The impact of GLRaV-3 on grapevine is irreversibly destructive for normal plant development and growth, and affects the overall yield of the crop. *Grapevine leafroll-associated virus 3* influences photosynthesis, berry cluster size, fruit ripening, cane lignification and lead to modifications of the levels of anthocyanins and phenolics [18–20, 82–85]. These disease responses are the result of the modulation of genes involved in a wide range of biological functions.

A complex network of mechanisms controls gene expression in plants to respond to biotic and abiotic stresses. Transcriptional or post-transcriptional gene silencing, also known as RNA silencing, is one mechanism that plants utilise to regulate gene expression. The role of small non-coding RNAs (sRNAs) in RNA silencing has been increasingly investigated to better understand their biogenesis and function in relation to the plant stress response [86–88]. Small RNAs regulate gene expression by interfering with mRNA translation or by the cleavage and degradation of mRNA [89–91]. In a study

by Alabi et al. [92], altered levels of sRNA species was found in GLRaV-3 infected plants, particularly sRNAs predicted to target transcription factors (Squamosa promoter binding protein-like and Auxin Response Factor), which can cause extensive developmental changes. The presence of viral sRNAs also suggested that virus replication can influence the small RNA profile of plants [92–95]. Studies by Singh et al. [96] and Pantaleo et al. [97] showed differential expression of sRNAs associated with viral infection in *Vitis vinifera*. More research into the interaction between sRNAs, host genes and the viral genome will contribute to understanding the GLRaV-3 specific disease response in plants.

2.4 Small RNAs

Small RNAs control the expression of target genes by binding to a complementary sequence for the regulation of gene expression during normal development, or in the plants' response to biotic or abiotic stress conditions. Small RNAs are generated from naturally formed double stranded RNA (dsRNA) precursors or a single-stranded RNA (ssRNA) molecule with a partial double-stranded conformation. Small RNAs are usually 17-30 nts in length and based on their mode of biogenesis and/or function, various types of sRNAs have been identified in plants, including microRNAs (miRNAs) and small interfering RNAs (siRNAs) (Figure 4) [90, 91, 98]. MicroRNAs are processed from dsRNA precursors transcribed from nuclear encoded *MIR* genes, while siRNAs are generated from either endogenous or exogenous dsRNA molecules (Figure 4). These siRNAs species include heterochromatic siRNAs or repeat-associated siRNAs (rasiRNAs), *trans*-acting siRNAs (tasiRNAs) or phased siRNAs (phasiRNAs), natural-antisense transcript-derived siRNAs (NatsiRNAs) and virus derived siRNAs (vsiRNA) (Figure 4). The generation of NGS data has also led to the discovery of new sRNAs that could be derived from existing RNA molecules such as transfer RNAs (tRNAs) (Figure 4) [99–104].

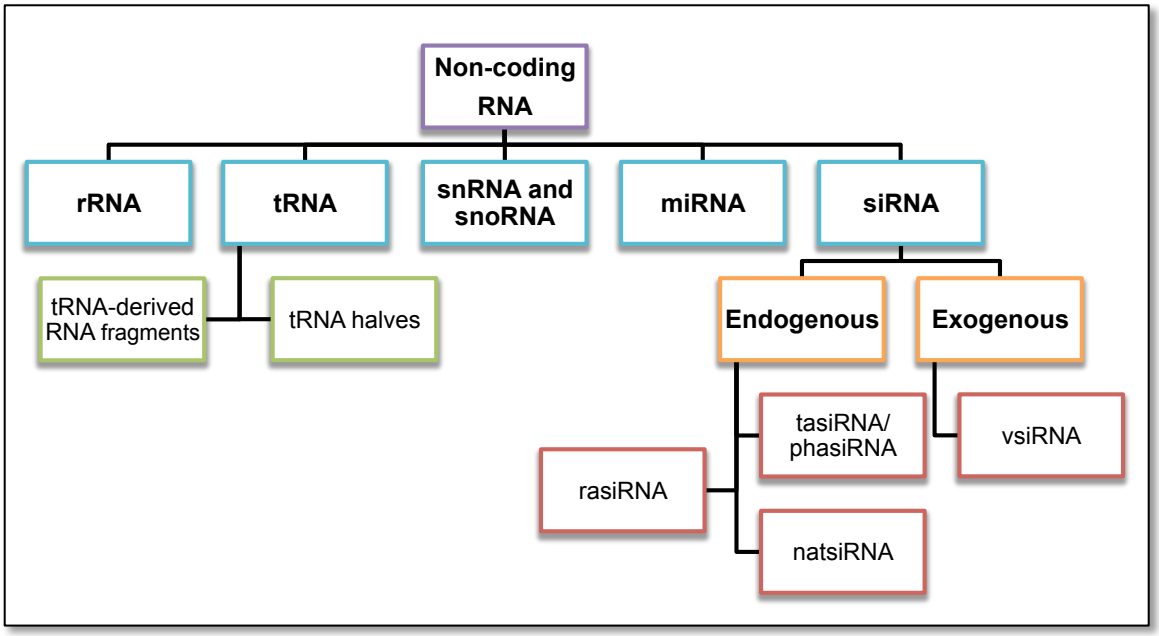


Figure 4. Classification of different non-coding small RNA species. Adapted from Guleria et al. [91], Axtell [98] and Garcia-Silva et al. [99] (rRNA: ribosomal RNA; tRNA: transfer RNA; snRNA: small nuclear RNA; snoRNA: small nucleolar RNA; miRNA: microRNA; siRNA: small interfering RNA; rasiRNA: repeat-associated siRNA; tasiRNA: *trans*-acting siRNA; phasiRNA: phased siRNA; natsiRNA: natural-antisense transcript-derived siRNA; vsiRNA:virus derived siRNA).

2.4.1 MicroRNAs

In plants, miRNAs are processed from single-stranded hairpin precursors that are transcribed from nuclear-encoded *MIR* genes by RNA polymerase II to form transcripts with a partial double-stranded stemloop structure (Figure 5) [89–91]. The processing of the precursor is catalysed in the nucleus by Dicer-like (DCL1), Hyponastic Leaves 1 (HYL1) and serrate (SE) proteins to yield a 20-22 nucleotide (nt) miRNA duplex [89–91]. The 3′ end of the duplex is then methylated and exported to the cytoplasm by HASTY (HST1). In the cytoplasm a helicase unwinds the duplex and the mature miRNA is available to associate with the RNA silencing complex (RISC). The mature miRNA then binds to the argonaute (AGO) protein, the catalytic site of RISC, which guides the RISC to the complementary target mRNA sequence (Figure 5). Cleavage of the target mRNA occurs between positions 10 and 11 of the miRNA alignment by the endonuclease activity of the AGO proteins [98]. Perfect complementarity is rare between a miRNA and its target, however not many examples exist with more than five mismatches [98]. The tolerance for mismatches from position 12 to the 3' end of the miRNA tends to be higher compared to mismatches at the 5′ end between positions 1-11 [105, 106].

MicroRNAs are involved in multiple biological and metabolic processes in plants, including regulation of development and responses to biotic and abiotic stresses. Environmental stresses can

modulate or initiate miRNA expression across most processes of plant growth, while some are developmental stage, tissue or environmental condition specific [107]. Differential expression of miRNAs has been shown for various abiotic stresses including draught and water stress [108, 109], extreme temperatures [110, 111], salinity [112, 113], hypoxia [114] and nutrient deficiencies [115, 116]. The involvement of miRNAs in plant-virus interactions has also been suggested in a range of plant species including *Nicotiana benthamiana* [117], *Arabidopsis thaliana* [118], *Hibiscus cannabinus* [119], tobacco [94], cotton [120], tomato [121, 122], rice [123], grapevine [92, 96] and citrus [124].

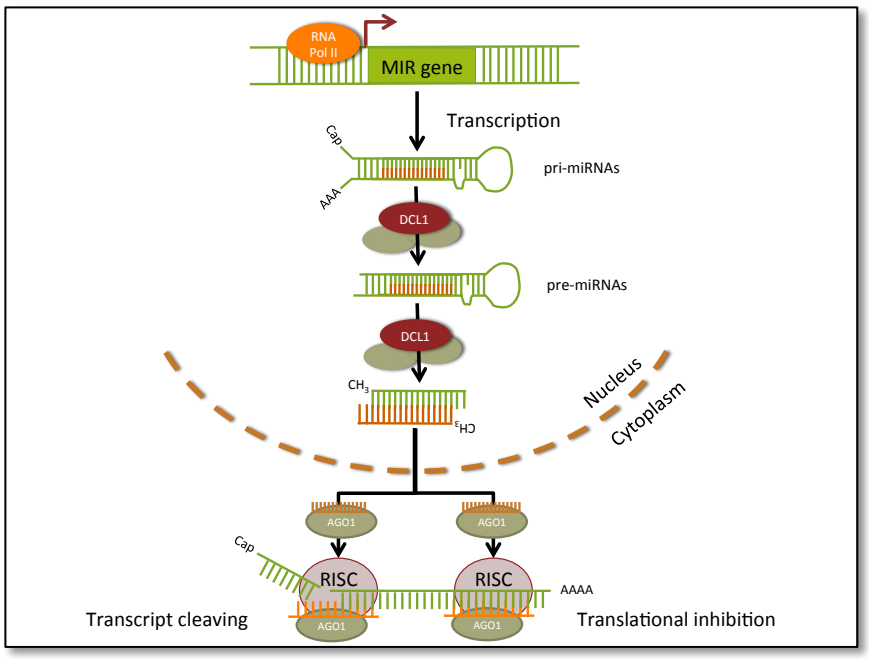


Figure 5. Biogenesis of miRNAs. Adapted from Khraiwesh et al. [90], Bartel et al. [89] and Jones-Rhoades et al. [125].

2.4.2 Small interfering RNAs

Small interfering RNAs differ from miRNAs in origin, structure and mode of action and can have either an endogenous or exogenous dsRNA precursor. The endogenous dsRNA molecules can originate from natural antisense transcript pairs from the plant genome, RNA transcribed from inverted repeats or from the conversion of single stranded RNA into dsRNA by RNA-dependent RNA polymerases (RDRs) [91]. Regions in the genome rich in retro-elements can also lead to dsRNA formation. The exogenous sources include transposons and viruses [91]. The dsRNA molecule can be cleaved by various DCL proteins into 21-24 nt siRNAs, depending on the catalytic activity of the specific DCL protein, irrespective of the dsRNA origin [126]. There are at least four classes of DICER-like enzymes, DCL1 to DCL4, of which DCL1 generates 18-21 nt miRNAs, and DCL2, DCL3 and DCL4 generates 22 nt, 24 nt and 21 nt siRNAs, respectively [126]. Similar to miRNAs, these different siRNAs are loaded into an AGO protein-containing RISC that guides the siRNA to the

target to regulate gene expression at a post-transcriptional or transcriptional level (Figure 6) [90, 91].

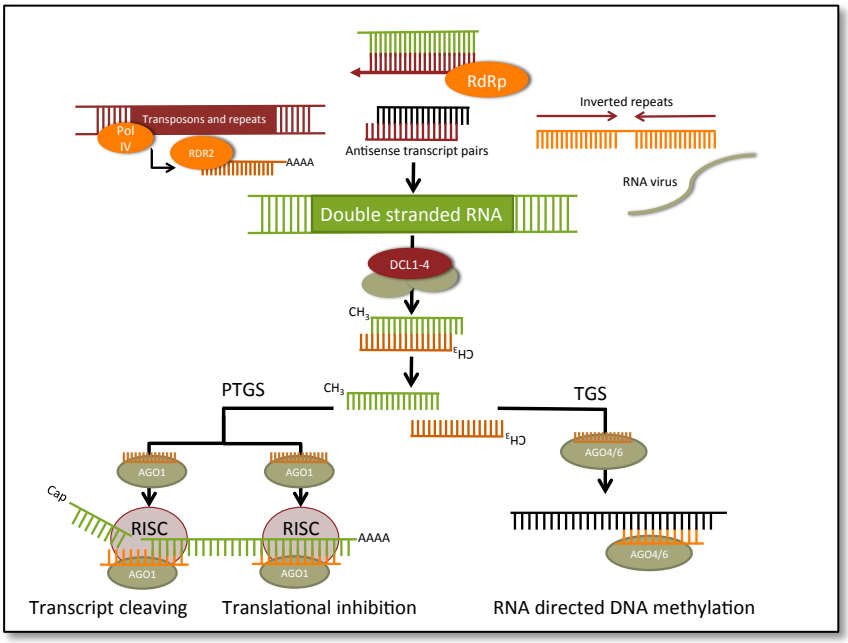


Figure 6. Biogenesis of siRNAs. Adapted from Khraiwesh et al. [90], Bartel et al. [89] and Jones-Rhoades et al. [125].

2.4.2.1 Trans-acting siRNA and phased siRNA

Trans-acting siRNAs (tasiRNAs) are sRNAs produced from *trans*-acting genes (*TAS*) that can regulate gene expression of targets other than their locus of origin. Previously they were considered to only function in *trans*, but since their *cis*-action was demonstrated, the term phased siRNAs (phasiRNAs) is used [127, 128]. In general, phasiRNAs refers to the siRNAs generated from a dsRNA precursor that is cleaved in precise increments from a primary sRNA cleavage site. Both protein-coding and non-coding genes have been identified as loci that can give rise to phasiRNAs. These are known as *PHAS* loci. The biogenesis of phasiRNAs involves the transcribing of a single-stranded RNA molecule from a *PHAS* locus, which is subsequently cleaved by a phase-initiator (either a miRNA or a phasiRNA) [127, 129, 130]. RNA-dependent RNA polymerase 6 converts the resulting fragments into dsRNA, after which it is cleaved by dicer-like 4 (DCL4) at every 21 nt relative to the cleavage site of the phase-initiator (Figure 7) [127, 129, 130]. This generates the 21 nt phasiRNAs. These phasiRNAs are known as tasiRNAs if they bind to AGO proteins to cause cleavage of a target mRNA away from the locus where they originated from.

Several *TAS* genes have been recognised in plant species, of which some are conserved and some species-specific [127, 131, 132]. The pentatricopeptide repeat genes are targeted by tasiRNAs produced from *TAS1* and *TAS2* with miR173 as phase-initiator [133]. The tasiRNAs of *TAS3*, a target of miR390, were shown to target the auxin response factor genes [133] and the *TAS4* tasiRNAs,

initiated by miRNA828, are believed to target MYB transcription factors [134]. The *TAS5* tasiRNAs are produced by the phase-initiator miR482, and targets resistance genes [135]. The *TAS6* locus was shown to be targeted by miR156 and miR529 and the resulting tasiRNA can target zinc finger proteins [136]. *TAS7* can also be initiated by miR828 and the *TAS7* tasiRNAs can target genes that codes for leucine-rich receptor protein kinase-like proteins and calcium-transporting ATPase [132].

Some *TAS* loci are flanked by two miRNA binding sites, like *TAS3*, which has two miR390 binding sites [133]. However, the transcript is only cleaved at the 3′ site and the tasiRNAs are synthesised from the 5′ fragment (Figure 7). This is known as the "2-hit" model and the initiator-sRNAs are usually 21 nts in length and involve the interaction with AGO7 [133]. The "1-hit" model includes *TAS* loci where only a single sRNA recognition site is present, the sRNA initiators are 22 nts in length and AGO1 proteins are involved in the miRNA cleavage [129].

Phased siRNAs were shown to be involved in normal plant development [137, 138] and in the plant response to biotic and abiotic stress. In cassava, several phased loci were either induced or repressed in response to cassava bacterial blight [139] and phased loci were also shown to be differentially regulated in response to phosphate deficiency, hypoxia and temperature stress [114, 115, 140].

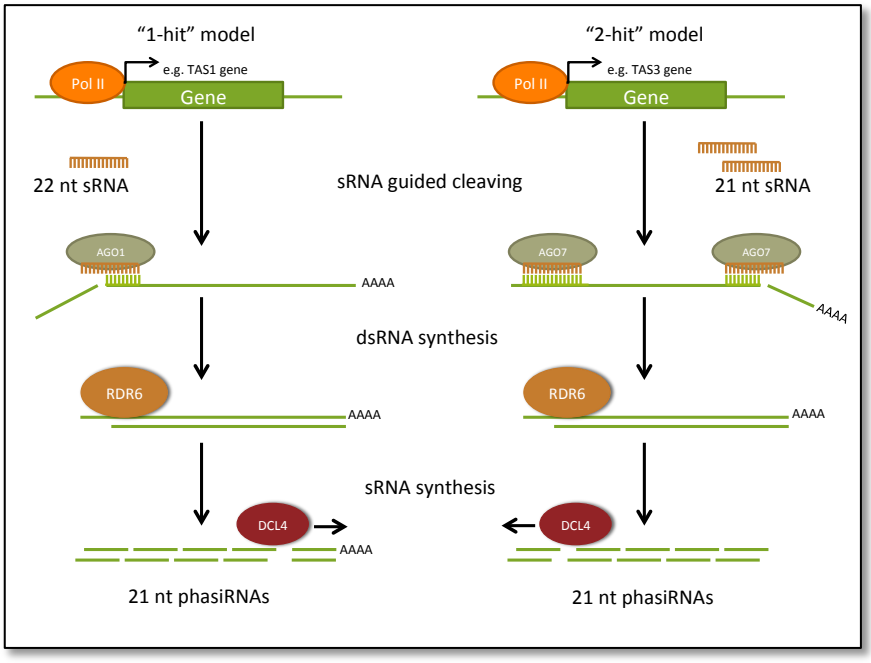


Figure 7. PhasiRNA biogenesis. Illustration of the "1-hit" and "2-hit" models.

2.4.2.2 Natural antisense transcripts siRNAs

Natural antisense transcripts (NATs) have been shown to play a key role in physiological and pathological processes [141]. NATs are endogenous RNA containing sequences complementary to

other transcripts [141–143] and can be classified into two groups based on whether they act in *cis* or *trans*. *Cis*-NAT pairs are transcribed from DNA strands at the same genomic locus but on opposite strands, whereas *trans*-NAT pairs are transcribed from different loci with partial complementarily between the two transcripts [141, 144, 145]. A transcript can either form a duplex with only one other transcript (one-to-one) or be part of multiple NAT pairs forming one-to-many or many-to-many hybridisations [145–148]. The underlying mechanism of NAT regulation are largely unknown, however NATs have been implicated in genomic imprinting, transcriptional interference, and RNA silencing [141, 143, 149]. The natural antisense transcript siRNAs (natsiRNA) originate from the overlapping regions of NATs and can down-regulate the expression of one of the transcripts of the NAT pair [150, 151]. Different mechanisms have been suggested for the biogenesis of plant natsiRNAs including the DCL2-dependent cleavage of a 24 nt natsiRNA from the double stranded duplex region [152] and the 22 nt DCL1-RNA-dependent RNA polymerase 6 cleavage of the duplex [151]. Down-regulation of the transcript due to processing into natsiRNAs can result in either degradation products or guide the cleavage of other copies of the original transcript not part of the NAT pair.

NatsiRNAs have been implicated in development processes, including double fertilisation in *Arabidopsis thaliana* [150], cytokinin regulation in *Petunia hybrida* [153] and regulation of cell-wall biosynthesis in barley [154]. These sRNA species also play a role in plant immunity [139, 151] and are believed to be involved in salt stress in *Arabidopsis thaliana* [152].

2.4.2.3 Heterochromatic siRNAs

Heterochromatic siRNAs or repeat-associated siRNAs (rasiRNAs) originates from intergenic and/or repetitive genomic regions and are mostly 24 nts in length [91]. These repetitive regions in the genome are usually satellite DNAs, integrated virus sequences, retrotransposons and DNA transposable elements. Certain 21 and 22 nt rasiRNAs have also been implicated in transposable element silencing [155–157]. In *Arabidopsis thaliana* it was shown that this species of siRNAs is dependent on DNA-dependent RNA polymerase IV, RNA-dependent RNA polymerase II, DCL3 and AGO4, to be produced from the heterochromatic locus [98, 158–161]. These rasiRNAs can regulate gene expression through DNA and histone modifications [162].

In wheat and *Arabidopsis* the importance of rasiRNAs was shown for optimal plant growth and seed development [163, 164]. The essential role of the siRNA pathway in restricting retrotransposition triggered by environmental stressors such as heat was shown in *Arabidopsis* [165] and phosphorous-deficient conditions lead to the differential expression of rasiRNAs in barley. In cotton plants infected with Cotton leafroll dwarf polerovirus (CLRDV), an overall alteration of transposon-derived small RNAs was observed [120].

2.4.2.4 tRNA-derived siRNAs

Transfer RNAs are well known for their role in translation, however the development of NGS has contributed to the identification of tRNA-derived siRNAs that can have a functional role beyond the direct involvement in translation. The biogenesis and function of these siRNAs have been documented for mammals, but in plants it remains to be elucidated. The classification of tRNA-derived siRNAs is based on the region of tRNA cleavage. Transfer RNAs cleaved in the anticodon loop are known as tRNA halves and are usually between 28-36 nts in length (Figure 8) [102]. Transfer RNA-derived RNA fragments (tRFs) are produced by either cleavage in the D loop (5′ tRFs) or in the T loop (3' CCA tRFs) and are approximately 20 nts in length (Figure 8) [101, 103, 166]. Studies have indicated that due to the precise sequence structure and specific expression patterns, these tRFs are not random tRNA degradation products, but sRNAs with biological function [103, 167].

In general, no correlation has been observed between the levels of tRNA-derived sRNAs and tRNAs, suggesting that the tRNA-derived sRNAs do not inhibit translation by decreasing the level of tRNAs *[103, 168–171]*. Translational repression can occur in a non-sequence-specific manner by the binding of tRFs to ribosomal subunits [172]. Similar regulation of gene expression to miRNAs is also suggested for tRNA-derived sRNAs due to the association of tRFs with AGO proteins in plants [101]. In plants, several reports exist of the association between tRNA-derived sRNAs and stress. These sRNAs have been linked with oxidative stress [171], drought conditions [101], phosphate deprivation [115, 173] and an increase in specific tRFs was observed in *Arabidopsis* infected with *Pseudomonas syringae* [101].

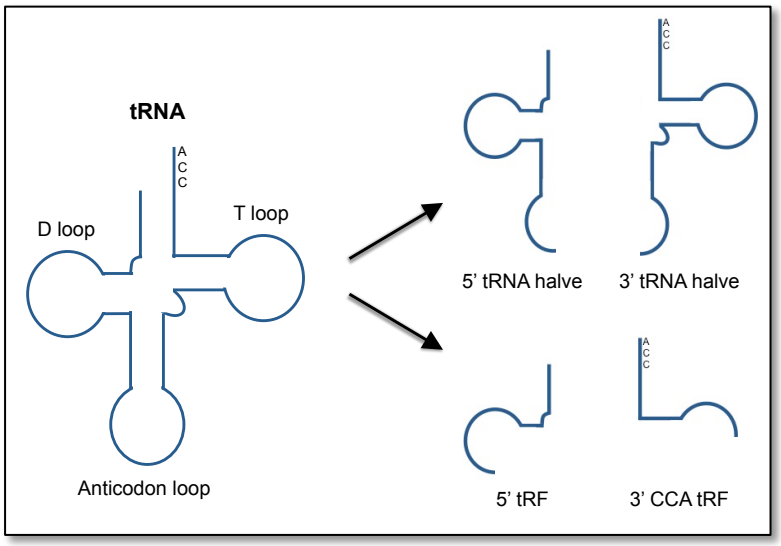


Figure 8. Illustration of tRNA-derived sRNAs. The tRNA halves and tRNA-derived RNA fragments (tRFs) are indicated.

2.4.2.5 Virus-derived siRNAs

Plants have developed RNA silencing as a mechanism to protect themselves against pathogen infections. This mechanism is triggered by the presence of dsRNA and can lead to complementary sequence-dependent target degradation. This host defence system involves the production of virus-derived siRNAs (vsiRNAs) to degrade the virus genome. The counter defence of the virus involves the expression of silencing suppressor proteins encoded by the viral genome or satellite or defective interfering RNA molecules [93, 174, 175]. These molecules can interfere with the plant's silencing machinery or be protected from degradation by secondary structures [93, 174, 175]. The dsRNA precursor of vsiRNAs can be produced from endogenous RNA-dependent RNA polymerase that converts single-stranded viral RNA into dsRNA or from the dsRNA intermediate of RNA virus replication, overlapping transcripts from DNA viruses or a secondary structure formed by the viral genome [174, 176, 177]. The majority of the vsiRNAs are 21 or 22 nts in length due to the processing by DCL4 and DCL2, respectively [176, 178, 179].

It has been hypothesised that it would be possible for host mRNAs to be degraded by vsiRNAs [180–184]. Through *in silico* analysis or experimental analysis it was shown that tobacco mosaic virus [180], cucumber mosaic virus [181, 182], sugarcane mosaic virus [183], grapevine fleck virus [184] and grapevine rupestris stem pitting-associated virus [184] produce vsiRNAs that resulted in the down-regulation of plant host genes.

2.5. Next-generation sequencing and validation

Comparative RNA profiling may lead to the identification of differentially expressed sRNAs and genes that will provide insight into how plants react to viral infections. Next-generation sequencing technologies are fast high-throughput techniques for sequencing DNA in a non-targeted manner without any prior sequence knowledge. The use of universal adaptors, instead of sequence-specific primers, makes NGS particularly suitable to determine the sequences of all the sRNAs and transcripts present in a defined sample [185–187]. During this process novel and rare sRNAs or genes can be identified. Several *in silico* software tools and algorithms have been developed for the extensive data analysis associated with NGS. These include command-line software for *de novo* and reference-based assemblies as well as commercial software packages that function with a user-friendly graphical interface. Specific tools for sRNA analysis or transcriptome assemblies are also available. These tools will enable sRNA detection, comparison of sRNA profiles, novel sRNA predictions and identifying sRNA gene targets or differentially expressed genes. However, the use of these tools will only lead to *in silico*-predicted results and requires experimental validation.

Microarrays are based on a fluorescent probe hybridisation platform for high throughput processing of

a large number of miRNAs or mRNAs. Expression profiling can be performed to assess the differential expression of sRNAs or genes in different samples [110, 188–193]. Northern blot analysis is another technique used to detect sRNA and gene expression using hybridisation probes [194–200]. In addition to miRNAs and northern blots, RT-qPCR can be used to validate differentially expressed sRNAs and gene transcripts [92, 121, 201, 202]. Due to the small size of sRNAs, amplification of these molecules through PCR can be challenging. This leads to the development of specific quantitative RT-PCR techniques optimised for miRNA detection, utilising either lock nucleic acid (LNA) technology or stemloop reverse transcription primers [121, 203–208]. The insertion of LNAs into the oligonucleotide primers increases the binding affinity of the primers and as a result a higher melting temperature can be used, resulting in increased specificity. The stemloop reverse transcription primers provide better specificity and sensitivity due to the base stacking and spatial constraint provided by the attached stemloop structure. Even though microarrays, northern blots and RT-qPCRs can be used for sRNA or gene profiling and therefore validation of NGS data, novel sRNAs or transcripts cannot be identified.

To validate potential sRNA targets through predicted cleavage sites, rapid amplification of cDNA ends (RACE) [201, 209, 210] or degradome sequencing can be performed. The 5′ sequence of the corresponding cleaved mRNA can be amplified using 5′ RACE, while degradome sequencing involves obtaining the 5′ ends of all RNA degradation products using NGS to identify patterns of RNA degradation [211–214].

2.5.1 Small RNA in silico analysis

MicroRNAs, being the best characterised sRNA species, resulted in the development of numerous software tools and algorithms for the analysis of NGS data. A few miRNA NGS data analysis tools include include miRDeep [215] or mirDeep-P (plant-specific version) [216], miR-PREFeR [217], MirPlex [218], miRSeqNovel [219], miRA [220], miReader [221], MIReNA [222], miR-BAG [223], omiRas [224], miRanalyzer [225, 226], ShortStack [227, 228] and the UEA sRNA Workbench utilising miRCat and miRProf [229]. These tools incorporate a range of functions including miRNA identification, prediction, differential expression and the prediction of miRNA targets. Some of these miRNA prediction tools utilise machine-learning methodology, resulting in predictions strongly dependent on the training data set used. These tools also differ in the level of user control and in some cases limiting the options for manual adjustment of parameters.

In order to identify novel miRNAs in NGS data, reads are mapped onto the host's reference genome to identify sRNA read clusters. These cluster regions will then be evaluated for their ability to form hairpin structures with the structural and thermodynamic properties associated with miRNA precursors. These criteria for plant miRNA annotation [230] have been well established and most

prediction tools abide by them. The Vienna RNA package is used by multiple programs to predict the secondary structure of a potential miRNA precursor [231].

The most popular alignment tools for short reads are Bowtie [232], Burrows-Wheeler Alignment tool (BWA) [233], MAQ [234] and SOAP [235]. Bowtie, MAQ and BWA implements the Burrows-Wheeler transform (BWT) algorithm for reference genome indexing and are optimised for fast alignment of large numbers of short reads with a low memory requirement. While SOAP can do both gapped and ungapped alignments [235], Bowtie is unable to account for indels. However Bowtie's sensitivity is comparable to that of SOAP and the lower memory footprint [232], together with the specificity required for the sRNA read mapping to the genome, makes Bowtie well-suited for sRNA analysis.

To identify known miRNAs, databases like the miRNA registry database (miRBase), can be queried to identify similar sequences [236]. The miRBase database (version 21) contains 35828 mature miRNA sequences of which 4828 sequences represent non-redundant plant miRNAs, including 119 unique *Vitis vinifera* miRNAs. These *Vitis vinifera* miRNAs were predominantly identified in cv. Pinot noir [237, 238].

Target prediction of miRNAs can be performed with *in silico* analysis of reverse complementary matching between the sRNA and target transcript or by combining the *in silico* analysis with degradome sequencing to validate miRNA cleavage sites. Tools for degradome analysis include CleaveLand [239], PAREsnip [240] and SeqTar [241]. psRNATarget is a degradome-independent miRNA prediction tool that can distinguish between translational and post-transcriptional inhibition and reports the number of small RNA/target site pairs by adhering to the miRNA target recognition criteria [242]. The present sRNA target prediction tools rely on miRNA-associated target recognition. Since the functioning of the other sRNA species is largely unknown; it remains to be demonstrated if these sRNAs has the same target recognition characteristics as miRNAs and if these tools will be applicable for all sRNA target predictions.

To facilitate the identification of non-miRNAs, tools like SeqCluster [243], segmentSeq [244], PhaseTank [245], ShortStack [227, 228] and the UEA sRNA Workbench [229] have been developed. However, these tools focus mostly on tasiRNAs or phasiRNAs identification and for sRNA species like natsiRNA, rasiRNA and tRNA-derived sRNAs, short-read alignment tools combined with custom scripts are mainly used for analyses.

2.5.2 Transcriptome in silico analysis

Next-generation sequencing of total RNA or mRNA is capable of discovering new genes and

transcripts, and measure the level of gene expression in a single assay. Likewise, cell- or condition-specific transcripts and a full collection of alternative splice isoforms are attainable. Additionally, since the number of reads produced from an RNA transcript is a function of transcript abundance, read density can be used for differential expression analysis. In order to convert raw sequence reads to transcripts either a reference-guided or a *de novo* assembly is necessary. The reference-guided assembly approach is suited for organisms with a well-assembled genome, while *de novo* assembly of reads allows transcriptome analysis without the need for a genome sequence, as is the case for many non-model organisms.

There are several considerations regarding transcriptome assembly that should be included in the experimental design. The choice of the library construction strategy will influence transcriptome assembly. Depletion methods, like the removal of ribosomal RNA (rRNA) or poly(A) selection, can bias the quantitation of genes towards highly abundant transcripts and rare transcripts may not be assembled [246]. Strand-specific RNA sequencing library preparation can aid in transcript assembly, facilitating the assembly of overlapping genes on opposite strands of the genome [247]. Additionally, the choice between single-end or paired-end data will influence the assembly strategy, as single-end data will be best suited for a reference-guided approach [246]. A poly(A) selected library of single-end reads can produce fragmented transcripts if a *de novo* assembly is attempted since only a small fraction of the transcript will be covered [248]. Longer sequencing reads will also reduce the complexity of the assembly [249]. A reference-based strategy can be very sensitive in assembling transcripts with a low expression level [250], however the quality of the reference genome will determine the success of the assembly. Reads that also mapped to multiple locations in the genome can lead to false positive results [249]. *De novo* assemblies have the advantage that it will be able to assemble unknown transcripts not included in the genome, although transcripts with high sequence similarity will probably assemble into one transcript [246].

The best known example of a reference-guided assembler is the tuxedo pipeline, including TopHat and Cufflinks [251]. TopHat utilises Bowtie2 [252] for alignment of reads to the reference genome, thereby addressing Bowtie's limitation of not being able to align reads that span introns. Subsequently, Cufflinks assemble the mapped reads into transcripts and Cuffdiff is used for differential expression analysis. Examples of *de novo* assembly tools include Trinity [248, 253], Oases [254] and SOAPdenovo-Trans [255]. Trinity is able to assemble transcripts over a broad range of expression levels and was specifically designed for transcriptome assembly, in contrast to other short-read assembly tools. Although Trinity can use both single- and paired-end reads, it was found that paired-end reads increased the distance at which it could resolve ambiguities [248]. Analysis with Oases showed that spliced variants are better detected by adjusting the assembly parameters to the different gene expression levels in a sample. It was shown to be beneficial to merge assemblies created from a

range of parameters for a more sensitive and more specific transcriptome assembly. [254]. Following the development of Trinity and Oases, SOAPdenovo-Trans was created to incorporate the novel ideas of previous transcriptome assemblers. It was found to provide a lower redundancy and faster execution than Oases and Trinity, except it is not yet able to perform strand-specific assemblies [255].

2.5.3 Statistical assessment of differential expression

Various tools have been developed to assess statistical significant differences between biological conditions analysed using NGS read count data. The majority of these tools are implemented in the R statistical computing environment and is available as open source software from Bioconducter. These tools mainly differ in their normalisation methods and the read count distribution assumption. The variation observed between biological replicates is often higher than expected by a poisson distribution due to overdispersion. To compensate for this variability, negative binomial models were introduced [256]. Examples of negative binomial model-associated tools include edgeR [257], DESeq [258] and baySeq [259] [6], non-parametric model examples are NOIseq [260] and SAMseq [261], limma is based on linear modeling [262, 263] and Cuffdiff [264] and EBSeq [265] represents tools based on a beta-negative binomial model and an empirical Bayesian model, respectively. However, the most frequently used tools are edgeR, DESeq2, limma and Cuffdiff.

The edgeR method uses empirical Bayes estimation and an exact test adapted for overdispersion to assess differential expression. It enables analysis with a low number of replicates and normalisation is performed using the Trimmed Mean of M value (TMM) normalisation procedure [257]. DESeq models the relationship between the mean and variance to estimate dispersion. DESeq first calculates the gene specific ratio per sample by dividing the read count by the geometric mean of a specific gene's read counts across all samples. Then a size factor is estimated for each sample by calculating the median of all the gene-specific ratios for a sample [258]. DESeq also allows for a small number of replicates. After comparing DESeq to other methods, it was considered to be too strict and as a result a new version, DESeq2, was developed to better balance sensitivity with the false positive rate [266]. Limma was originally designed for microarray analysis, but was extended for NGS read count data. Limma also recommends a TMM normalisation after which the normalised counts is transformed to a logarithmic scale and a weight determined for each observation by estimating the mean–variance relationship [262, 263]. Linear modelling is applied for statistical testing. Cuffdiff controls for both read mapping ambiguity and variability. To account for the different sequencing depths of the different samples, a similar scaling factor method as DESeq is applied [264].

The identification of differentially expressed genes relies on the execution of numerous statistical tests and therefore it is important to control for multiple testing to assess the significance of the difference observed. The family-wise error rate correction is often too conservative in biological scenarios and

therefore the false discovery rate (FDR) is controlled by correcting the p-values using for example the Benjamini–Hochberg method [267]. The Benjamini–Hochberg method is applied by default in the edgeR, DESeq2, limma and Cuffdiff tools.

The accurate identification of differentially expressed RNA molecules is strongly dependent on the number of replicate samples, thus an increase in the number of replicates, rather than sequencing depth, will add power to the analysis [268]. The choice of statistical method will depend on different variables, including the number of replicates, multi-factored studies, and whether detection included alternative spliced transcripts [256, 269]. The source of the RNA counts will also determine the choice of statistical method. [269]. Quantile normalisation is a non-scaling method that assumes most RNA species are not differentially expressed, and that the expression level distribution is equal across all the samples being normalised [269]. This approach is frequently used in microarray data analysis [270], though not always applicable for sRNA NGS data [269, 271].

2.6 Conclusion

Grapevine leafroll disease is detrimental to normal plant development and can cause significant yield losses that will negatively impact on the sustainability of the grapevine industry. Research in GLRaV-3 has mainly focussed on detection and epidemiology, with limited studies on the host-virus interaction. Plants have developed mechanisms to regulate gene expression in response to biotic stress, of which the production of sRNAs is one. Characterising the differentially expressed sRNA species and (target) genes in healthy and diseased plants will aid the unravelling of the grapevine stress response to viral infection.

The aim of this study was to use NGS and bioinformatic analysis tools to analyse sRNA and gene expression profiles to contribute to our understanding of the molecular mechanisms behind the virus-triggered changes in grapevine physiology.

2.7 References

1. Janick J, Moore JN (1997) Fruit Breeding, 3 Volume Set. Wiley
2. Martelli GP (2014) Directory of virus and virus-like diseases of the grapevine and their agents. J Plant Pathol 96:1–136.
3. Al Rwahnih M, Daubert S, Golino D, Rowhani A (2009) Deep sequencing analysis of RNAs from a grapevine showing Syrah decline symptoms reveals a multiple virus infection that includes a novel virus. Virology 387:395–401. doi: 10.1016/j.virol.2009.02.028
4. Zhang Y, Singh K, Kaur R, Qiu W (2011) Association of a novel DNA virus with the grapevine vein-clearing and vine decline syndrome. Phytopathology 101:1081–1090. doi: 10.1094/PHYTO-02-11-0034
5. Giampetruzzi A, Roumi V, Roberto R, et al (2012) A new grapevine virus discovered by deep sequencing of virus- and viroid-derived small RNAs in Cv Pinot gris. Virus Res 163:262–268. doi: 10.1016/j.virusres.2011.10.010

6. Al Rwahnih M, Sudarshana MR, Uyemoto JK, Rowhani A (2012) Complete Genome Sequence of a Novel Vitivirus Isolated from Grapevine. J Virol 86:9545–9545. doi: 10.1128/JVI.01444-12

7. Molenaar N, Burger JT, Maree HJ (2015) Detection of a divergent variant of grapevine virus F by next-generation sequencing. Arch Virol 160:2125–2127. doi: 10.1007/s00705-015-2466-3

8. Al Rwahnih M, Dave A, Anderson MM, et al (2013) Association of a DNA virus with grapevines affected by red blotch disease in California. Phytopathology 103:1069–1076. doi: 10.1094/PHYTO-10-12-0253-R

9. Maree HJ, Almeida RPP, Bester R, et al (2013) Grapevine leafroll-associated virus 3. Front Microbiol 4:82. doi: 10.3389/fmicb.2013.00082

10. Naidu R, Rowhani A, Fuchs M, et al (2014) Grapevine Leafroll: A Complex Viral Disease Affecting a High-Value Fruit Crop. Plant Dis 98:1172–1185. doi: 10.1094/PDIS-08-13-0880-FE

11. Almeida RPP, Daane KM, Bell VA, et al (2013) Ecology and management of grapevine leafroll disease. Front Microbiol 4:94. doi: 10.3389/fmicb.2013.00094

12. Atallah SS, Gomez MI, Fuchs MF, Martinson TE (2012) Economic Impact of Grapevine Leafroll Disease on Vitis vinifera cv. Cabernet franc in Finger Lakes Vineyards of New York. Am J Enol Vitic 63:73–79. doi: 10.5344/ajev.2011.11055

13. Woodham RC, Antcliff A., Krake LR, Taylor R. (1984) Yield differences between Sultana clones related to virus status and genetic factors. Vitis 23:73–83.

14. Wolpert JA, Vilas EP (1992) Effect of Mild Leafroll Disease on Growth, Yield, and Fruit Maturity Indices of Riesling and Zinfandel. Am J Enol Vitic 43:367–369.

15. Gutha LR, Casassa LF, Harbertson JF, Naidu RA (2010) Modulation of flavonoid biosynthetic pathway genes and anthocyanins due to virus infection in grapevine (Vitis vinifera L.) leaves. BMC Plant Biol 10:187. doi: 10.1186/1471-2229-10-187

16. Mannini F, Mollo A, Credi R (2011) Field Performance and Wine Quality Modification in a Clone of Vitis vinifera cv. Dolcetto after GLRaV-3 Elimination. Am J Enol Vitic 63:144–147. doi: 10.5344/ajev.2011.11020

17. Cabaleiro C, Segura A, García-Berrios JJ (1999) Effects of Grapevine Leafroll-Associated Virus 3 on the Physiology and Must of Vitis vinifera L. cv. Albariño Following Contamination in the Field. Am J Enol Vitic 50:40–44.

18. Espinoza C, Medina C, Somerville S, Arce-Johnson P (2007) Senescence-associated genes induced during compatible viral interactions with grapevine and Arabidopsis. J Exp Bot 58:3197–3212. doi: 10.1093/jxb/erm165

19. Vega A, Gutierrez RA, Pena-Neira A, et al (2011) Compatible GLRaV-3 viral infections affect berry ripening decreasing sugar accumulation and anthocyanin biosynthesis in Vitis vinifera. Plant Mol Biol 77:261–274. doi: 10.1007/s11103-011-9807-8

20. Endeshaw ST, Sabbatini P, Romanazzi G, et al (2014) Effects of grapevine leafroll associated virus 3 infection on growth, leaf gas exchange, yield and basic fruit chemistry of Vitis vinifera L. cv. Cabernet Franc. Sci Hortic 170:228–236. doi: 10.1016/j.scienta.2014.03.021

21. Charles JG, Cohen D, Walker JTS, et al (2006) A review of the ecology of grapevine leafroll associated virus type 3 (GLRaV-3). N Z Plant Prot 59:330–337.

22. Charles J, Froud K, van den B, Allan D (2009) Mealybugs and the spread of grapevine leafroll-associated virus 3 (GLRaV-3) in a New Zealand vineyard. Australas Plant Pathol 38:576–583. doi: 10.1071/AP09042

23. Tsai C-W, Chau J, Fernandez L, et al (2008) Transmission of grapevine leafroll-associated virus 3 by the vine mealybug (Planococcus ficus). Phytopathology 98:1093–1098. doi: 10.1094/PHYTO-98-10-1093

24. Tsai C-W, Rowhani A, Golino DA, et al (2010) Mealybug transmission of Grapevine leafroll viruses: an analysis of virus-vector specificity. Phytopathology 100:830–834. doi: 10.1094/PHYTO-100-8-0830

25. Martelli GP, Agranovsky AA, Bar-Joseph M, et al (2002) The family Closteroviridae revised. Arch Virol 147:2039–2044. doi: 10.1007/s007050200048

26. Bell VA, Bonfiglioli RGE, Walker JTS, et al (2009) Grapevine leafroll-associated virus 3 persistence in Vitis vinifera remnant roots. J Plant Pathol 91:527–533. doi: 10.4454/jpp.v91i3.543

27. Martelli GP, Abou Ghanem-Sabanadzovic N, Agranovsky AA, et al (2012) Taxonomic revision of the family Closteroviridae with special reference to the grapevine leafroll-associated members of the genus Ampelovirus and the putative species unassigned to the family. J Plant Pathol 94:7–19. doi: 10.4454/jpp.fa.2012.022

28. Al Rwahnih M, Dolja VV, Daubert S, et al (2012) Genomic and biological analysis of Grapevine leafroll-associated virus 7 reveals a possible new genus within the family Closteroviridae. Virus Res 163:302–309. doi: 10.1016/j.virusres.2011.10.018

29. Ling KS, Zhu HY, Alvizo H, et al (1997) The coat protein gene of grapevine leafroll associated closterovirus-3: cloning, nucleotide sequencing and expression in transgenic plants. Arch Virol 142:1101–1116. doi: 10.1007/s007050050145

30. Ling K-S (2004) Complete nucleotide sequence and genome organization of Grapevine leafroll-associated virus 3, type member of the genus Ampelovirus. J Gen Virol 85:2099–2102. doi: 10.1099/vir.0.80007-0

31. Maree H, Freeborough M, Burger J (2008) Complete nucleotide sequence of a South African isolate of grapevine leafroll-associated virus 3 reveals a 5'UTR of 737 nucleotides. Arch Virol 153:755–757. doi: 10.1007/s00705-008-0040-y

32. Agranovsky AA, Koonin EV, Boyko VP, et al (1994) Beet yellows closterovirus: Complete genome structure and identification of a leader papain-like thiol protease. Virology 198:311 – 324. doi: http://dx.doi.org/10.1006/viro.1994.1034

33. Ling KS, Zhu HY, Drong RF, et al (1998) Nucleotide sequence of the 3'-terminal two-thirds of the grapevine leafroll-associated virus-3 genome reveals a typical monopartite closterovirus. J Gen Virol 79:1299–1307.

34. King AMQ, Lefkowitz E, Adams MJ, Carstens EB (2011) Virus Taxonomy: Ninth Report of the International Committee on Taxonomy of Viruses. Elsevier Science

35. Koonin EV, Dolja VV, Morris TJ (1993) Evolution and Taxonomy of Positive-Strand RNA Viruses: Implications of Comparative Analysis of Amino Acid Sequences. Crit Rev Biochem Mol Biol 28:375–430. doi: 10.3109/10409239309078440

36. Dolja VV, Kreuze JF, Valkonen JPT (2006) Comparative and functional genomics of closteroviruses. Virus Res 117:38–51. doi: 10.1016/j.virusres.2006.02.002

37. Ling K-S, Zhu H-Y, Gonsalves D (2004) Complete nucleotide sequence and genome organization of Grapevine leafroll-associated virus 3, type member of the genus Ampelovirus. J Gen Virol 85:2099–2102. doi: 10.1099/vir.0.80007-0

38. Klaassen VA, Mayhew D, Fisher D, Falk BW (1996) In Vitro Transcripts from Cloned cDNAs of the Lettuce Infectious Yellows Closterovirus Bipartite Genomic RNAs Are Competent for Replication inNicotiana benthamianaProtoplasts. Virology 222:169–175. doi: 10.1006/viro.1996.0407

39. Peremyslov VV, Hagiwara Y, Dolja VV (1998) Genes required for replication of the 15.5-kilobase RNA genome of a plant closterovirus. J Virol 72:5870–5876.

40. Peng C-W, Dolja VV (2000) Leader proteinase of the beet yellows closterovirus: mutation analysis of the function in genome amplification. J Virol 74:9766–9770. doi: 10.1128/JVI.74.20.9766-9770.2000

41. Peng C-W, Napuli AJ, Dolja VV (2003) Leader Proteinase of Beet Yellows Virus Functions in Long-Distance Transport. J Virol 77:2843–2849. doi: 10.1128/JVI.77.5.2843-2849.2003

42. Liu Y-P, Peremyslov VV, Medina V, Dolja VV (2009) Tandem leader proteases of Grapevine leafroll-associated virus-2: Host-specific functions in the infection cycle. Virology 383:291–299. doi: 10.1016/j.virol.2008.09.035

43. Van den Born E, Omelchenko MV, Bekkelund A, et al (2008) Viral AlkB proteins repair RNA damage by oxidative demethylation. Nucleic Acids Res 36:5451–5461. doi: 10.1093/nar/gkn519

44. Peremyslov VV, Pan Y-W, Dolja VV (2004) Movement Protein of a Closterovirus Is a Type III Integral Transmembrane Protein Localized to the Endoplasmic Reticulum. J Virol 78:3704–3709. doi: 10.1128/JVI.78.7.3704-3709.2004

45. Peremyslov VV, Hagiwara Y, Dolja VV (1999) HSP70 homolog functions in cell-to-cell movement of a plant virus. Proc Natl Acad Sci 96:14771–14776. doi: 10.1073/pnas.96.26.14771

46. Alzhanova DV, Prokhnevsky AI, Peremyslov VV, Dolja VV (2007) Virion tails of Beet yellows virus: coordinated assembly by three structural proteins. Virology 359:220–226. doi: 10.1016/j.virol.2006.09.007

47. Alzhanova DV, Hagiwara Y, Peremyslov VV, Dolja VV (2000) Genetic Analysis of the Cell-to-Cell Movement of Beet Yellows Closterovirus. Virology 268:192–200. doi: 10.1006/viro.1999.0155

48. Agranovsky AA, Lesemann DE, Maiss E, et al (1995) "Rattlesnake" structure of a filamentous plant RNA virus built of two capsid proteins. Proc Natl Acad Sci 92:2470–2473.

49. Satyanarayana T, Gowda S, Ayllón MA, Dawson WO (2004) Closterovirus bipolar virion: evidence for initiation of assembly by minor coat protein and its restriction to the genomic RNA 5′ region. Proc Natl Acad Sci U S A 101:799–804. doi: 10.1073/pnas.0307747100

50. Reed JC, Kasschau KD, Prokhnevsky AI, et al (2003) Suppressor of RNA silencing encoded by Beet yellows virus. Virology 306:203–209. doi: 10.1006/S0042-6822(02)00051-X

51. Lu R, Folimonov A, Shintaku M, et al (2004) Three distinct suppressors of RNA silencing encoded by a 20-kb viral RNA genome. Proc Natl Acad Sci U S A 101:15742–15747. doi: 10.1073/pnas.0404940101

52. Chiba M, Reed JC, Prokhnevsky AI, et al (2006) Diverse suppressors of RNA silencing enhance agroinfection by a viral replicon. Virology 346:7–14. doi: 10.1016/j.virol.2005.09.068

53. Prokhnevsky AI, Peremyslov VV, Napuli AJ, Dolja VV (2002) Interaction between Long-Distance Transport Factor and Hsp70-Related Movement Protein of Beet Yellows Virus. J Virol 76:11003–11011. doi: 10.1128/JVI.76.21.11003-11011.2002

54. Gouveia P, Nolasco G (2012) The p19.7 RNA silencing suppressor from Grapevine leafroll-associated virus 3 shows different levels of activity across phylogenetic groups. Virus Genes 45:333–339. doi: 10.1007/s11262-012-0772-3

55. Bustamante PI, Hull R (1998) Plant virus gene expression strategies. Electron J Biotechnol 1:13–24.

56. Karasev AV, Agranovsky AA, Rogov VV, et al (1989) Virion RNA of Beet Yellows Closterovirus: Cell-free Translation and Some Properties. J Gen Virol 70:241–245. doi: 10.1099/0022-1317-70-1-241

57. Maree HJ, Gardner HFJ, Freeborough M-J, Burger JT (2010) Mapping of the 5' terminal nucleotides of Grapevine leafroll-associated virus 3 sgRNAs. Virus Res 151:252–255. doi: 10.1016/j.virusres.2010.05.011

58. Jarugula S, Gowda S, Dawson WO, Naidu RA (2010) 3'-coterminal subgenomic RNAs and putative cis-acting elements of Grapevine leafroll-associated virus 3 reveals "unique" features of gene expression strategy in the genus Ampelovirus. Virol J 7:180. doi: 10.1186/1743-422X-7-180

59. Miller WA, Koev G (2000) Synthesis of Subgenomic RNAs by Positive-Strand RNA Viruses. Virology 273:1–8. doi: 10.1006/viro.2000.0421

60. Chooi KM, Cohen D, Pearson MN (2013) Molecular characterisation of two divergent variants of grapevine leafroll-associated virus 3 in New Zealand. Arch Virol 158:1597–1602. doi: 10.1007/s00705-013-1631-9

61. Jooste AEC, Maree HJ, Bellstedt DU, et al (2010) Three genetic grapevine leafroll-associated virus 3 variants identified from South African vineyards show high variability in their 5'UTR. Arch Virol 155:1997–2006. doi: 10.1007/s00705-010-0793-y

62. Bester R, Maree HJ, Burger JT (2012) Complete nucleotide sequence of a new strain of grapevine leafroll-associated virus 3 in South Africa. Arch Virol 157:1815–1819. doi: 10.1007/s00705-012-1333-8

63. Seah Y, Sharma AM, Zhang S, Almeida RP (2012) A divergent variant of Grapevine leafroll-associated virus 3 is present in California. Virol J 9:235. doi: 10.1186/1743-422X-9-235

64. Fei F, Lyu M-D, Li J, et al (2013) Complete nucleotide sequence of a Chinese isolate of Grapevine leafroll-associated virus 3 reveals a 5' UTR of 802 nucleotides. Virus Genes 46:182–185. doi: 10.1007/s11262-012-0823-9

65. Maree HJ, Pirie MD, Oosthuizen K, et al (2015) Phylogenomic Analysis Reveals Deep Divergence and Recombination in an Economically Important Grapevine Virus. PLOS ONE 10:e0126819. doi: 10.1371/journal.pone.0126819

66. Gouveia P, Santos MT, Eiras-Dias JE, Nolasco G (2011) Five phylogenetic groups identified in the coat protein gene of grapevine leafroll-associated virus 3 obtained from Portuguese grapevine varieties. Arch Virol 156:413–420. doi: 10.1007/s00705-010-0878-7

67. Jooste AEC, Molenaar N, Maree HJ, et al (2015) Identification and distribution of multiple virus infections in Grapevine leafroll diseased vineyards. Eur J Plant Pathol 142:363–375. doi: 10.1007/s10658-015-0620-0

68. Ling K, Zhu H, Jiang Z, Gonsalves D (2000) Effective Application of DAS-ELISA for Detection of Grapevine Leafroll Associated Closterovirus-3 Using a Polyclonal Antiserum Developed from Recombinant Coat Protein. Eur J Plant Pathol 106:301–309. doi: 10.1023/A:1008713319294

69. Ling K-S, Zhu H-Y, Petrovic N, Gonsalves D (2001) Comparative Effectiveness of ELISA and RT-PCR for Detecting Grapevine Leafroll-Associated Closterovirus-3 in Field Samples. Am J Enol Vitic 52:21–27.

70. Bester R, Jooste AEC, Maree HJ, Burger JT (2012) Real-time RT-PCR high-resolution melting curve analysis and multiplex RT-PCR to detect and differentiate grapevine leafroll-associated virus 3 variant groups I, II, III and VI. Virol J 9:219. doi: 10.1186/1743-422X-9-219

71. Chooi KM, Cohen D, Pearson MN (2013) Generic and sequence-variant specific molecular assays for the detection of the highly variable Grapevine leafroll-associated virus 3. J Virol Methods 189:20–29. doi: 10.1016/j.jviromet.2012.12.018

72. Dovas CI, Katis NI (2003) A spot multiplex nested RT-PCR for the simultaneous and generic detection of viruses involved in the aetiology of grapevine leafroll and rugose wood of grapevine. J Virol Methods 109:217–226. doi: 10.1016/S0166-0934(03)00074-0

73. Osman F, Leutenegger C, Golino D, Rowhani A (2007) Real-time RT-PCR (TaqMan®) assays for the detection of Grapevine Leafroll associated viruses 1-5 and 9. J Virol Methods 141:22–29. doi: 10.1016/j.jviromet.2006.11.035

74. Jooste AEC, Goszczynski DE (2005) Single-strand conformation polymorphism (SSCP), cloning and sequencing reveal two major groups of divergent molecular variants of grapevine leafroll-associated virus 3 (GLRaV-3). Vitis 44:39–43.

75. Engel EA, Escobar PF, Rojas LA, et al (2010) A diagnostic oligonucleotide microarray for simultaneous detection of grapevine viruses. J Virol Methods 163:445–451. doi: 10.1016/j.jviromet.2009.11.009

76. Thompson JR, Fuchs M, Perry KL (2012) Genomic analysis of grapevine leafroll associated virus-5 and related viruses. Virus Res. doi: 10.1016/j.virusres.2011.08.006

77. Bester R, Pepler PT, Burger JT, Maree HJ (2014) Relative quantitation goes viral: An RT-qPCR assay for a grapevine virus. J Virol Methods 210:67–75. doi: 10.1016/j.jviromet.2014.09.022

78. Coetzee B, Freeborough M-J, Maree HJ, et al (2010) Deep sequencing analysis of viruses infecting grapevines: Virome of a vineyard. Virology 400:157–163. doi: 10.1016/j.virol.2010.01.023

79. Al Rwahnih M, Daubert S, Úrbez-Torres JR, et al (2010) Deep sequencing evidence from single grapevine plants reveals a virome dominated by mycoviruses. Arch Virol 156:397–403. doi: 10.1007/s00705-010-0869-8

80. Coetzee B, Maree HJ, Stephan D, et al (2010) The first complete nucleotide sequence of a grapevine virus E variant. Arch Virol 155:1357–1360. doi: 10.1007/s00705-010-0685-1

81. Kreuze JF, Perez A, Untiveros M, et al (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. doi: 10.1016/j.virol.2009.03.024

82. Bertamini M, Muthuchelian K, Nedunchezhian N (2004) Effect of Grapevine Leafroll on the Photosynthesis of Field Grown Grapevine Plants (Vitis vinifera L. cv. Lagrein). J Phytopathol 152:145–152. doi: 10.1111/j.1439-0434.2004.00815.x

83. Cretazzo E, Padilla C, Carambula C, et al (2010) Comparison of the effects of different virus infections on performance of three Majorcan grapevine cultivars in field conditions. Ann Appl Biol 156:1–12. doi: 10.1111/j.1744-7348.2009.00355.x

84. Espinoza C, Vega A, Medina C, et al (2007) Gene expression associated with compatible viral diseases in grapevine cultivars. Funct Integr Genomics 7:95–110. doi: 10.1007/s10142-006-0031-6

85. Alabi OJ, Casassa LF, Gutha LR, et al (2016) Impacts of Grapevine Leafroll Disease on Fruit Yield and Grape and Wine Chemistry in a Wine Grape (Vitis vinifera L.) Cultivar. PLOS ONE 11:e0149666. doi: 10.1371/journal.pone.0149666

86. Kang H-C, Yoon S-H, Lee C-M, Koo B-S (2012) Non-coding RNAs Associated with Biotic and Abiotic Stresses in Plants. J Appl Biol Chem 55:71–77. doi: 10.3839/jabc.2011.062

87. Kruszka K, Pieczynski M, Windels D, et al (2012) Role of microRNAs and other sRNAs of plants in their changing environments. J Plant Physiol 169:1664–1672. doi: 10.1016/j.jplph.2012.03.009

88. Sunkar R, Li Y-F, Jagadeeswaran G (2012) Functions of microRNAs in plant stress responses. Trends Plant Sci 17:196–203. doi: 10.1016/j.tplants.2012.01.010

89. Bartel DP (2004) MicroRNAs: genomics, biogenesis, mechanism, and function. Cell 116:281–297. doi: 10.1016/S0092-8674(04)00045-5

90. Khraiwesh B, Zhu J-K, Zhu J (2012) Role of miRNAs and siRNAs in biotic and abiotic stress responses of plants. Biochim Biophys Acta BBA - Gene Regul Mech 1819:137–148. doi: 10.1016/j.bbagrm.2011.05.001

91. Guleria P, Mahajan M, Bhardwaj J, Yadav SK (2011) Plant Small RNAs: Biogenesis, Mode of Action and Their Roles in Abiotic Stresses. Genomics Proteomics Bioinformatics 9:183–199. doi: 10.1016/S1672-0229(11)60022-3

92. Alabi OJ, Zheng Y, Jagadeeswaran G, et al (2012) High-throughput sequence analysis of small RNAs in grapevine Vitis vinifera affected by grapevine leafroll disease: Small RNAs in leafroll disease-infected grapevine. Mol Plant Pathol 13:1060–1076. doi: 10.1111/j.1364-3703.2012.00815.x

93. Chapman EJ (2004) Viral RNA silencing suppressors inhibit the microRNA pathway at an intermediate step. Genes Dev 18:1179–1186. doi: 10.1101/gad.1201204

94. Bazzini AA, Hopp HE, Beachy RN, Asurmendi S (2007) Infection and coaccumulation of tobacco mosaic virus proteins alter microRNA levels, correlating with symptom and plant development. Proc Natl Acad Sci 104:12157–12162. doi: 10.1073/pnas.0705114104

95. Pantaleo V, Saldarelli P, Miozzi L, et al (2010) Deep sequencing analysis of viral short RNAs from an infected Pinot Noir grapevine. Virology 408:49–56. doi: 10.1016/j.virol.2010.09.001

96. Singh K, Talla A, Qiu W (2012) Small RNA profiling of virus-infected grapevines: evidences for virus infection-associated and variety-specific miRNAs. Funct Integr Genomics 12:659–669. doi: 10.1007/s10142-012-0292-1

97. Pantaleo V, Vitali M, Boccacci P, et al (2016) Novel functional microRNAs from virus-free and infected Vitis vinifera plants under water stress. Sci Rep 6:20167. doi: 10.1038/srep20167

98. Axtell MJ (2013) Classification and comparison of small RNAs from plants. Annu Rev Plant Biol 64:137–159. doi: 10.1146/annurev-arplant-050312-120043

99. Garcia-Silva MR, Cabrera-Cabrera F, Güida MC, Cayota A (2012) Hints of tRNA-Derived Small RNAs Role in RNA Silencing Mechanisms. Genes 3:603–614. doi: 10.3390/genes3040603

100. Lambertz U, Oviedo Ovando ME, Vasconcelos E, et al (2015) Small RNAs derived from tRNAs and rRNAs are highly enriched in exosomes from both old and new world Leishmania providing evidence for conserved exosomal RNA Packaging. BMC Genomics 16:151. doi: 10.1186/s12864-015-1260-7

101. Loss-Morais G, Waterhouse PM, Margis R (2013) Description of plant tRNA-derived RNA fragments (tRFs) associated with argonaute and identification of their putative targets. Biol Direct 8:1–5. doi: 10.1186/1745-6150-8-6

102. Lee SR, Collins K (2005) Starvation-induced Cleavage of the tRNA Anticodon Loop in Tetrahymena thermophila. J Biol Chem 280:42744–42749. doi: 10.1074/jbc.M510356200

103. Lee YS, Shibata Y, Malhotra A, Dutta A (2009) A novel class of small RNAs: tRNA-derived RNA fragments (tRFs). Genes Dev 23:2639–2649. doi: 10.1101/gad.1837609

104. Cole C, Sobala A, Lu C, et al (2009) Filtering of deep sequencing data reveals the existence of abundant Dicer-dependent small RNAs derived from tRNAs. RNA 15:2147–2160. doi: 10.1261/rna.1738409

105. Mallory AC, Reinhart BJ, Jones-Rhoades MW, et al (2004) MicroRNA control of PHABULOSA in leaf development: importance of pairing to the microRNA 5' region. EMBO J 23:3356. doi: 10.1038/sj.emboj.7600340

106. Schwab R, Palatnik JF, Riester M, et al (2005) Specific Effects of MicroRNAs on the Plant Transcriptome. Dev Cell 8:517–527. doi: 10.1016/j.devcel.2005.01.018

107. Jin D, Wang Y, Zhao Y, Chen M (2013) MicroRNAs and Their Cross-Talks in Plant Development. J Genet Genomics 40:161–170. doi: 10.1016/j.jgg.2013.02.003

108. Bertolini E, Verelst W, Horner DS, et al (2013) Addressing the Role of microRNAs in Reprogramming Leaf Growth during Drought Stress in Brachypodium distachyon. Mol Plant 6:423–443. doi: 10.1093/mp/sss160

109. Zhou M, Li D, Li Z, et al (2013) Constitutive Expression of a miR319 Gene Alters Plant Development and Enhances Salt and Drought Tolerance in Transgenic Creeping Bentgrass. Plant Physiol 161:1375–1391. doi: 10.1104/pp.112.208702

110. Lv D-K, Bai X, Li Y, et al (2010) Profiling of cold-stress-responsive miRNAs in rice by microarrays. Gene 459:39–47. doi: 10.1016/j.gene.2010.03.011

111. Guan Q, Lu X, Zeng H, et al (2013) Heat stress induction of *miR398* triggers a regulatory loop that is critical for thermotolerance in Arabidopsis. Plant J 74:840–851. doi: 10.1111/tpj.12169

112. Macovei A, Tuteja N (2012) microRNAs targeting DEAD-box helicases are involved in salinity stress response in rice (Oryza sativa L.). BMC Plant Biol 12:183. doi: 10.1186/1471-2229-12-183

113. Wang M, Wang Q, Zhang B (2013) Response of miRNAs and their targets to salt and drought stresses in cotton (Gossypium hirsutum L.). Gene 530:26–32. doi: 10.1016/j.gene.2013.08.009

114. Moldovan D, Spriggs A, Yang J, et al (2010) Hypoxia-responsive microRNAs and trans-acting small interfering RNAs in Arabidopsis. J Exp Bot 61:165–177. doi: 10.1093/jxb/erp296

115. Hsieh L-C, Lin S-I, Shih AC-C, et al (2009) Uncovering Small RNA-Mediated Responses to Phosphate Deficiency in Arabidopsis by Deep Sequencing. Plant Physiol 151:2120–2132. doi: 10.1104/pp.109.147280

116. Zhao M, Ding H, Zhu J-K, et al (2011) Involvement of miR169 in the nitrogen-starvation responses in Arabidopsis. New Phytol 190:906–915. doi: 10.1111/j.1469-8137.2011.03647.x

117. Xiao B, Yang X, Ye C-Y, et al (2014) A diverse set of miRNAs responsive to begomovirus-associated betasatellite in Nicotiana benthamiana. BMC Plant Biol 14:60. doi: 10.1186/1471-2229-14-60

118. Tagami Y, Inaba N, Kutsuna N, et al (2007) Specific Enrichment of miRNAs in Arabidopsis thaliana Infected with Tobacco mosaic virus. DNA Res 14:227–233. doi: 10.1093/dnares/dsm022

119. Gao R, Wan ZY, Wong S-M (2013) Plant Growth Retardation and Conserved miRNAs Are Correlated to Hibiscus Chlorotic Ringspot Virus Infection. PLoS ONE 8:e85476. doi: 10.1371/journal.pone.0085476

120. Romanel E, Silva TF, Corrêa RL, et al (2012) Global alteration of microRNAs and transposon-derived small RNAs in cotton (Gossypium hirsutum) during Cotton leafroll dwarf polerovirus (CLRDV) infection. Plant Mol Biol 80:443–460. doi: 10.1007/s11103-012-9959-1

121. Feng J, Wang K, Liu X, et al (2009) The quantification of tomato microRNAs response to viral infection by stem-loop real-time RT-PCR. Gene 437:14–21. doi: 10.1016/j.gene.2009.01.017

122. Feng J, Lin R, Chen J (2013) Alteration of tomato microRNAs expression during fruit development upon Cucumber mosaic virus and Tomato aspermy virus infection. Mol Biol Rep 40:3713–3722. doi: 10.1007/s11033-012-2447-5

123. Du P, Wu J, Zhang J, et al (2011) Viral Infection Induces Expression of Novel Phased MicroRNAs from Conserved Cellular MicroRNA Precursors. PLoS Pathog 7:e1002176. doi: 10.1371/journal.ppat.1002176

124. Ruiz-Ruiz S, Navarro B, Gisel A, et al (2011) Citrus tristeza virus infection induces the accumulation of viral small RNAs (21–24-nt) mapping preferentially at the 3′-terminal region of the genomic RNA and affects the host small RNA profile. Plant Mol Biol 75:607–619. doi: 10.1007/s11103-011-9754-4

125. Jones-Rhoades MW, Bartel DP, Bartel B (2006) MicroRNAs and their regulatroy role in plants. Annu Rev Plant Biol 57:19–53. doi: 10.1146/annurev.arplant.57.032905.105218

126. Margis R, Fusaro AF, Smith NA, et al (2006) The evolution and diversification of Dicers in plants. FEBS Lett 580:2442–2450. doi: 10.1016/j.febslet.2006.03.072

127. Allen E, Xie Z, Gustafson AM, Carrington JC (2005) microRNA-Directed Phasing during Trans-Acting siRNA Biogenesis in Plants. Cell 121:207–221. doi: 10.1016/j.cell.2005.04.004

128. Zhai J, Jeong D-H, De Paoli E, et al (2011) MicroRNAs as master regulators of the plant NB-LRR defense gene family via the production of phased, trans-acting siRNAs. Genes Dev 25:2540–2553. doi: 10.1101/gad.177527.111

129. Zhang C, Li G, Wang J, et al (2013) Cascading cis-Cleavage on Transcript from trans-Acting siRNA-Producing Locus 3. Int J Mol Sci 14:14689–14699. doi: 10.3390/ijms140714689

130. Yoshikawa M (2005) A pathway for the biogenesis of trans-acting siRNAs in Arabidopsis. Genes Dev 19:2164–2175. doi: 10.1101/gad.1352605

131. Howell MD, Fahlgren N, Chapman EJ, et al (2007) Genome-Wide Analysis of the RNA-DEPENDENT RNA POLYMERASE6/DICER-LIKE4 Pathway in Arabidopsis Reveals Dependency on miRNA- and tasiRNA-Directed Targeting. Plant Cell Online 19:926–942. doi: 10.1105/tpc.107.050062

132. Zhang C, Li G, Wang J, Fang J (2012) Identification of trans-acting siRNAs and their regulatory cascades in grapevine. Bioinformatics 28:2561–2568. doi: 10.1093/bioinformatics/bts500

133. Axtell MJ, Jan C, Rajagopalan R, Bartel DP (2006) A Two-Hit Trigger for siRNA Biogenesis in Plants. Cell 127:565–577. doi: 10.1016/j.cell.2006.09.032

134. Rajagopalan R, Vaucheret H, Trejo J, Bartel DP (2006) A diverse and evolutionarily fluid set of microRNAs in Arabidopsis thaliana. Genes Dev 20:3407–3425. doi: 10.1101/gad.1476406

135. Li F, Orban R, Baker B (2012) SoMART: a web server for plant miRNA, tasiRNA and target gene analysis: web tools for miRNA and tasiRNA analysis. Plant J 70:891–901. doi: 10.1111/j.1365-313X.2012.04922.x

136. Arif MA, Fattash I, Ma Z, et al (2012) DICER-LIKE3 activity in Physcomitrella patens DICER-LIKE4 mutants causes severe developmental dysfunction and sterility. Mol Plant 5:1281–1294. doi: 1100.1.1093//mp//sss036

137. Cho SH, Coruh C, Axtell MJ (2012) miR156 and miR390 Regulate tasiRNA Accumulation and Developmental Timing in Physcomitrella patens. Plant Cell 24:4837–4849. doi: 10.1105/tpc.112.103176

138. Johnson C, Kasprzewska A, Tennessen K, et al (2009) Clusters and superclusters of phased small RNAs in the developing inflorescence of rice. Genome Res 19:1429–1440. doi: 10.1101/gr.089854.108

139. Quintero A, Pérez-Quintero AL, López C (2013) Identification of ta-siRNAs and Cis-nat-siRNAs in Cassava and Their Roles in Response to Cassava Bacterial Blight. Genomics Proteomics Bioinformatics 11:172–181. doi: 10.1016/j.gpb.2013.03.001

140. Kume K, Tsutsumi K, Saitoh Y (2010) TAS1 trans-Acting siRNA Targets Are Differentially Regulated at Low Temperature, and TAS1 trans- Acting siRNA Mediates Temperature-Controlled At1g51670 Expression. Biosci Biotechnol Biochem 74:1435–1440. doi: 10.1271/bbb.100111

141. Lapidot M, Pilpel Y (2006) Genome-wide natural antisense transcription: coupling its regulation to its different regulatory mechanisms. EMBO Rep 7:1216–1222. doi: 10.1038/sj.embor.7400857

142. Vanhée-Brossollet C, Vaquero C (1998) Do natural antisense transcripts make sense in eukaryotes? Gene 211:1–9. doi: 10.1016/S0378-1119(98)00093-6

143. Lavorgna G, Dahary D, Lehner B, et al (2004) In search of antisense. Trends Biochem Sci 29:88–94. doi: 10.1016/j.tibs.2003.12.002

144. Wang X-J, Gaasterland T, Chua N-H (2005) Genome-wide prediction and identification of cis-natural antisense transcripts in Arabidopsis thaliana. Genome Biol 6:R30. doi: 10.1186/gb-2005-6-4-r30

145. Wang H, Chua N-H, Wang X-J (2006) Prediction of trans-antisense transcripts in Arabidopsis thaliana. Genome Biol 7:R92. doi: 10.1186/gb-2006-7-10-r92

146. Zhou X, Sunkar R, Jin H, et al (2008) Genome-wide identification and analysis of small RNAs originated from natural antisense transcripts in Oryza sativa. Genome Res 19:70–78. doi: 10.1101/gr.084806.108

147. Zheng H, Qiyan J, Zhiyong N, Hui Z (2013) Prediction and identification of natural antisense transcripts and their small RNAs in soybean (Glycine max). BMC Genomics 14:280.

148. Lu T, Zhu C, Lu G, et al (2012) Strand-specific RNA-seq reveals widespread occurrence of novel cis-natural antisense transcripts in rice. BMC Genomics 13:721. doi: 10.1186/1471-2164-13-721

149. Faghihi MA, Wahlestedt C (2009) Regulatory roles of natural antisense transcripts. Nat Rev Mol Cell Biol 10:637–643. doi: 10.1038/nrm2738

150. Ron M, Alandete Saez M, Eshed Williams L, et al (2010) Proper regulation of a sperm-specific cis-nat-siRNA is essential for double fertilization in Arabidopsis. Genes Dev 24:1010–1021. doi: 10.1101/gad.1882810

151. Katiyar-Agarwal S, Morgan R, Dahlbeck D, et al (2006) A pathogen-inducible endogenous siRNA in plant immunity. Proc Natl Acad Sci 103:18002–18007. doi: 10.1073/pnas.0608258103

152. Borsani O, Zhu J, Verslues PE, et al (2005) Endogenous siRNAs Derived from a Pair of Natural cis-Antisense Transcripts Regulate Salt Tolerance in Arabidopsis. Cell 123:1279–1291. doi: 10.1016/j.cell.2005.11.035

153. Zubko E, Meyer P (2007) A natural antisense transcript of the Petunia hybrida Sho gene suggests a role for an antisense mechanism in cytokinin regulation. Plant J 52:1131–1139. doi: 10.1111/j.1365-313X.2007.03309.x

154. Held MA, Penning B, Brandt AS, et al (2008) Small-interfering RNAs from natural antisense transcripts derived from a cellulose synthase gene modulate cell wall biosynthesis in barley. Proc Natl Acad Sci 105:20534–20539. doi: 0.1073/pnas.0809408105

155. Klevebring D, Street NR, Fahlgren N, et al (2009) Genome-wide profiling of Populus small RNAs. BMC Genomics 10:620. doi: 10.1186/1471-2164-10-620

156. Slotkin RK, Vaughn M, Borges F, et al (2009) Epigenetic Reprogramming and Small RNA Silencing of Transposable Elements in Pollen. Cell 136:461–472. doi: 10.1016/j.cell.2008.12.038

157. Wang X, Elling AA, Li X, et al (2009) Genome-Wide and Organ-Specific Landscapes of Epigenetic Modifications and Their Relationships to mRNA and Small RNA Transcriptomes in Maize. Plant Cell Online 21:1053–1069. doi: 10.1105/tpc.109.065714

158. Lu C, Kulkarni K, Souret FF, et al (2006) MicroRNAs and other small RNAs enriched in the Arabidopsis RNA-dependent RNA polymerase-2 mutant. Genome Res 16:1276–1288. doi: 10.1101/gr.5530106

159. Kasschau KD, Fahlgren N, Chapman EJ, et al (2007) Genome-wide profiling and analysis of Arabidopsis siRNAs. PLoS Biol 5:e57. doi: 10. 1371/journal.pbio.0050057

160. Mosher RA, Schwach F, Studholme D, Baulcombe DC (2008) PolIVb influences RNA-directed DNA methylation independently of its role in siRNA biogenesis. Proc Natl Acad Sci 105:3145–3150. doi: 10.1073/pnas.0709632105

161. Havecker ER, Wallbridge LM, Hardcastle TJ, et al (2010) The Arabidopsis RNA-Directed DNA Methylation Argonautes Functionally Diverge Based on Their Expression and Interaction with Target Loci. Plant Cell 22:321–334. doi: 10.1105/tpc.109.072199

162. Zilberman D, Cao X, Jacobsen SE (2003) ARGONAUTE4 Control of Locus-Specific siRNA Accumulation and DNA and Histone Methylation. Science 299:716. doi: 10.1126/science.1079695

163. Wei L, Gu L, Song X, et al (2014) Dicer-like 3 produces transposable element-associated 24-nt siRNAs that control agricultural traits in rice. Proc Natl Acad Sci 111:3877–3882. doi: 10.1073/pnas.1318131111

164. Sun F, Guo W, Du J, et al (2013) Widespread, abundant, and diverse TE-associated siRNAs in developing wheat grain. Gene 522:1–7. doi: 10.1016/j.gene.2013.03.101

165. Ito H, Gaubert H, Bucher E, et al (2011) An siRNA pathway prevents transgenerational retrotransposition in plants subjected to stress. Nature 472:115–119. doi: 10.1038/nature09861

166. Sobala A, Hutvagner G (2011) Transfer RNA-derived fragments: origins, processing, and functions. Wiley Interdiscip Rev RNA 2:853–862. doi: 10.1002/wrna.96

167. Wang Q, Lee I, Ren J, et al (2013) Identification and Functional Characterization of tRNA-derived RNA Fragments (tRFs) in Respiratory Syncytial Virus Infection. Mol Ther 21:368–379. doi: 10.1038/mt.2012.237

168. Fu H, Feng J, Liu Q, et al (2009) Stress induces tRNA cleavage by angiogenin in mammalian cells. FEBS Lett 583:437–442. doi: 10.1016/j.febslet.2008.12.043

169. Yamasaki S, Ivanov P, Hu G, Anderson P (2009) Angiogenin cleaves tRNA and promotes stress-induced translational repression. J Cell Biol 185:35–42. doi: 10.1083/jcb.200811106

170. Zhang S, Sun L, Kragler F (2009) The Phloem-Delivered RNA Pool Contains Small Noncoding RNAs and Interferes with Translation. Plant Physiol 150:378–387. doi: 10.1104/pp.108.134767

171. Thompson DM, Lu C, Green PJ, Parker R (2008) tRNA cleavage is a conserved response to oxidative stress in eukaryotes. RNA 14:2095–2103. doi: 10.1261/rna.1232808

172. Gebetsberger J, Zywicki M, Künzi A, Polacek N (2012) tRNA-Derived Fragments Target the Ribosome and Function as Regulatory Non-Coding RNA in Haloferax volcanii. Archaea 2012:1–11. doi: 10.1155/2012/260909

173. Hackenberg M, Huang P-J, Huang C-Y, et al (2013) A Comprehensive Expression Profile of MicroRNAs and Other Classes of Non-Coding Small RNAs in Barley Under Phosphorous-Deficient and -Sufficient Conditions. DNA Res 20:109–125. doi: 10.1093/dnares/dss037

174. Baulcombe D (2004) RNA silencing in plants. Nature 431:356–363. doi: 10.1038/nature02874

175. Wang M-B, Bian X-Y, Wu L-M, et al (2004) On the role of RNA silencing in the pathogenicity and evolution of viroids and viral satellites. Proc Natl Acad Sci U S A 101:3275–3280. doi: 10.1073/pnas.0400104101

176. Blevins T, Rajeswaran R, Shivaprasad PV, et al (2006) Four plant Dicers mediate viral small RNA biogenesis and DNA virus induced silencing. Nucleic Acids Res 34:6233–6246. doi: 10.1093/nar/gkl886

177. Chellappan P, Vanitharani R, Fauquet CM (2004) Short Interfering RNA Accumulation Correlates with Host Recovery in DNA Virus-Infected Hosts, and Gene Silencing Targets Specific Viral Sequences. J Virol 78:7465–7477. doi: 10.1128/JVI.78.14.7465-7477.2004

178. Bouché N, Lauressergues D, Gasciolli V, Vaucheret H (2006) An antagonistic function for Arabidopsis DCL2 in development and a new function for DCL4 in generating viral siRNAs. EMBO J 25:3347–3356. doi: 10.1038/sj.emboj.7601217

179. Deleris A, Gallego-Bartolome J, Bao J, et al (2006) Hierarchical Action and Inhibition of Plant Dicer-Like Proteins in Antiviral Defense. Science 313:68. doi: 10.1126/science.1128214

180. Qi X, Bao FS, Xie Z (2009) Small RNA Deep Sequencing Reveals Role for Arabidopsis thaliana RNA-Dependent RNA Polymerases in Viral siRNA Biogenesis. PLoS ONE 4:e4971. doi: 10.1371/journal.pone.0004971

181. Smith NA, Eamens AL, Wang M-B (2011) Viral Small Interfering RNAs Target Host Genes to Mediate Disease Symptoms in Plants. PLoS Pathog 7:e1002022. doi: 10.1371/journal.ppat.1002022

182. Shimura H, Pantaleo V, Ishihara T, et al (2011) A Viral Satellite RNA Induces Yellow Symptoms on Tobacco by Targeting a Gene Involved in Chlorophyll Biosynthesis using the RNA Silencing Machinery. PLoS Pathog 7:e1002021. doi: 10.1371/journal.ppat.1002021

183. Xia Z, Peng J, Li Y, et al (2014) Characterization of Small Interfering RNAs Derived from Sugarcane Mosaic Virus in Infected Maize Plants by Deep Sequencing. PLoS ONE 9:e97013. doi: 10.1371/journal.pone.0097013

184. Miozzi L, Gambino G, Burgyan J, Pantaleo V (2013) Genome-wide identification of viral and host transcripts targeted by viral siRNAs in Vitis vinifera. Mol Plant Pathol 14:30–43. doi: 10.1111/j.1364-3703.2012.00828.x

185. Hall N (2007) Advanced sequencing technologies and their wider impact in microbiology. J Exp Biol 210:1518–1525. doi: 10.1242/jeb.001370

186. Mardis ER (2008) The impact of next-generation sequencing technology on genetics. Trends Genet 24:133–141. doi: 10.1016/j.tig.2007.12.007

187. Tucker T, Marra M, Friedman JM (2009) Massively Parallel Sequencing: The Next Big Thing in Genetic Medicine. Am J Hum Genet 85:142–154. doi: 10.1016/j.ajhg.2009.06.022

188. Pasini L, Bergonti M, Fracasso A, et al (2014) Microarray analysis of differentially expressed mRNAs and miRNAs in young leaves of sorghum under dry-down conditions. J Plant Physiol 171:537–548. doi: http://dx.doi.org/10.1016/j.jplph.2013.12.014

189. Liu H-H, Tian X, Li Y-J, et al (2008) Microarray-based analysis of stress-regulated microRNAs in Arabidopsis thaliana. RNA 14:836–843. doi: 10.1261/rna.895308

190. Liu H, Shen D, Jia S, et al (2013) Microarray-based screening of the microRNAs associated with caryopsis development in Oryza sativa. Biol Plant 57:255–261. doi: 10.1007/s10535-012-0270-4

191. Liu H, Qin C, Chen Z, et al (2014) Identification of miRNAs and their target genes in developing maize ears by combined small RNA and degradome sequencing. BMC Genomics 15:25. doi: 10.1186/1471-2164-15-25

192. Jin W, Wu F, Xiao L, et al (2012) Microarray-based Analysis of Tomato miRNA Regulated by Botrytis cinerea. J Plant Growth Regul 31:38–46. doi: 10.1007/s00344-011-9217-9

193. Mascarell-Creus A, Cañizares J, Vilarrasa-Blasi J, et al (2009) An oligo-based microarray offers novel transcriptomic approaches for the analysis of pathogen resistance and fruit quality traits in melon (Cucumis melo L.). BMC Genomics 10:467. doi: 10.1186/1471-2164-10-467

194. Valoczi A (2004) Sensitive and specific detection of microRNAs by northern blot analysis using LNA-modified oligonucleotide probes. Nucleic Acids Res 32:e175–e175. doi: 10.1093/nar/gnh171

195. Sandmann T, Cohen SM (2007) Identification of novel Drosophila melanogaster microRNAs. PLoS One 2:e1265. doi: 10.1371/ journal.pone.0001265

196. Várallyay É, Burgyán J, Havelda Z (2007) Detection of microRNAs by Northern blot analyses using LNA probes. Methods 43:140–145. doi: 10.1016/j.ymeth.2007.04.004

197. Torres AG, Fabani MM, Vigorito E, Gait MJ (2011) MicroRNA fate upon targeting with anti-miRNA oligonucleotides as revealed by an improved Northern-blot-based method for miRNA detection. RNA 17:933–943. doi: 10.1261/rna.2533811

198. Hagiwara Y, Peremyslov VV, Dolja VV (1999) Regulation of closterovirus gene expression examined by insertion of a self-processing reporter and by Northern hybridization. J Virol 73:7988–7993.

199. Moustafa K, Cross J (2016) Genetic Approaches to Study Plant Responses to Environmental Stresses: An Overview. Biology 5:20. doi: 10.3390/biology5020020

200. Rossel JB (2002) Global Changes in Gene Expression in Response to High Light in Arabidopsis. Plant Physiol 130:1109–1120. doi: 10.1104/pp.005595

201. Wang C, Han J, Liu C, et al (2012) Identification of microRNAs from Amur grape (Vitis amurensis Rupr.) by deep sequencing and analysis of microRNA variations with bioinformatics. BMC Genomics 13:122. doi: 10.1186/1471-2164-13-122

202. Chambers C, Shuai B (2009) Profiling microRNA expression in Arabidopsis pollen using microRNA array and real-time PCR. BMC Plant Biol 9:87. doi: 10.1186/1471-2229-9-87

203. Raymond CK (2005) Simple, quantitative primer-extension PCR assay for direct monitoring of microRNAs and short-interfering RNAs. RNA 11:1737–1744. doi: 10.1261/rna.2148705

204. Balcells I, Cirera S, Busk PK (2011) Specific and sensitive quantitative RT-PCR of miRNAs with DNA primers. BMC Biotechnol 11:70. doi: 10.1186/1472-6750-11-70

205. Jensen SG, Lamy P, Rasmussen MH, et al (2011) Evaluation of two commercial global miRNA expression profiling platforms for detection of less abundant miRNAs. BMC Genomics 12:435. doi: 10.1186/1471-2164-12-435

206. Varkonyi-Gasic E, Wu R, Wood M, et al (2007) Protocol: a highly sensitive RT-PCR method for detection and quantification of microRNAs. Plant Methods 3:12. doi: 10.1186/1746-4811-3-12

207. Chen C, Ridzon DA, Broomer AJ, et al (2005) Real-time quantification of microRNAs by stem–loop RT–PCR. Nucleic Acids Res 33:e179–e179. doi: 10.1093/nar/gni178

208. Czimmerer Z, Hulvely J, Simandi Z, et al (2013) A Versatile Method to Design Stem-Loop Primer-Based Quantitative PCR Assays for Detecting Small Regulatory RNA Molecules. PLoS ONE 8:e55168. doi: 10.1371/journal.pone.0055168

209. Sun X, Korir NK, Han J, et al (2012) Characterization of grapevine microR164 and its target genes. Mol Biol Rep 39:9463–9472. doi: 10.1007/s11033-012-1811-9

210. Wang C, Shangguan L, Kibet KN, et al (2011) Characterization of microRNAs Identified in a Table Grapevine Cultivar with Validation of Computationally Predicted Grapevine miRNAs by miR-RACE. PLoS ONE 6:e21259. doi: 10.1371/journal.pone.0021259

211. Pantaleo V, Szittya G, Moxon S, et al (2010) Identification of grapevine microRNAs and their targets using high-throughput sequencing and degradome analysis. Plant J 62:960–976. doi: 10.1111/j.1365-313X.2010.04208.x

212. Meng Y, Gou L, Chen D, et al (2010) High-throughput degradome sequencing can be used to gain insights into microRNA precursor metabolism. J Exp Bot 61:3833–3837. doi: 10.1093/jxb/erq209

213. Zhou M, Gu L, Li P, et al (2010) Degradome sequencing reveals endogenous small RNA targets in rice (Oryza sativa L. ssp. indica). Front Biol 5:67–90. doi: 10.1007/s11515-010-0007-8

214. Yang X, Wang L, Yuan D, et al (2013) Small RNA and degradome sequencing reveal complex miRNA regulation during cotton somatic embryogenesis. J Exp Bot 64:1521–1536. doi: 10.1093/jxb/ert013

215. Friedländer MR, Chen W, Adamidi C, et al (2008) Discovering microRNAs from deep sequencing data using miRDeep. Nat Biotechnol 26:407–415. doi: 10.1038/nbt1394

216. Yang X, Li L (2011) miRDeep-P: a computational tool for analyzing the microRNA transcriptome in plants. Bioinformatics 27:2614–2615. doi: 10.1093/bioinformatics/btr430

217. Lei J, Sun Y (2014) miR-PREFeR: an accurate, fast and easy-to-use plant miRNA prediction tool using small RNA-Seq data. Bioinformatics 30:2837–2839. doi: 10.1093/bioinformatics/btu380

218. Mapleson D, Moxon S, Dalmay T, Moulton V (2013) MirPlex: A Tool for Identifying miRNAs in High-Throughput sRNA Datasets Without a Genome. J Exp Zoolog B Mol Dev Evol 320:47–56. doi: 10.1002/jez.b.22483

219. Qian K, Auvinen E, Greco D, Auvinen P (2012) miRSeqNovel: An R based workflow for analyzing miRNA sequencing data. Mol Cell Probes 26:208–211. doi: 10.1016/j.mcp.2012.05.002

220. Evers M, Huttner M, Dueck A, et al (2015) miRA: adaptable novel miRNA identification in plants using small RNA sequencing data. BMC Bioinformatics 16:370. doi: 10.1186/s12859-015-0798-3

221. Jha A, Shankar R (2013) miReader: Discovering novel miRNAs in species without sequenced genome. PloS One 8:e66857. doi: 10.1371/journal.pone.0066857

222. Mathelier A, Carbone A (2010) MIReNA: finding microRNAs with high accuracy and no learning at genome scale and from deep sequencing data. Bioinformatics 26:2226–2234. doi: 10.1093/bioinformatics/btq329

223. Jha A, Chauhan R, Mehra M, et al (2012) miR-BAG: Bagging Based Identification of MicroRNA Precursors. PLoS ONE 7:e45782. doi: 10.1371/journal.pone.0045782

224. Muller S, Rycak L, Winter P, et al (2013) omiRas: a Web server for differential expression analysis of miRNAs derived from small RNA-Seq data. Bioinformatics 29:2651–2652. doi: 10.1093/bioinformatics/btt457

225. Hackenberg M, Sturm M, Langenberger D, et al (2009) miRanalyzer: a microRNA detection and analysis tool for next-generation sequencing experiments. Nucleic Acids Res 37:W68–W76. doi: 10.1093/nar/gkp347

226. Hackenberg M, Rodriguez-Ezpeleta N, Aransay AM (2011) miRanalyzer: an update on the detection and analysis of microRNAs in high-throughput sequencing experiments. Nucleic Acids Res 39:W132–W138. doi: 10.1093/nar/gkr247

227. Axtell MJ (2013) ShortStack: Comprehensive annotation and quantification of small RNA genes. RNA 19:740–751. doi: 10.1261/rna.035279.112

228. Shahid S, Axtell MJ (2014) Identification and annotation of small RNA genes using ShortStack. Methods 67:20–27. doi: 10.1016/j.ymeth.2013.10.004

229. Stocks MB, Moxon S, Mapleson D, et al (2012) The UEA sRNA workbench: a suite of tools for analysing and visualizing next generation sequencing microRNA and small RNA datasets. Bioinformatics 28:2059–2061. doi: 10.1093/bioinformatics/bts311

230. Meyers BC, Axtell MJ, Bartel B, et al (2008) Criteria for Annotation of Plant MicroRNAs. Plant Cell 20:3186–3190. doi: 10.1105/tpc.108.064311

231. Hofacker IL (2003) Vienna RNA secondary structure server. Nucleic Acids Res 31:3429–3431. doi: 10.1093/nar/gkg599

232. Langmead B, Trapnell C, Pop M, Salzberg SL (2009) Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. Genome Biol 10:R25. doi: 10.1186/gb-2009-10-3-r25

233. Li H, Durbin R (2009) Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics 25:1754–1760. doi: 10.1093/bioinformatics/btp324

234. Li H, Ruan J, Durbin R (2008) Mapping short DNA sequencing reads and calling variants using mapping quality scores. Genome Res 18:1851–1858. doi: 10.1101/gr.078212.108

235. Li R, Li Y, Kristiansen K, Wang J (2008) SOAP: short oligonucleotide alignment program. Bioinformatics 24:713–714. doi: 10.1093/bioinformatics/btn025

236. Kozomara A, Griffiths-Jones S (2014) miRBase: annotating high confidence microRNAs using deep sequencing data. Nucleic Acids Res 42:D68–D73. doi: 10.1093/nar/gkt1181

237. Jaillon O, Aury J-M, Noel B, et al (2007) The grapevine genome sequence suggests ancestral hexaploidization in major angiosperm phyla. Nature 449:463–467. doi: 10.1038/nature06148

238. Mica E, Piccolo V, Delledonne M, et al (2009) High throughput approaches reveal splicing of primary microRNA transcripts and tissue specific expression of mature microRNAs in Vitis vinifera. BMC Genomics 10:558. doi: 10.1186/1471-2164-10-558

239. Addo-Quaye C, Miller W, Axtell MJ (2008) CleaveLand: a pipeline for using degradome data to find cleaved small RNA targets. Bioinformatics 25:130–131. doi: 10.1093/bioinformatics/btn604

240. Folkes L, Moxon S, Woolfenden HC, et al (2012) PAREsnip: a tool for rapid genome-wide discovery of small RNA/target interactions evidenced through degradome sequencing. Nucleic Acids Res 40:e103–e103. doi: 10.1093/nar/gks277

241. Zheng Y, Li Y-F, Sunkar R, Zhang W (2012) SeqTar: an effective method for identifying microRNA guided cleavage sites from degradome of polyadenylated transcripts in plants. Nucleic Acids Res 40:e28–e28. doi: 10.1093/nar/gkr1092

242. Dai X, Zhao PX (2011) psRNATarget: a plant small RNA target analysis server. Nucleic Acids Res 39:W155–W159. doi: 10.1093/nar/gkr319

243. Pantano L, Estivill X, Marti E (2011) A non-biased framework for the annotation and classification of the non-miRNA small RNA transcriptome. Bioinformatics 27:3202–3203. doi: 10.1093/bioinformatics/btr527

244. Hardcastle TJ, Kelly KA, Baulcombe DC (2012) Identifying small interfering RNA loci from high-throughput sequencing data. Bioinformatics 28:457–463. doi: 10.1093/bioinformatics/btr687

245. Guo Q, Qu X, Jin W (2015) PhaseTank: genome-wide computational identification of phasiRNAs and their regulatory cascades. Bioinformatics 31:284–286. doi: 10.1093/bioinformatics/btu628

246. Martin JA, Wang Z (2011) Next-generation transcriptome assembly. Nat Rev Genet 12:671–682. doi: 10.1038/nrg3068

247. Dominic Mills J, Kawahara Y, Janitz M (2013) Strand-specific RNA-seq provides greater resolution of transcriptome profiling. Curr Genomics 14:173–181. doi: 10.2174/1389202911314030003

248. Grabherr MG, Haas BJ, Yassour M, et al (2011) Full-length transcriptome assembly from RNA-Seq data without a reference genome. Nat Biotechnol 29:644–652. doi: 10.1038/nbt.1883

249. Ward JA, Ponnala L, Weber CA (2012) Strategies for transcriptome analysis in nonmodel plants. Am J Bot 99:267–276. doi: 10.3732/ajb.1100334

250. Trapnell C, Williams BA, Pertea G, et al (2010) Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. Nat Biotechnol 28:511–515. doi: 10.1038/nbt.1621

251. Trapnell C, Roberts A, Goff L, et al (2012) Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and Cufflinks. Nat Protoc 7:562–578. doi: 10.1038/nprot.2012.016

252. Langmead B, Salzberg SL (2012) Fast gapped-read alignment with Bowtie 2. Nat Methods 9:357–359. doi: 10.1038/nmeth.1923

253. Haas BJ, Papanicolaou A, Yassour M, et al (2013) De novo transcript sequence reconstruction from RNA-seq using the Trinity platform for reference generation and analysis. Nat Protoc 8:1494–1512. doi: 10.1038/nprot.2013.084

254. Schulz MH, Zerbino DR, Vingron M, Birney E (2012) Oases: robust de novo RNA-seq assembly across the dynamic range of expression levels. Bioinformatics 28:1086–1092. doi: 10.1093/bioinformatics/bts094

255. Xie Y, Wu G, Tang J, et al (2014) SOAPdenovo-Trans: de novo transcriptome assembly with short RNA-Seq reads. Bioinformatics 30:1660–1666. doi: 10.1093/bioinformatics/btu077

256. Seyednasrollah F, Laiho A, Elo LL (2015) Comparison of software packages for detecting differential expression in RNA-seq studies. Brief Bioinform 16:59–70. doi: 10.1093/bib/bbt086

257. Robinson MD, McCarthy DJ, Smyth GK (2010) edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. Bioinformatics 26:139–140. doi: 10.1093/bioinformatics/btp616

258. Anders S, Huber W (2010) Differential expression analysis for sequence count data. Genome Biol 11:R106. doi: 10.1186/gb-2010-11-10-r106

259. Hardcastle TJ, Kelly KA (2010) baySeq: empirical Bayesian methods for identifying differential expression in sequence count data. BMC Bioinformatics 11:422. doi: 10.1186/1471-2105-11-422

260. Tarazona S, García-Alcalde F, Dopazo J, et al (2011) Differential expression in RNA-seq: a matter of depth. Genome Res 21:2213–2223. doi: 10.1101/gr.124321.111

261. Li J, Tibshirani R (2013) Finding consistent patterns: A nonparametric approach for identifying differential expression in RNA-Seq data. Stat Methods Med Res 22:519–536. doi: 10.1177/0962280211428386

262. Smyth GK (2005) Limma: linear models for microarray data. In: Bioinforma. Comput. Biol. Solut. Using R Bioconductor. Springer, pp 397–420

263. Smyth Gordon K (2004) Linear Models and Empirical Bayes Methods for Assessing Differential Expression in Microarray Experiments. Stat Appl Genet Mol Biol 3:1. doi: 10.2202/1544-6115.1027

264. Trapnell C, Hendrickson DG, Sauvageau M, et al (2013) Differential analysis of gene regulation at transcript resolution with RNA-seq. Nat Biotech 31:46–53. doi: 10.1038/nbt.2450

265. Leng N, Dawson JA, Thomson JA, et al (2013) EBSeq: an empirical Bayes hierarchical model for inference in RNA-seq experiments. Bioinformatics 29:1035–1043. doi: 10.1093/bioinformatics/btt087

266. Love MI, Huber W, Anders S (2014) Moderated estimation of fold change and dispersion for RNA-Seq data with DESeq2.

267. Benjamini Y, Hochberg Y (1995) Controlling the false discovery rate: a practical and powerful approach to multiple testing. J R Stat Soc Ser B Methodol 57:289–300.

268. Robles JA, Qureshi SE, Stephen SJ, et al (2012) Efficient experimental design and analysis strategies for the detection of differential expression using RNA-Sequencing. BMC Genomics 13:1–14. doi: 10.1186/1471-2164-13-484

269. Rapaport F, Khanin R, Liang Y, et al (2013) Comprehensive evaluation of differential gene expression analysis methods for RNA-seq data. Genome Biol 14:R95. doi: 10.1186/gb-2013-14-9-r95

270. Bolstad BM, Irizarry RA, Astrand M, Speed TP (2003) A comparison of normalization methods for high density oligonucleotide array data based on variance and bias. Bioinformatics 19:185–193. doi: 10.1093/bioinformatics/19.2.185

271. McCormick KP, Willmann MR, Meyers BC (2011) Experimental design, preprocessing, normalization and differential expression analysis of small RNA sequencing experiments. Silence 2:2. doi: 10.1186/1758-907X-2-2

Chapter 3: Differential expression of miRNAs and associated gene targets in grapevine leafroll-associated virus 3 infected plants.

3.1 Introduction

Small non-coding RNAs (sRNAs) are universal components of plant transcriptomes and can be categorised into several groups based on their biogenesis and function [1–3]. The most common sRNA species in plants are microRNAs (miRNAs) and small interfering RNAs (siRNAs). Although miRNAs are not the most abundant sRNA group they are the best characterised. A number of miRNAs have been reported to regulate genes associated with plant development, and in the defence responses to abiotic and biotic stresses, which include viruses [4–7]. In plants, miRNAs are processed from single-stranded hairpin precursors that are transcribed from nuclear encoded *MIR* genes [8]. The processing of the miRNA duplex is catalysed in the nucleus after which the mature miRNA can bind to the catalytic site of the RNA silencing complex (RISC) in the cytoplasm. A complementary target mRNA sequence is then cleaved to cause post-transcriptional gene silencing [3].

Comparative miRNA profiling of virus infected plants may lead to the identification of differentially expressed sRNAs, and combined with knowledge of the corresponding target genes, may provide insight into plant defence responses. Different high-throughput techniques are available to measure the expression levels of miRNAs, these include microarray hybridisation [9, 10] and next-generation sequencing (NGS) [11]. Microarrays provide a standardised genome-wide assay that can be used for the detection of known miRNAs, while NGS of sRNA can be used to detect known miRNAs and has the advantage of being able to predict unknown miRNAs.

Grapevine is one of the most widely grown perennial fruit crops and is exposed to a variety of pests and pathogens that threaten the viability of the viticulture industry. Grapevine leafroll disease (GLD) is one of the most important viral diseases affecting grape cultivars and *Grapevine leafroll-associated virus 3* (GLRaV-3) (family *Closteroviridae*, genus *Ampelovirus*) is considered the main etiological agent [12]. The importance of the grapevine industry and the magnitude of the problems caused by pathogens, have been the main motivation behind extensive research programmes focused at disease prevention. To combat pathogens there is a need for resistant cultivars or control measures to prevent the spread of diseases. To date no natural resistance to GLRaV-3 has been demonstrated in *Vitis vinifera*.

In this study microarray analysis and NGS were used to generate miRNA and gene expression profiles to characterise the response of grapevine plants to GLRaV-3 infection in order to glean a better understanding of host-pathogen interactions. Through understanding the molecular interaction between GLRaV-3 and grapevine, potential targets for engineering viral tolerance can be identified.

3.2 Materials and methods

3.2.1 Plant material and sample collection

Three *Vitis vinifera* cv. Cabernet Sauvignon own-rooted plants, singly infected with GLRaV-3 variant group II (isolate GP18, Accession No. EU259806), and three healthy plants, were established in a greenhouse. The plants were established from cuttings made from a naturally infected Cabernet Sauvignon plant and rooted in a greenhouse. The healthy Cabernet Sauvignon control plants were established from cuttings obtained from a nursery that provides certified virus-free plant material. Plants were maintained under natural light with temperature ranging from 22 °C to 28 °C and plants did not undergo winter dormancy. Only one shoot was allowed to grow and all side shoots were constantly removed. Maintenance included pruning back the plants every six months. Phloem material was collected from the plants twice, 18 months apart. The developmental stage of the plants was similar for both sampling times to negate possible variation imparted by physiological growth stage. Phloem material was sampled as soon as the shoot material reached lignification and GLRaV-3 symptoms, reddening of the interveinal areas and downward curling of the leaf margins, were observed in the infected plants. High quality total RNA was extracted from phloem material using a modified cetyltrimethylammonium bromide (CTAB) protocol [13, 14]. The virus status of these plants was confirmed using end-point RT-PCRs for frequently occurring grapevine viruses (Appendix A3)[i] [15]. The genetic variant of GLRaV-3 was confirmed by real-time RT-PCR high-resolution melting curve analysis and multiplex RT-PCR [16].

3.2.2 Microarrays

To obtain an overview of the known miRNAs present in each of the samples, microarray experiments were conducted on the first set of RNA extracts. The analyses were performed by poly(A) tailing 10 µg of total RNA, followed by a ligation step to generate biotin-labelled RNA using the FlashTag™

---

[i] Appendix A3. Jooste AEC, Molenaar N, Maree HJ, Bester R, Morey L, De Koker WC, Burger JT (2015) Identification and distribution of multiple virus infections in Grapevine leafroll diseased vineyards. Eur J Plant Pathol 142:363–375.

A survey of viruses infecting grapevine in the wine regions of the Western Cape Province in South Africa was conducted. The survey determined the relative abundance of five different grapevine leafroll-associated virus 3 (GLRaV-3) variants. Virus profiles were also determined for individual vines. A total of 315 plants were sampled and analysed over two growing seasons. The complexity of virus populations detected in this study, highlights the need for detection methods able to identify all viruses and their variants in vineyards.

Biotin HSR RNA Labelling Kit (Affymetrix). The prepared targets were hybridised to MicroRNA GeneChip 1.0 arrays (Affymetrix), washed and stained using the GeneChip Fluidics Station 450 (Affymetrix) and scanned using the GeneChip® Scanner 3000 7G (Affymetrix). All fluorescence intensity data were processed by performing background correction (Robust Multi-array Average), normalisation (quantile normalisation) and summarisation of probe set intensities (median polish) using the R package, Affy [17]. The normalised expression values for *Vitis vinifera* and other *Viridiplantae* miRNA probes were extracted and differential expression analysis was performed using the R package, Limma [18]. The false discovery rate correction was used to correct for multiple testing. MicroRNAs with an adjusted p-value < 0.05 were regarded as differentially expressed. Targets for differentially expressed miRNAs were predicted using psRNATarget [19] with default parameters. BLAST sequence similarity searches were performed using Blast2GO [20].

3.2.3 Next-generation sequencing

Next-generation sequencing was performed on the second set of RNA extracts. An sRNA sequencing library was prepared for each plant sample using the Illumina Small RNA TruSeq kit and sequenced (1x50bp) on an Illumina HiSeq instrument (Fasteris, Switzerland). The same total RNA sample was used to prepare a transcriptome library with the Illumina mRNA stranded RNA kit. Single-end NGS (1x125bp) was performed on an Illumina HiSeq instrument (Fasteris, Switzerland). For the sRNA NGS data set, adapter sequences were removed and reads were filtered for quality (phred score > 20 over 100% of the read) using FASTX-toolkit (http://hannonlab.cshl.edu/fastx_toolkit/index.html). For the transcriptome NGS data set, adapter trimming was performed using cutadapt [21] and quality trimming, using Trimmomatic [22] (HEADCROP of 12 nucleotides (nts), SLIDINGWINDOW of 3 nts with Q20, MINLEN of 20 nts).

3.2.4 sRNA NGS data analysis

Only reads 18-26 nts in length from the sRNA NGS data were used for miRNA analysis. Known *Vitis vinifera* miRNAs (vvi-miRNAs), as well as sRNAs identical to known miRNAs of other plant species, were identified using an in-house Python script, allowing no mismatches to entries in the miRNA Registry Database (miRBase) version 21 [23]. ShortStack (v3.3) [24] was used to perform novel miRNA prediction from sRNA reads mapped with one mismatch to the *Vitis vinifera* reference genome [25]. To determine variation in sRNA expression levels between healthy and GLRaV-3-infected samples, the R package, DESeq2 [26] was used. The false discovery rate correction was used to correct for multiple testing. MicroRNAs with an adjusted p-value < 0.05 were regarded as differentially expressed. Targets for differentially expressed miRNAs were predicted using

psRNATarget [19] with default parameters. BLAST sequence similarity searches were performed using Blast2GO [20].

3.2.5 NGS transcriptome data analysis

The high quality sequence reads were mapped to the *Vitis vinifera* reference genome [25] using TopHat version 2.0.14 [27]. TopHat identifies splice junctions between exons by using the short-read aligner Bowtie [28]. Reference-based assembly of the reads was performed using Cufflinks and Cuffmerge version 2.2.1 [27]. The expression level of each transcript was expressed as reads per transcript kilobase per million reads mapped (RPKM) value, calculated based on the number of mapped reads. Cuffdiff version 2.2.1 was used to detect differentially expressed genes [27]. Transcripts with an adjusted p-value < 0.05 were regarded as differentially expressed.

3.2.6 Stemloop RT-qPCR miRNA validation

Differentially expressed miRNAs were validated using stemloop RT-qPCR assays [29]. Complementary DNA (cDNA) was synthesised from 1 µg of total RNA using 1 µM of stemloop primer (IDT) (Appendix B1), 0.5 mM dNTPs (Thermo Scientific), 100 U Maxima reverse transcriptase (Thermo Scientific) and 20 U Ribolock (Thermo Scientific) in a final volume of 20 µl. Incubation for 30 minutes at 16 °C was performed, followed by a pulsed reverse transcription of 60 cycles at 30 °C for 30 seconds, 42 °C for 30 seconds and 50 °C for 1 second. Five µl of each cDNA sample was pooled and a 5-fold dilution series was prepared to construct a representative standard curve for each primer sets. The remaining cDNA was diluted 1:24 to quantify each sample separately using the miRNA-specific and the reference miRNA primer sets. All cDNA dilutions were stored at −20 °C. The RT-qPCRs were performed using the Rotor-Gene Q thermal cycler (Qiagen). Reactions contained 1x FastStart Universal probe master (ROX) (Roche), 0.1 µM Universal ProbeLibrary probe #21 (Roche), 3.3 µl Milli-Q H_2O and 0.6 µM specific forward and universal reverse primers (IDT) (Appendix B1). One µl cDNA was added to each reaction to a final reaction volume of 10 µl. The "no-template" and "no-reverse transcriptase" controls were included in all runs. All reactions were performed in triplicate in Rotor-Gene Q 0.1 ml tube-and-cap strips (Qiagen). Cycling parameters included an initial activation of 95 °C for 10 minutes and 45 cycles of 95 °C for 10 seconds and 60 °C for 60 seconds. Acquisition on the green channel was recorded at the end of the extension step.

The Rotor-gene Q software version 2.3.1 (Qiagen) was used to calculate primer efficiencies, Cq values and gene quantitation values for all targets. The relative concentration ratio (CR) were calculated as

previously shown in Bester et al. [14] (Appendix A4)[ii] using a reference gene index, calculated using the geometric mean of the concentration of two stable expressed miRNAs (vvi-miR159c and vvi-miR167a). The non-parametric Wilcoxon signed-rank test was used to evaluate differential expression. A p-value < 0.05 was regarded as significantly differentially expressed. All calculations were performed on the web-based application, Harbin (https://rbester.shinyapps.io/Harbin/) (Appendix A5)[iii].

3.2.7 RT-qPCR target validation

Complementary DNA was synthesised from 1 µg of total RNA using 0.15 µg random primers (Promega), 0.5 mM dNTPs (Thermo Scientific), 100 U Maxima reverse transcriptase (Thermo Scientific) and 20 U Ribolock (Thermo Scientific) in a final volume of 20 µl. Five µl of each cDNA sample was pooled and a 5-fold dilution series was prepared to construct a representative standard curve for each primer sets. The remaining cDNA was diluted 1:24 to quantify each sample separately using the gene of interest-specific and the reference genes' primer sets. All cDNA dilutions were stored at −20 °C. The RT-qPCRs were performed using the Rotor-Gene Q thermal cycler (Qiagen) and the SensiMixTM SYBR No-ROX Kit (Bioline). Reactions contained 1x SensiMixTM SYBR (Bioline) No-ROX, Milli-Q H_2O and 0.4 µM forward and reverse primers (IDT) (Appendix B1). Two µl cDNA was added to each reaction to a final reaction volume of 12.5 µl. The same cDNA dilution series was used to construct all primer-specific standard curves and the same 1:24 dilution of each "unknown" sample was screened with all primer sets for quantitation. The "no-template" and "no-reverse transcriptase" control reactions were included in all runs. All reactions were performed in triplicate in Qiagen Rotor-Gene Q 0.1 ml tube-and-cap strips. Cycling parameters included an initial activation of 95 °C for 10 minutes and 45 cycles of 95 °C for 15 seconds, 58 °C for 15 seconds (55 °C for 15 seconds for the reference genes) and 72 °C for 15 seconds. Acquisition on the green channel was

[ii] Appendix A4. Bester R, Pepler PT, Burger JT, Maree HJ (2014) Relative quantitation goes viral: An RT-qPCR assay for a grapevine virus. J Virol Methods 210:67–75.

Three genomic regions (ORF1a, coat protein and 3′UTR) were targeted to quantitate GLRaV-3 relative to three stably expressed reference genes (actin, GAPDH and alpha-tubulin). These assays were able to detect all known variant groups of GLRaV-3, including the divergent group VI, with equal efficiency. No link could be established between the concentration ratios of the different genomic regions and subgenomic RNA (sgRNA) expression. However, a significant lower virus concentration ratio for plants infected with variant group VI compared to variant group II was observed for the ORF1a, coat protein and the 3′ UTR. Significant higher accumulation of the virus in the growth tip was also detected for both variant groups.

[iii] Appendix A5. Bester R, Pepler PT, Aldrich DJ, Maree HJ (2016). Harbin: A quantitation PCR analysis tool. Biotechnol Lett. DOI 10.1007/s10529-016-2221-1.

To enable comparisons of different relative quantitation experiments, a web-browser application called Harbin was created for a dynamic interaction with qPCR data. A quantile-based scoring system is proposed that will allow for comparison of samples at different time points and between experiments. Harbin simplifies the analysis of high-density qPCR assays, either for individual experiments or across sets of replicates and biological conditions. The application uses the standard curve method for relative quantitation of genes with normalisation using a reference gene index to calculate a concentration ratio (CR). This Harbin quantile bootstrap test for evaluating if different data sets can be combined, was shown to be less conservative than the Kolmogorov-Smirnov test, and therefore more sensitive in detecting distributional differences between data sets. Statistical significance testing for CRs across biological conditions is also possible with Harbin.

recorded at the end of the extension step. Melting curve analysis of PCR amplicons was performed with temperatures ranging from 65 °C to 95 °C with a 1 °C increase in temperature every 5 seconds to identify primer-dimers and non-specific amplification. The relative CR were calculated as previously shown in Bester et al. [14] (Appendix A4) using a reference gene index, calculated using the geometric mean of the concentration of three reference genes (GAPDH, α-tubulin and actin) previously shown to be constitutively expressed in *Vitis vinifera* phloem material [14, 30]. The non-parametric Wilcoxon signed-rank test was used to evaluate differential expression. A p-value < 0.05 was regarded as significantly differentially expressed. All calculations were performed on the web-based application, Harbin (https://rbester.shinyapps.io/Harbin/) (Appendix A5).

3.3. Results and discussion

3.3.1 Microarrays

Due to the high redundancy of miRNA sequences in miRBase, which results in similar redundancy on the microarray, the differentially expressed miRNA sequences identified were collapsed to identify only the non-redundant sequences. Differential expression analysis of the microarray data, using Limma, identified 41 non-redundant vvi-miRNAs to be differentially expressed, of which nine were down-regulated and 32 were up-regulated in infected samples. Additionally, 67 down-regulated and 122 up-regulated miRNAs from other plant species were identified when all plant probes were included in the analysis. Of these 230 differentially expressed miRNAs identified, 157 could be classified as isomiRs (sequence variants) of other miRNAs.

3.3.2 sRNA NGS

The six sRNA NGS libraries produced 47927477 high quality sequence reads of 18-26 nts in length (7337021 to 9178033 reads per library). To identify known miRNAs present in the data set, only perfect matches to miRBase version 21 [23] were allowed. For both infected and control samples, 12 % of the reads mapped to *Vitis vinifera* miRNAs (vvi-miRNAs) detecting 79 of the 119 non-redundant vvi-miRNAs in miRBase. In addition to the known vvi-miRNAs, 163 non-redundant miRNA-associated reads (2,130,603 redundant miRNA-associated reads) with 100 % sequence similarity to miRNAs from other plant species were also identified. Differential expression analysis, using DESeq2, identified eight miRNAs (vvi-miR477b-3p, vvi-miR398b-c, vvi-miR395a-m, vvi-miR408, vvi-miR3634-3p, vvi-miR477b-5p, gma-miR408d, cca-miR408) to be up-regulated and two miRNAs (vvi-miR2950-5p, ath-miR858b) to be down-regulated in infected samples compared to healthy samples. MicroRNAs gma-miR408d and cca-miR408 were identified as isomiRs of vvi-miR408. Four of the differentially expressed miRNAs (vvi-miR398b-c, vvi-miR395a-m, vvi-miR408 and cca-

miR408) identified in the sRNA NGS data set were validated by the microarray data set (Figure. 1). Probes for the remaining six miRNAs were not available on the array, since the microarray only included probes for miRNAs in miRBase version 11.

Novel miRNAs were predicted using the sRNA NGS data set. Fourteen putative miRNA hairpins, not present in miRBase were predicted. Both the major and minor (previously named miRNA star) mature miRNA products were present in the data set (Appendix B2). Four of the putative mature miRNA sequences have previously been identified in grapevine [31–33]. Differential expression analysis identified three of the novel miRNAs to be up-regulated in GLRaV-3 infected samples. The predicted targets of these three novel miRNAs had high sequence similarities to a LRR receptor-like serine/threonine-protein kinase, a nitrate transporter (NRT1/PTR family protein), DNA binding protein, major facilitator superfamily protein and a nuclear fusion defective protein.

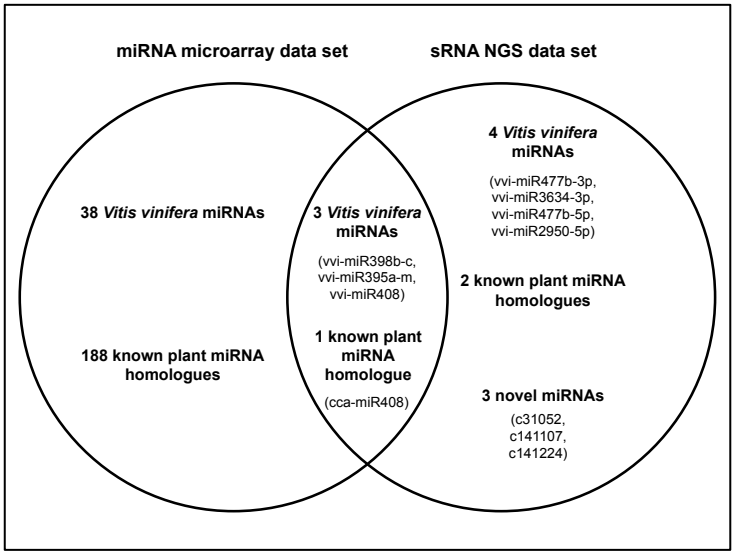


Figure 1. Venn diagram displaying overlap of the differentially expressed miRNAs identified in the microarray and sRNA NGS data set. The miRNAs that were validated with qPCR are listed in parenthesis.

3.3.3 NGS transcriptome

In the transcriptome NGS data set, differential expression analysis identified 2801 genes to be modulated in infected samples (1515 up-regulated and 1286 down-regulated genes). The four differentially expressed miRNAs common to the microarray and sRNA NGS data were predicted to target proteins with high sequence similarity to a serine threonine-protein kinase, an ATP sulfurylase, an rRNA processing isoform and a phagocyte signalling-impaired protein. Three of these targets were differentially expressed in the NGS transcriptome data set. Two of these genes (GSVIVT01000937001

and GSVIVT01018057001) were significantly down-regulated and GSVIVT01024634001 were up-regulated (Table 1). Additionally, the putative targets of 26 of the differentially expressed miRNAs of the microarray data set were differentially expressed in the GLRaV-3 infected samples. These putative targets are involved in various biological processes including the regulation of transcription, protein phosphorylation and xylan biosynthesis (Figure 2, Appendix B3).

A lower number than expected of the predicted miRNA targets were identified as differentially expressed in the NGS transcriptome data. This can be as a result of the miRNA target prediction that relied on complementary sequence searches of the predicted Pinot noir reference transcripts [25]. Cultivar sequence differences and novel genes not present in the Pinot noir reference transcriptome can result in different transcripts predicted as possible miRNA targets. However, a large number of other genes were identified in the NGS transcriptome data set to be differentially expressed in the GLRaV-3 infected samples. These can potentially be involved in the plant's response to pathogens and specifically GLRaV-3, though not regulated by miRNAs.

Table 1. Differentially expressed miRNAs and predicted targets identified in both microarray and next-generation sequencing data.

		Differentially expressed miRNAs[a]		
		vvi-miR398b-c	vvi-miR395a-m	cca-miR408
sRNA next-generation sequencing	log_2(fold change)	1.88	1.67	1.87
	p-value	1.05E-08	0.52E-03	2.93E-06
	Adjusted p-value	3.74E-07*	0.92E-02*	0.15E-03*
Microarray	log_2(fold change)	9.38	5.44	9.19
	p-value	3.75E-05	0.51E-03	3.81E-06
	Adjusted p-value	0.10E-02*	0.22E-02*	0.56E-03*
psRNAtarget	Predicted target	GSVIVT01000937001	GSVIVT01018057001	GSVIVT01024634001
Blast2GO	Target description	Serine threonine-protein kinase	ATP sulfurylase	Phagocyte signalling-impaired protein
	GO annotation[b]	P: signal transduction; F:ion binding; P:cellular protein modification process; F:kinase activity; C:plasma membrane; P:response to stress; P:immune system process	F:nucleotidyltransferase activity; P:sulfur compound metabolic process	P:secondary metabolic process; F:oxidoreductase activity; P:catabolic process; F:ion binding; C:extracellular region
Transcriptome next-generation sequencing	log_2(fold change)	-0.73	-1.77	1.01
	p-value	0.75E-03	5.00E-05	5.00E-05
	Adjusted p-value	0.82E-02*	0.82E-03*	0.82E-03*

[a]Only three of the four differentially expressed miRNAs common to the microarray and sRNA NGS data had differentially expressed targets in the transcriptome data.
[b]GO annotations if available: P=Biological process, F=Molecular function, C=Cellular component
*adjusted p-value < 0.05 was regarded as significantly differentially expressed

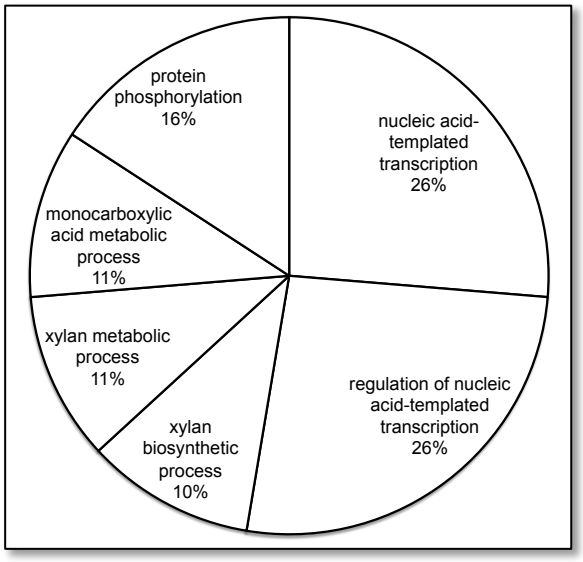


Figure 2. Gene ontology (GO) annotation. Biological processes of the modulated putative targets of the differentially expressed miRNAs identified in the microarray data that was differentially expressed in the NGS transcriptome data set (Appendix B3).

3.3.4 qPCR validation

Stemloop RT-qPCR assays were designed for the four miRNAs (vvi-miR398b-c, vvi-miR395a-m, vvi-miR408, cca-miR408) identified to be differentially expressed in both the microarray and sRNA NGS data sets. Additionally, assays were designed for four known (vvi-miR477b-3p, vvi-miR3634-3p, vvi-miR477b-5p, vvi-miR2950-5p) and three novel (c31052, c141107, c141224) *Vitis vinifera* miRNAs differentially expressed in the sRNA NGS data set (Figure 1). The up-regulation of vvi-miR398b-c, vvi-miR395a-m, vvi-miR408, cca-miR408, c31052, c141107 and c141224 was validated with stemloop RT-qPCR assays and a $\log_2$(fold change) > 1.5 was observed for all the miRNAs in GLRaV-3 infected samples (Figure 3).

The predicted target of the novel miRNA c141224 (GSVIVT01033079001) was also up-regulated in the NGS transcriptome data set. This transcript had a high sequence similarity to a major facilitator superfamily transporter. These transporters are responsible for the uptake and secretion of essential nutrients and ions to regulate plant development [34]. It is possible that the up-regulation of miRNA c141224 can regulate the transcription of the transporter gene as a plant stress response.

The differential expression of the four additional *Vitis vinifera* miRNAs of the sRNA NGS data set (vvi-miR477b-3p, vvi-miR3634-3p, vvi-miR477b-5p, vvi-miR2950-5p) was also validated with qPCR (data shown in Chapter 5).

The predicted targets of the three miRNAs identified in both the microarray and NGS sRNA data sets with differentially expressed targets in the NGS transcriptome data set were assessed with RT-qPCR assays. The anti-correlation of two of these targets were validated with a $\log_2$(fold change) of -1.17 and -2.68 for GSVIVT01000937001 and GSVIVT01018057001, respectively (Fig. 3). Even though GSVIVT01024634001 was up-regulated in the qPCR data, as was found in the NGS transcriptome data set, the variation between biological replicates were too high to be considered statistically significant. The putative up-regulated target of cca-miR408 (GSVIVT01024634001) had a high sequence similarity to a phagocyte signalling-impaired protein. This protein plays a role in actin cytoskeleton organisation, which supports cellular processes linked to vesicle transport, endocytosis, spatial distribution of organelles and plant innate immunity [35–37]. The lower $\log_2$(fold change) and the lower read count per sample in the NGS transcriptome data set for the target of cca-miR408 compared to the other two targets assayed, could indicate that the concentration of the target was too low for accurate quantitation.

The anti-correlation of miRNA expression (up-regulation of vvi-miR398b-c and vvi-miR395a-m) and putative target expression (down-regulation of serine threonine-protein kinase and ATP sulfurylase), that were confirmed with the microarray, sRNA NGS, transcriptome NGS and qPCR assays, can provide the first insight into the complex host-pathogen interactions in GLD. Serine threonine-protein kinases play a central role in signalling during pathogen recognition, activation of plant defence mechanisms and developmental control [38]. In *Arabidopsis thaliana*, miR398 targets superoxide dismutases that function as reactive oxygen species (ROS) scavengers for stress resistance and survival in plants [39]. However, in *Paulownia tomentosa,* pau-miR398, was shown to target a gene encoding serine/threonine-protein kinase that play a role in regulating cell proliferation, programmed cell death and cell differentiation [40]. The down-regulation of serine/threonine-protein kinase can therefore impact on plant growth and the ability of the plant to activate plant defence responses. Adenosine triphosphate (ATP) sulfurylase is involved in plant-tolerance to several biotic stresses and can initiate plant-pathogen responses [41]. In *Arabidopsis thaliana*, miR395 was shown to regulate sulfate accumulation and that the accumulation of miR395 can be triggered by reduced internal sulfate levels [42, 43]. In both *Nicotiana tabacum* infected with tobacco mosaic virus and *Hibiscus cannabinus* infected with hibiscus chlorotic ringspot virus, a correlation between higher sulphate levels and plant resistance was observed [44, 45]. Therefore, it is plausible that the up-regulation of miR395 can be involved in protecting the plant against virus infections.

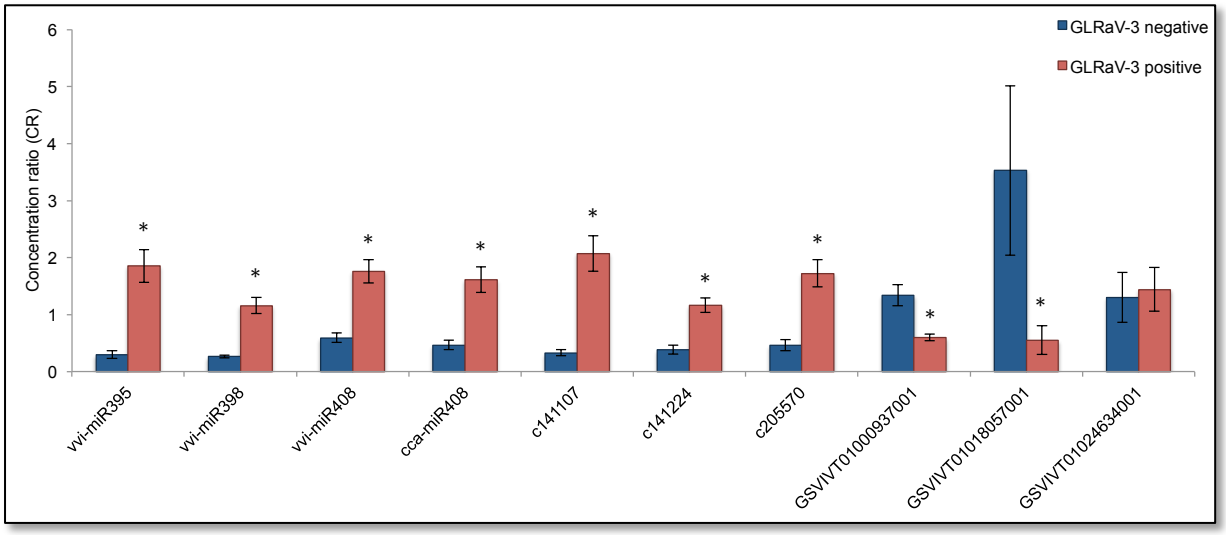


Figure 3. Differential expression of miRNAs and target genes assessed using RT-qPCR assays. The mean concentration ratio (CR) ± standard error (SE) of three biological replicates, with each replicate an average of three technical replicates is displayed. Statistically significant differences between GLRaV-3 negative and positive samples, determined by the Wilcoxon rank sum test, are indicated by asterisks (*p-value < 0.05).

3.4 Conclusion

In this study two high-throughput techniques were used to identify differentially expressed miRNA associated with a GLRaV-3 infection. Four miRNAs were up-regulated in response to a GLRaV-3 infection in both the microarray and sRNA NGS data sets. The modulated expression of three of the predicted targets of these miRNAs was validated in the NGS transcriptome data set. The expression patterns of the four miRNAs were confirmed with stemloop RT-qPCR assays and two of the predicted miRNA targets were validated using RT-qPCR assays.

A surprising result was the unusually large number of differentially expressed miRNAs identified with the microarray analysis compared to the sRNA NGS. While the presence of a large number of isomiRs due to the cross-hybridisation of similar sequences [46–48] can explain this to an extent, we believe that these numbers also represent false positives, emphasising the need for a second technology for validation. Even though prior sequence knowledge is necessary for microarray design, it can complement NGS data for validations, be applied for a genome wide screen of a large numbers of samples, and be used to focus deep sequencing by identifying regions/genes of interest to enrich genomic fractions before NGS. The similarities between the NGS data sets and the available information provided by the microarray data set, regardless of the fact that the RNA of the microarray and NGS experiments were extracted 18 months apart, provides a strong base for further investigation to better understand the molecular interaction between the plant and the virus.

This study also identified novel miRNAs not present in miRBase version 21. The putative miRNAs identified in *Vitis vinifera* cv. Cabernet Sauvignon will add to the existing databases and provide a platform for future studies. The presence and expression patterns of three of these miRNAs were confirmed with qPCR assays.

These results are compatible with the growing evidence that virus infection could alter plant miRNA biogenesis and be correlated with developmental changes and disease symptoms. The negative effect of GLRaV-3 on plant growth and longevity can be linked to the modulated gene targets identified in the study as the down-regulation of these targets can result in reduced plant growth and lower resistance to biotic stress. The differentially expressed miRNAs and associated targets identified in GLRaV-3 infected own-rooted Cabernet Sauvignon plants can be useful in elucidating the regulatory mechanisms of miRNAs in various aspects of GLD. The knowledge can also be translated into possible management strategies to control the disease.

3.5 Supplementary material

Appendix B1: Primers for miRNA stemloop RT-qPCR and miRNA target RT-qPCR assays.

Appendix B2: Putative miRNAs predicted using ShortStack.

Appendix B3: Predicted targets of the differentially expressed miRNAs in the microarray data set that were differentially expressed in the NGS transcriptome data set.

3.6 References

1. Khraiwesh B, Zhu J-K, Zhu J (2012) Role of miRNAs and siRNAs in biotic and abiotic stress responses of plants. Biochim Biophys Acta BBA - Gene Regul Mech 1819:137–148. doi: 10.1016/j.bbagrm.2011.05.001
2. Axtell MJ (2013) Classification and comparison of small RNAs from plants. Annu Rev Plant Biol 64:137–159. doi: 10.1146/annurev-arplant-050312-120043
3. Bartel DP (2004) MicroRNAs: genomics, biogenesis, mechanism, and function. Cell 116:281–297. doi: 10.1016/S0092-8674(04)00045-5
4. Alabi OJ, Zheng Y, Jagadeeswaran G, et al (2012) High-throughput sequence analysis of small RNAs in grapevine Vitis vinifera affected by grapevine leafroll disease: Small RNAs in leafroll disease-infected grapevine. Mol Plant Pathol 13:1060–1076. doi: 10.1111/j.1364-3703.2012.00815.x
5. Singh K, Talla A, Qiu W (2012) Small RNA profiling of virus-infected grapevines: evidences for virus infection-associated and variety-specific miRNAs. Funct Integr Genomics 12:659–669. doi: 10.1007/s10142-012-0292-1
6. Bazzini AA, Hopp HE, Beachy RN, Asurmendi S (2007) Infection and coaccumulation of tobacco mosaic virus proteins alter microRNA levels, correlating with symptom and plant development. Proc Natl Acad Sci 104:12157–12162. doi: 10.1073/pnas.0705114104
7. Tagami Y, Inaba N, Kutsuna N, et al (2007) Specific Enrichment of miRNAs in Arabidopsis thaliana Infected with Tobacco mosaic virus. DNA Res 14:227–233. doi: 10.1093/dnares/dsm022

8. Guleria P, Mahajan M, Bhardwaj J, Yadav SK (2011) Plant Small RNAs: Biogenesis, Mode of Action and Their Roles in Abiotic Stresses. Genomics Proteomics Bioinformatics 9:183–199. doi: 10.1016/S1672-0229(11)60022-3

9. Pasini L, Bergonti M, Fracasso A, et al (2014) Microarray analysis of differentially expressed mRNAs and miRNAs in young leaves of sorghum under dry-down conditions. J Plant Physiol 171:537–548. doi: http://dx.doi.org/10.1016/j.jplph.2013.12.014

10. Liu H, Shen D, Jia S, et al (2013) Microarray-based screening of the microRNAs associated with caryopsis development in Oryza sativa. Biol Plant 57:255–261. doi: 10.1007/s10535-012-0270-4

11. Pantaleo V, Saldarelli P, Miozzi L, et al (2010) Deep sequencing analysis of viral short RNAs from an infected Pinot Noir grapevine. Virology 408:49–56. doi: 10.1016/j.virol.2010.09.001

12. Maree HJ, Almeida RPP, Bester R, et al (2013) Grapevine leafroll-associated virus 3. Front Microbiol 4:82. doi: 10.3389/fmicb.2013.00082

13. Carra A, Gambino G, Schubert A (2007) A cetyltrimethylammonium bromide-based method to extract low-molecular-weight RNA from polysaccharide-rich plant tissues. Anal Biochem 360:318–320. doi: 10.1016/j.ab.2006.09.022

14. Bester R, Pepler PT, Burger JT, Maree HJ (2014) Relative quantitation goes viral: An RT-qPCR assay for a grapevine virus. J Virol Methods 210:67–75. doi: 10.1016/j.jviromet.2014.09.022

15. Jooste AEC, Molenaar N, Maree HJ, et al (2015) Identification and distribution of multiple virus infections in Grapevine leafroll diseased vineyards. Eur J Plant Pathol 142:363–375. doi: 10.1007/s10658-015-0620-0

16. Bester R, Jooste AE, Maree HJ, Burger JT (2012) Real-time RT-PCR high-resolution melting curve analysis and multiplex RT-PCR to detect and differentiate grapevine leafroll-associated associated virus 3 variant groups I, II, III and VI. Virol J 9:219. doi: 10.1186/1743-422X-9-219.

17. Gautier L, Cope L, Bolstad BM, Irizarry RA (2004) affy--analysis of Affymetrix GeneChip data at the probe level. Bioinformatics 20:307–315. doi: 10.1093/bioinformatics/btg405

18. Smyth GK (2005) Limma: linear models for microarray data. In: Bioinforma. Comput. Biol. Solut. Using R Bioconductor. Springer, pp 397–420

19. Dai X, Zhao PX (2011) psRNATarget: a plant small RNA target analysis server. Nucleic Acids Res 39:W155–W159. doi: 10.1093/nar/gkr319

20. Conesa A, Götz S (2008) Blast2GO: A Comprehensive Suite for Functional Analysis in Plant Genomics. Int J Plant Genomics 2008:1–12. doi: 10.1155/2008/619832

21. Martin M (2011) Cutadapt removes adapter sequences from high-throughput sequencing reads. EMBnet J 17:10–12. doi: 10.14806/ej.17.1.200

22. Bolger AM, Lohse M, Usadel B (2014) Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics 30:2114–2120. doi: 10.1093/bioinformatics/btu170

23. Kozomara A, Griffiths-Jones S (2014) miRBase: annotating high confidence microRNAs using deep sequencing data. Nucleic Acids Res 42:D68–D73. doi: 10.1093/nar/gkt1181

24. Axtell MJ (2013) ShortStack: Comprehensive annotation and quantification of small RNA genes. RNA 19:740–751. doi: 10.1261/rna.035279.112

25. Jaillon O, Aury J-M, Noel B, et al (2007) The grapevine genome sequence suggests ancestral hexaploidization in major angiosperm phyla. Nature 449:463–467. doi: 10.1038/nature06148

26. Love MI, Huber W, Anders S (2014) Moderated estimation of fold change and dispersion for RNA-Seq data with DESeq2.

27. Trapnell C, Roberts A, Goff L, et al (2012) Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and Cufflinks. Nat Protoc 7:562–578. doi: 10.1038/nprot.2012.016

28. Langmead B, Trapnell C, Pop M, Salzberg SL (2009) Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. Genome Biol 10:R25. doi: 10.1186/gb-2009-10-3-r25

29. Varkonyi-Gasic E, Wu R, Wood M, et al (2007) Protocol: a highly sensitive RT-PCR method for detection and quantification of microRNAs. Plant Methods 3:12. doi: 10.1186/1746-4811-3-12

30. Reid K, Olsson N, Schlosser J, et al (2006) An optimized grapevine RNA isolation procedure and statistical determination of reference genes for real-time RT-PCR during berry development. BMC Plant Biol 6:1–11. doi: 10.1186/1471-2229-6-27

31. Wang C, Wang X, Kibet NK, et al (2011) Deep sequencing of grapevine flower and berry short RNA library for discovery of novel microRNAs and validation of precise sequences of grapevine microRNAs deposited in miRBase. Physiol Plant 143:64–81. doi: 10.1111/j.1399-3054.2011.01481.x

32. Han J, Fang J, Wang C, et al (2014) Grapevine microRNAs responsive to exogenous gibberellin. BMC Genomics 15:111. doi: 10.1186/1471-2164-15-111

33. Belli Kullan J, Lopes Paim Pinto D, Bertolini E, et al (2015) miRVine: a microRNA expression atlas of grapevine based on small RNA sequencing. BMC Genomics 16:393. doi: 10.1186/s12864-015-1610-5

34. Pao SS, Paulsen IT, Saier MH (1998) Major facilitator superfamily. Microbiol Mol Biol Rev 62:1–34.

35. Samaj J (2004) Endocytosis, Actin Cytoskeleton, and Signaling. PLANT Physiol 135:1150–1161. doi: 10.1104/pp.104.040683

36. Romanenko AS, Rifel' AA, Salyaev the CM of the RR (2002) Endocytosis of exopolysaccharides of the potato ring rot causal agent by host-plant cells. In: Dokl. Biol. Sci. Springer, pp 451–453

37. Henty-Ridilla JL, Shimono M, Li J, et al (2013) The Plant Actin Cytoskeleton Responds to Signals from Microbe-Associated Molecular Patterns. PLoS Pathog 9:e1003290. doi: 10.1371/journal.ppat.1003290

38. Afzal AJ, Wood AJ, Lightfoot DA (2008) Plant receptor-like serine threonine kinases: roles in signaling and plant defense. Mol Plant Microbe Interact 21:507–517. doi: 10.1094 / MPMI -21-5-0507

39. Zhu C, Ding Y, Liu H (2011) MiR398 and plant stress responses. Physiol Plant 143:1–9. doi: 10.1111/j.1399-3054.2011.01477.x

40. Fan G, Niu S, Xu T, et al (2015) Plant–Pathogen Interaction-Related MicroRNAs and Their Targets Provide Indicators of Phytoplasma Infection in Paulownia tomentosa × Paulownia fortunei. PLOS ONE 10:e0140590. doi: 10.1371/journal.pone.0140590

41. Anjum NA, Gill R, Kaushik M, et al (2015) ATP-sulfurylase, sulfur-compounds, and plant stress tolerance. Front Plant Sci 6:210. doi: 10.3389/fpls.2015.00210

42. Matthewman CA, Kawashima CG, Húska D, et al (2012) miR395 is a general component of the sulfate assimilation regulatory network in Arabidopsis. FEBS Lett 586:3242–3248. doi: 10.1016/j.febslet.2012.06.044

43. Liang G, Yang F, Yu D (2010) MicroRNA395 mediates regulation of sulfate accumulation and allocation in Arabidopsis thaliana: miRNA395 and sulfate homeostasis. Plant J 62:1046–1057. doi: 10.1111/j.1365-313X.2010.04216.x

44. Höller K, Király L, Künstler A, et al (2010) Enhanced glutathione metabolism is correlated with sulfur-induced resistance in tobacco mosaic virus-infected genetically susceptible Nicotiana tabacum plants. Mol Plant Microbe Interact 23:1448–1459. doi: 10.1094 / MPMI -05-10-0117

45. Zhang X, Wong S-M (2009) Hibiscus chlorotic ringspot virus upregulates plant sulfite oxidase transcripts and increases sulfate levels in kenaf (Hibiscus cannabinus L.). J Gen Virol 90:3042–3050. doi: 10.1099/vir.0.012112-0

46. Hurd PJ, Nelson CJ (2009) Advantages of next-generation sequencing versus the microarray in epigenetic research. Brief Funct Genomic Proteomic 8:174–183. doi: 10.1093/bfgp/elp013

47. Okoniewski MJ, Miller CJ (2006) Hybridization interactions between probesets in short oligo microarrays lead to spurious correlations. BMC Bioinformatics 7:276. doi: 10.1186/1471-2105-7-276

48. Pradervand S, Weber J, Lemoine F, et al (2010) Concordance among digital gene expression, microarrays, and qPCR when measuring differential expression of microRNAs. BioTechniques 48:219–222. doi: 10.2144/000113367

Chapter 4: The small RNA repertoire of three *Vitis vinifera* cultivars

4.1 Introduction

Grapevine (*Vitis* spp) is one of the most widely cultivated fruit crops and an economically important commodity. More than seven million hectares of the world surface area are under vines to produce wine, table grapes, raisins, juice, vinegar and distilled spirits (http://www.oiv.int). The generation of genomic resources can add significant value to extending the knowledge of grapevine physiology, cultivar-specific characteristics and understanding how diseases and environmental conditions affect the plant. The availability of the annotated genome sequence of *Vitis vinifera* cultivar Pinot noir (PN40024) [1] has contributed genetic information to various databases [2–4] and transcriptomic studies have led to the description of gene regulatory networks associated with grapevine development [5–8]. However, less is known about the grapevine's interaction with the environment and how biotic and abiotic stresses impact on plant physiology. The extension of genetic resources will aid future research in the development of cultivars resistant to adverse biotic and abiotic stresses.

Many studies have shown how small non-coding RNAs (sRNAs) can regulate gene expression by interfering with mRNA translation or by cleavage and subsequent degradation of target mRNAs [9–13]. These sRNA regulators can influence normal development and/or responses to environmental stimuli. Small RNAs are a class of double stranded RNAs of 20–30 nucleotides (nts) in length. Based on their mode of biogenesis and/or function, various types of sRNAs have been identified in plants, including microRNAs (miRNAs) and small interfering RNAs (siRNAs).

In plants, miRNAs are processed from single-stranded hairpin precursors to produce 20-22 nucleotide (nt) regulating RNAs. The primary miRNAs (pri-miRNAs) are transcribed from nuclear encoded miRNA genes (*MIR* genes) by RNA polymerase II to form transcripts with a hairpin structure [14]. The processing of the pri-miRNAs, is catalysed in the nucleus by Dicer-like (DCL) proteins to release mature double-stranded molecules containing a small number of mismatches between the miRNA and its antisense strand [15, 16]. In the cytoplasm a helicase unwinds the miRNA double-stranded duplex and the mature miRNA is exposed to the RNA silencing complex (RISC). The mature miRNA can then bind to the argonaute (AGO) protein, the catalytic component of RISC, which guides the RISC to the complementary target mRNA sequence to cleave and suppress the translation of the gene at a post-transcriptional level [9, 11, 12]. In addition to posttranscriptional gene silencing, miRNAs also regulate gene expression associated with epigenetic changes such as DNA and histone methylation [17, 18].

Small interfering RNAs differ from miRNAs in that the precursors of siRNAs are usually long double stranded RNA and can either be of endogenous or exogenous origin. The endogenous precursors originate from RNA transcribed from natural *cis*-antisense transcript pairs, inverted repeats, genome regions rich in retro-elements, or RNA-dependent RNA polymerases that converts single-stranded RNA into double stranded RNA [11]. Exogenous precursors are generated from transposons or the transcripts of replicating viruses [19, 20]. Irrespective of its origin, dsRNA is cleaved into 21-24 nt siRNAs by multiple DCL proteins and like miRNAs, siRNAs are loaded into AGO protein-containing RISC for target regulation at a posttranscriptional or a transcriptional level.

Endogenous sRNAs can be divided into heterochromatic siRNA, secondary siRNA and natural antisense siRNA (natsiRNA) [10]. Heterochromatic siRNA or repeat-associated siRNA (rasiRNA) originates from intergenic and/or repetitive genomic regions and are mostly 23-24 nts long [10, 21]. The production of secondary siRNAs is initiated by one or more sRNAs that target an initial primary transcript. This leads to the synthesis of a dsRNA molecule that is processed into secondary siRNAs [10]. *Trans*-acting siRNAs (tasiRNAs) are secondary siRNAs that can repress mRNA targets distinct from their locus of origin. They can also be phased, in that they originate from successive DCL-catalysed processing events from a single dsRNA terminus [10, 22, 23]. Natural antisense siRNA (natsiRNA) can arise from the hybridisation of separately transcribed, complementary RNAs [24]. They can be *cis*-natsiRNA if they were transcribed from opposite strands of the same locus or be *trans*-natsiRNA if hybridisation occurred between different genes.

Even though sRNAs have been well described in a number of organisms, recent advances in high-throughput sequencing approaches accelerated the identification of sRNAs. The generation of next-generation sequencing (NGS) data has led to the discovery of new sRNAs. These new sRNAs could be derived from existing molecules such as ribosomal RNAs (rRNAs) or transfer RNAs (tRNAs), which could extend this group of RNAs' functional roll beyond that which they were originally characterised for [25, 26]. Small RNAs associated with tRNAs can be divided into tRNA halves or RNA-derived RNA fragments (tRFs) based on the tRNA region they originate from [27–29].

In this study the aim was to characterise the sRNA species of three *Vitis vinifera* cultivars (Chardonnay, Chenin blanc and Cabernet Sauvignon) by using NGS. Computational analysis of sRNA data can provide a resource to deepen our understanding of the biological function of specific sRNAs in grapevine development and establish leads for targeted functional studies associated with sRNA regulation.

4.2 Material and methods

4.2.1 Next-generation sequencing and data quality control

Three young *Vitis vinifera* plants of each cultivar Chardonnay (CY), Chenin blanc (CB), Cabernet Sauvignon (CS) and own-rooted Cabernet Sauvignon (Cab) were collected from a certified virus-free nursery and re-established in a greenhouse. Plants were maintained under natural light with temperature ranging from 22 °C to 28 °C. Plants were vertically trained allowing only one shoot. All side shoots were constantly removed and the plants pruned back every 6 months. Phloem material from shoots was sampled from all plants in the same physiological growth stage, as soon as the shoot material reached lignification. High quality total RNA (A260/A280 above 2, A260/A230 above 2 and RNA integrity number above 6.5) was extracted from phloem material using a modified cetyltrimethylammonium bromide (CTAB) protocol [30, 31]. RNA quality was assessed using spectrophotometry, gel electrophoresis and Agilent Bioanalyzer analysis. Plants were tested for frequently occurring grapevine viruses using end-point RT-PCRs (Appendix A3)[i] [32].

An sRNA sequencing library was prepared from total RNA by polyacrylamide gel size selection of the 18-30 nt fraction from each sample. The Illumina Small RNA TruSeq kit was used for library preparation and sequencing (1 x 50 bp) was performed on an Illumina HiSeq instrument (Fasteris, Switzerland). Adapter sequences were removed using cutadapt [33] and reads were filtered for quality (phred score above 20 over 100% of the read) using FASTX-toolkit (http://hannonlab.cshl.edu/fastx_toolkit/index.html). Only reads 18-26 nts in length were used for miRNA, phased small interfering RNA, natsiRNA and rasiRNA analysis. The tRNA-derived small RNA analysis was performed using the 17-44 nt read fraction. To identify the level of reads associated with ribosomal RNA, the high quality filtered 18-26 nt reads were mapped to all the *Viridiplantae* rRNA sequences available in Genbank using Bowtie (1.1.2) [34] allowing no mismatches and reporting only the best alignment (Bowtie reporting parameter: --best). All bioinformatics analyses were performed on the high-performance computer (HPC) of the Central Analytical Facility at Stellenbosch University (http://www.sun.ac.za/hpc). Optimised parameters were used and changes to critical parameters were stated.

[i] Appendix A3. Jooste AEC, Molenaar N, Maree HJ, Bester R, Morey L, De Koker WC, Burger JT (2015) Identification and distribution of multiple virus infections in Grapevine leafroll diseased vineyards. Eur J Plant Pathol 142:363–375.

A survey of viruses infecting grapevine in the wine regions of the Western Cape Province in South Africa was conducted. The survey determined the relative abundance of five different grapevine leafroll-associated virus 3 (GLRaV-3) variants. Virus profiles were also determined for individual vines. A total of 315 plants were sampled and analysed over two growing seasons. The complexity of virus populations detected in this study, highlights the need for detection methods able to identify all viruses and their variants in vineyards.

4.2.2 miRNA analysis and novel miRNA stemloop RT-qPCR validation

Known *Vitis vinifera* miRNAs (vvi-miRNAs) were identified using an in-house Python script, allowing no mismatches to entries in the miRNA Registry Database (miRBase) version 21 [3]. ShortStack (v3.3) [35, 36] was used to perform novel miRNA prediction from sRNAs mapped with one mismatch to the *Vitis vinifera* reference genome (PN40024) [1]. ShortStack identifies miRNA precursors by filtering the predicted hairpin structures using the criteria set by Meyers et al. [37]. MicroRNA targets were predicted using the web-based tool psRNATarget [38], applying optimised parameters. BLAST sequence similarity searches were performed using Blast2GO (Blastx algorithm with e-value threshold of 0.001) [39].

Novel miRNA validations were performed using stemloop RT-qPCR assays [40]. Complementary DNA (cDNA) was synthesised from 1 µg of total RNA using 1 µM of stemloop primer (IDT) (Appendix C1), 0.5 mM dNTPs (Thermo Scientific), 100 U Maxima reverse transcriptase (Thermo Scientific) and 20 U Ribolock (Thermo Scientific) in a final volume of 20 µl. Incubation for 30 minutes at 16 °C was performed, followed by a pulsed reverse transcription of 60 cycles at 30 °C for 30 seconds, 42 °C for 30 seconds and 50 °C for 1 second. Five µl of each cDNA sample was pooled and a 5-fold dilution series was prepared to construct a representative standard curve for each primer sets. The remaining cDNA was diluted 1:24 to quantify each sample separately using the miRNA-specific and the reference miRNA primer sets. All cDNA dilutions were stored at −20 °C. The RT-qPCRs were performed using the Rotor-Gene Q thermal cycler (Qiagen). Reactions contained 1x FastStart Universal probe master (ROX) (Roche), 0.1 µM Universal probe library probe #21 (Roche), 3.3 µl Milli-Q H_2O and 0.6 µM specific forward and universal reverse primers (IDT) (Appendix C1). One µl cDNA was added to each reaction to a final reaction volume of 10 µl. The "no-template" and "no-reverse transcriptase" controls were included in all runs. All reactions were performed in triplicate in Rotor-Gene Q 0.1 ml tube-and-cap strips (Qiagen). Cycling parameters included an initial activation of 95 °C for 10 minutes and 45 cycles of 95 °C for 10 seconds and 60 °C for 60 seconds. Acquisition on the green channel was recorded at the end of the extension step. The Rotor-gene Q software version 2.3.1 (Qiagen) was used to calculate primer efficiencies, Cq values and gene quantitation values for all targets. The relative concentration ratio (CR) were calculated as previously shown in Bester et al. [31] (Appendix A4)[ii] using a reference gene index, calculated using the geometric mean of the

[ii] Appendix A4. Bester R, Pepler PT, Burger JT, Maree HJ (2014) Relative quantitation goes viral: An RT-qPCR assay for a grapevine virus. J Virol Methods 210:67–75.

Three genomic regions (ORF1a, coat protein and 3′ UTR) were targeted to quantitate GLRaV-3 relative to three stably expressed reference genes (actin, GAPDH and alpha-tubulin). These assays were able to detect all known variant groups of GLRaV-3, including the divergent group VI, with equal efficiency. No link could be established between the concentration ratios of the different genomic regions and subgenomic RNA (sgRNA) expression. However, a significant lower virus concentration ratio for plants infected with variant group VI compared to variant group II was observed for the ORF1a, coat protein and the 3′ UTR. Significant higher accumulation of the virus in the growth tip was also detected for both variant groups.

concentration of two stable expressed miRNAs (vvi-miR159c and vvi-miR167a). The non-parametric Wilcoxon signed-rank test was used to evaluate differential expression. A p-value < 0.05 was regarded as significantly differentially expressed. All calculations were performed on the web-based application, Harbin (https://rbester.shinyapps.io/Harbin/) (Appendix A5)[iii].

4.2.3 Phased cluster and phasiRNA identification

ShortStack (v3.3) [35, 36] were used to identify phased regions. For valid phased loci, ShortStack uses the formula described by Guo et al. [41]. The reads were mapped with one mismatch to the *Vitis vinifera* reference genome (PN40024). Potential phasiRNA clusters were defined as regions that contain at least four phased reads with a maximum separation distance of 84 nts. Each region was evaluated by calculating the number of reads that mapped to each possible phasing register. The phasing registers determined had a window size of 21 nts and a step size of 1 nt. A phased score was calculated for each region based on phased ratio, number and abundance [41, 42]. Higher phasing scores indicated more phasing signature. ShortStack calculated a phase score in each possible phase size and returned the best score. The register that contained the most reads in a cluster was identified as a potential phasiRNA region and the associated reads as phasiRNAs. Phased regions with a phased score above 100 were considered as putative phased loci. Potential phase-initiating miRNAs were identified by extending each phased locus with 100 nts to the 5′ and 3′ ends. The known and putative miRNAs identified per cultivar was used to predict miRNA cleavage sites on both strands of the phased loci with the 100 nt extensions using psRNATarget [38]. To identify in-phase cleavage sites, the main registry (registry with the highest read count) had to be in 21 nt (or the dicer phase size for the specific locus) increments from the cleavage site.

4.2.4 NatsiRNA identification

The *cis*- and *trans*-natural antisense transcripts (NATs) were identified following the workflow described by Visser et al. and Zhou et al. [24, 43]. The *Vitis vinifera* genome annotation was used to identify putative *cis*-NATs. A *cis*-NAT pair was identified if a pair of overlapping genes was located on opposite strands at the same locus and the overlap was equal or longer than 50 nts. To identify *trans*-NATs, *Vitis vinifera* transcript sequences were obtained from the Grape Genome Browser (http://www.genoscope.cns.fr/externe/Download/Projets/Projet_ML/data/12X/annotation/). Transcript

[iii] Appendix A5. Bester R, Pepler PT, Aldrich DJ, Maree HJ (2016). Harbin: A quantitation PCR analysis tool. Biotechnol Lett. DOI 10.1007/s10529-016-2221-1.

To enable comparisons of different relative quantitation experiments, a web-browser application called Harbin was created for a dynamic interaction with qPCR data. A quantile-based scoring system is proposed that will allow for comparison of samples at different time points and between experiments. Harbin simplifies the analysis of high-density qPCR assays, either for individual experiments or across sets of replicates and biological conditions. The application uses the standard curve method for relative quantitation of genes with normalisation using a reference gene index to calculate a concentration ratio (CR). This Harbin quantile bootstrap test for evaluating if different datasets can be combined, was shown to be less conservative than the Kolmogorov-Smirnov test, and therefore more sensitive in detecting distributional differences between data sets. Statistical significance testing for CRs across biological conditions is also possible with Harbin.

sequences included all the coding sequences (CDSs) predicted by Jaillon et al. [1]. The untranslated sequences (UTRs) available for each transcript were concatenated to the transcripts.

Pairwise alignments of these transcripts was performed using standalone BLAST (V 2.2.31+) (100% identity and e-value threshold of 0.00001) [44] to find overlaps with high sequence complementary. A pair of transcripts from different genomic loci was considered as *trans*-NATs if the transcripts had a continuous pairing region equal or longer than 100 nts and if the overlapping region were able to form an RNA–RNA duplex. Duplex formation was validated *in silico* using UNAfold [45]. In order to identify if the overlapping regions are enriched for sRNAs, the density of sRNAs mappings on the overlapping and non-overlapping regions of the NATs was compared. Reads were mapped to the individual transcripts and the overlap sequences with Bowtie (1.1.2) [34] allowing one mismatch and reporting only the best alignment (Bowtie reporting parameter: --best). The number of reads per kilobase of overlapping or non-overlapping NAT regions were calculated and the significance of the enrichment of small RNAs in the overlapping regions was assessed by means of a Wilcoxon rank-sum test.

4.2.5 Repeat-associated siRNA identification

Repeat-associated siRNAs were identified by mapping the reads to previously identified repeat sequences in the *Vitis vinifera* genome. All reads with perfect matches to known vvi-miRNAs were removed from the data sets. Bowtie (1.1.2) [34] was used to map reads to the *Vitis vinifera* repeat sequences present in Repbase Update 21.07 [46] to identify putative rasiRNAs. A single mismatch was allowed between the sRNA read and the repeat sequence to compensate for cultivar differences and only the best alignment was reported per read (Bowtie reporting parameter: --best).

4.2.6 tRNA-derived siRNAs

Reads 17-44 nts in length were used for the tRNA analysis. A tRNA database was created using the mature tRNA sequences of five angiosperms (*Arabidopsis thaliana, Brachypodium distachyon, Medicago truncatula, Oryza sativa and Populus trichocarpa*) available in the PlantRNA database [47]. Bowtie (1.1.2) [34] were used to map reads to the mature tRNA sequences to identify putative tRFs and the tRNA halves allowing two mismatches and reporting only the best alignment (Bowtie reporting parameter: --best).

4.3 Results and discussion

4.3.1 sRNA sequencing data

Twelve sRNA NGS libraries were sequenced and on average 12 million reads were generated per library. Each library consisted of 70 % sRNA reads in the range of 18-26 nts (Table 1). After quality filtering, 80 % of the 18-26 nt reads were retained (Table 1). The most high quality reads in the 18-26 nt sRNA fraction were 21 nts (27.2 - 30.2 %) and 24 nts (26.3 - 27.6 %) in length (Figure 1). The 21 nt size reads are associated with miRNAs and phased siRNAs and the abundance of this size groups observed in all cultivars evaluated, can point to their association with regulating biological functions in plants. The 24 nt reads showed the greatest redundancy (Figure 1). This was also observed by Arikit et al. [48] and can possible be ascribed to heterochromatic siRNAs that fall within the 24 nt size range that originates from a wide set of genomic repeats. Small RNAs associated with ribosomal RNA was identified in all cultivar groups, representing 25 - 27 % of the high quality sequencing reads (Table 2).

Table 1. Small RNA NGS data statistics. Read count per sequencing library before and after quality filtering.

Sample	Library	Total reads	18-26 nt reads	18-26 nt reads after QC	18-26 nt non-redundant reads after QC
Chardonnay (CY7)	HUS1	9795775	7069867	5682351	1553819
Chardonnay (CY8)	HUS2	10271935	6987162	5561984	1376471
Chardonnay (CY10)	HUS3	11241033	7741779	6182702	1588126
Chenin blanc (CB7)	HUS7	12742772	8714933	6929352	1795048
Chenin blanc (CB8)	HUS8	12042720	7887448	6273976	1770949
Chenin blanc (CB10)	HUS9	13625850	10025676	7950515	2127098
Cabernet Sauvignon (CS7)	HUS13	14675941	10911194	8634821	2154639
Cabernet Sauvignon (CS8)	HUS14	12736929	9457986	7546111	1718277
Cabernet Sauvignon (CS10)	HUS15	14372921	10453027	8379796	2068858
Own-rooted Cabernet Sauvignon (Cab1)	HUS19	13390466	9589029	7661851	1951781
Own-rooted Cabernet Sauvignon (Cab2)	HUS20	12683151	9142351	7337021	1629271
Own-rooted Cabernet Sauvignon (Cab6)	HUS21	14744256	10176066	8168847	2098718
Total		152323749	108156518	86309327	21833055
Average		12693646	9013043	7192444	1819421
Mimimum		9795775	6987162	5561984	1376471
Maximum		14744256	10911194	8634821	2154639

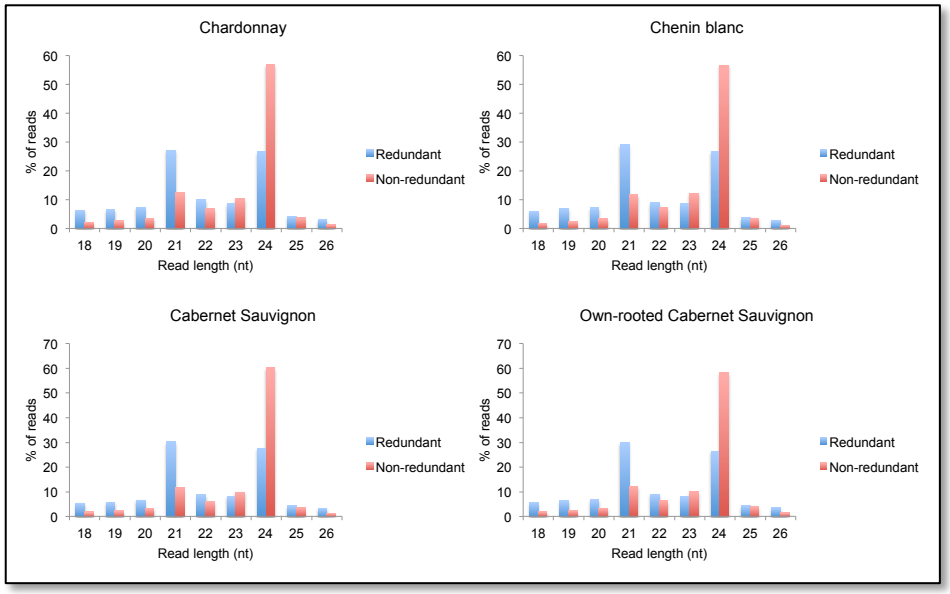


Figure 1. Read length distribution per cultivar. Histogram illustrating size distribution of the number of reads per read length as a percentage of the total number of 18-26 nt sized reads.

Table 2. Summary of read counts identified per sRNA species.

	Chardonnay	Chenin blanc	Cabernet Sauvignon	Own-rooted Cabernet Sauvignon
High quality reads (18-26 nts)	17427037	21153843	24560728	23167719
Known vvi-miRNAs	1611797	2275750	2801523	2435313
Novel miRNAs	512133	803691	1086498	941117
phasiRNA	1019398	1046989	1479948	1247486
natsiRNA	391	713	656	614
rasiRNA	637664	736388	1010002	740761
tRNA-derived (18-26 nts)	674052	1357318	822468	1045242
tRNA-derived (17, 27-44 nts)	284501	332450	327417	383385
tRNA-derived (17-44 nts)	958553	1689768	1149885	1428627
rRNA	4809084	5496168	6172077	6293125

4.3.2 Known miRNAs

Reads were compared to the publicly available miRBase version 21 [3]. MiRBase version 21 contains 186 mature miRNA sequences that can be classified into 35 families based on the miRNA hairpin precursor. Of these 186 mature sequences, only 119 are unique sequences. To identify known vvi-miRNAs, only perfect matches were allowed, and more than 1.6 million reads mapped to miRBase entries for Chardonnay, Chenin blanc, Cabernet Sauvignon and own-rooted Cabernet Sauvignon (Table 2). Seventy of the 119 unique vvi-miRNA entries were present in all cultivars evaluated, with 76 (33 families) in Chardonnay, 80 (34 families) in Chenin blanc, 76 in Cabernet Sauvignon (34 families) and 77 (34 families) in own-rooted Cabernet Sauvignon (Appendix C2:A). No members of

the MIR158 family were detected in cultivars Chardonnay and Chenin blanc, while miR828a was only detected in one sample of cultivar Chenin blanc with a very low read count. More than 70 % of the reads that mapped to known vvi-miRNAs had a perfect match with the vvi-miR166c-h cluster. The second largest miRNA family was vvi-miR3634-3p, followed by vvi-miR159c. The vvi-miR166 family are predicted to target homeobox-leucine zipper proteins that play an important role in plant growth and development [49, 50]. This high level of vvi-miR166 was also previously observed in grapevine [51] and apple [43]. Previous studies showed that miR166 regulates a range of plant developmental processes, including shoot meristem formation, floral and vascular development and leaf polarity. Results from mutation studies of the miRNA complementary site on homeobox-leucine zipper genes suggested that the binding of miR166 to these genes for negative regulation is required for normal plant development [52, 53].

4.3.3 Novel miRNA prediction

The majority of the vvi-miRNAs in the miRBase registry was identified in cv. Pinot noir [1, 54]. To expand the miRNA knowledge base and to identify potential cultivar specific miRNAs, a miRBase-independent analysis was performed to predict putative miRNAs. MicroRNA precursors were predicted for Chardonnay, Chenin blanc, Cabernet Sauvignon and own-rooted Cabernet Sauvignon (Table 3). The most abundant mature miRNA sequences were extracted for each locus, together with its complement (previously named miRNA star) to compare between cultivars (Table 3). Only loci with a mature miRNA sequence represented by at least 10 reads were analysed further (Figure 2, Appendix C2:B-E). The number of known miRNAs identified using ShortStack was low compared to the 100 % identity analysis performed. Only 22, 25, 26 and 25 known miRNA sequences were detected compared to the 76, 80, 76 and 77 reads identified with 100 % identity to miRBase sequences for Chardonnay, Chenin blanc, Cabernet Sauvignon and own-rooted Cabernet Sauvignon, respectively (Appendix C2). This can be expected, since miRNA annotation by the latest version of ShortStack significantly reduces false positives at the cost of an increased false negative rate. Predicted miRNA precursor loci that overlapped with loci of known vvi-miRNAs were identified. However, the mature miRNA sequence with the highest read count was different to the known miRNA sequence. IsomiRs (sequence variants) of known vvi-miRNA, as well as miRNAs identical to miRNAs from other plant species were identified (Appendix C2:B-E). The differences in the mature miRNA sequences identified, compared to known miRBase entries, can be due to different isomiRs being expressed at different levels relative to each other as a result of cultivar, tissue type or environmental differences. Putative novel precursor miRNA loci were also identified in all cultivars (Figure 3). A few of the mature miRNAs with the most abundant read count predicted for these novel precursors were the same as known vvi-miRNAs. Some of the novel miRNA loci, with novel mature sequences, were previously identified in other studies, but not submitted to miRBase [55–59] (Appendix C2:B-E). Between the

different cultivars used in this study, 45 novel mature miRNAs were predicted. Even though plant miRNAs mainly originate from non-coding intergenic regions, miRNA loci were predicted that overlapped with *Vitis vinifera* transcript sequences (Appendix C2:B-E, Table 4). This can indicate that grapevine may also use exons and spliced introns as a source of miRNAs as was shown for potato and rice [60, 61].

In silico target prediction with psRNAtarget identified putative targets for more than 88 % of the novel mature miRNAs in the different cultivar groups (Appendix C2:F-I). Multiple targets were identified for several miRNAs resulting in more than 190 different *Vitis vinifera* transcripts predicted as targets in the different cultivar groups (Appendix C2:F-I).

Ten putative mature miRNAs were selected based on either expression in all cultivars or cultivar specificity and validated with stemloop RT-qPCR. Four were predicted in all cultivars, one in Chardonnay and Chenin blanc, two in Chenin blanc and three in Cabernet Sauvignon. Stemloop RT-qPCRs validated all ten the mature miRNAs in the samples analysed (Figure 4). No cultivar specific miRNA was identified amongst the 10 evaluated. Even though ShortStack predicted cultivar-specific miRNAs amongst the 10 selected for qPCR analysis, reads for the predicted miRNAs were present in all the samples. ShortStack will only predict a miRNA locus if there is 100 % evidence that supports the annotation of a miRNA. The absence of the predictions for the miRNAs in the other cultivars can therefore be false negatives. The 10 validated novel miRNAs are predicted to target 50 *Vitis vinifera* transcripts (Appendix C2:J) that could be linked to gene ontology (GO) terms mainly associated with metabolic processes, cyclic compound binding and intracellular cellular components (Figure 5).

Table 3. Number of putative miRNAs per cultivar predicted using ShortStack.

	Chardonnay	Chenin blanc	Cabernet Sauvignon	Own-rooted Cabernet Sauvignon
Predicted genomic loci	66	79	74	72
Putative mature miRNAs with read count of at least 10	54	59	57	58
Putative mature miRNAs with read count of at least 10	54	55	54	56

Table 4. Pairwise comparison to identify number of overlaps between predicted hairpin sequences. Comparisons between cultivars and to predicted *Vitis vinifera* gene sequences are shown.

	Predicted hairpin overlaps			Overlap with gene sequences	
	Chardonnay	Chenin blanc	Cabernet Sauvignon	Genes with introns	Exons
Chardonnay				5	0
Chenin blanc	53			10	2
Cabernet Sauvignon	48	55		5	2
Own-rooted Cabernet Sauvignon	45	54	56	6	0

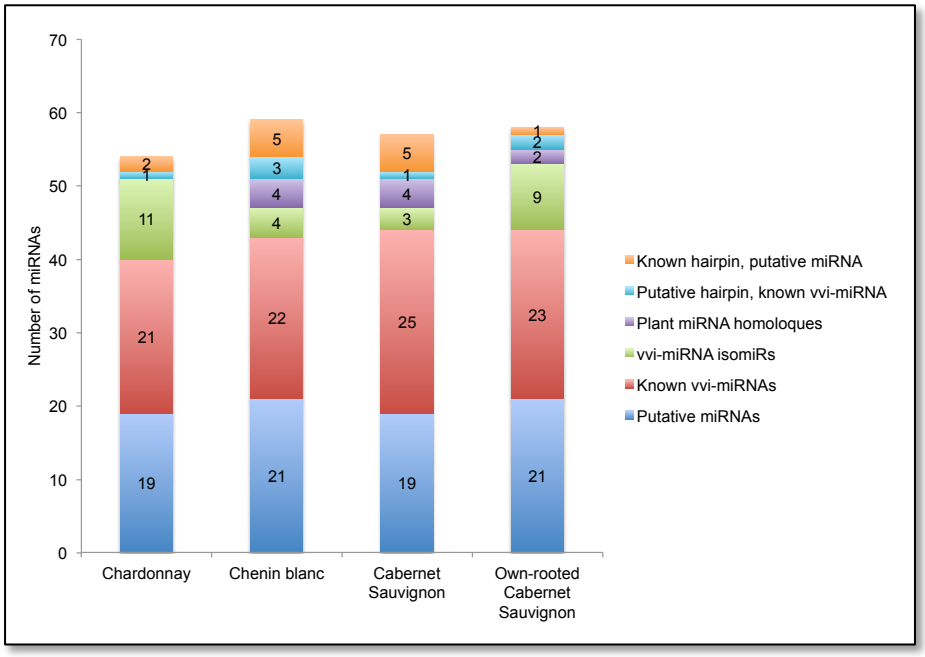


Figure 2. Known and novel miRNA predictions. Stacked histogram displaying the number of mature miRNAs predicted per category.

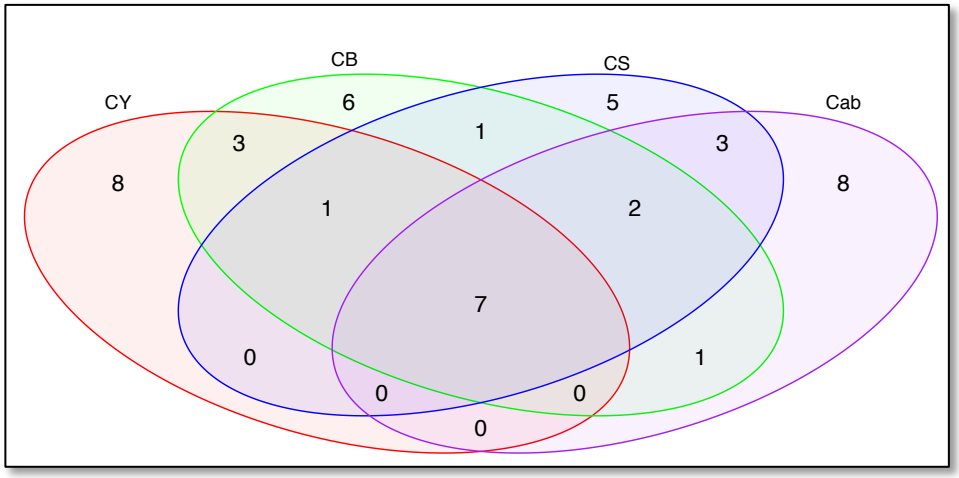


Figure 3. Venn diagram displaying overlaps between the different cultivars' putative mature miRNA sequences. CY: Chardonnay, CB: Chenin blanc, CS: Cabernet Sauvignon, Cab: Own-rooted Cabernet Sauvignon.

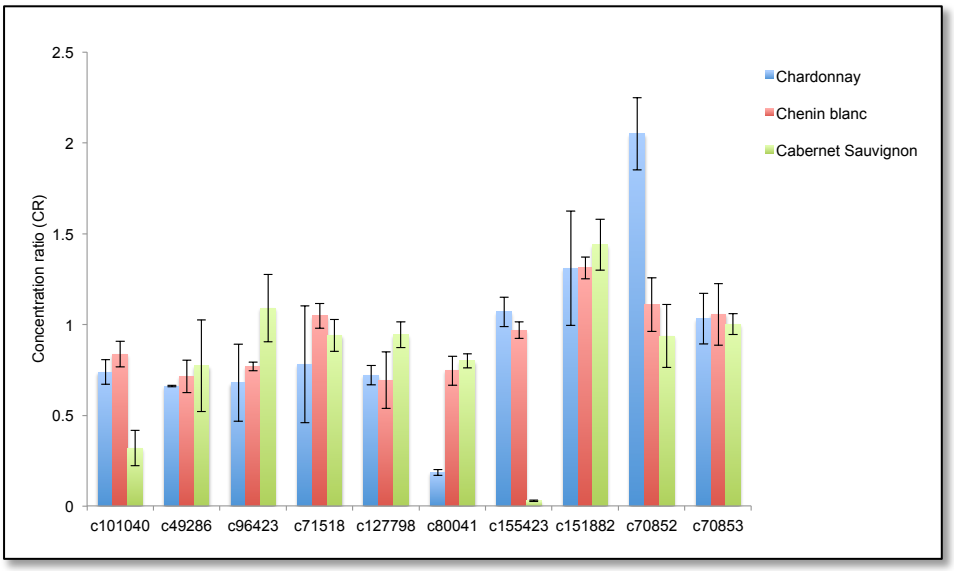


Figure 4. Predicted miRNAs validated using stemloop RT-qPCR. The mean concentration ratio (CR) ± standard error (SE) of three biological replicates, with each replicate an average of three technical replicates is displayed.

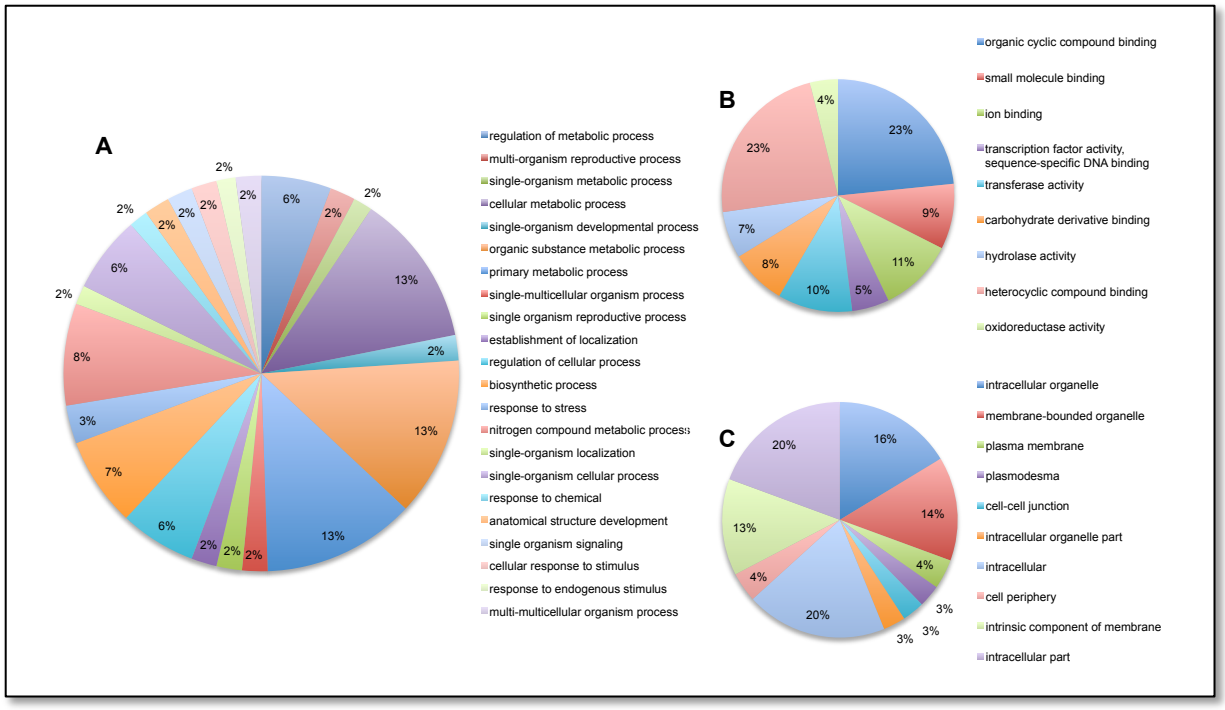


Figure 5. Gene ontology terms associated with the predicted target genes of the validated novel miRNAs. A: Biological processes; B: Molecular function; C: Cellular component.

4.3.4 Phased loci analysis

Trans-acting small interfering RNAs differ from miRNAs in that they arise from double-stranded RNA that requires the activity of RNA-dependent RNA polymerases. A primary transcript is cleaved

by an sRNA molecule (i.e. miRNA), after which the RNA product is then transcribed by an RNA-dependent RNA polymerase. This leads to dsRNA that is processed by Dicer-like proteins to yield sRNAs capable of acting in *trans* to regulate a distinct mRNA target. The dsRNA precursor can also be cleaved sequentially to produce phased siRNA in a 21 nt register with the primary sRNA cleavage site [62–64]. Not all of these siRNAs function in *trans* and the term phased siRNA (phasiRNA) was introduced for these siRNAs irrespective if they target other transcripts or their transcript of origin (*PHAS* genes) [65]. The phasing phenomenon can be guided by miRNAs through either one or two miRNA binding sites. The one-hit miRNA trigger is typically 22 nts in length [63, 66] and the two-hit model requires two 21-nucleotide miRNA target sites to trigger phasiRNAs formation [64].

The ShortStack analysis predicted phased loci with a phased score above 100 for Chardonnay, Chenin blanc, Cabernet Sauvignon and own-rooted Cabernet Sauvignon (Table 5, Appendix C3:A-D). Eighty-three percent of these loci had a dicer phase size of 21 nts. Cultivar-specific loci were also identified. In plants, phasiRNAs were shown to be generated from both protein-coding and intergenic loci [67]. On average, 56 % of the loci overlapped with *Vitis vinifera* predicted transcripts per cultivar (Table 5). The loci that overlapped with predicted transcripts were annotated using Blast2GO (Appendix C3:A-D). More than 60 % of these loci were identified to have a high similarity to disease resistance proteins. Some of the loci also had a high similarity to ankyrin repeat-containing and pentatricopeptide repeat-containing proteins, as well as NAC transcription factors, MYB transcription factors and auxin signalling F-box proteins. Previous studies have reported these proteins to not only be *PHAS* loci in plants, but also targets of phasiRNAs [48, 65, 68–71]. All these proteins play an important role during gene expression, emphasising the significance of the production of phasiRNA for the regulation of plant development.

More than 75 % of the loci identified in the four groups overlapped with loci previously identified as *PHAS* genes in *Vitis vinifera* (Appendix C3:A-D) [72]. Four non-coding *trans*-acting siRNA loci (*TAS* genes) were first identified in *Arabidopsis thaliana* [22]. The *TAS1* and *TAS2* loci are unique to *Arabidopsis thaliana*, however *TAS3* and *TAS4* were also found in other plants [70, 72–74]. *TAS3* is flanked by dual miR390 complementary sites [64], following the two-hit model for phasing, while the one-hit miRNA initiator of *TAS4* is miR828 [75]. A phased locus with significant overlap to a *Vitis vinifera TAS3* locus [74] was only predicted in the own-rooted Cabernet Sauvignon libraries (Appendix C3:D). Loci with overlap to the *Vitis vinifera TAS*7 (Chardonnay and Chenin blanc) and *TAS8* (Chardonnay, Chenin blanc, Cabernet Sauvignon and own-rooted Cabernet Sauvignon) loci (Appendix C3:A-D) [74] were also predicted. In both Chardonnay and Chenin blanc, a predicted locus was identified with significant overlap to the *TAS4* locus on chromosome 14, previously identified in grapevine [72, 73]. The phasiRNAs generated from this locus can potentially target the MYB transcription factor gene family and regulate the anthocyanin biosynthesis pathway [75–77] (Appendix

C3:A-B). The absence of the *TAS4* predicted locus from the Cabernet Sauvignon libraries could suggest a cultivar difference in red-berried cultivars compared to white-berried cultivars. Even though the *TAS4* locus was not predicted in Cabernet Sauvignon in this study, it has been reported in red-berried cultivars like Pinot noir and Merlot [73]. Therefore, the cultivar differences are likely due to a phasiRNA concentration difference rather that the complete absence of this locus from certain cultivars. The initiator (miR828) of the *TAS4* phasiRNAs was only identified at a very low level in the Chenin blanc libraries, indicating that the miR828 level was too low to be detected in the other sRNA libraries.

Five, 12, 17 and 13 of the remaining predicted phased loci for Chardonnay, Chenin blanc, Cabernet Sauvignon and own-rooted Cabernet Sauvignon, respectively, had a high phasing signature (phased score above 200) and are potential novel *Vitis vinifera* phased loci. The study of these newly identified loci will contribute to unravelling the function of phasiRNAs in normal development as well as defence-associated regulatory networks in plants.

Eighteen potential phase-initiating miRNAs, targeting the different phased loci predicted for Chardonnay, Chenin blanc, Cabernet Sauvignon and own-rooted Cabernet Sauvignon, were identified using psRNAtarget (Appendix C3:A-D). All the predicted initiator-miRNAs complied with the single (one-hit) phased model. Only one miRNA target cleavage site (miR3634-3p) fell into the phased register with the majority of the reads for one of the phased loci in Chardonnay (cluster_75653), Cabernet Sauvignon (cluster_96093) and own-rooted Cabernet Sauvignon (cluster_93554) (Figure 6). This locus has a high sequence similarity to an Ankyrin repeat-containing protein. Proteins containing Ankyrin repeats are involved in regulating transcription, signal transduction in chloroplasts and systemic resistance to pathogens in plants [78]. The low number of in-phase cleavage sites can possibly be attributed to a different sRNA initiator, other than a miRNA, or phase-drift [71, 79]. Phase-drift can occur after several DCL4 processing cycles of phasiRNA precursors, or if the phased siRNAs are initiated by a second cleavage event in a different region of the same transcript. The reads associated with the main phased registry for each locus, with a phased score above 100 (phasiRNAs) accounted for 4.95 % - 6.03 % of the total library read count after quality control for the different cultivar groups (Table 2).

Table 5. Number of predicted phased loci per cultivar. Number of overlaps between cultivars and overlaps with transcript loci is shown.

	Predicted phased loci with phased score above 100	Loci with transcript overlap	Cultivar loci overlap			
			Chardonnay	Chenin blanc	Cabernet Sauvignon	Own-rooted Cabernet Sauvignon
Chardonnay	65	36		53	53	50
Chenin blanc	68	39			54	53
Cabernet Sauvignon	88	49				73
Own-rooted Cabernet Sauvignon	87	51				

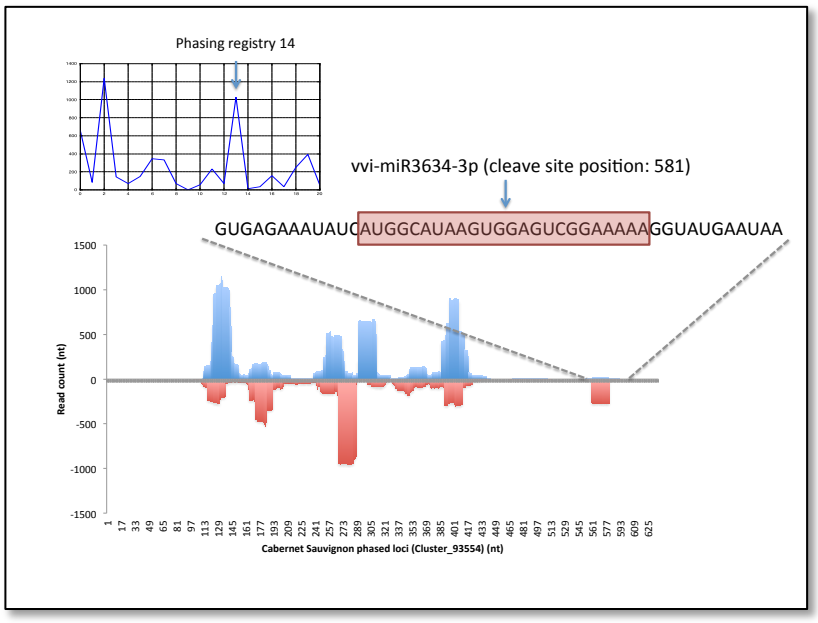


Figure 6. Illustration of the mapping of siRNA reads on the sense and antisense strands of the phased locus Cluster_93554. Two phasing registries with high numbers of siRNAs were identified, with the 14th phased registry in exact 21 nt increments from the vvi-miR3634-3p cleavage site.

4.3.5 NatsiRNAs

Natural antisense transcript siRNAs are involved in the regulation of various developmental processes in plants. The expression of NATs can occur in response to environmental stimuli like pathogen infections, temperature and salt stress [24, 80–82], and can be developmental stage or tissue-specific [24]. Antisense transcription has also been linked to the control of cytokinin levels in plants [83], cell wall biosynthesis [84], reproductive function [85] and disease resistance [82, 86].

In this study, the genomic loci of all annotated transcript were compared in order to search for transcript pairs that overlap in an antiparallel manner, to identify potential *cis*-NATs. The *Vitis vinifera* genome contains 26 346 genes based on the 12X PN40024 annotation from the Grape Genome Browser on Genoscope [1]. Eleven overlapping transcripts were identified on opposite strands at the same locus, however none of the overlaps were equal or longer than 50 nts.

The *trans*-NATs were identified through pairwise alignments of the UTRs and coding sequences of each transcript. Twenty-six transcript pairs, with a continuous pairing region equal or longer than 100 nts, were identified. Twenty-two transcript pairs were predicted to form RNA-RNA duplexes using UNAfold. Three pairs contained two overlapping regions and three transcripts were involved in two pairings, resulting in 25 *trans*-NAT pairs (Appendix C4). Previous studies have shown that *trans*-NATs can form one-to-one, one-to-many or many-to-many relationships, indicating the complex regulatory networks where NATs may be involved in [24, 43].

The computational analysis predicted the NATs that have the potential to hybridise in plant cells, however to identify NATs that are expressed in the same cellular location for the duplexes to form, the sRNA libraries were searched for potential NAT-associated siRNAs. The sRNA reads were mapped to the overlap regions, and the whole transcript, to determine if the overlapping regions are enriched for sRNAs. Only *trans*-NATs that gave rise to small RNAs from their overlapping regions were analysed for the different cultivar groups (Table 6). If two out of the three samples per cultivar group had no read mappings to the overlap region, the NAT was removed from the statistical analysis for the specific variant group. The mean density of sRNAs mapping on the overlapping regions, compared to the non-overlapping regions of the NATs, was higher (Figure 7), however no significant enrichment of small RNAs in the overlapping regions overall was found for the Chardonnay (p-value = 0.4462), Chenin blanc (p-value = 0.4869), Cabernet Sauvignon (p-value = 0.6934) or own-rooted Cabernet Sauvignon (p-value = 0.5445) samples. Overlapping regions of individual *trans*-NATs were enriched in the different cultivar groups (Chardonnay = 2, Chenin blanc = 4, Cabernet Sauvignon = 2, own-rooted Cabernet Sauvignon = 4). The overlapping regions of *trans*-NAT pairs GSVIVT01010800001 and GSVIVT01011363001 were enriched for sRNAs in all four cultivar groups (Figure 8). In *Vitis vinifera,* GSVIVT01011363001 has high sequence similarity to a polyvinylalcohol dehydrogenase-like gene with an LTR-retrotransposon-like element (gag-pol). Plant alcohol dehydrogenases have been shown to be involved in lignin biosynthesis and defence-related functions [87, 88], and its down-regulation can influence plant secondary metabolism and response to stress. A strand-bias in the natsiRNAs of this *trans*-NAT pair was also observed (Figure 8). A 1.3-2.5-fold increase was observed for the different cultivar groups indicating that the siRNAs were derived predominantly from one of the NATs and can suggest the possible down-regulation of this transcript.

The NAT pair analysis may underestimate the number of NATs in *Vitis vinifera*, since the current genome annotation, which is based on the Pinot noir genome, may lack transcripts specific to other cultivars. The *cis*-NAT analysis also focussed on transcripts only and no pseudogenes and transposons were retained in the analysis. Only transcripts located on opposite strands at the same locus were included to identify *cis*-NAT and not transcripts located at adjacent genomic loci, as was done in a previous study to identify *Vitis vinifera cis*-NATs [89]. The *trans*-NAT analysis was based on conservative BLAST parameters (100 % identity and e-value of 0.00001) that can also explain the low number of *trans*-NATs identified.

Table 6. Number of *trans*-natural antisense transcripts identified per cultivar. Average density and overlap enrichment score is shown.

	trans-NATs			
	Chardonnay	Chenin blanc	Cabernet Sauvignon	own-rooted Cabernet Sauvignon
Pairs	14	13	12	13
Portion of total transcripts (%)	0.11	0.10	0.05	0.10
Overlap median length (nt)	184	151	184	201
One-to-one	13	12	11	12
One-to-many	1	1	1	1
Average density (Reads/kb) in overlap (non-overlapping region)	5.51 (4.81)	4.31 (6.19)	5.49 (7.43)	4.61 (7.96)
Overlap enrichment (p-value)	0.4462	0.4869	0.6934	0.5445
> 2-fold strand bias	3	4	2	5

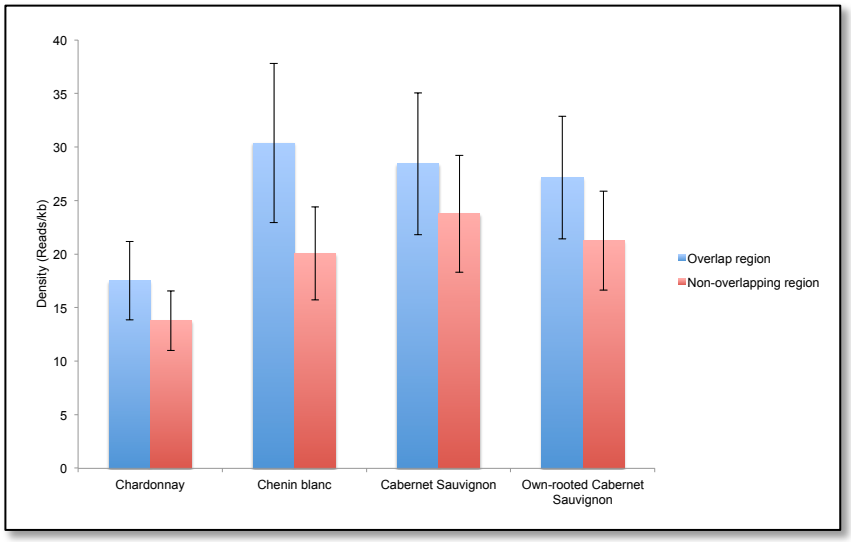


Figure 7. Histogram displaying the number of reads per kilobase of overlapping or non-overlapping NAT regions. The mean density for each cultivar group is displayed with standard error bars.

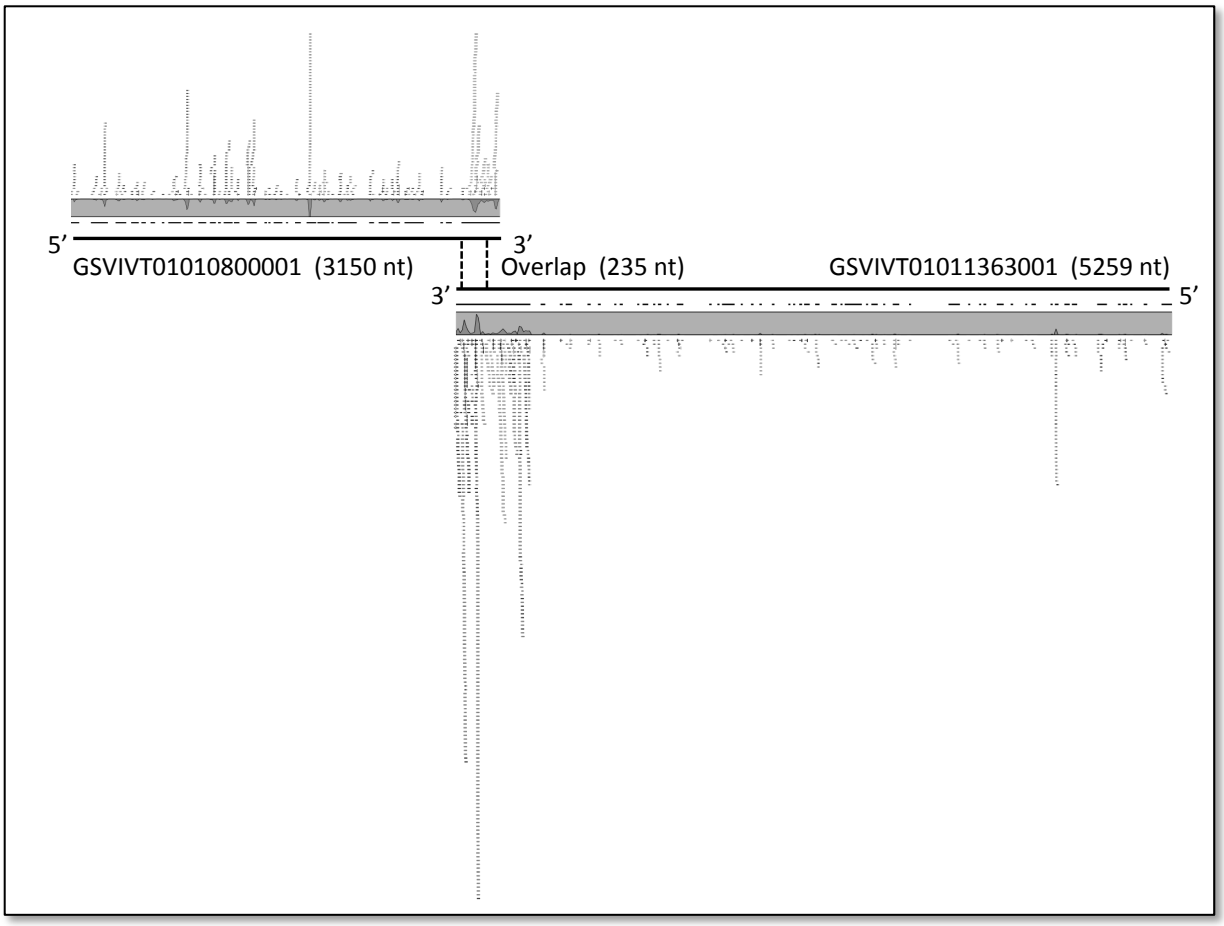


Figure 8. Illustration showing the *trans*-NAT pair GSVIVT01010800001/GSVIVT01011363001. The siRNA enrichment in the overlapping region and the strand bias in the overlapping region of GSVIVT01011363001 is shown.

4.3.6 Repeat-associated siRNA identification

Heterochromatic siRNAs are a diverse type of endogenous sRNAs, which are mainly 24 nts long and are involved in the repression of transposable elements (TEs), other repetitive DNA sequences and DNA methylation in some sequences [90, 91]. These TEs can influence the size, organisation and genetic diversity of the host's genome. In cotton it was found that altered levels of 24 nt heterochromatin-associated siRNAs were correlated with reactivation of transposable elements in virus-infected plants [92], while in the pollen of *Arabidopsis* it was shown that the 21 nt siRNAs from a class of retrotransposons can activate TEs in the vegetative nucleus, and target silencing in gametes [93]. A hybridisation study in maize found different levels of 21–22-nt retrotransposon-derived siRNAs accumulating in different maize lines, indicating the ability of these siRNAs to contribute to genetic variation in plant species [21].

Repeat-associated siRNAs were identified through mapping the sRNA reads to the Repbase Update 21.07. The Repbase Update 21.07 contains 362 unique *Vitis vinifera* repeat sequences that can be classified into DNA transposons, integrated virus sequences, LTR and non-LTR retrotransposons. This study showed that all the repeat sequences produced sRNAs with at least 94 % of the 362 repeats having a read count of at least 10 for the different cultivar groups (Appendix C5). The largest cluster of reads (550,108 reads) mapped to LTR retrotransposons Copia and Gypsy, followed by the integrated virus sequences of Caulimoviridae (Table 7). These two LTR retrotransposon superfamilies are the major components of intergenic regions in all the plant genomes that has been sequenced [94]. The single repeat sequences with the highest read count were CAULIV1, hAT-13 and Gypsy-26 for all cultivar groups, except for Cabernet Sauvignon. This group had much higher levels of V1_I_Gypsy, V1_LTR_Gypsy and Gypsy-9 in one of the biological replicates. In the remaining two biological replicates the same repeat sequences had the highest read counts, similar to the other cultivar groups. This was the first suggestion of variation between the three biological replicates of Cabernet Sauvignon and inferences regarding the statistical analysis of these samples will be made with caution. The siRNAs associated with the Caulimoviridae repeat sequences can play a role in viral immunity since the hypothesis is that the presence of these endogenous viral sequences can be involved in heritable virus resistance in plants through small RNA-mediated methylation or degradation of the viral RNA [95, 96]. The majority of the rasiRNAs identified were 24 nts in length (Figure 9). This size-group has repeatedly been linked to heterochromatin-associated siRNAs with the second largest size of 21 nts (Figure 9) also associated with TE silencing [21, 92, 93]. A strand bias (more than 2-fold difference) were observe in 15 % of the rasiRNAs mapping to the single repeat sequences in at least one cultivar group, and 5 % had a higher than 2-fold difference in all cultivar groups. The majority of these repeat sequences were Copia LTR retrotransposons. A Copia repeat in the Apple genome was also previously reported to have a siRNA strand bias [43].

Table 7. Total read count for each transposable element superfamily/clade.

		Read count			
Superfamily/clade	Group	Chardonnay	Chenin blanc	Cabernet Sauvignon	Own-rooted Cabernet Sauvignon
Copia	LTR retrotransposon	232334	266256	306015	256199
Gypsy	LTR retrotransposon	200625	208644	415608	222188
Caulimoviridae	Integrated virus sequences	74552	78736	109842	100158
hAT	DNA transposon	54083	80505	67022	64763
MuDR	DNA transposon	33110	45835	47872	42519
L1	Non-LTR retrotransposon	12971	16382	19542	15317
EnSpm/CACTA	DNA transposon	10665	13596	14147	13149
Harbinger	DNA transposon	8766	12129	13250	12637
Helitron	DNA transposon	19	34	30	30

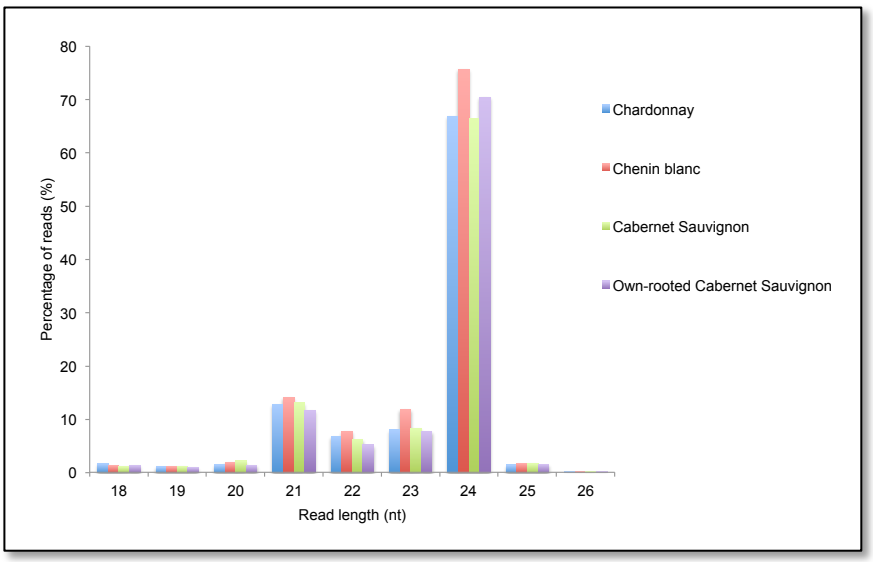


Figure 9. Size distribution of repeat-associated siRNA. Histogram displaying the number of repeat-associated siRNA reads per read length as a percentage of the total number of repeat-associated siRNA reads in the18-26 nt size-range.

4.3.7 tRNA-derived siRNAs

Transfer RNA-derived siRNAs have been identified in many organisms and are a growing, not well-understood class of non-coding RNAs. They are separated into two main classes according to their length and biogenesis. The tRNA halves are usually 28-36 nts in length, due to cleavage in the anticodon loop [25] and the second group are 14-22 nts in length and termed tRNA fragments (tRFs) [25]. The tRFs can further be divided in those originating from the 5′ end of tRNAs (cleaved in the D-loop), the 3′ end of mature tRNAs (cleaved in the T-loop) and those generated from the 3′ end of pre-tRNAs [27, 28].

Due to their precise sequence structure and size, specific expression patterns and associated biological function [27], tRFs are believed not to be tRNA degradation by-products. The asymmetric generation of preferentially either the 5′ or 3′ end fragments, the anti-correlation in abundance of the different types of tRF compared to the number of parent tRNA and the precise cleavage at specific bases, suggest that they originate from tRNAs in a non-random manner and that they can have a regulatory function similar to other sRNA species [29, 97].

In this study, the tRNA-derived sRNAs varied in size from 17-44 nts in length, representing putative tRNA-derived sRNAs originating from tRNA cleavage in the D-, T- or anticodon-loop. Small RNA reads associated with tRNA sequences from 561 unique nuclear tRNAs, 32 unique mitochondrial tRNAs and 97 unique chloroplast tRNAs, with a read count of at least 10, were identified in all

cultivar groups (Appendix C6). The high quality 18-26 nt reads included 3.3 - 6.4 % of the tRNA-derived siRNAs for the different variant groups (Table 2).

The majority of tRNA-derived sRNAs were 18 and 19 nts in length (Figure 10). The size range of the tRNA-derived sRNAs was limited by the polyacrylamide gel size selection before library preparation and can explain why a lower fraction of tRNA-derived sRNAs was identified compared to a previous study that reported tRFs as second in abundance only to miRNAs [27]. Even though reads longer than 30 nts were obtained, it is possible that more tRNA-derived sRNAs longer than 30 nts are present in these plants, and that the current fraction is not necessarily an absolute representation of the larger sRNA species. The obtaining of these longer sRNAs can be as a result of specific secondary structures that modified the movement of the sRNAs through the polyacrylamide gel.

The most abundant class of tRNA-derived sRNAs originated from the 5′ end of the tRNAs for all cultivar groups (Table 8), as was reported previously [98, 99]. Based on the size of the tRNA-derived sRNAs (longer than 28 nts), 11.0 %, 6.8 %, 10.6 %, 11.7 % of the Chardonnay, Chenin blanc, Cabernet Sauvignon and own-rooted Cabernet Sauvignon tRNA-derived sRNA reads were classified as tRNA-halves. The tRNA-derived sRNAs displayed a higher than 96 % sequence redundancy and 4600, 4865, 4837, 5388 of the unique tRNA-derived sRNAs had a read count of at least 10 for Chardonnay, Chenin blanc, Cabernet Sauvignon and own-rooted Cabernet Sauvignon, respectively. More than 98% of the tRNA-derived sRNAs identified were present in all cultivar groups (Appendix C7). The tRNA-derived sRNAs unique to each cultivar had very low read counts with only four and seven having a read count of at least 100 in Chardonnay and Chenin blanc, respectively. The most abundant single tRNA-derived sRNA in all cultivar groups was a 5′ tRF of 18 nts originating from tRNA-Arg^{CCT} (Appendix C7). The consistent pattern of tRNA-derived sRNAs identified in each cultivar group suggests that these sRNAs are functional molecules with specific biogenesis rather than random degradation products.

The function of these tRNA-derived sRNAs remains to be elucidated. However, it is believed that tRFs can result in the down-regulation of gene expression. The increased generation of tRNA fragments have been linked to stress, including changes to environmental conditions and pathogen stress [99–103]. In *Arabidopsis thaliana*, specific tRFs were also found to be overexpressed in root tissues treated with phosphate deprivation [104]; in rice, differential expression of tRFs was found in callus and leaves [105]; and in barley, a tRF was the most abundant sRNA identified [106]. Furthermore, in the phloem sap of pumpkin, tRNA-derived sRNAs were linked to the long-distance signalling system observed in plants [107]. Transfer RNA-derived fragments were also shown to be associated to members of the RNA silencing mechanism [28], and to directly bind to key enzymes

during protein synthesis [101]. These tRFs were found to be associated with Argonaute proteins, and *trans*-silencing activity similar to miRNAs and siRNA were shown [28, 98].

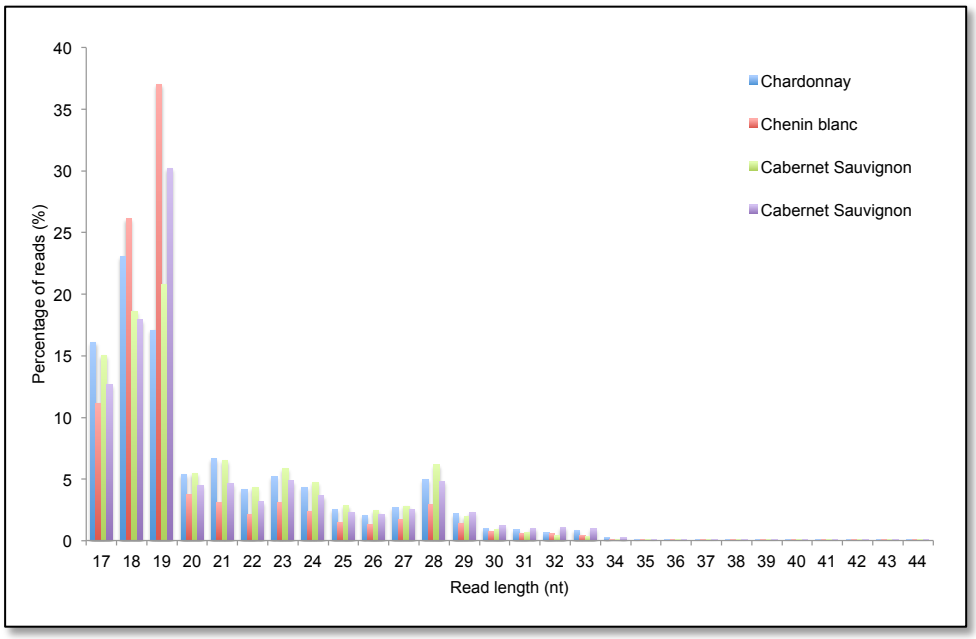


Figure 10. Size-distribution of tRNA-derived sRNA reads. Histogram displaying the number of tRNA-derived sRNA reads per read length as a percentage of the total number of tRNA-derived sRNA reads in the 17-44 nt size-range.

Table 8. Read count of tRNA-derived siRNAs per cultivar. Classification of tRNA-derived sRNAs based on first nucleotide position on mature tRNA is shown.

	Read count (%)				
	5′ fragment	Anticodon fragment	3′ fragment	Variable region	Total
Chardonnay	500419 (52.4)	40187 (4.2)	366162 (38.4)	47376 (5.0)	954144
Chenin blanc	1293959 (76.9)	52034 (3.1)	292319 (17.4)	45117 (2.7)	1683429
Cabernet Sauvignon	596221 (52.1)	37332 (3.3)	423121 (37.0)	69690 (6.1)	1144503
Own-rooted Cabernet Sauvignon	939443 (66.0)	54054 (3.8)	366875 (25.8)	62836 (4.4)	1423208

4.4 Conclusion

In this chapter, the sRNA profiles of three *Vitis vinifera* cultivars were characterised utilising NGS. The grapevine miRNA knowledge base was extended through the identification of 45 putative mature *Vitis vinifera* miRNAs and the presence of ten putative miRNAs was validated using stemloop RT-qPCRs. Both non-coding and protein-coding gene regions were identified as putative phased loci, producing phasiRNAs. The majority of the protein-coding gene loci were identified as disease resistance proteins, indicating a possible stress response mechanism in grapevine. Potential phase-initiating miRNAs were identified from both the known and novel predicted miRNA pools and one

miRNA target cleavage site fell into the main phasing register for three of the cultivar groups. This phased locus was annotated as an Ankyrin protein involved in transcription, signal transduction and pathogen resistance. The absence of this phasing signature in Chenin blanc plants can suggest a cultivar-specific response. One of the *in silico* predicted *trans*-NATs had a significant enrichment for sRNAs in the overlap region for all cultivar groups, indicating a central role for these natsiRNAs in regulating plant development. The rasiRNAs identified, extend the sRNA profile of grapevine, displaying the high number of sRNAs originating from almost all the repeat sequences in the *Vitis vinifera* genome. This study also confirmed the non-random manner in which tRNA-derived sRNAs originate. The number of tRNA-halves identified in Chenin blanc was lower compared to the other cultivar groups. Since these tRNA-derived sRNAs are predominantly linked to stress, this finding may suggest a possible difference in the Chenin blanc defence response compared to the other cultivars.

This study is the most extensive characterisation of the sRNA profiles of grapevine to date and contributes significantly to establishing an sRNA reference database for unravelling sRNA regulation in plants, potentially creating tools for grapevine functional studies.

4.5 Supplementary material

Appendix C1: Primers for novel *Vitis vinifera* miRNA stemloop RT-qPCRs.

Appendix C2: Known and novel miRNAs.

Appendix C3: Phased genomic loci.

Appendix C4: sRNA reads associated with *Vitis vinifera trans*-natural antisense transcripts.

Appendix C5: sRNA reads associated with *Vitis vinifera* repeat sequences.

Appendix C6: sRNA reads associated with tRNA sequences.

Appendix C7: tRNA-derived siRNA.

4.6 References

1. Jaillon O, Aury J-M, Noel B, et al (2007) The grapevine genome sequence suggests ancestral hexaploidization in major angiosperm phyla. Nature 449:463–467. doi: 10.1038/nature06148
2. Naithani S, Raja R, Waddell EN, et al (2014) VitisCyc: a metabolic pathway knowledgebase for grapevine (Vitis vinifera). Front Plant Sci 5:644. doi: 10.3389/fpls.2014.00644
3. Kozomara A, Griffiths-Jones S (2014) miRBase: annotating high confidence microRNAs using deep sequencing data. Nucleic Acids Res 42:D68–D73. doi: 10.1093/nar/gkt1181
4. Vitulo N, Forcato C, Carpinelli EC, et al (2014) A deep survey of alternative splicing in grape reveals changes in the splicing machinery related to tissue, stress condition and genotype. BMC Plant Biol 14:1–16. doi: 10.1186/1471-2229-14-99
5. Zenoni S, Ferrarini A, Giacomelli E, et al (2010) Characterization of Transcriptional Complexity during Berry Development in Vitis vinifera Using RNA-Seq. PLANT Physiol 152:1787–1795. doi: 10.1104/pp.109.149716

6. Zamboni A, Di Carli M, Guzzo F, et al (2010) Identification of Putative Stage-Specific Grapevine Berry Biomarkers and Omics Data Integration into Networks. PLANT Physiol 154:1439–1459. doi: 10.1104/pp.110.160275

7. Fasoli M, Dal Santo S, Zenoni S, et al (2012) The Grapevine Expression Atlas Reveals a Deep Transcriptome Shift Driving the Entire Plant into a Maturation Program. Plant Cell 24:3489–3505. doi: 10.1105/tpc.112.100230

8. Dal Santo S, Tornielli GB, Zenoni S, et al (2013) The plasticity of the grapevine berry transcriptome. Genome Biol 14:R54. doi: 10.1186/gb-2013-14-6-r54

9. Bartel DP (2004) MicroRNAs: genomics, biogenesis, mechanism, and function. Cell 116:281–297. doi: 10.1016/S0092-8674(04)00045-5

10. Axtell MJ (2013) Classification and comparison of small RNAs from plants. Annu Rev Plant Biol 64:137–159. doi: 10.1146/annurev-arplant-050312-120043

11. Khraiwesh B, Zhu J-K, Zhu J (2012) Role of miRNAs and siRNAs in biotic and abiotic stress responses of plants. Biochim Biophys Acta BBA - Gene Regul Mech 1819:137–148. doi: 10.1016/j.bbagrm.2011.05.001

12. Guleria P, Mahajan M, Bhardwaj J, Yadav SK (2011) Plant Small RNAs: Biogenesis, Mode of Action and Their Roles in Abiotic Stresses. Genomics Proteomics Bioinformatics 9:183–199. doi: 10.1016/S1672-0229(11)60022-3

13. Kruszka K, Pieczynski M, Windels D, et al (2012) Role of microRNAs and other sRNAs of plants in their changing environments. J Plant Physiol 169:1664–1672. doi: 10.1016/j.jplph.2012.03.009

14. Lee Y, Kim M, Han J, et al (2004) MicroRNA genes are transcribed by RNA polymerase II. EMBO J 23:4051–4060. doi: 10.1038/sj.emboj.7600385

15. Reinhart BJ, Weinstein EG, Rhoades MW, et al (2002) MicroRNAs in plants. Genes Dev 16:1616–1626. doi: 10.1101/gad.1004402

16. Kurihara Y, Watanabe Y (2004) Arabidopsis micro-RNA biogenesis through Dicer-like 1 protein functions. Proc Natl Acad Sci U S A 101:12753–12758. doi: 10.1073/pnas.0403115101

17. Wu L, Zhou H, Zhang Q, et al (2010) DNA Methylation Mediated by a MicroRNA Pathway. Mol Cell 38:465–475. doi: 10.1016/j.molcel.2010.03.008

18. Khraiwesh B, Arif MA, Seumel GI, et al (2010) Transcriptional Control of Gene Expression by MicroRNAs. Cell 140:111–122. doi: 10.1016/j.cell.2009.12.023

19. Obbard DJ, Gordon KH., Buck AH, Jiggins FM (2009) The evolution of RNAi as a defence against viruses and transposable elements. Philos Trans R Soc B Biol Sci 364:99–115. doi: 10.1098/rstb.2008.0168

20. Kreuze JF, Perez A, Untiveros M, et al (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. doi: 10.1016/j.virol.2009.03.024

21. Barber WT, Zhang W, Win H, et al (2012) Repeat associated small RNAs vary among parents and following hybridization in maize. Proc Natl Acad Sci 109:10444–10449. doi: 10.1073/pnas.1202073109

22. Allen E, Xie Z, Gustafson AM, Carrington JC (2005) microRNA-Directed Phasing during Trans-Acting siRNA Biogenesis in Plants. Cell 121:207–221. doi: 10.1016/j.cell.2005.04.004

23. Liu Y, Wang Y, Zhu Q-H, Fan L (2013) Identification of phasiRNAs in wild rice (Oryza rufipogon). Plant Signal Behav 8:e25079. doi: 10.4161/psb.25079

24. Zhou X, Sunkar R, Jin H, et al (2008) Genome-wide identification and analysis of small RNAs originated from natural antisense transcripts in Oryza sativa. Genome Res 19:70–78. doi: 10.1101/gr.084806.108

25. Garcia-Silva MR, Cabrera-Cabrera F, Güida MC, Cayota A (2012) Hints of tRNA-Derived Small RNAs Role in RNA Silencing Mechanisms. Genes 3:603–614. doi: 10.3390/genes3040603

26. Lambertz U, Oviedo Ovando ME, Vasconcelos E, et al (2015) Small RNAs derived from tRNAs and rRNAs are highly enriched in exosomes from both old and new world Leishmania providing evidence for conserved exosomal RNA Packaging. BMC Genomics 16:151. doi: 10.1186/s12864-015-1260-7

27. Lee YS, Shibata Y, Malhotra A, Dutta A (2009) A novel class of small RNAs: tRNA-derived RNA fragments (tRFs). Genes Dev 23:2639–2649. doi: 10.1101/gad.1837609

28. Haussecker D, Huang Y, Lau A, et al (2010) Human tRNA-derived small RNAs in the global regulation of RNA silencing. RNA 16:673–695. doi: 10.1261/rna.2000810

29. Sobala A, Hutvagner G (2011) Transfer RNA-derived fragments: origins, processing, and functions. Wiley Interdiscip Rev RNA 2:853–862. doi: 10.1002/wrna.96

30. Carra A, Gambino G, Schubert A (2007) A cetyltrimethylammonium bromide-based method to extract low-molecular-weight RNA from polysaccharide-rich plant tissues. Anal Biochem 360:318–320. doi: 10.1016/j.ab.2006.09.022

31. Bester R, Pepler PT, Burger JT, Maree HJ (2014) Relative quantitation goes viral: An RT-qPCR assay for a grapevine virus. J Virol Methods 210:67–75. doi: 10.1016/j.jviromet.2014.09.022

32. Jooste AEC, Molenaar N, Maree HJ, et al (2015) Identification and distribution of multiple virus infections in Grapevine leafroll diseased vineyards. Eur J Plant Pathol 142:363–375. doi: 10.1007/s10658-015-0620-0

33. Martin M (2011) Cutadapt removes adapter sequences from high-throughput sequencing reads. EMBnet J 17:10–12. doi: 10.14806/ej.17.1.200

34. Langmead B, Trapnell C, Pop M, Salzberg SL (2009) Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. Genome Biol 10:R25. doi: 10.1186/gb-2009-10-3-r25

35. Axtell MJ (2013) ShortStack: Comprehensive annotation and quantification of small RNA genes. RNA 19:740–751. doi: 10.1261/rna.035279.112

36. Shahid S, Axtell MJ (2014) Identification and annotation of small RNA genes using ShortStack. Methods 67:20–27. doi: 10.1016/j.ymeth.2013.10.004

37. Meyers BC, Axtell MJ, Bartel B, et al (2008) Criteria for Annotation of Plant MicroRNAs. Plant Cell 20:3186–3190. doi: 10.1105/tpc.108.064311

38. Dai X, Zhao PX (2011) psRNATarget: a plant small RNA target analysis server. Nucleic Acids Res 39:W155–W159. doi: 10.1093/nar/gkr319

39. Conesa A, Götz S (2008) Blast2GO: A Comprehensive Suite for Functional Analysis in Plant Genomics. Int J Plant Genomics 2008:1–12. doi: 10.1155/2008/619832

40. Varkonyi-Gasic E, Wu R, Wood M, et al (2007) Protocol: a highly sensitive RT-PCR method for detection and quantification of microRNAs. Plant Methods 3:12. doi: 10.1186/1746-4811-3-12

41. Guo Q, Qu X, Jin W (2015) PhaseTank: genome-wide computational identification of phasiRNAs and their regulatory cascades. Bioinformatics 31:284–286. doi: 10.1093/bioinformatics/btu628

42. Axtell MJ (2010) A Method to Discover Phased siRNA Loci. In: Meyers CB, Green JP (eds) Plant MicroRNAs Methods Protoc. Humana Press, Totowa, NJ, pp 59–70

43. Visser M, van der Walt AP, Maree HJ, et al (2014) Extending the sRNAome of Apple by Next-Generation Sequencing. PLoS ONE 9:e95782. doi: 10.1371/journal.pone.0095782

44. Altschul SF, Gish W, Miller W, et al (1990) Basic local alignment search tool. J Mol Biol 215:403–410.

45. Markham NR, Zuker M (2008) UNAFold. In: Keith JM (ed) Bioinforma. Struct. Funct. Appl. Humana Press, Totowa, NJ, pp 3–31

46. Bao W, Kojima KK, Kohany O (2015) Repbase Update, a database of repetitive elements in eukaryotic genomes. Mob DNA 6:11. doi: 10.1186/s13100-015-0041-9

47. Cognat V, Pawlak G, Duchene A-M, et al (2013) PlantRNA, a database for tRNAs of photosynthetic eukaryotes. Nucleic Acids Res 41:D273–D279. doi: 10.1093/nar/gks935

48. Arikit S, Xia R, Kakrana A, et al (2014) An Atlas of Soybean Small RNAs Identifies Phased siRNAs from Hundreds of Coding Genes. Plant Cell Online 26:4584–4601. doi: 10.1105/tpc.114.131847

49. Elhiti M, Stasolla C (2009) Structure and function of homodomain-leucine zipper (HD-Zip) proteins. Plant Signal Behav 4:86–88.

50. Singh A, Singh S, Panigrahi KCS, et al (2014) Balanced activity of microRNA166/165 and its target transcripts from the class III homeodomain-leucine zipper family regulates root growth in Arabidopsis thaliana. Plant Cell Rep 33:945–953. doi: 10.1007/s00299-014-1573-z

51. Pantaleo V, Vitali M, Boccacci P, et al (2016) Novel functional microRNAs from virus-free and infected Vitis vinifera plants under water stress. Sci Rep 6:20167. doi: 10.1038/srep20167

52. Sakaguchi J, Watanabe Y (2012) miR165/166 and the development of land plants: miR165/166 and land plant development. Dev Growth Differ 54:93–99. doi: 10.1111/j.1440-169X.2011.01318.x

53. 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. doi: 10.1007/s00425-006-0439-1

54. Mica E, Piccolo V, Delledonne M, et al (2009) High throughput approaches reveal splicing of primary microRNA transcripts and tissue specific expression of mature microRNAs in Vitis vinifera. BMC Genomics 10:558. doi: 10.1186/1471-2164-10-558

55. Belli Kullan J, Lopes Paim Pinto D, Bertolini E, et al (2015) miRVine: a microRNA expression atlas of grapevine based on small RNA sequencing. BMC Genomics 16:393. doi: 10.1186/s12864-015-1610-5

56. Han J, Fang J, Wang C, et al (2014) Grapevine microRNAs responsive to exogenous gibberellin. BMC Genomics 15:111. doi: 10.1186/1471-2164-15-111

57. Alabi OJ, Zheng Y, Jagadeeswaran G, et al (2012) High-throughput sequence analysis of small RNAs in grapevine (*Vitis vinifera* L.) affected by grapevine leafroll disease: Small RNAs in leafroll disease-infected grapevine. Mol Plant Pathol 13:1060–1076. doi: 10.1111/j.1364-3703.2012.00815.x

58. Pantaleo V, Szittya G, Moxon S, et al (2010) Identification of grapevine microRNAs and their targets using high-throughput sequencing and degradome analysis: Grapevine microRNAs and their targets. Plant J 960–976. doi: 10.1111/j.1365-313X.2010.04208.x

59. Wang C, Leng X, Zhang Y, et al (2014) Transcriptome-wide analysis of dynamic variations in regulation modes of grapevine microRNAs on their target genes during grapevine development. Plant Mol Biol 84:269–285. doi: 10.1007/s11103-013-0132-2

60. Zhang R, Marshall D, Bryan GJ, Hornyik C (2013) Identification and Characterization of miRNA Transcriptome in Potato by High-Throughput Sequencing. PLoS ONE 8:e57233. doi: 10.1371/journal.pone.0057233

61. Zhu Q-H, Spriggs A, Matthew L, et al (2008) A diverse set of microRNAs and microRNA-like small RNAs in developing rice grains. Genome Res 18:1456–1465. doi: 10.1101/gr.075572.107

62. Yoshikawa M (2005) A pathway for the biogenesis of trans-acting siRNAs in Arabidopsis. Genes Dev 19:2164–2175. doi: 10.1101/gad.1352605

63. Chen H-M, Chen L-T, Patel K, et al (2010) 22-nucleotide RNAs trigger secondary siRNA biogenesis in plants. Proc Natl Acad Sci 107:15269–15274. doi: 10.1073/pnas.1001738107

64. Axtell MJ, Jan C, Rajagopalan R, Bartel DP (2006) A Two-Hit Trigger for siRNA Biogenesis in Plants. Cell 127:565–577. doi: 10.1016/j.cell.2006.09.032

65. Zhai J, Jeong D-H, De Paoli E, et al (2011) MicroRNAs as master regulators of the plant NB-LRR defense gene family via the production of phased, trans-acting siRNAs. Genes Dev 25:2540–2553. doi: 10.1101/gad.177527.111

66. Cuperus JT, Carbonell A, Fahlgren N, et al (2010) Unique functionality of 22-nt miRNAs in triggering RDR6-dependent siRNA biogenesis from target transcripts in Arabidopsis. Nat Struct Mol Biol 17:997–1003. doi: 10.1038/nsmb.1866

67. Fei Q, Xia R, Meyers BC (2013) Phased, Secondary, Small Interfering RNAs in Posttranscriptional Regulatory Networks. Plant Cell 25:2400–2415. doi: 10.1105/tpc.113.114652

68. Xia R, Zhu H, An Y, et al (2012) Apple miRNAs and tasiRNAs with novel regulatory networks. Genome Biol 13:R47. doi: 10.1186/gb-2012-13-6-r47

69. Wang X-B, Jovel J, Udomporn P, et al (2011) The 21-Nucleotide, but Not 22-Nucleotide, Viral Secondary Small Interfering RNAs Direct Potent Antiviral Defense by Two Cooperative Argonautes in Arabidopsis thaliana. Plant Cell 23:1625–1638. doi: 10.1105/tpc.110.082305

70. Zheng Y, Wang S, Sunkar R (2014) Genome-Wide Discovery and Analysis of Phased Small Interfering RNAs in Chinese Sacred Lotus. PLoS ONE 9:e113790. doi: 10.1371/journal.pone.0113790

71. Howell MD, Fahlgren N, Chapman EJ, et al (2007) Genome-Wide Analysis of the RNA-DEPENDENT RNA POLYMERASE6/DICER-LIKE4 Pathway in Arabidopsis Reveals Dependency on miRNA- and tasiRNA-Directed Targeting. Plant Cell Online 19:926–942. doi: 10.1105/tpc.107.050062

72. Zheng Y, Wang Y, Wu J, et al (2015) A dynamic evolutionary and functional landscape of plant phased small interfering RNAs. BMC Biol 13:32. doi: 10.1186/s12915-015-0142-4

73. Rock CD (2013) Trans-acting small interfering RNA4: key to nutraceutical synthesis in grape development? Trends Plant Sci 18:601–610. doi: 10.1016/j.tplants.2013.07.006

74. Zhang C, Li G, Wang J, Fang J (2012) Identification of trans-acting siRNAs and their regulatory cascades in grapevine. Bioinformatics 28:2561–2568. doi: 10.1093/bioinformatics/bts500

75. Rajagopalan R, Vaucheret H, Trejo J, Bartel DP (2006) A diverse and evolutionarily fluid set of microRNAs in Arabidopsis thaliana. Genes Dev 20:3407–3425. doi: 10.1101/gad.1476406

76. Bogs J, Jaffe FW, Takos AM, et al (2007) The Grapevine Transcription Factor VvMYBPA1 Regulates Proanthocyanidin Synthesis during Fruit Development. Plant Physiol 143:1347–1361. doi: 10.1104/pp.106.093203

77. Luo Q-J, Mittal A, Jia F, Rock CD (2012) An autoregulatory feedback loop involving PAP1 and TAS4 in response to sugars in Arabidopsis. Plant Mol Biol 80:117–129. doi: 10.1007/s11103-011-9778-9

78. Voronin DA, Kiseleva EV (2008) Functional role of proteins containing ankyrin repeats. Cell Tissue Biol 2:1–12. doi: 10.1134/S1990519X0801001X

79. De Paoli E, Dorantes-Acosta A, Zhai J, et al (2009) Distinct extremely abundant siRNAs associated with cosuppression in petunia. RNA 15:1965–1970. doi: 10.1261/rna.1706109

80. Smith CA, Robertson D, Yates B, et al (2008) The effect of temperature on Natural Antisense Transcript (NAT) expression in Aspergillus flavus. Curr Genet 54:241–269. doi: 10.1007/s00294-008-0215-9

81. Borsani O, Zhu J, Verslues PE, et al (2005) Endogenous siRNAs Derived from a Pair of Natural cis-Antisense Transcripts Regulate Salt Tolerance in Arabidopsis. Cell 123:1279–1291. doi: 10.1016/j.cell.2005.11.035

82. Katiyar-Agarwal S, Jin H (2007) Discovery of Pathogen-Regulated Small RNAs in Plants. In: Rossi JJ, Hannon GJ (eds) MicroRNA Methods. Academic Press, pp 215 – 227

83. Zubko E, Meyer P (2007) A natural antisense transcript of the Petunia hybrida Sho gene suggests a role for an antisense mechanism in cytokinin regulation. Plant J 52:1131–1139. doi: 10.1111/j.1365-313X.2007.03309.x

84. Held MA, Penning B, Brandt AS, et al (2008) Small-interfering RNAs from natural antisense transcripts derived from a cellulose synthase gene modulate cell wall biosynthesis in barley. Proc Natl Acad Sci 105:20534–20539. doi: 0.1073/pnas.0809408105

85. Ron M, Alandete Saez M, Eshed Williams L, et al (2010) Proper regulation of a sperm-specific cis-nat-siRNA is essential for double fertilization in Arabidopsis. Genes Dev 24:1010–1021. doi: 10.1101/gad.1882810

86. Quintero A, Pérez-Quintero AL, López C (2013) Identification of ta-siRNAs and Cis-nat-siRNAs in Cassava and Their Roles in Response to Cassava Bacterial Blight. Genomics Proteomics Bioinformatics 11:172–181. doi: 10.1016/j.gpb.2013.03.001

87. Strommer J (2011) The plant ADH gene family: The plant ADH gene family. Plant J 66:128–142. doi: 10.1111/j.1365-313X.2010.04458.x

88. Umezawa T (2010) The cinnamate/monolignol pathway. Phytochem Rev 9:1–17. doi: 10.1007/s11101-009-9155-3

89. Chen D, Yuan C, Zhang J, et al (2012) PlantNATsDB: a comprehensive database of plant natural antisense transcripts. Nucleic Acids Res 40:D1187–D1193. doi: 10.1093/nar/gkr823

90. Chan SW-L, Zhang X, Bernatavichute YV, Jacobsen SE (2006) Two-Step Recruitment of RNA-Directed DNA Methylation to Tandem Repeats. PLoS Biol 4:e363. doi: 10.1371/journal.pbio.0040363

91. Grewal SIS, Elgin SCR (2007) Transcription and RNA interference in the formation of heterochromatin. Nature 447:399–406. doi: 10.1038/nature05914

92. Romanel E, Silva TF, Corrêa RL, et al (2012) Global alteration of microRNAs and transposon-derived small RNAs in cotton (Gossypium hirsutum) during Cotton leafroll dwarf polerovirus (CLRDV) infection. Plant Mol Biol 80:443–460. doi: 10.1007/s11103-012-9959-1

93. Slotkin RK, Vaughn M, Borges F, et al (2009) Epigenetic Reprogramming and Small RNA Silencing of Transposable Elements in Pollen. Cell 136:461–472. doi: 10.1016/j.cell.2008.12.038

94. Moisy C, Garrison K, Meredith CP, Pelsy F (2008) Characterization of ten novel Ty1 copia-like retrotransposon families of the grapevine genome. BMC Genomics 9:469. doi: 10.1186/1471-2164-9-469

95. Mette M, Kanno T, Aufsatz W, et al (2002) Endogenous viral sequences and their potential contribution to heritable virus resistance in plants. EMBO J 21:461–469. doi: 10.1093/emboj/21.3.461

96. Bertsch C, Beuve M, Dolja VV, et al (2009) Retention of the virus-derived sequences in the nuclear genome of grapevine as a potential pathway to virus resistance. Biol Direct 4:21. doi: 10.1186/1745-6150-4-21

97. Li Z, Ender C, Meister G, et al (2012) Extensive terminal and asymmetric processing of small RNAs from rRNAs, snoRNAs, snRNAs, and tRNAs. Nucleic Acids Res 40:6787–6799. doi: 10.1093/nar/gks307

98. Loss-Morais G, Waterhouse PM, Margis R (2013) Description of plant tRNA-derived RNA fragments (tRFs) associated with argonaute and identification of their putative targets. Biol Direct 8:1–5. doi: 10.1186/1745-6150-8-6

99. Wang Q, Lee I, Ren J, et al (2013) Identification and Functional Characterization of tRNA-derived RNA Fragments (tRFs) in Respiratory Syncytial Virus Infection. Mol Ther 21:368–379. doi: 10.1038/mt.2012.237

100. Fu H, Feng J, Liu Q, et al (2009) Stress induces tRNA cleavage by angiogenin in mammalian cells. FEBS Lett 583:437–442. doi: 10.1016/j.febslet.2008.12.043

101. Gebetsberger J, Zywicki M, Künzi A, Polacek N (2012) tRNA-Derived Fragments Target the Ribosome and Function as Regulatory Non-Coding RNA in Haloferax volcanii. Archaea 2012:1–11. doi: 10.1155/2012/260909

102. Thompson DM, Lu C, Green PJ, Parker R (2008) tRNA cleavage is a conserved response to oxidative stress in eukaryotes. RNA 14:2095–2103. doi: 10.1261/rna.1232808

103. Levitz R, Chapman D, Amitsur M, et al (1990) The optional E. coli prr locus encodes a latent form of phage T4-induced anticodon nuclease. EMBO J 9:1383–1389.

104. Hsieh L-C, Lin S-I, Shih AC-C, et al (2009) Uncovering Small RNA-Mediated Responses to Phosphate Deficiency in Arabidopsis by Deep Sequencing. Plant Physiol 151:2120–2132. doi: 10.1104/pp.109.147280

105. Chen C-J, liu Q, Zhang Y-C, et al (2011) Genome-wide discovery and analysis of microRNAs and other small RNAs from rice embryogenic callus. RNA Biol 8:538–547. doi: 10.4161/rna.8.3.15199

106. Hackenberg M, Huang P-J, Huang C-Y, et al (2013) A Comprehensive Expression Profile of MicroRNAs and Other Classes of Non-Coding Small RNAs in Barley Under Phosphorous-Deficient and -Sufficient Conditions. DNA Res 20:109–125. doi: 10.1093/dnares/dss037

107. Zhang S, Sun L, Kragler F (2009) The Phloem-Delivered RNA Pool Contains Small Noncoding RNAs and Interferes with Translation. Plant Physiol 150:378–387. doi: 10.1104/pp.108.134767

Chapter 5: Characterisation of virus disease-associated plant responses by transcriptome analysis

5.1 Introduction

Cellular differentiation, growth and adaptability to environmental stresses are controlled by the regulation of gene expression, which include the modulation of transcription, RNA splicing, translation and post-translational modifications. Biotic stresses from viral pathogens are a major constraint to the production of high quality agricultural crops and research into plant-pathogen interactions can lead to the identification of genes involved in pathogen tolerance or resistance, or the identification of plant defence response triggers.

Grapevines are exposed to a variety of pests and pathogens that threaten the development and health of the world's viticulture industry [1]. No natural resistance to viruses has been identified; once infected, plants develop disease and remain chronically infected. However, even though plants are unable to stop viral replication and systemic infection, susceptible hosts are not passive against viruses. The plant symptoms observed are the visual representation of the host defence responses, including the molecular, cellular and physiological changes associated with the virus infection. The plant response involves changes in the expression of defence and stress-associated genes and an antiviral defence system based on RNA silencing has been implicated in the host-pathogen interaction. RNA silencing mediated by small RNAs (sRNAs) is a potential defence response of plants to attempt prevention of virus replication and inducing pathogenesis. Small RNAs have been shown to be involved in normal plant development and plant response to biotic and abiotic stresses [2–14].

Grapevine leafroll disease (GLD) is an economically important disease affecting wine and table grape cultivars, as well as rootstocks. The phenotypic symptoms associated with the disease were described extensively [15–17], however the molecular plant response elicited is still poorly understood. The disease does not only negatively affect vine growth but also has a detrimental effect on grape yield and juice quality. The main causative agent, *Grapevine leafroll-associated virus 3* (GLRaV-3) [15], is a phloem-limited virus and degeneration of phloem cells in leaves, stems and petioles, associated with GLRaV-3 infection, have been reported. It was shown that GLRaV-3 causes a drastic reduction in leaf photosynthesis [18, 19], anthocyanin biosynthesis and sugar levels in berries [18, 20]. *Grapevine leafroll-associated virus 3* infection also induced a significant reduction in CO_2 assimilation, yield, vine size and cane lignification [21] and an up-regulation of sugar transporters and senescence-related gene expression was observed in GLRaV-3 infected leaves [22].

Gene expression profiling provides a method to analyse the response to stresses and during viral

infection, the virus-associated sRNAs and regulated genes can be identified by their altered expression levels. In this chapter, the aim was to follow an integrated sRNA and mRNA next-generation sequencing approach to identify genes and sRNAs associated with GLD in three *Vitis vinifera* cultivars (Chardonnay, Chenin blanc and Cabernet Sauvignon). The molecular characterisation of the interaction between the grapevine host and the virus pathogen will provide insight into the plant host-pathogen response that can contribute to disease control or prevention.

5.2 Materials and methods

5.2.1 Plant material and sample collection

Three Cabernet Sauvignon own-rooted plants, singly infected with GLRaV-3 variant group II (isolate GP18, Accession No. EU259806), were established from cuttings made from a naturally infected Cabernet Sauvignon plant and rooted in the greenhouse. Three healthy Cabernet Sauvignon control plants were established from cuttings collected from a certified virus-free plant obtained from a grapevine nursery and rooted in the greenhouse. Additionally, six young virus-free certified *Vitis vinifera* plants of cultivars Chardonnay (rootstock: 101-14), Chenin blanc (rootstock: Richter 99) and Cabernet Sauvignon (rootstock: Richter 110) were collected from a nursery. These plants were re-established in the greenhouse in five litre bags containing a mixture of sand (45 %), palm peat (45 %) and vermiculite (10 %). Grapevine leafroll-associated virus 3 variant group II (isolate GP18, Genbank accession No. EU259806) was graft inoculated onto three plants from each cultivar using infected buds from the own-rooted Cabernet Sauvignon plants. Plants were maintained under natural light with temperatures ranging from 22°C to 28°C. Only one shoot was allowed to grow and all side shoots were constantly removed. Phloem material was sampled from all plants in the same physiological growth stage, as soon as the shoot material reached lignification and GLRaV-3 symptoms (reddening of the interveinal areas and downward curling of the leaf margins) were observed in the infected plants. High quality total RNA (A260/A280 above 2, A260/A230 above 2 and RNA integrity number above 6.5) was extracted from phloem material using a modified cetyltrimethylammonium bromide (CTAB) protocol [23, 24]. RNA quality was assessed using spectrophotometry, gel electrophoresis and Agilent Bioanalyzer analysis. The virus status of these plants was confirmed using end-point RT-PCRs for frequently-occurring grapevine viruses (Appendix A3)[i] [25]. The genetic variant of GLRaV-3 was confirmed by real-time RT-PCR high-resolution melting curve analysis and multiplex RT-PCR [26]. The GLRaV-3 virus concentration ratio (VCR) was determined using a relative quantitation RT-

[i]Appendix A3. Jooste AEC, Molenaar N, Maree HJ, Bester R, Morey L, De Koker WC, Burger JT (2015) Identification and distribution of multiple virus infections in Grapevine leafroll diseased vineyards. Eur J Plant Pathol 142:363–375.

A survey of viruses infecting grapevine in the wine regions of the Western Cape Province in South Africa was conducted. The survey determined the relative abundance of five different grapevine leafroll-associated virus 3 (GLRaV-3) variants. Virus profiles were also determined for individual vines. A total of 315 plants were sampled and analysed over two growing seasons. The complexity of virus populations detected in this study, highlights the need for detection methods able to identify all viruses and their variants in vineyards.

qPCR assay targeting open reading frame 1a (ORF1a) of the virus genome (Appendix A4)[ii] [24].

5.2.2 Small RNA next-generation sequencing

An sRNA sequencing library was prepared from total RNA by polyacrylamide gel electrophoresis size selection of the 18-30 nucleotide (nt) fraction from each plant sample. The Illumina Small RNA TruSeq kit was used for library preparation and sequencing (1 x 50 bp) was performed on an Illumina HiSeq instrument (Fasteris, Switzerland). Adapter sequences were removed using cutadapt [27] and reads were filtered for quality (phred score > 20 over 100% of the reads) using FASTX-toolkit (http://hannonlab.cshl.edu/fastx_toolkit/index.html). Only reads 18-26 nucleotides (nts) in length were used for virus-derived small interfering RNA (vsiRNA), microRNA (miRNA), phased small interfering RNA (phasiRNA), natural antisense transcript small interfering RNA (natsiRNA) and repeat-associated small interfering RNA (rasiRNA) analysis. The tRNA-derived small RNA analysis was performed using the 17-44 nt read fraction.

De novo assemblies were performed with CLC genomic workbench 8 (Qiagen) to confirm virus and virus variant status of all plants. A bubble size of 50, a word size of 20 and a minimum contig length of 50 were selected as *de novo* assembly parameters. BLAST sequence similarity searches were performed to identify contigs using Blast2GO (Blastx algorithm with e-value threshold of 0.001).

The command-line bioinformatic analysis was performed on the high-performance computer (HPC) of the Central Analytical Facility (CAF) at Stellenbosch University (http://www.sun.ac.za/hpc). Optimised parameters were used and changes to critical parameters were stated.

5.2.3 Transcriptome next-generation sequencing

The same total RNA extracts used for the construction of the sRNA libraries were used to prepare the transcriptome libraries with the Illumina mRNA stranded RNA kit. Single-end NGS (1 x 125bp) was performed on an Illumina HiSeq instrument (Fasteris, Switzerland). Adapter sequences were removed using cutadapt [27] and Trimmomatic [28] was used for quality trimming (HEADCROP of 12 nts, SLIDINGWINDOW of 3 nts with Q20, MINLEN of 20 nts).

[ii] Appendix A4. Bester R, Pepler PT, Burger JT, Maree HJ (2014) Relative quantitation goes viral: An RT-qPCR assay for a grapevine virus. J Virol Methods 210:67–75.

Three genomic regions (ORF1a, coat protein and 3′UTR) were targeted to quantitate GLRaV-3 relative to three stably expressed reference genes (actin, GAPDH and alpha-tubulin). These assays were able to detect all known variant groups of GLRaV-3, including the divergent group VI, with equal efficiency. No link could be established between the concentration ratios of the different genomic regions and subgenomic RNA (sgRNA) expression. However, a significant lower virus concentration ratio for plants infected with variant group VI compared to variant group II was observed for the ORF1a, coat protein and the 3'UTR. Significant higher accumulation of the virus in the growth tip was also detected for both variant groups. The primer set targeting ORF1a was selected for quantitation in this study due to the lower detection limit and to eliminate the possible influence of sgRNAs.

5.2.4 Identification of virus-derived siRNAs

Bowtie (1.1.2) [29] was used to map the high quality reads of 18-26 nts in length to the *Vitis vinifera* nuclear [30], chloroplast (Genbank accession No. NC_007957.1) and mitochondrial (Genbank accession No. NC012119) genomes. A single mismatch was allowed to compensate for cultivar differences and only the best alignment was reported per read (Bowtie reporting parameter: --best). The successfully mapped reads were removed from the sRNA libraries and the remaining reads were mapped (using Bowtie) onto the GLRaV-3 variant group II isolate GP18 genome (Genbank accession No. EU259806) to identify vsiRNAs. Only perfect matches between the sRNA read and the genome were allowed and only the best alignment was reported per read (Bowtie reporting parameter: --best).

5.2.5 Differential expression of sRNA species

Variation in sRNA expression levels between the GLRaV-3 negative and GLRaV-3 positive samples were assessed using the R package, DESeq2 [31]. The false discovery rate correction was used to correct for multiple testing. MicroRNAs with an adjusted p-value < 0.05 were regarded as differentially expressed. Small RNA expression levels were investigated for miRNAs, phasiRNAs, natsiRNAs, rasiRNAs and tRNA-derived siRNAs. These sRNA species analysed were identified as previously described in Chapter 4. Targets for differentially expressed miRNAs were predicted using psRNATarget [32]. BLAST sequence similarity searches of the predicted targets were performed using Blast2GO (Blastx algorithm with e-value threshold of 0.001) [33].

5.2.6 Stemloop RT-qPCR sRNA validation

Differentially expressed known and novel sRNAs were validated using stemloop RT-qPCR assays [34]. Complementary DNA (cDNA) was synthesised from 1 µg of the same total RNA extract used for NGS with 1 µM of stemloop primer (IDT) (Appendix D1), 0.5 mM dNTPs (Thermo Scientific), 100 U Maxima reverse transcriptase (Thermo Scientific) and 20 U Ribolock (Thermo Scientific) in a final volume of 20 µl. Incubation for 30 minutes at 16 °C was performed, followed by a pulsed reverse transcription of 60 cycles at 30 °C for 30 seconds, 42 °C for 30 seconds and 50 °C for 1 second. Five µl of each cDNA sample was pooled and a 5-fold dilution series was prepared to construct a representative standard curve for each primer sets. The remaining cDNA was diluted 1:24 to quantify each sample separately using the miRNA-specific and the reference miRNA primer sets. All cDNA dilutions were stored at −20 °C. The RT-qPCRs were performed using the Rotor-Gene Q thermal cycler (Qiagen). Reactions contained 1x FastStart Universal probe master (ROX) (Roche), 0.1 µM Universal probe library probe #21 (Roche), 3.3 µl Milli-Q H_2O and 0.6 µM specific forward and universal reverse primers (IDT) (Appendix D1). One µl cDNA was added to each reaction to a final

reaction volume of 10 µl. The "no-template" and "no-reverse transcriptase" controls were included in all runs. All reactions were performed in triplicate in Rotor-Gene Q 0.1 ml tube-and-cap strips (Qiagen). Cycling parameters included an initial activation of 95 °C for 10 minutes and 45 cycles of 95 °C for 10 seconds and 60 °C for 60 seconds. Acquisition on the green channel was recorded at the end of the extension step. The Rotor-gene Q software version 2.3.1 (Qiagen) was used to calculate primer efficiencies, Cq values and gene quantitation values for all targets. The relative concentration ratio (CR) were calculated as previously shown in Bester et al. [24] (Appendix A4) using a reference gene index, calculated using the geometric mean of the concentration of two stable expressed miRNAs (vvi-miR159c and vvi-miR167a). The non-parametric Wilcoxon signed-rank test was used to evaluate differential expression. A p-value < 0.05 was regarded as significantly differentially expressed. All calculations were performed on the web-based application, Harbin (https://rbester.shinyapps.io/Harbin/) (Appendix A5)[iii].

5.2.7 Repeat sequence validation

Due to sequence similarity between certain rasiRNAs and vsiRNAs, the origin of these siRNA was investigated by sequencing a fraction of the relevant repeat sequence in each of the cultivar groups. The cultivar diversity in repeat sequences was assessed with PCR and Sanger sequencing. A fraction of the EnSpm-3 VV EnSpm/CACTA repeat was amplified from all cultivar groups in a 25 µl PCR reaction mixture containing 1× KAPA Taq buffer A (KAPA Biosystems), 0.4 mM dNTP mix (Thermo Scientific), 0.4 µM forward and reverse primers (IDT) (Appendix D1) and 0.08 U/µl KAPA Taq DNA polymerase (KAPA Biosystems). Cycle conditions included an initial denaturation step at 94 °C for 5 minutes, followed by 35 cycles of 94 °C for 30 seconds, 53 °C for 30 seconds and elongation at 72 °C for 60 seconds. Final extension was at 72 °C for 7 min. Amplicons were visualised on an ethidium bromide-stained 1 % TAE agarose gel (2 M Tris, 1 M glacial acetic acid, 0.05 M Na_2EDTA, pH 8) and send for bidirectional Sanger sequencing at the Central Analytical Facility (CAF) at Stellenbosch University.

5.2.8 Differential expression of sRNA reads

The high quality 18-26 nt reads of all six libraries of each cultivar were concatenated using FASTX-toolkit (http://hannonlab.cshl.edu/fastx_toolkit/index.html), all identical sequences were collapsed into

[iii] Appendix A5. Bester R, Pepler PT, Aldrich DJ, Maree HJ (2016). Harbin: A quantitation PCR analysis tool. Biotechnol Lett. DOI 10.1007/s10529-016-2221-1.

To enable comparisons of different relative quantitation experiments, a web-browser application called Harbin was created for a dynamic interaction with qPCR data. A quantile-based scoring system is proposed that will allow for comparison of samples at different time points and between experiments. Harbin simplifies the analysis of high-density qPCR assays, either for individual experiments or across sets of replicates and biological conditions. The application uses the standard curve method for relative quantitation of genes with normalisation using a reference gene index to calculate a concentration ratio (CR). This Harbin quantile bootstrap test for evaluating if different datasets can be combined, was shown to be less conservative than the Kolmogorov-Smirnov test, and therefore more sensitive in detecting distributional differences between data sets. Statistical significance testing for CRs across biological conditions is also possible with Harbin.

a single sequence to create a non-redundant database of all the reads. The unique sequence list of each cultivar was used to calculate a count for each sequence in each of the individual libraries, using an in-house Python script. Differential expression of the cultivar-unique sequences was assessed using the R package, DESeq2 [31]. The false discovery rate correction was used to correct for multiple testing. MicroRNAs with an adjusted p-value < 0.05 were regarded as differentially expressed. Mapping the differentially expressed sequences to the different databases created previously to identify the sRNA species (Chapter 4), identified the origin of the differentially expressed sequences. The vsiRNA sequences were identified by mapping to the GLRaV-3 variant group II isolate GP18 genome (Genbank accession No. EU259806); miRNAs were identified by mapping to the miRNA Registry Database (miRBase version 21) [35], and the novel miRNAs predicted earlier (Chapter 4). The rasiRNAs were identified by mapping to the unique *Vitis vinifera* repeat sequences present in Repbase Update 21.07 [36]. The *in silico* predicted natural antisense transcripts (NATs) were used to identify sequences that can possibly be natsiRNAs; phasiRNAs were identified by mapping to the phased loci predicted, and tRNA-derived siRNAs were identified by mapping to the mature tRNA sequences of five angiosperms (*Arabidopsis thaliana, Brachypodium distachyon, Medicago truncatula, Oryza sativa and Populus trichocarpa*) available in the PlantRNA database [37]. Bowtie (1.1.2) [29] was used to perform all read-mapping analyses.

5.2.9 NGS transcriptome data analysis

The high quality transcriptome sequence reads were mapped to the *Vitis vinifera* reference genome [30] using TopHat version 2.0.14 [38]. Tophat identified splice junctions between exons by using the short-read aligner Bowtie 2 version 2.2.6 [39]. Reference-based assembly of the reads was performed using Cufflinks and Cuffmerge version 2.2.1 [38], applying the bias detection algorithm and the multi-read correction to improve transcript abundance estimates, and accurate weighting of the reads mapping to multiple locations in the genome, respectively. The expression level of each transcript was expressed as reads per transcript kilobase per million reads mapped (RPKM), calculated based on the number of mapped reads. Cuffdiff version 2.2.1 was used to detect differentially expressed genes [38]. Transcripts with a false discovery rate adjusted p-value < 0.05 were regarded as differentially expressed.

5.2.10 RT-qPCR target validation

Primers for differentially expressed genes were designed to span an intron in each transcript to eliminate possible amplification from genomic DNA. Complementary DNA was synthesised from 1 µg of the same total RNA extract sent for NGS, using 0.15 µg random primers (Promega), 0.5 mM dNTPs (Thermo Scientific), 100 U Maxima reverse transcriptase (Thermo Scientific) and 20 U Ribolock (Thermo Scientific) in a final volume of 20 µl. Five µl of each cDNA sample was pooled

and a 5-fold dilution series was prepared to construct a representative standard curve for each primer sets. The remaining cDNA was diluted 1:24 to quantify each sample separately using the gene of interest-specific and the reference genes' primer sets. All cDNA dilutions were stored at −20 °C. The RT-qPCRs were performed using the Rotor-Gene Q thermal cycler (Qiagen) and the SensiMixTM SYBR No-ROX Kit (Bioline). Reactions contained 1x SensiMixTM SYBR (Bioline) No-ROX, Milli-Q H2O and 0.4 µM forward and reverse primers (IDT) (Appendix D1). Two µl cDNA was added to each reaction to a final reaction volume of 12.5 µl. The same cDNA dilution series was used to construct all primer-specific standard curves and the same 1:24 dilution of each "unknown" sample was screened with all primer sets for quantitation. The "no-template" and "no-reverse transcriptase" control reactions were included in all runs. All reactions were performed in triplicate in Qiagen Rotor-Gene Q 0.1 ml tube-and-cap strips. Cycling parameters included an initial activation of 95 °C for 10 minutes and 45 cycles of 95 °C for 15 seconds, 58 °C for 15 seconds (55 °C for 15 seconds for the reference genes) and 72 °C for 15 seconds. Acquisition on the green channel was recorded at the end of the extension step. Melting curve analysis of PCR amplicons was performed with temperatures ranging from 65 °C to 95 °C with a 1 °C increase in temperature every 5 seconds to identify primer-dimers and non-specific amplification. The relative CR were calculated as previously shown in Bester et al. [24] using a reference gene index, calculated using the geometric mean of the concentration of three reference genes (GAPDH, α-tubulin and actin) previously shown to be constitutively expressed in *Vitis vinifera* phloem material [24, 40]. The non-parametric Wilcoxon signed-rank test was used to evaluate differential expression. A p-value < 0.05 was regarded as significantly differentially expressed. All calculations were performed on the web-based application, Harbin (https://rbester.shinyapps.io/Harbin/) (Appendix A5).

5.3 Results and discussion

5.3.1 sRNA sequencing data

An average of nine million high quality reads was generated per sRNA library for both the GLRaV-3 positive and negative samples (Table 1). No significant read length differences were observed between the GLRaV-3 positive and negative samples, with the most abundant read lengths being 21 and 24 nts (Figure 1). This sRNA size pattern reflects a typical population dominated by miRNAs (21 nts), phasiRNA (21 nts) and repeat-associated siRNAs (24 nts). Under the experimental conditions of this study, GLRaV-3 did not seem to have an effect on the overall sRNA population structure of infected samples (Figure 1). On average, 25.5% of reads of the different cultivar groups aligned to ribosomal RNA (rRNA) sequences obtained from Genbank. The virus status of the different samples was confirmed with the *de novo* assemblies. The contigs identified as GLRaV-3 sequences had the highest similarity to GLRaV-3 variant group II. Viroid sequences were also detected in all samples, with Grapevine yellow speckle viroid only in Chardonnay and Chenin blanc samples. Hop stunt viroid

sequences were detected in all samples except one Chenin blanc sample. A low level of reads assembled into contigs with a high similarity to Grapevine virus A (GVA) in one of the biological replicates of GLRaV-3 infected Cabernet Sauvignon. After extracting the reads mapping to the nuclear, mitochondria and chloroplast genomes of *Vitis vinifera*, 7.66% of the reads mapped to GVA and 64.05% of the reads mapped to GLRaV-3 for this sample.

Table 1. Small RNA NGS data statistics. Read count per sequencing library before and after quality filtering. Read counts associated with GLRaV-3 are also indicated.

Sample	Library	18-26 nt reads	18-26 nt reads after QC	18-26 nt non-redundant reads after QC	GLRaV-3 vsiRNA	GLRaV-3 vsiRNA %
Chardonnay (CY7)	HUS1	7069867	5682351	1553819	42	
Chardonnay (CY8)	HUS2	6987162	5561984	1376471	31	
Chardonnay (CY10)	HUS3	7741779	6182702	1588126	44	
Chardonnay (CY1)	HUS4	9120830	7324586	1610429	606140	8.28
Chardonnay (CY4)	HUS5	9738644	7804757	2005979	322087	4.13
Chardonnay (CY5)	HUS6	9616316	7632139	1839570	415168	5.44
Chenin blanc (CB7)	HUS7	8714933	6929352	1795048	36	
Chenin blanc (CB8)	HUS8	7887448	6273976	1770949	32	
Chenin blanc (CB10)	HUS9	10025676	7950515	2127098	38	
Chenin blanc (CB2)	HUS10	8742406	6933327	1564252	461121	6.65
Chenin blanc (CB3)	HUS11	8009127	6410121	1648958	329351	5.14
Chenin blanc (CB4)	HUS12	7945247	6305822	1429912	330234	5.24
Cabernet Sauvignon (CS7)	HUS13	10911194	8634821	2154639	65	
Cabernet Sauvignon (CS8)	HUS14	9457986	7546111	1718277	41	
Cabernet Sauvignon (CS10)	HUS15	10453027	8379796	2068858	64	
Cabernet Sauvignon (CS1)	HUS16	9452286	7567653	1829539	545655	7.21
Cabernet Sauvignon (CS3)	HUS17	8313232	6608686	1643279	480874	7.28
Cabernet Sauvignon (CS5)	HUS18	9327510	7495685	2149302	518589	6.92
Own-rooted Cabernet Sauvignon (Cab1)	HUS19	9589029	7661851	1951781	39	
Own-rooted Cabernet Sauvignon (Cab2)	HUS20	9142351	7337021	1629271	33	
Own-rooted Cabernet Sauvignon (Cab6)	HUS21	10176066	8168847	2098718	51	
Own-rooted Cabernet Sauvignon (GH33)	HUS22	9511171	7594624	1878818	495102	6.52
Own-rooted Cabernet Sauvignon (GH34)	HUS23	11427138	9178033	1801876	524508	5.71
Own-rooted Cabernet Sauvignon (GH36)	HUS24	10041711	7987101	1755663	538247	6.74
Total		219402136	175151861	42990632	5567592	
Average		9141756	7297994	1791276	231983	
Mimimum		6987162	5561984	1376471	31	
Maximum		11427138	9178033	2154639	606140	
Average GLRaV-3 positive samples		9270468	7403544.5	1763131	463923	
Average GLRaV-3 negative samples		9013043	7192444	1819421	43	

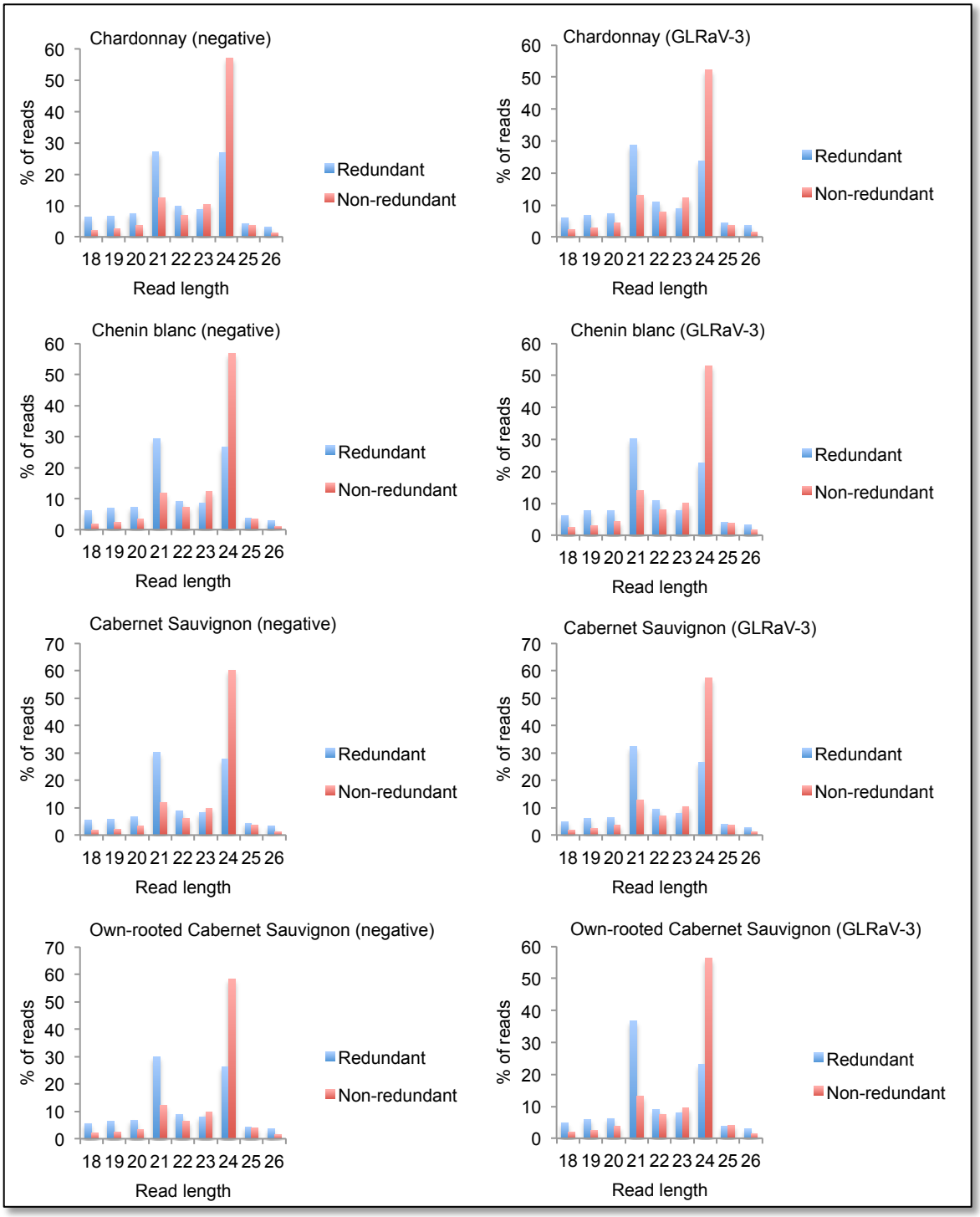


Figure 1. Read length distribution per cultivar. Histogram illustrating size distribution of the number of reads per read length as a percentage of the total number of 18-26 nt sized redundant and non-redundant reads for GLRaV-3 negative and positive samples.

5.3.2 Virus concentration ratio (VCR)

The GLRaV-3 VCR was quantitated in all samples using a relative quantitation RT-qPCR assay with an efficiency correction. The GLRaV-3 VCR was measured using a primer set targeting ORF1a of the GLRaV-3 genome. Normalisation of the virus concentration was performed using the reference genes,

actin, GAPDH and alpha-tubulin. A significant higher VCR was detected in the infected Chenin blanc samples compared to the other cultivars (Wilcoxon rank sum test p-value = 0.00004114) (Figure 2). However, no significant increase in the read counts of the Chenin blanc samples was detected, suggesting that there is no direct correlation between the number of GLRaV-3 virus genomes and the sRNA response of the plant against the virus infection. This can suggests that vsiRNAs will accumulate to a specific level, irrespective of VCR and warrants further investigation. In this experiment, the Chardonnay, Cabernet Sauvignon and own-rooted Cabernet Sauvignon samples displayed the same trend for read counts compared to VCRs, while for Chenin blanc, higher VCRs were observed with the same number of vsiRNAs compared to the other cultivars. Based on sRNA biogenesis in these cultivars, our results seem to suggest that cv. Chenin blanc responds differently to the virus infection.

Figure 2. Virus concentration ratios (VCRs) and sRNA read counts per sample. Histogram displaying the VCRs determined by RT-qPCR compared to the redundant and non-redundant reads mapped per library million on GLRaV-3 isolate GP18.

5.3.3 GLRaV-3-associated vsiRNAs

The libraries from GLRaV-3 infected samples were used to analyse the production of vsiRNAs. Reads, which did not align to the *Vitis vinifera* nuclear, chloroplast or mitochondrial genomes, were mapped onto the complete genome of GLRaV-3 isolate GP18 (Genbank accession No. EU259806). No mismatches were allowed, as the same isolate was used to infect the plants. On average, 5.9%, 5.7%, 7.1% and 6.3% of the high quality 18-26 nt reads of the infected samples mapped onto the virus genome with 100 % coverage (Figure 3) for Chardonnay, Chenin blanc, Cabernet Sauvignon and own-

rooted Cabernet Sauvignon, respectively. A significant low number of reads from the uninfected samples mapped to the GLRaV-3 genome, confirming the absence of the virus from these samples (Table 1). The majority of the vsiRNA reads were 21 nts in length, followed by the 22 nt reads for all cultivar groups (Figure 4). Similar distributions were previously observed for single-stranded RNA viruses [41–45]. The library preparation utilised an RNA strand-specific protocol that allowed the investigation of sRNA mappings on both strands. For the different cultivar groups, 1.3-2 fold more positive-strand vsiRNAs compared to negative-strand vsiRNAs were observed (Figure 3). Since positive strand RNA viruses produce excess positive- over negative-strand RNAs [46], this strand bias can suggest that a fraction of these vsiRNA reads probably are traces of GLRaV-3 genome degradation. The non-redundant mapping (Figure 3B) showed multiple unique vsiRNAs associated with the same genomic region, indicating the probability of different Dicer-like (DCL) cleavage sites in the same vicinity on the genome [45]. In all cultivars, a high number of reads mapped to ORF10 compared to the rest of the genome (Table 2). Open reading frame 10 encodes a protein that are believed to be involved in suppression of the host RNA silencing mechanism [47, 48]. The higher number of vsiRNAs associated with this genomic region can be the result of the higher template available due to the presence of subgenomic RNAs (sgRNAs) [49], or be linked to a host-pathogen interaction to suppress the plant's antiviral response.

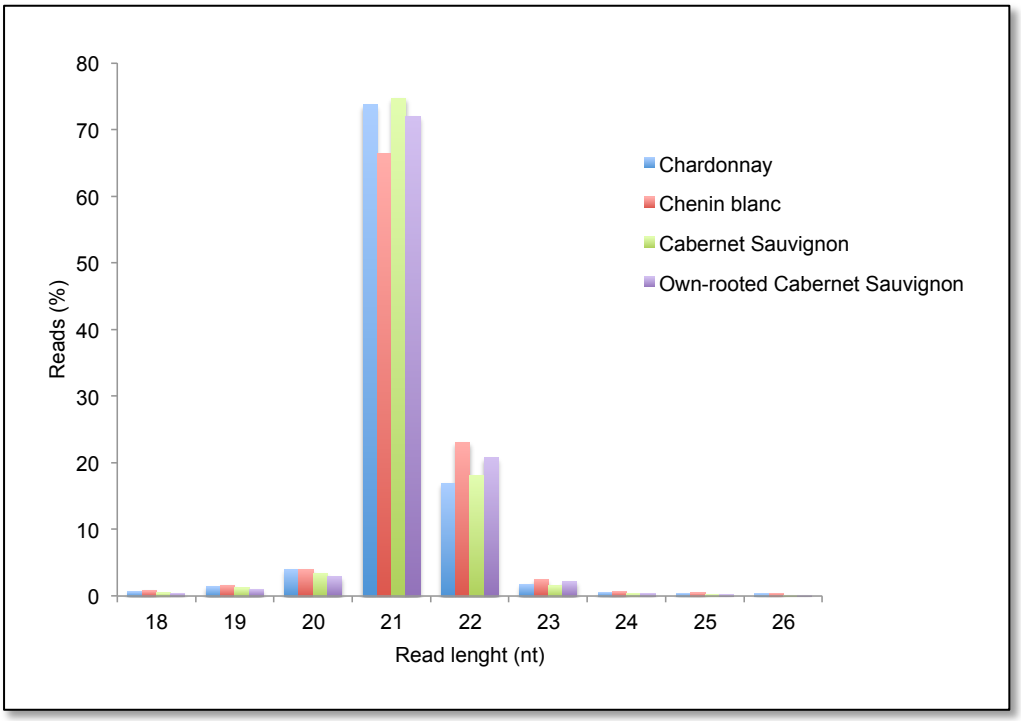


Figure 4. Size distribution of virus-derived siRNA reads. Histogram displaying the number of vsiRNA reads per read length as a percentage of the total number of reads in the 18-26 nt size-range.

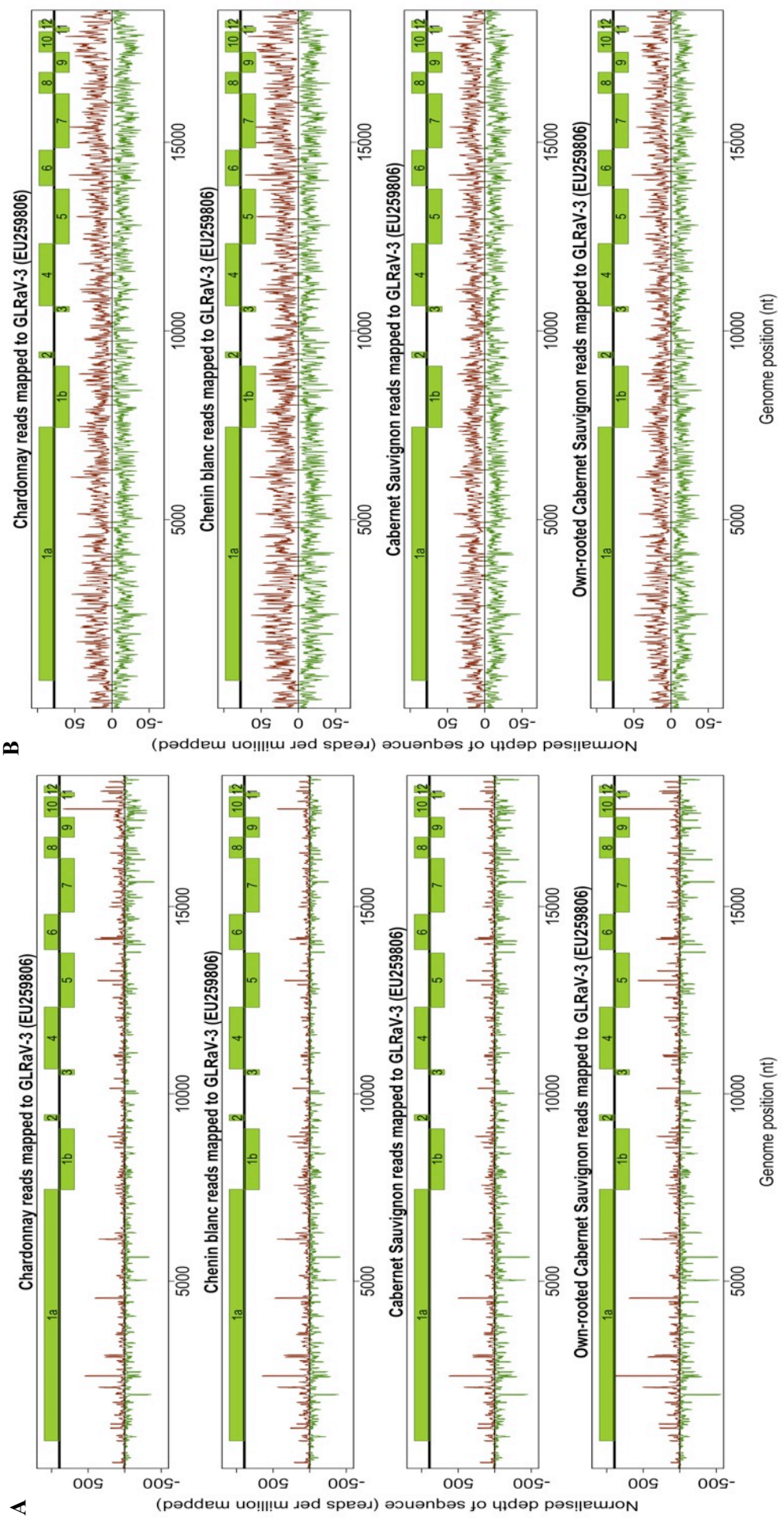


Figure 3. Distribution of sRNA reads along the GLRaV-3 genome. The mapping of redundant (A) and non-redundant (B) reads on GLRaV-3 isolate GP18 (Genbank accession No. EU259806) per cultivar.

Table 2. Normalised read count per GLRaV-3 open reading frame (ORF). Read counts were normalised with library size and expressed as reads per million mapped divided by ORF size.

Open reading frame (ORF)	Chardonnay	Chenin blanc	Cabernet Sauvignon	Own-rooted Cabernet Sauvignon
5'UTR	29.49	36.65	34.09	28.92
ORF1a	66.32	82.78	81.21	83.83
ORF1b	53.03	72.24	68.29	68.64
ORF2	43.15	55.97	54.54	50.45
ORF3	33.97	29.87	31.79	32.67
ORF4	47.58	40.39	50.19	53.64
ORF5	58.95	47.97	61.36	69.24
ORF6	90.71	62.73	79.61	72.16
ORF7	95.41	64.61	83.02	76.78
ORF8	71.00	50.77	63.03	55.10
ORF9	98.52	59.43	78.78	80.84
ORF10	183.59	104.77	128.92	134.97
ORF11	98.82	49.73	60.24	71.90
ORF12	89.79	46.58	51.43	62.99
3'UTR	55.82	40.01	56.75	57.21

5.3.4 Differentially expressed sRNAs

5.3.4.1 miRNAs

Differentially expressed miRNAs were identified through allowing no mismatches between the sRNA reads and the *Vitis vinifera* miRNAs present in miRBase, while the differentially expressed novel miRNAs were identified by mapping to the novel miRNA precursors identified earlier (Chapter 4).

Three known and three novel miRNAs were identified as differentially expressed in Chardonnay GLRaV-3 infected samples. The up and down regulation of these miRNAs were validated using stemloop RT-qPCRs (Figure 5). The correlation coefficient for the sRNA NGS $\log_2$(fold change) and the RT-qPCR $\log_2$(fold change) was 0.94, providing credibility to the validation approach selected. The three known miRNAs were up-regulated and predicted to target six genes with a high sequence similarity to a transcription factor pif4-like (miR3633a-5p), chloroplastic gamma aminobutyrate transaminase (miR3633a-5p), methyltransferase-like protein (miR3633a-5p), RNA-binding protein fus isoform (miR3633a-5p), l-ascorbate oxidase homolog (miR3633a-5p), electron isoform (miR3633a-5p) and serine threonine-protein kinase (miR398b-c) (Appendix D2). The three novel miRNAs were down-regulated and predicted to target five genes with a high sequence similarity to a DNA ion isoform (c187937), nrt1 ptr family-like protein (c130253), universal stress protein (c130253), wrky transcription factor (c130253) and a major facilitator superfamily protein isoform (c130381) (Appendix D2). One of the targets predicted for the down-regulated miRNA c130253 (universal stress

protein) was shown to be differentially up-regulated in the NGS transcriptome data set (Appendix D2).

In Chenin blanc GLRaV-3 infected samples, three miRNAs were identified as differentially expressed Although the up-regulation trend of all three miRNAs was validated with stemloop RT-qPCR, only one novel miRNA were significantly up-regulated in the RT-qPCR data (Figure 6, Appendix D2). The predicted targets of these miRNAs were annotated as receptor-like protein kinase (miR396a), resistance protein (miR396a, c40118), growth-regulating factors (miR396a), nuclear ribonucleoprotein (miR396a), pentatricopeptide repeat-containing proteins (miR396a, c40118), homeobox-leucine zipper protein (c40118) and a GAMYB transcription factor (c134686) (Appendix D2).

All 12 miRNAs identified as differentially expressed in the sRNA data of infected own-rooted Cabernet Sauvignon samples were validated with stemloop RT-qPCR (Figure 7). Multiple targets were predicted for these miRNAs with high sequence similarity to serine threonine-protein kinase (miR398b-c), flowering-promoting factor protein (miR477b-3p), chloroplastic ATP sulfurylase (miR395a-m), F-box protein (miR2950-5p), n-acetyl-beta-glucosaminyl asparagine amidase (miR477b-5p), oxalate ligase (c31052) and a major facilitator superfamily protein isoform (c141224). The predicted targets of seven of these miRNAs were differentially expressed in the NGS transcriptome data (Appendix D2) and three of these targets were selected for validation using RT-qPCR (Chapter 3). The NGS transcriptome data confirmed the anti-correlation expression mechanism between miRNA and target for five out of the eight miRNA/target pairs identified in both sRNA and transcriptome NGS data sets.

No differentially expressed miRNAs were identified in the Cabernet Sauvignon infected samples, which was an unexpected result. As a consequence, the hypothesis is that the sRNA GLRaV-3 response can be growth stage specific and differential expression only detectable in specific developmental stages of the plant. Even though phloem material was sampled with caution to ensure sampling at the same physiological growth stage, the grafted Cabernet Sauvignon samples required seven months to reach the same lignified growth stage compared to five (Chardonnay and Chenin blanc) and six (own-rooted Cabernet Sauvignon) months for the other cultivar groups. The own-rooted Cabernet Sauvignon samples were also established before the grafted plants, and harboured the original GLRaV-3 isolate used to infect the other cultivars. The difference between the own-rooted and grafted Cabernet Sauvignon samples can therefore be as a result of a more established infection status in the own-rooted plants.

Even though no universal miRNA response was identified between cultivars as a response to GLRaV-3, miR398b-c was up-regulated in both Chardonnay and own-rooted Cabernet Sauvignon plants. This miRNA was predicted to target a serine threonine-protein kinase, which play a key role in signalling

during pathogen recognition and the activation of plant defence mechanisms [50]. These defence responses can include the generation of nitric oxide and superoxide, antimicrobial compound production and programmed cell death [50]. The down-regulation of these receptor-like kinases will therefore influence the plant's normal development and ability to activate plant defence responses. Since both Chardonnay and Cabernet Sauvignon is symptomatic cultivars with regards to GLRaV-3 infection, the absence of the modulated miRNA expression in infected Chenin blanc plants can hint towards host specificity for viral pathogenicity. As described in Chapter 3, the down-regulation of the predicted target of miR398b-c were validated with stemloop RT-qPCR in own-rooted Cabernet Sauvignon and confirmed the anti-correlation between this miRNA and the predicted target.

Three genes (DNA ion isoform 3, NRT1/PTR family protein, major facilitator superfamily protein isoform) were predicted as targets for three differentially expressed miRNAs in both infected Chardonnay and own-rooted Cabernet Sauvignon samples. These miRNAs, identified in the two different cultivars, were identical in sequence i.e. isomiRs of each other. Novel miRNA c187937 (Chardonnay) was identical to c205570 (own-rooted Cabernet Sauvignon) and c130381 (Chardonnay) was identical to c141224 (own-rooted Cabernet Sauvignon). Novel miRNA c130253 (Chardonnay) and c141107 (own-rooted Cabernet Sauvignon) were isomiRs of each other. These miRNAs were down-regulated in Chardonnay, however up-regulated in own-rooted Cabernet Sauvignon, implying a complex defence response with cultivar specificity. This also suggests that different isomiRs can be expressed at different levels relative to each other in different cultivars.

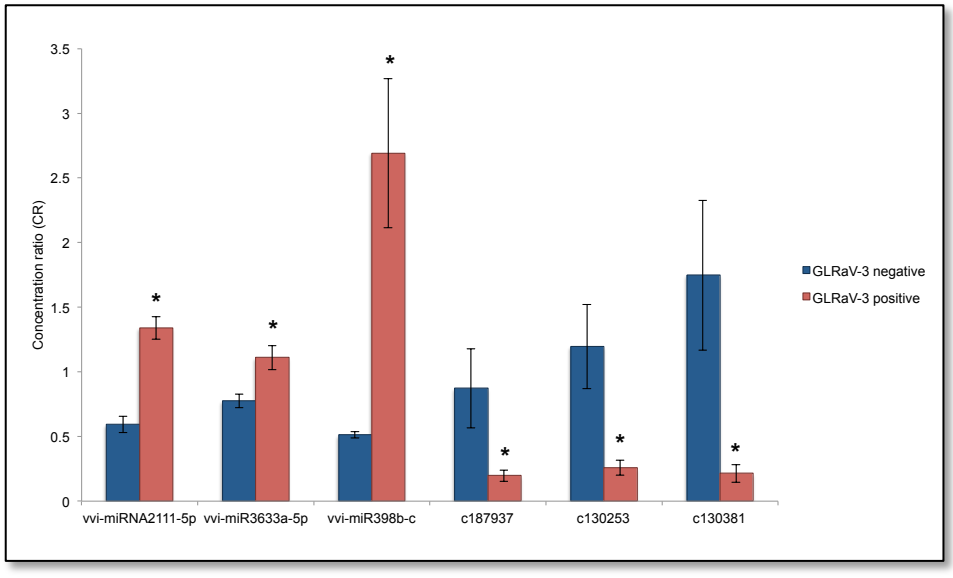


Figure 5. Differentially expressed miRNAs identified in infected Chardonnay samples validated using stemloop RT-qPCR. The mean concentration ratio (CR) ± standard error (SE) of three biological replicates, with each replicate an average of three technical replicates is displayed. Statistically significant differences between GLRaV-3 negative and positive samples are indicated by asterisks (*p-value < 0.05).

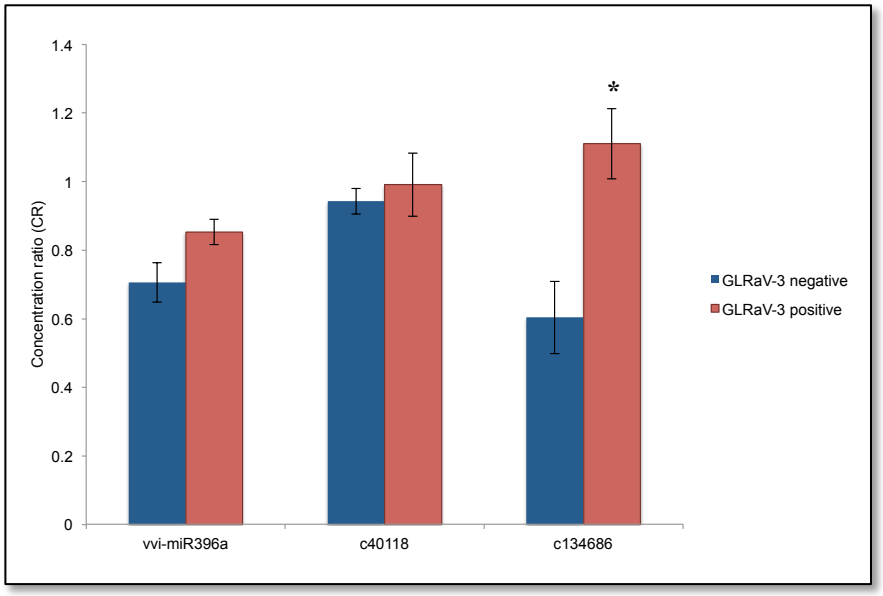


Figure 6. Differentially expressed miRNAs identified in infected Chenin blanc samples validated using stemloop RT-qPCR. The mean concentration ratio (CR) ± standard error (SE) of three biological replicates, with each replicate an average of three technical replicates is displayed. Statistically significant differences between GLRaV-3 negative and positive samples are indicated by asterisks (*p-value < 0.05).

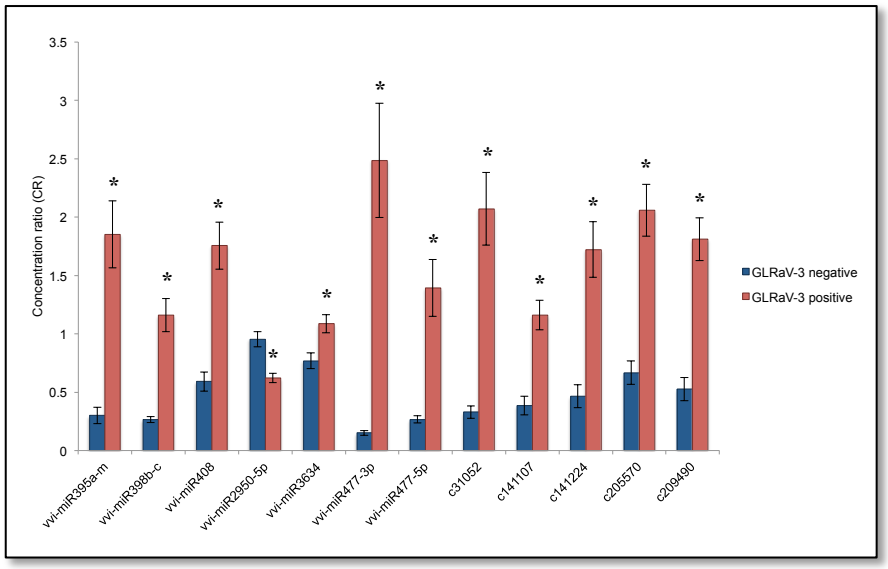


Figure 7. Differentially expressed miRNAs identified in infected own-rooted Cabernet Sauvignon samples validated using stemloop RT-qPCR. The mean concentration ratio (CR) ± standard error (SE) of three biological replicates, with each replicate an average of three technical replicates is displayed. Statistically significant differences between GLRaV-3 negative and positive samples are indicated by asterisks (*p-value < 0.05).

5.3.4.2 phasiRNAs

Phased loci with a differential number of associated sRNAs were identified by mapping the sRNA reads to the phased loci identified earlier (Chapter 4). More sRNA reads mapped to the phased loci of GLRaV-3 infected Chardonnay, Chenin blanc and own-rooted Cabernet Sauvignon compared to GLRaV-3 negative samples. Twenty-four, 12 and 44 phased loci had differential numbers of sRNAs associated with each locus for the respective cultivars, with an adjusted p-value < 0.05 and a $\log_2$(fold change) between -0.7 and 0.5 (Appendix D3, Table 3). As with the miRNA analysis, no differential results were identified for grafted Cabernet Sauvignon. Forty-five percent of the characterised differential loci were annotated as disease resistance genes using Blast2GO, indicating a possible host response towards the virus infection (Appendix D3:A). Four loci, differentially enriched for siRNAs, overlapped between Chardonnay, Chenin blanc and own-rooted Cabernet Sauvignon, of which two loci had an anti-correlated $\log_2$(fold change) between the white and red cultivars (Appendix D3:A). Two of these loci had more phased-associated siRNA reads in Chardonnay and Chenin blanc infected samples compared to the GLRaV-3 negative samples, while the opposite was observed in infected samples of own-rooted Cabernet Sauvignon. Sixteen, five and 15 individual phased loci associated siRNAs (phasiRNAs) were differentially expressed in Chardonnay, Chenin blanc and own-rooted Cabernet Sauvignon GLRaV-3 infected samples, respectively (Appendix D3:B). One of these phasiRNAs was up-regulated in Chardonnay and down-regulated in own-rooted Cabernet Sauvignon samples. The link between phasiRNA expression and a potential red-white cultivar-specific defence response will need to be confirmed using additional cultivars and more biological replicates.

Table 3. Phased loci differentially enriched for siRNAs in the different cultivar groups.

		Homologue loci in other cultivars			
	Phased loci	Chardonnay	Chenin blanc	Cabernet Sauvignon	Own-rooted Cabernet Sauvignon
Chardonnay	24		6	0	9
Chenin blanc	12	6		0	5
Cabernet Sauvignon	0	0	0		0
Own-rooted Cabernet Sauvignon	44	9	5	0	

5.3.4.3 natsiRNAs

Natural antisense transcripts (NATs) with a differential number of associated sRNAs, were identified by mapping the sRNA reads to the NAT overlap regions identified earlier (Chapter 4). The overlapping region of *trans*-NAT pair GSVIVT01010800001 and GSVIVT01011363001 was enriched for sRNAs in both GLRaV-3 negative and GLRaV-3 positive samples (Figure 8). A difference in the strand bias ratio was observed in both Chardonnay and Chenin blanc GLRaV-3 infected samples compared to the GLRaV-3 negative samples. Differences in read counts associated with one or the

other genomic DNA strand were noticed between GLRaV-3 infected and healthy samples. In the case of GLRaV-3 infected samples, 2.6 and 1.9 fold more siRNA reads were associated with the one strand, compared to 1.4 and 2.5 fold more siRNAs on the same strand in GLRaV-3 negative samples. This implies an increase in the production of siRNAs deriving predominantly from one of the NATs for Chardonnay and Chenin blanc; with an increase in siRNA biogenesis in infected Chardonnay samples, and a decrease in infected Chenin blanc samples. This can suggest a mechanism of transcript down-regulation in normal plant development, while uniquely adapted for different cultivars in response to GLRaV-3 infection.

Even though a trend was observed regarding the strand bias in GLRaV-3 infected samples, none of the 25 NAT overlap regions had a significant enrichment for siRNAs when GLRaV-3 positive and negative samples were compared, and no individual sRNA associated with the NATs displayed significant differential expression in any of the cultivar groups. This result implies that the NAT-siRNA mechanism in plants is either part of a complex network of regulatory processes, so that the effect was not visible using the experimental approach applied here, or that natsiRNA biogenesis is not altered as a response to pathogen infection.

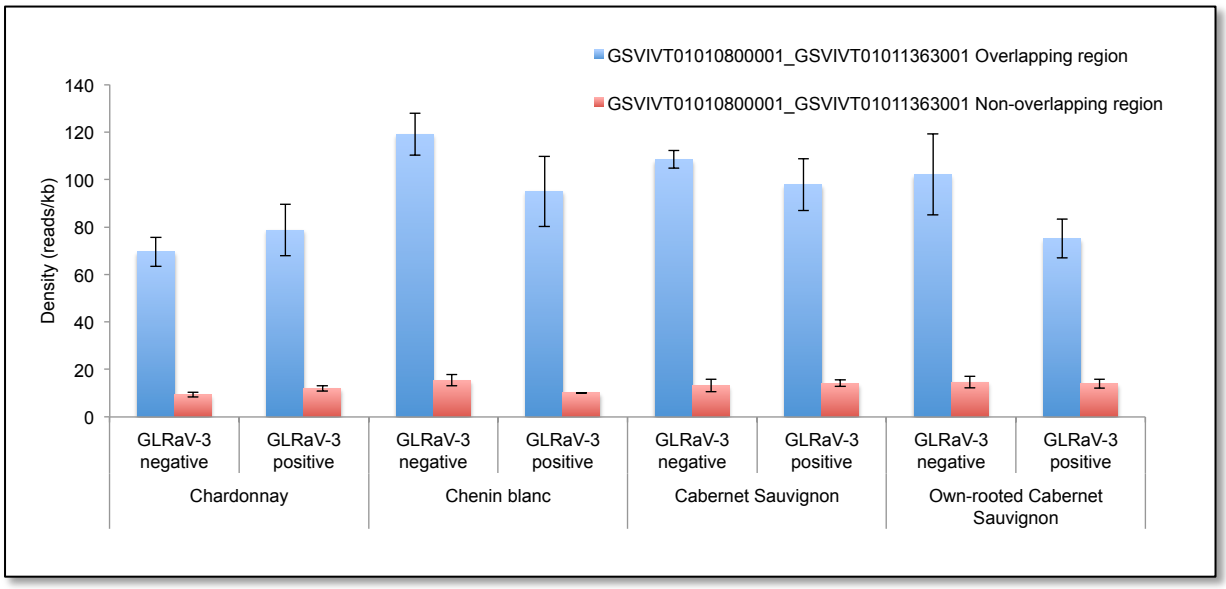


Figure 8. The number of reads per kilobase of overlapping or non-overlapping region for NAT pair GSVIVT01010800001/GSVIVT01011363001. The mean density ± standard error (SE) of three biological replicates is displayed.

5.3.4.4 rasiRNAs

Repeat sequences with a differential number of associated sRNAs, were identified by mapping the sRNA reads to the *Vitis vinifera* repeat sequences present in Repbase. Overall, a lower number of

sRNA reads mapped to the repeat sequences in GLRaV-3 infected samples compared to the uninfected samples after normalising for total sequencing library size (Table 4). However, significantly more sRNAs mapped to the repeat sequence EnSpm-3 VV EnSpm/CACTA in GLRaV-3 infected samples (Appendix D4:A). Focussing on individual siRNAs associated with each repeat, revealed differential expression of five siRNAs associated with the same EnSpm-3 VV EnSpm/CACTA repeat in all cultivars (Appendix D4:B). Grapevine CACTA elements can range in size from 10 to 25 kilobases (kb) and account for 0.34 % of the grapevine genome [51]. The EnSpm-3 VV EnSpm/CACTA repeat sequence was subjected to the NCBI open reading frame finder tool and a large ORF of 2469 nts was predicted. Blast analysis of this ORF showed two transposase-associated domains, suggesting that this transposable element is probably not defective and has retained the capacity to be transcribed. In the past, transposable elements have been linked to regulation of anthocyanin biosynthesis in grapevine, where the skin colour of white berries are the outcome of a retrotransposon insertion in the promoter of a Myb-related gene [52]. Therefore, the up-regulation of rasiRNAs can potentially play a role in the regulation of the gene's expression in which the repeat sequence is located. The rasiRNA reads, associated with the EnSpm-3 VV EnSpm/CACTA repeat sequences, mapped to five different *Vitis vinifera* chromosomes on the sense strand and seven chromosomes on the antisense strand, indicating potential origin or target loci. These loci were all intergenic regions of the genome, signifying potential regulation sites of the up- and downstream genes. One of these loci was upstream of transcript GSVIVT01020514001 on chromosome 19; and in both the GLRaV-3 infected Chardonnay and own-rooted Cabernet Sauvignon transcriptome NGS data, this transcript's down-regulation was statistically significant. It has high sequence identity to a leucine-rich repeat receptor-like serine/threonine-protein kinase, which is believed to play an important role in signalling during pathogen recognition [50].

Additional repeat sequences displayed significant enrichment for siRNAs in Chardonnay (5), Chenin blanc (17) and own-rooted Cabernet Sauvignon (1) (Appendix D4:A). Additional individual rasiRNAs displayed differential expression in GLRaV-3 infected plants; one each in Chardonnay and own-rooted Cabernet Sauvignon plants and 12 in Chenin blanc plants (Appendix D4:B). Stemloop RT-qPCRs were used to validate the differentially expressed EnSpm-3 VV EnSpm/CACTA-associated rasiRNAs with the largest $\log_2$(fold change) in all cultivars (415333_EnSpm-3_VV_EnSpm/CACTA). The presence and the up-regulation of this rasiRNA in GLRaV-3 infected samples were confirmed (Figure 9).

This universal rasiRNA signal observed was subsequently further scrutinised with alignments to the GLRaV-3 genome, as well as representative sequences of the repeat sequences obtained from the different cultivar groups through Sanger sequencing of EnSpm-3 VV EnSpm/CACTA repeat amplicons. The differential rasiRNA sequence had one mismatch with the respective repeat sequences

and no mismatches to the viral genome (Figure 10). This questions the origin of this siRNA, as it also conforms to the criteria set out for the identification of vsiRNAs. The original vsiRNA pool was identified by first extracting the reads mapping to the *Vitis vinifera* nuclear, chloroplast or mitochondrial genomes, permitting one mismatch to allow for cultivar differences, which explains why these siRNAs were not mapped to the GLRaV-3 genome initially. Based on sequence identity, these siRNAs are most likely virus derived. However, whether the presence of this short viral sequence in the *Vitis vinifera* genome is a coincidence or represents a possible interaction between the virus and the host genome remains to be determined.

Table 4. Read counts of sRNAs associated with the different *Vitis vinifera* repeat superfamilies/clades present in Repbase. Read counts are indicated separately for GLRaV-3 positive and negative per cultivar, and each represent the sum of three biological replicates.

	Read count							
	Chardonnay		Chenin blanc		Cabernet Sauvignon		Own-rooted Cabernet Sauvignon	
Superfamily/clade	GLRaV-3 negative	GLRaV-3 positive	GLRaV-3 negative	GLRaV-3 positive	GLRaV-3 negative	GLRaV-3 positive	GLRaV-3 negative	GLRaV-3 positive
Copia	232334	286595	266256	225635	306015	267566	256199	229748
Gypsy	200625	203107	208644	199362	415608	268682	222188	214136
Caulimoviridae	74552	87893	78736	69749	109842	95160	100158	82279
hAT	54083	59812	80505	63126	67022	59249	64763	57117
MuDR	33110	41198	45835	32459	47872	42276	42519	38029
L1	12971	16882	16382	14019	19542	17386	15317	14564
EnSpm/CACTA	10665	14308	13596	11680	14147	14439	13149	13288
Harbinger	8766	10260	12129	7983	13250	11924	12637	11138
Helitron	19	21	34	31	30	32	30	28
Total rasiRNA reads	637664	770937	736388	593467	1010002	809369	740761	661164
Normalised read count (reads per million mapped of total library size)	36591	33870	34811	30203	41123	37346	31974	26703

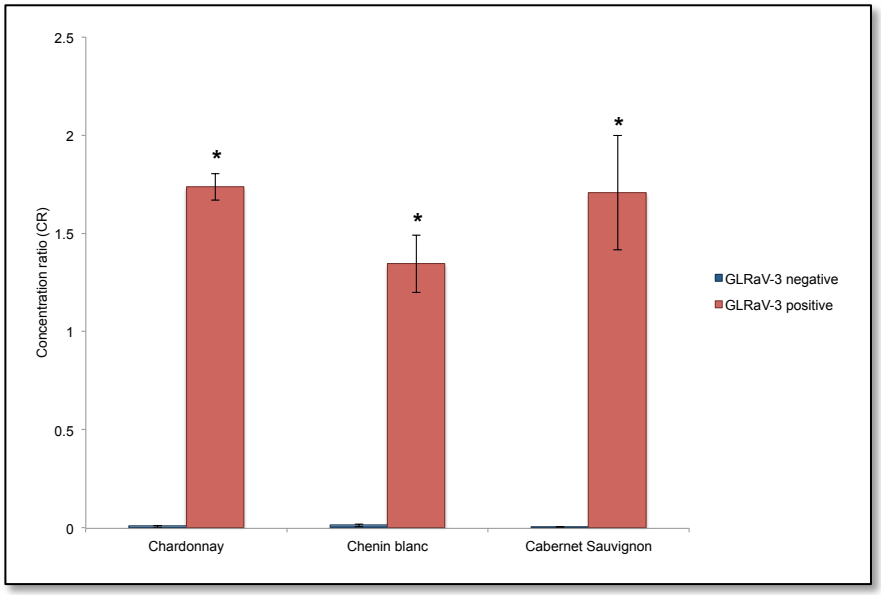


Figure 9. Histogram displaying differential expression of rasiRNA 415333_EnSpm-3_VV_EnSpm/CACTA identified in infected Chardonnay, Chenin blanc, Cabernet Sauvignon and own-rooted Cabernet Sauvignon samples using stemloop RT-qPCR. The mean concentration ratio (CR) ± standard error (SE) of three biological replicates, with each replicate an average of three technical replicates is displayed. Statistically significant differences between GLRaV-3 negative and positive samples are indicated by asterisks (*p-value < 0.05).

```
EnSpm-3_VV_EnSpm/CACTA_Vitis_vinifera          CAAGAACGAAATGAATGTGTTTGAGGATGATTTTGAAACATGGAGTCTTAA

415333_EnSpm-3_VV_EnSpm/CACTA_Vitis_vinifera                    TTTGAGGATGATTTTGAAACT
421072_EnSpm-3_VV_EnSpm/CACTA_Vitis_vinifera                   TTTTGAGGATGATTTTGAAAC
422874_EnSpm-3_VV_EnSpm/CACTA_Vitis_vinifera                  TTTTTGAGGATGATTTTGAAA
398721_EnSpm-3_VV_EnSpm/CACTA_Vitis_vinifera                    TTGAGGATGATTTTGAAACTT
422875_EnSpm-3_VV_EnSpm/CACTA_Vitis_vinifera                  TTTTTGAGGATGATTTTGAAAC

Chardonnay                                     CNAGAACGAAATGAATGTGNTTGAGGATGATTTTNAAACANGGAGTCTTAA
Chenin blanc                                   CAAGAACGAAATGAATGTGTTTGAGGATGATTTTGAAACATGGAGTCTTAA
Cabernet Sauvignon                             CAAGAACGAAATGAATGTGTTTGAGGATGATTTTGAAACATGGAGTCTTAA
Own-rooted Cabernet Sauvignon                  CAAGAACGAAATGAATGTGTTTGAGGATGATTTTGAAACATGGAGTNTTAA

GLRaV-3 Isolate GP18 (7559 nt - 7612 nt)       GGCTTTCGATTATGACCTTTTTGAGGATGATTTTGAAACTTCAGATCAGTC
```

Figure 10. Sequence identity between repeat EnSpm-3 VV EnSpm/CACTA, GLRaV-3 and sRNA reads. Sequence alignment displaying the identities between the differentially expressed rasiRNAs, EnSpm-3 VV EnSpm/CACTA repeat sequence and GLRaV-3 isolate GP18. Red boxes indicate the sequence differences between the plant and virus genome.

5.3.4.5 tRNA-derived siRNAs

Transfer RNAs with a differential number of associated sRNAs, were identified by mapping the sRNA reads to the tRNAs present in the plantRNA database. Read mapping analysis showed no overall

difference in the read counts of tRNA-derived sRNAs in the GLRaV-3 infected samples compared to the negative samples. However, the number of sRNAs associated with specific tRNAs was found to significantly vary between the GLRaV-3 positive and negative samples of the different cultivars (Appendix D5:A-C). Not only was a differential number of sRNAs associated with specific tRNAs, but several individual tRNA-derived siRNAs showed significant variation in read counts between GLRaV-3 positive and negative samples (Appendix D5:D-F). The infected Chardonnay plants had the most tRNA-derived siRNAs with a differential read count (Table 5). Different tRNAs with differential sRNA read counts were identified in each cultivar (Appendix D5, Table 5). In Chardonnay, 83 % of the tRNAs with a differential sRNA read count, encoded Arginine, while in Chenin blanc it was mainly tRNAs coding for Asparagine (75 %) (Appendix D5:A). The tRNAs with differential sRNA read counts in own-rooted Cabernet Sauvignon, were in 31 % of cases tRNA-Gly^{TCC}. The up-regulation of tRNA fragments deriving from Gly^{TCC} have previously been shown to be linked to phosphate deprivation and drought conditions [53, 54], and is therefore potentially linked to a plant stress-response mechanism. The increase in tRNA fragments associated with Ala^{AGC}, Arg^{CCT}, Arg^{TCG} and Gly^{TCC} were also shown in response to biotic stress in *Arabidopsis thaliana* infected with *Pseudomonas syringae* [54].

The same tRNA-derived fragments with a differential sRNA read count were observed in Chardonnay and Chenin blanc (Gly^{GCC}) and in Chenin blanc and own-rooted Cabernet Sauvignon (Lys^{TTT}), though the rest of the differentially expressed tRNA-derived fragments were unique to each cultivar. This suggests a cultivar-specific tRNA-mediated regulation mechanism in response to stress. The predominant (53 %) 5' terminal nt of the tRNA-derived siRNAs was a Guanine and can possibly play a role in the loading of the sRNA into a specific argonaute protein for post-transcriptional gene silencing [54]. Even though previous studies have speculated that these siRNAs can bind to ribosomes and cause down-regulation of genes, the specific biogenesis and function of the tRNA-derived sRNAs remains to be elucidated. The tRNA-derived sRNAs identified in this study provide support for the involvement of these sRNA species in biotic stress and tRNA-mediated gene regulation.

Table 5. Number of tRNAs with differential read count of associated sRNAs. The number of tRNAs with a differential read count of sRNAs, the number of individual differential tRNA-derived siRNAs with a differential read count, and the tRNA with the highest read count is shown per cultivar.

	tRNAs						Individual tRNA-derived siRNA					
	Nuclear		Mitochondria		Chloroplast		Nuclear		Mitochondria		Chloroplast	
		Major tRNA		Major tRNA		Major tRNA		Major tRNA		Major tRNA		Major tRNA
Chardonnay	29	Arg^{ACG} and Arg^{CCT}	1	His^{GTG}	5	His^{GTG}	85	Gly^{GCC} and Arg^{ACG}	9	His^{GTG}	17	His^{GTG}
Chenin blanc	12	Asp^{GTC}	0		0		22	His^{GTG} and Gly^{CCC}	3	Asn^{GTT}	7	Asn^{GTT}
Cabernet Sauvignon	0		0		0		0		0		0	
Own-rooted Cabernet Sauvignon	13	Pro^{AGG} and Gly^{TCC}	0		0		15	Pro^{AGG} and Lys^{TTT}	0		0	

5.3.4.6 sRNA reads

In an attempt to identify additional differentially expressed sRNAs in GLRaV-3 infected plants, a reversed strategy was followed. Differential expression analysis was performed on read counts before sRNA species identification, by using a non-redundant cultivar-specific list of read sequences to count the number of reads specific to each unique sequence in each plant sample. More than 8000 sequences were differentially expressed in GLRaV-3 infected samples (Table 6) with an adjusted p-value < 0.05, however less than 500 had an absolute $\log_2$(fold change) > 1 and an average read count > 100 (Table 6, Appendix D6:A-D). The 181 sequences identical in all cultivars were identified as vsiRNAs. A disproportion was observed between the numbers of down-regulated sequences compared to up-regulated sequences, thought this could be explained by the high percentage of differentially vsiRNA reads present in all cultivar groups. The number of differentially expressed vsiRNA reads was much higher in Cabernet Sauvignon and own-rooted Cabernet Sauvignon, with 98% of the differentially expressed reads in Cabernet Sauvignon annotated as vsiRNAs. This confirms the abovementioned analyses where no other sRNAs was identified as differentially expressed in own-rooted Cabernet Sauvignon. The same differential sRNAs were identified in the other cultivars as with the abovementioned analyses with the addition of 20, 25, 2 and 27 unknown *Vitis vinifera* genome-derived sRNAs for Chardonnay, Chenin blanc, Cabernet Sauvignon and own-rooted Cabernet Sauvignon, respectively. These sRNA reads mapped to non-coding regions of the genome. Between 8.8 and 19.6 % of the differential sRNAs aligned to rRNA.

One of the differentially expressed sRNAs in Chenin blanc and own-rooted Cabernet Sauvignon aligned with perfect sequence similarity to the *Vitis vinifera* transcript GSVIVT01014787001. This sRNA did not conform to the criteria set out in the bioinformatics analyses for the identification of the

other sRNA species and the biogenesis of this sRNA remains unknown. Even though this transcript codes for an unknown protein, the transcript was also down-regulated in infected Chenin blanc plants in the NGS transcriptome data, suggesting a possible regulation mechanism between this sRNA and transcript GSVIVT01014787001.

Next-generation sequencing identified a high number of sRNAs, however a large fraction of the reads remains not identified. This approach attempted to identify differentially expressed sRNAs not characterised as a specific sRNA species. Although no additional differentially expressed sRNAs was identified, the remaining reads can be uncharacterised sRNA species not regulated by biotic stress.

Table 6. Number of differentially expressed sRNA reads identified in the different cultivar groups. The significant number of sRNA reads is displayed by setting a threshold first with average read count and then with $\log_2$(fold change).

	p-value < 0.05				
		Average read count > 100			
			Absolute $\log_2$(Fold change) > 1 (down and up)		
				Cultivar specific	vsiRNA
Chardonnay	8231	521	350 (51 and 299)	81	254 (49 %)
Chenin blanc	8568	510	341 (82 and 259)	78	242 (47 %)
Cabernet Sauvignon	10339	449	446 (2 and 444)	46	442 (98 %)
Own-rooted Cabernet Sauvignon	8156	588	499 (59 and 440)	115	412 (70 %)

5.3.5 Transcriptome NGS data

Differential gene expression was investigated by using Tophat to map the high quality transcriptome sequence reads to the *Vitis vinifera* reference genome. An average of twelve million high quality reads was generated per library for the GLRaV-3 positive and negative samples (Table 7). More than 92 % of the reads was mapped to the reference genome for all samples. The majority of the remaining reads of the GLRaV-3 infected samples could be assembled into GLRaV-3 and viroid sequences, as was shown in a previous study (Appendix A6)[iv] [55]. The Illumina mRNA stranded RNA kit selected for poly(A)-tailed mRNAs in order to sequence actively expressed genes. Since GLRaV-3 is a non-polyadenylated virus, the presence of viral sequences was unexpected. However, the 3′ bias observed after mapping the reads to the GLRaV-3 genome may be ascribed to the expression of subgenomic

[iv] Appendix A6. Visser M, Bester R, Burger JT, Maree HJ (2016) Next-generation sequencing for virus detection: covering all the bases. Virol J. 13:85.

In this study, genome coverage at different sequencing depths was determined for a number of viruses, viroids, hosts and sequencing library types, using both read-mapping and *de novo* assembly-based approaches. The results highlighted the strength of ribo-depleted RNA and sRNA in obtaining saturated genome coverage with the least amount of data, while even though the poly(A)-selected RNA yielded virus-derived reads, it was insufficient to cover the complete genome of a non-polyadenylated virus. The ribo-depleted RNA data also outperformed the sRNA data in terms of the percentage of coverage that could be obtained particularly with the *de novo* assembled contigs.

RNAs (sgRNAs). An increase in genome coverage was observed around the sgRNA initiation point for GLRaV-3 ORF6 (coat protein) suggesting the presence of sgRNA 3′ poly(A)-tails [56, 57].

The reference-based assembly of the mapped reads resulted in 27270 assembled transcripts of which 65 - 67 % had RPKM of at least 1 in either the GLRaV-3 positive or GLRaV-3 negative samples per cultivar (Table 8, Figure 11A). Novel transcripts not present in the reference transcriptome, available on the Grape Genome Browser (Genoscope), were identified in all cultivar groups. Seven percent of the assembled transcripts with RPKM of at least 1 in either the GLRaV-3 positive or GLRaV-3 negative samples, were novel in the different cultivar groups (Table 8). Of these novel transcripts, 1604 were predicted in all cultivar groups, 86 in the more severe GLRaV-3 symptomatic cultivars (Chardonnay, Cabernet Sauvignon and own-rooted Cabernet Sauvignon), 91 in the white-berried cultivars (Chardonnay and Chenin blanc) and 116 in the red-berried cultivar (Cabernet Sauvignon and own-rooted Cabernet Sauvignon) (Figure 11C).

Table 7. Read count statistics for the transcriptome NGS libraries. Read counts per sequencing library before and after quality filtering. The percentage of reads mapped to the reference genome is also indicated.

Sample	Library	Total reads	Reads after QC	*Vitis vinifera* genome read mapping rate (%)
Chardonnay (CY7)	HUS-25	14882406	12183673	93.6
Chardonnay (CY8)	HUS-26	12185799	9880509	93.4
Chardonnay (CY10)	HUS-27	12508150	10156868	93.6
Chardonnay (CY1)	HUS-28	10942878	8908862	93.1
Chardonnay (CY4)	HUS-29	10847767	8780395	93.2
Chardonnay (CY5)	HUS-30	9871543	7985506	92.1
Chenin blanc (CB7)	HUS-31	10061801	8149766	93.1
Chenin blanc (CB8)	HUS-32	11755681	9527621	93.4
Chenin blanc (CB10)	HUS-33	15211179	12441698	93.5
Chenin blanc (CB2)	HUS-34	13743800	11164104	93.2
Chenin blanc (CB3)	HUS-35	15445557	12686181	93.5
Chenin blanc (CB4)	HUS-36	10633008	8508910	92.9
Cabernet Sauvignon (CS7)	HUS-37	11076575	8840243	92.2
Cabernet Sauvignon (CS8)	HUS-38	11003709	8755420	92.7
Cabernet Sauvignon (CS10)	HUS-39	13478061	10802118	92.6
Cabernet Sauvignon (CS1)	HUS-40	10733827	8448266	92.8
Cabernet Sauvignon (CS3)	HUS-41	11625507	9366581	92.7
Cabernet Sauvignon (CS5)	HUS-42	15135238	12402986	92.8
Own-rooted Cabernet Sauvignon (Cab1)	HUS-43	12949969	10469946	93.0
Own-rooted Cabernet Sauvignon (Cab2)	HUS-44	11555318	9374230	92.7
Own-rooted Cabernet Sauvignon (Cab6)	HUS-45	13245653	10413553	92.1
Own-rooted Cabernet Sauvignon (GH33)	HUS-46	11855789	9585021	92.3
Own-rooted Cabernet Sauvignon (GH34)	HUS-47	12253656	9926372	92.3
Own-rooted Cabernet Sauvignon (GH36)	HUS-48	11965306	9564419	92.5
Total		294968177	238323248	
Average		12290341	9930135	
Mimimum		9871543	7985506	
Maximum		15445557	12686181	
Average GLRaV-3 positive samples		145053876	117327603	
Average GLRaV-3 negative samples		149914301	120995645	

Table 8. Predicted transcripts assembled with Cufflinks. The number of transcripts, and transcripts differentially expressed are shown, including the number of novel transcripts identified.

	All transcripts (RPKM > 1)		Novel predicted transcripts (RPKM >1)	
	Total	Differentially expressed	Total	Differentially expressed
Chardonnay	17773	924	1947	87
Chenin blanc	18184	915	2062	131
Cabernet Sauvignon	17821	181	2013	21
Own-rooted Cabernet Sauvignon	18305	2801	2056	259

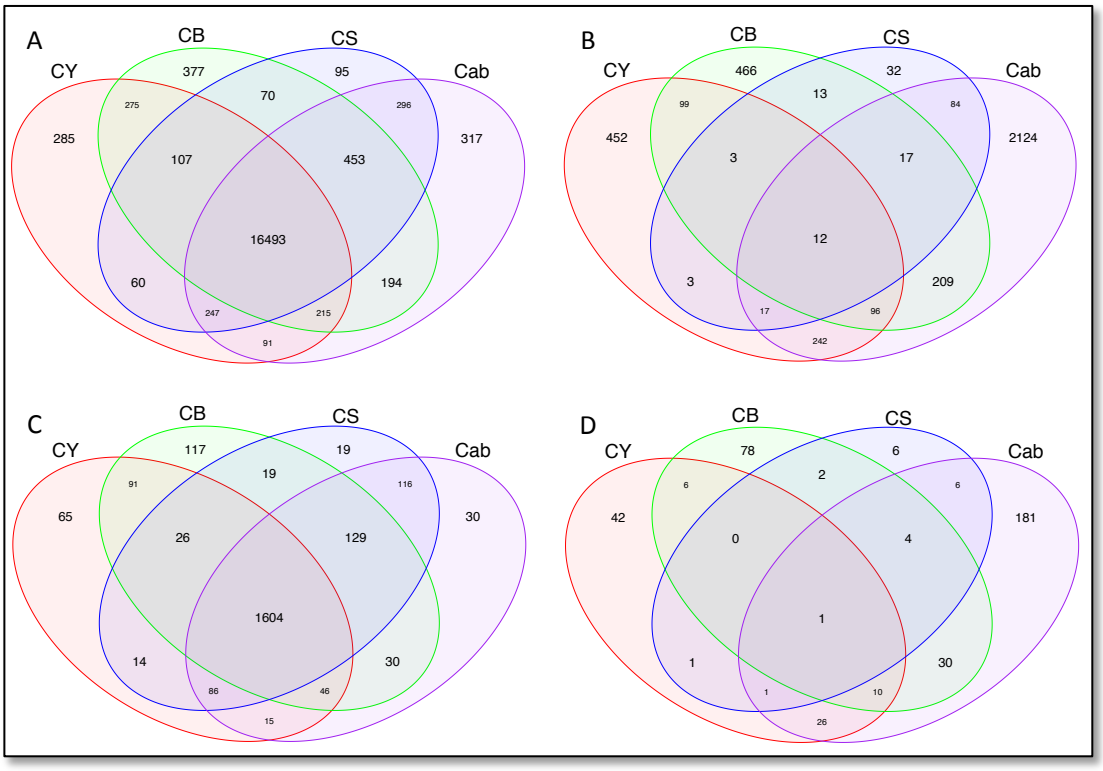


Figure 11. Venn diagram displaying the assembled transcripts shared between the different cultivars. A: All transcripts with RPKM > 1; B: Differentially expressed transcripts with RPKM > 1; C: Novel predicted transcripts with RPKM > 1; D: Differentially expressed novel transcripts with RPKM > 1. (CY: Chardonnay, CB: Chenin blanc, CS: Cabernet Sauvignon, Cab: Own-rooted Cabernet Sauvignon)

5.3.6 Differentially expressed genes

Cuffdiff was used to detect differentially expressed genes in GLRaV-3 infected samples. In order to identify a universal plant response to the virus, the differentially expressed genes were compared amongst the different cultivars. Twelve genes, including one novel predicted transcript, were identified as differentially expressed in all GLRaV-3 infected samples (Figure 11B and 11D, Appendix D7). Cultivar-specific genes were identified, and 17 genes were only differentially expressed in the severe symptomatic cultivars (Figure 11B, Appendix D8). Nine of these genes were up-regulated in all three cultivar groups, while eight genes were down-regulated in infected Chardonnay, but up-regulated in both the infected Cabernet Sauvignon groups. This suggests an interesting phenomenon where white cultivars respond different to the stress of a GLRaV-3 infection than red cultivars, but still in the same metabolic pathway. The nine up-regulated genes were annotated to encode proteins with sequence identity to MIZU-KUSSEI, E6 protein, pectate, leucine-rich repeat receptor kinase, polyol transporter, aquaporin, alpha-expansin and beta-D-xylosidase. These genes play important roles during root development [58], cell wall synthesis and integrity [59–

64], cellular signalling [65] and substrate transport [66–68]. The genes down-regulated in Chardonnay had a high sequence identity to beta-amyrin-oxidase, tubulin, squalene mono-oxygenase, proline-rich protein and a laccase-like protein. These genes are involved in the formation of triterpenoid saponins for antioxidant activity [69, 70], development of cell walls for plant growth and elongation [71–75] and sterol biosynthesis as a membrane constituent or for regulation of plant metabolism [76–78]. The majority of these genes play an important role in normal development and plant growth, specifically with regards to the formation of cell walls. The modulation of these genes in response to GLRaV-3 can potentially be correlated with the severe symptom expression observed in Chardonnay and Cabernet Sauvignon.

The 12 genes identified as differentially expressed in all cultivar groups, including Chenin blanc, were annotated as NAC transcription factors, proline-rich-like protein, GTPase-activating protein, glucan endo-1,3-beta-glucosidase, WAT1-related protein, expansin, thaumatin, fidgetin and lipid-transfer DIR1 protein. The functions of these proteins can all be linked to plant stress responses. The up- or down-regulation of these genes can either enhance the plant's ability to defend against the virus or lead to the negative outcomes of the disease. NAC transcription factors play an important role in transcriptional reprogramming coupled with plant immune responses, and are a large family of transcriptional regulators in plants [79]. Plant structural proteins are usually rich in proline amino acids and the altered expression of genes encoding for these proteins, can cause cell wall defects [72]. GTPase-activating proteins can function as stress signalling molecules where they usually initiate the production of NADPH oxidase for a primary response against a pathogen attack [80]. Glucan endo-1,3-beta-glucosidase was shown to be induced in the presence of viruses, and also plays a key role in development, including microsporogenesis and pollen germination [81]. Secondary cell wall thickness and auxin export is controlled by WAT1 proteins [82, 83]; the modulation of these will influence the plant's normal growth cycle. Expansin proteins are important for cell elongation, cell structure, intercellular communication and plant-microbe interactions [63, 64, 84–86]. Studies showed that the suppression of expansion genes can enhance the protection of plants against pathogens through preventing plant cell walls from loosening [86]. Thaumatin-like proteins, also known as pathogenesis-related proteins, were shown to have antifungal activity by interfering with cell wall components [87] and have glucan binding and glucanase activities that are also linked to biotic stress in plants [88]. Fidgetin-like proteins regulate the interaction between centrosomes and spindle fibres and influence meiotic crossovers. The modulation of Fidgetin-like genes can therefore have an effect on genomic stability [89]. Proteins encoded by the *DIR1* gene have been linked to the long-distance signalling associated with systemic-acquired resistance [90, 91]. Upon interactions with pathogens, plants respond through many mechanisms of which systemic acquired resistance can be one. Lipid-transfer DIR1 proteins were shown to be associated with signal transmission from infected to healthy cells, which is essential for systemic acquired resistance to be effective [90, 91].

The differential expression of 12 genes in all cultivars was assessed with RT-qPCRs. For cv. Chardonnay the expression of six genes were confirmed of which two was down-regulated and four up-regulated (Appendix D9). The down-regulation of one gene and the up-regulation of three genes were confirmed in cv Chenin blanc (Appendix D9). The down-regulation of two genes and the up-regulation of one gene was validated in Cabernet Sauvignon, and in own-rooted Cabernet Sauvignon, one gene was confirmed to be down-regulated and six genes up-regulated (Appendix D9). The correlation coefficient for the transcriptome NGS $\log_2$(fold change) and the RT-qPCR $\log_2$(fold change), for the genes validated, was 0.81. The direction of the modulation of the 12 genes was confirmed with the RT-qPCR assays, except for one gene in Chenin blanc and two genes in own-rooted Cabernet Sauvignon. However, the variation between biological replicates resulted in not all being statistically significant. The differential genes that could not be validated with RT-qPCR can be the result of the lower RPKM values or the lower $\log_2$(fold change) in certain cultivar groups, which indicate that the concentration of the target was too low for accurate quantitation. The success of RT-qPCR as an NGS validation method depends strongly on the balance between $\log_2$(fold change), RPKM, and variation between biological replicates. Both the NGS and RT-qPCR data (XLOC_011044) (Figure 12) showed that one gene was consistently up-regulated (p-value < 0.05) in all GLRaV-3 infected samples, in all cultivars. This gene has high sequence identity to a gene in the expansin family, therefore its up-regulation in infected plants can signify the plant's lowered resistance to the virus. This finding provides a possible mechanism to understand the molecular interaction between the virus and the plant host.

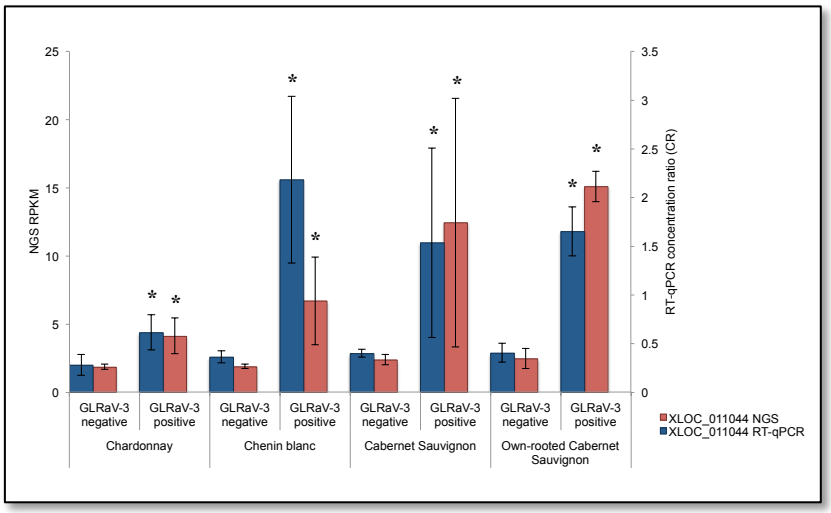


Figure 12. Histogram displaying the up-regulation of transcript XLOC_011044 in GLRaV-3 infected samples. The mean RPKM and concentration ratio (CR) ± standard error (SE) of three biological replicates, with each replicate an average of three technical replicates is displayed. Statistically significant differences between GLRaV-3 negative and positive samples are indicated by asterisks (*p-value < 0.05).

5.4 Conclusion

This chapter describes an investigation of the differential expression of sRNAs and genes in GLRaV-3 infected grapevine, in order to identify a universal GLRaV-3-associated stress response. Even though differential expression of different sRNA species was identified, these sRNAs were cultivar-specific and, except for the vsiRNAs, no sRNA was consistently modulated in response to GLRaV-3 infection in the cultivars investigated. The anti-correlation between miRNAs and their predicted targets was validated in Chardonnay and own-rooted Cabernet Sauvignon NGS data sets. The lack of additional correlations between the sRNAs and targets can possibly be due to *in silico* false positive predictions [92–94]. Since phased loci was identified in both intergenic regions and in transcripts, the absence of differential expression of these associated transcripts in the transcriptome NGS data, does not necessarily indicate that phasiRNA production is not regulated by viral infection. It is possible that the phasiRNAs originating from these transcripts, target intergenic regions that might influence gene expression of the up- or downstream genes. Alternatively, these phasiRNAs may target transcripts different to their loci of origin. The phased loci found in intergenic regions also complicate the predictions of targets, since they do not necessarily target the closest gene up- or downstream of their loci of origin. Currently, sRNA target prediction tools rely on miRNA-associated target recognition and since the functioning of the other sRNA species is largely unknown; it remains to be demonstrated if these sRNAs have the same target recognition characteristics as miRNAs.

It is well known that grapevine does respond to GLRaV-3 in the form of visual symptom expression and other cellular and physiological changes. Since RNA silencing has been implicated in the antiviral defence system and the vsiRNAs were the only sRNA species identified as differentially expressed in all cultivars investigated, it remains plausible that the vsiRNAs are either produced as a defence response against the virus or to down-regulate plant host genes [95–99].

Absence of a universal sRNA response, other than vsiRNAs, would suggest that the different sRNA species are all part of the same complex network, and that each cultivar will respond with different sRNA expression at different times. It can also indicate that the sRNA response is not the main regulator of the GLRaV-3 physiological symptoms observed in the disease. For this reason, differential gene expression, independent of sRNAs, was also investigated.

Cultivar-specific responses, as well as a universal response in the cultivars investigated, were identified using transcriptome NGS data. Genes specific to the severe symptomatic cultivars provide the first insight into the molecular interactions involved in the differences observed between symptomatic and non-symptomatic cultivars infected with the virus. The differentially expressed genes identified in all the cultivars investigated provide plausible reasons for the degeneration

associated with GLD. The transcript with sequence similarity to an expansin gene, which was identified as differentially expressed in all GLRaV-3 infected plants and validated with RT-qPCR, provides a strong lead into a mechanism to counteract the detrimental effects of GLD. If the cell wall structure can be manipulated by the down-regulation of one of the genes in the expansin family, it can lead to the possible protection against the pathogen.

Even though the main goal of this study was to identify a universal sRNA and associated target signal, it has shown that grapevine plants respond on a transcriptome level to the virus infection, and that the sRNA response is most likely cultivar specific. However, whether these responses are specific to GLRaV-3 or stress-dependent in general, remains to be confirmed. The cultivar-specific and universal viral responses identified here, contribute to elucidating the molecular mechanisms underlying the GLRaV-3 stress response in grapevine and can be utilised as targets for engineering viral tolerance to lead to the control or prevention of the disease.

5.5 Supplementary material

Appendix D1: Primers for small RNA stemloop RT-qPCR and transcript RT-qPCR assays.
Appendix D2: Differentially expressed miRNAs and their predicted targets.
Appendix D3: Differentially expressed phased loci and individual differentially expressed phasiRNA.
Appendix D4: Differentially expressed repeat sequences and individual differentially expressed rasiRNAs.
Appendix D5: Differential expression of siRNAs associated with complete tRNA and individual differentially expressed tRNA-derived siRNAs.
Appendix D6: Differentially expressed sRNA reads.
Appendix D7: Differentially expressed genes identified in all GLRaV-3 infected samples.
Appendix D8: Differentially expressed genes identified in severe symptomatic GLRaV-3 infected samples.
Appendix D9: Comparison between NGS and RT-qPCR results of differentially expressed genes identified in all GLRaV-3 infected samples.

5.6 References

1. Martelli GP (2014) Directory of virus and virus-like diseases of the grapevine and their agents. J Plant Pathol 96:1–136.
2. Alabi OJ, Zheng Y, Jagadeeswaran G, et al (2012) High-throughput sequence analysis of small RNAs in grapevine Vitis vinifera affected by grapevine leafroll disease: Small RNAs in leafroll disease-infected grapevine. Mol Plant Pathol 13:1060–1076. doi: 10.1111/j.1364-3703.2012.00815.x
3. Singh K, Talla A, Qiu W (2012) Small RNA profiling of virus-infected grapevines: evidences for virus infection-associated and variety-specific miRNAs. Funct Integr Genomics 12:659–669. doi: 10.1007/s10142-012-0292-1

4. Bazzini AA, Hopp HE, Beachy RN, Asurmendi S (2007) Infection and coaccumulation of tobacco mosaic virus proteins alter microRNA levels, correlating with symptom and plant development. Proc Natl Acad Sci 104:12157–12162. doi: 10.1073/pnas.0705114104

5. Tagami Y, Inaba N, Kutsuna N, et al (2007) Specific Enrichment of miRNAs in Arabidopsis thaliana Infected with Tobacco mosaic virus. DNA Res 14:227–233. doi: 10.1093/dnares/dsm022

6. Pantaleo V, Saldarelli P, Miozzi L, et al (2010) Deep sequencing analysis of viral short RNAs from an infected Pinot Noir grapevine. Virology 408:49–56. doi: 10.1016/j.virol.2010.09.001

7. Guleria P, Mahajan M, Bhardwaj J, Yadav SK (2011) Plant Small RNAs: Biogenesis, Mode of Action and Their Roles in Abiotic Stresses. Genomics Proteomics Bioinformatics 9:183–199. doi: 10.1016/S1672-0229(11)60022-3

8. Pasini L, Bergonti M, Fracasso A, et al (2014) Microarray analysis of differentially expressed mRNAs and miRNAs in young leaves of sorghum under dry-down conditions. J Plant Physiol 171:537–548. doi: http://dx.doi.org/10.1016/j.jplph.2013.12.014

9. Kumar R (2014) Role of MicroRNAs in Biotic and Abiotic Stress Responses in Crop Plants. Appl Biochem Biotechnol 174:93–115. doi: 10.1007/s12010-014-0914-2

10. Pantaleo V, Vitali M, Boccacci P, et al (2016) Novel functional microRNAs from virus-free and infected Vitis vinifera plants under water stress. Sci Rep 6:20167. doi: 10.1038/srep20167

11. Sunkar R (2010) MicroRNAs with macro-effects on plant stress responses. Semin Cell Dev Biol 21:805–811. doi: 10.1016/j.semcdb.2010.04.001

12. Macovei A, Tuteja N (2012) microRNAs targeting DEAD-box helicases are involved in salinity stress response in rice (Oryza sativa L.). BMC Plant Biol 12:183. doi: 10.1186/1471-2229-12-183

13. Hwang E-W (2011) Identification of MicroRNAs and Their Putative Targets that Respond to Drought Stress in Solanum tuberosum. J Korean Soc Appl Biol Chem 54:317–324. doi: 10.3839/jksabc.2011.051

14. Jin Q, Xue Z, Dong C, et al (2015) Identification and Characterization of MicroRNAs from Tree Peony (Paeonia ostii) and Their Response to Copper Stress. PLOS ONE 10:e0117584. doi: 10.1371/journal.pone.0117584

15. Maree HJ, Almeida RPP, Bester R, et al (2013) Grapevine leafroll-associated virus 3. Front Microbiol 4:82. doi: 10.3389/fmicb.2013.00082

16. Almeida RPP, Daane KM, Bell VA, et al (2013) Ecology and management of grapevine leafroll disease. Front Microbiol 4:94. doi: 10.3389/fmicb.2013.00094

17. Naidu R, Rowhani A, Fuchs M, et al (2014) Grapevine Leafroll: A Complex Viral Disease Affecting a High-Value Fruit Crop. Plant Dis 98:1172–1185. doi: 10.1094/PDIS-08-13-0880-FE

18. Gutha LR, Casassa LF, Harbertson JF, Naidu RA (2010) Modulation of flavonoid biosynthetic pathway genes and anthocyanins due to virus infection in grapevine (Vitis vinifera L.) leaves. BMC Plant Biol 10:187. doi: 10.1186/1471-2229-10-187

19. Mannini F, Mollo A, Credi R (2011) Field Performance and Wine Quality Modification in a Clone of Vitis vinifera cv. Dolcetto after GLRaV-3 Elimination. Am J Enol Vitic 63:144–147. doi: 10.5344/ajev.2011.11020

20. Vega A, Gutierrez RA, Pena-Neira A, et al (2011) Compatible GLRaV-3 viral infections affect berry ripening decreasing sugar accumulation and anthocyanin biosynthesis in Vitis vinifera. Plant Mol Biol 77:261–274. doi: 10.1007/s11103-011-9807-8

21. Endeshaw ST, Sabbatini P, Romanazzi G, et al (2014) Effects of grapevine leafroll associated virus 3 infection on growth, leaf gas exchange, yield and basic fruit chemistry of Vitis vinifera L. cv. Cabernet Franc. Sci Hortic 170:228–236. doi: 10.1016/j.scienta.2014.03.021

22. Espinoza C, Medina C, Somerville S, Arce-Johnson P (2007) Senescence-associated genes induced during compatible viral interactions with grapevine and Arabidopsis. J Exp Bot 58:3197–3212. doi: 10.1093/jxb/erm165

23. Carra A, Gambino G, Schubert A (2007) A cetyltrimethylammonium bromide-based method to extract low-molecular-weight RNA from polysaccharide-rich plant tissues. Anal Biochem 360:318–320. doi: 10.1016/j.ab.2006.09.022

24. Bester R, Pepler PT, Burger JT, Maree HJ (2014) Relative quantitation goes viral: An RT-qPCR assay for a grapevine virus. J Virol Methods 210:67–75. doi: 10.1016/j.jviromet.2014.09.022

25. Jooste AEC, Molenaar N, Maree HJ, et al (2015) Identification and distribution of multiple virus infections in Grapevine leafroll diseased vineyards. Eur J Plant Pathol 142:363–375. doi: 10.1007/s10658-015-0620-0

26. Bester R, Jooste AE, Maree HJ, Burger JT (2012) Real-time RT-PCR high-resolution melting curve analysis and multiplex RT-PCR to detect and differentiate grapevine leafroll-associated associated virus 3 variant groups I, II, III and VI. Virol J 9:219. doi: 10.1186/1743-422X-9-219.

27. Martin M (2011) Cutadapt removes adapter sequences from high-throughput sequencing reads. EMBnet J 17:10–12. doi: 10.14806/ej.17.1.200

28. Bolger AM, Lohse M, Usadel B (2014) Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics 30:2114–2120. doi: 10.1093/bioinformatics/btu170

29. Langmead B, Trapnell C, Pop M, Salzberg SL (2009) Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. Genome Biol 10:R25. doi: 10.1186/gb-2009-10-3-r25

30. Jaillon O, Aury J-M, Noel B, et al (2007) The grapevine genome sequence suggests ancestral hexaploidization in major angiosperm phyla. Nature 449:463–467. doi: 10.1038/nature06148

31. Love MI, Huber W, Anders S (2014) Moderated estimation of fold change and dispersion for RNA-Seq data with DESeq2.

32. Dai X, Zhao PX (2011) psRNATarget: a plant small RNA target analysis server. Nucleic Acids Res 39:W155–W159. doi: 10.1093/nar/gkr319

33. Conesa A, Götz S (2008) Blast2GO: A Comprehensive Suite for Functional Analysis in Plant Genomics. Int J Plant Genomics 2008:1–12. doi: 10.1155/2008/619832

34. Varkonyi-Gasic E, Wu R, Wood M, et al (2007) Protocol: a highly sensitive RT-PCR method for detection and quantification of microRNAs. Plant Methods 3:12. doi: 10.1186/1746-4811-3-12

35. Kozomara A, Griffiths-Jones S (2014) miRBase: annotating high confidence microRNAs using deep sequencing data. Nucleic Acids Res 42:D68–D73. doi: 10.1093/nar/gkt1181

36. Bao W, Kojima KK, Kohany O (2015) Repbase Update, a database of repetitive elements in eukaryotic genomes. Mob DNA 6:11. doi: 10.1186/s13100-015-0041-9

37. Cognat V, Pawlak G, Duchene A-M, et al (2013) PlantRNA, a database for tRNAs of photosynthetic eukaryotes. Nucleic Acids Res 41:D273–D279. doi: 10.1093/nar/gks935

38. Trapnell C, Roberts A, Goff L, et al (2012) Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and Cufflinks. Nat Protoc 7:562–578. doi: 10.1038/nprot.2012.016

39. Langmead B, Salzberg SL (2012) Fast gapped-read alignment with Bowtie 2. Nat Methods 9:357–359. doi: 10.1038/nmeth.1923

40. Reid K, Olsson N, Schlosser J, et al (2006) An optimized grapevine RNA isolation procedure and statistical determination of reference genes for real-time RT-PCR during berry development. BMC Plant Biol 6:1–11. doi: 10.1186/1471-2229-6-27

41. Wang X-B, Wu Q, Ito T, et al (2010) RNAi-mediated viral immunity requires amplification of virus-derived siRNAs in Arabidopsis thaliana. Proc Natl Acad Sci 107:484–489. doi: 10.1073/pnas.0904086107

42. Jakubiec A, Yang SW, Chua N-H (2012) Arabidopsis DRB4 protein in antiviral defense against Turnip yellow mosaic virus infection: DRB4 function in antiviral defense. Plant J 69:14–25. doi: 10.1111/j.1365-313X.2011.04765.x

43. Donaire L, Wang Y, Gonzalez-Ibeas D, et al (2009) Deep-sequencing of plant viral small RNAs reveals effective and widespread targeting of viral genomes. Virology 392:203–214. doi: 10.1016/j.virol.2009.07.005

44. Wang X-B, Jovel J, Udomporn P, et al (2011) The 21-Nucleotide, but Not 22-Nucleotide, Viral Secondary Small Interfering RNAs Direct Potent Antiviral Defense by Two Cooperative Argonautes in Arabidopsis thaliana. Plant Cell 23:1625–1638. doi: 10.1105/tpc.110.082305

45. Visser M, Maree HJ, Rees DJ, Burger JT (2014) High-throughput sequencing reveals small RNAs involved in ASGV infection. BMC Genomics 15:568. doi: 10.1186/1471-2164-15-568

46. Ahlquist P, Noueiry AO, Lee W-M, et al (2003) Host Factors in Positive-Strand RNA Virus Genome Replication. J Virol 77:8181–8186. doi: 10.1128/JVI.77.15.8181-8186.2003

47. Gouveia P, Nolasco G (2012) The p19.7 RNA silencing suppressor from Grapevine leafroll-associated virus 3 shows different levels of activity across phylogenetic groups. Virus Genes 45:333–339. doi: 10.1007/s11262-012-0772-3

48. Gouveia P, Dandlen S, Costa Â, et al (2012) Identification of an RNA silencing suppressor encoded by Grapevine leafroll-associated virus 3. Eur J Plant Pathol 133:237–245. doi: 10.1007/s10658-011-9876-1

49. Maree HJ, Gardner HFJ, Freeborough M-J, Burger JT (2010) Mapping of the 5' terminal nucleotides of Grapevine leafroll-associated virus 3 sgRNAs. Virus Res 151:252–255. doi: 10.1016/j.virusres.2010.05.011

50. Afzal AJ, Wood AJ, Lightfoot DA (2008) Plant receptor-like serine threonine kinases: roles in signaling and plant defense. Mol Plant Microbe Interact 21:507–517. doi: 10.1094 / MPMI -21-5-0507

51. Benjak A, Forneck A, Casacuberta JM (2008) Genome-Wide Analysis of the "Cut-and-Paste" Transposons of Grapevine. PLoS ONE 3:e3107. doi: 10.1371/journal.pone.0003107

52. Kobayashi S, Goto-Yamamoto N, Hirochika H (2004) Retrotransposon-Induced Mutations in Grape Skin Color. Science 304:982–982. doi: 10.1126/science.1095011

53. Hsieh L-C, Lin S-I, Shih AC-C, et al (2009) Uncovering Small RNA-Mediated Responses to Phosphate Deficiency in Arabidopsis by Deep Sequencing. Plant Physiol 151:2120–2132. doi: 10.1104/pp.109.147280

54. Loss-Morais G, Waterhouse PM, Margis R (2013) Description of plant tRNA-derived RNA fragments (tRFs) associated with argonaute and identification of their putative targets. Biol Direct 8:1–5. doi: 10.1186/1745-6150-8-6

55. Visser M, Bester R, Burger JT, Maree HJ (2016) Next-generation sequencing for virus detection: covering all the bases. Virol J. doi: 10.1186/s12985-016-0539-x

56. Sztuba-Solińska J, Stollar V, Bujarski JJ (2011) Subgenomic messenger RNAs: Mastering regulation of (+)-strand RNA virus life cycle. Virology 412:245–255. doi: 10.1016/j.virol.2011.02.007

57. Wierzchoslawski R, Dzianott A, Bujarski J (2004) Dissecting the Requirement for Subgenomic Promoter Sequences by RNA Recombination of Brome Mosaic Virus In Vivo: Evidence for Functional Separation of Transcription and Recombination. J Virol 78:8552–8564. doi: 10.1128/JVI.78.16.8552-8564.2004

58. Miyazawa Y, Moriwaki T, Uchida M, et al (2012) Overexpression of MIZU-KUSSEI1 Enhances the Root Hydrotropic Response by Retaining Cell Viability Under Hydrostimulated Conditions in Arabidopsis thaliana. Plant Cell Physiol 53:1926–1933. doi: 10.1093/pcp/pcs129

59. John ME, Crow LJ (1992) Gene expression in cotton (Gossypium hirsutum L.) fiber: cloning of the mRNAs. Proc Natl Acad Sci 89:5769–5773.

60. Scavetta RD, Herron SR, Hotchkiss AT, et al (1999) Structure of a plant cell wall fragment complexed to pectate lyase C. Plant Cell 11:1081–1092. doi: 10.1105/tpc.11.6.1081

61. Marin-Rodriguez MC (2002) Pectate lyases, cell wall degradation and fruit softening. J Exp Bot 53:2115–2119. doi: 10.1093/jxb/erf089

62. Goujon T, Minic Z, El Amrani A, et al (2003) AtBXL1, a novel higher plant (Arabidopsis thaliana) putative beta-xylosidase gene, is involved in secondary cell wall metabolism and plant development. Plant J 33:677–690. doi: 10.1046/j.1365-313X.2003.01654.x

63. Sampedro J, Cosgrove DJ (2005) The expansin superfamily. Genome Biol 6:242. doi: 10.1186/gb-2005-6-12-242

64. Cho H-T, Kende H (1997) Expression of expansin genes is correlated with growth in deepwater rice. Plant Cell 9:1661–1671. doi: 10.1105/tpc.9.9.1661

65. Zhang X (1998) Leucine-rich repeat receptor-like kinases in plants. Plant Mol Biol Report 16:301–311. doi: 10.1016/S0074-7696(04)34001-5

66. Klepek Y-S (2005) Arabidopsis POLYOL TRANSPORTER5, a New Member of the Monosaccharide Transporter-Like Superfamily, Mediates H+-Symport of Numerous Substrates, Including myo-Inositol, Glycerol, and Ribose. Plant Cell 17:204–218. doi: 10.1105/tpc.104.026641

67. Maurel C, Boursiac Y, Luu D-T, et al (2015) Aquaporins in Plants. Physiol Rev 95:1321–1358. doi: 10.1152/physrev.00008.2015

68. Maurel C, Verdoucq L, Luu D-T, Santoni V (2008) Plant Aquaporins: Membrane Channels with Multiple Integrated Functions. Annu Rev Plant Biol 59:595–624. doi: 10.1146/annurev.arplant.59.032607.092734

69. Marica Bakovic NH (2015) Biologically Active Triterpenoids and Their Cardioprotective and Anti-Inflammatory Effects. J Bioanal Biomed S12:005. doi: 10.4172/1948-593X.S12-005

70. Hamberger B, Bak S (2013) Plant P450s as versatile drivers for evolution of species-specific chemical diversity. Philos Trans R Soc B Biol Sci 368:20120426. doi: 10.1098/rstb.2012.0426

71. Ludwig SR, Oppenheimer DG, Silflow CD, Snustad DP (1987) Characterization of the alpha-tubulin gene family of Arabidopsis thaliana. Proc Natl Acad Sci 84:5833–5837.

72. Kavi Kishor PB (2015) Role of proline in cell wall synthesis and plant development and its implications in plant ontogeny. Front Plant Sci 6:544. doi: 10.3389/fpls.2015.00544

73. Zhan X, Wang B, Li H, et al (2012) Arabidopsis proline-rich protein important for development and abiotic stress tolerance is involved in microRNA biogenesis. Proc Natl Acad Sci 109:18198–18203. doi: 10.1073/pnas.1216199109

74. Wyatt RE, Nagao RT, Key JL (1992) Patterns of soybean proline-rich protein gene expression. Plant Cell 4:99–110. doi: 10.1105/tpc.4.1.99

75. Ranocha P (2002) Laccase Down-Regulation Causes Alterations in Phenolic Metabolism and Cell Wall Structure in Poplar. Plant Physiol 129:145–155. doi: 10.1104/pp.010988

76. Wentzinger LF (2002) Inhibition of Squalene Synthase and Squalene Epoxidase in Tobacco Cells Triggers an Up-Regulation of 3-Hydroxy-3-Methylglutaryl Coenzyme A Reductase. Plant Physiol 130:334–346. doi: 10.1104/pp.004655

77. Schäfer UA, Reed DW, Hunter DG, et al (1999) An example of intron junctional sliding in the gene families encoding squalene monooxygenase homologues in Arabidopsis thaliana and Brassica napus. Plant Mol Biol 39:721–728.

78. Burden RS, Cooke DT, Carter GA (1989) Inhibitors of sterol biosynthesis and growth in plants and fungi. Phytochemistry 28:1791–1804. doi: 10.1016/S0031-9422(00)97862-2

79. Nuruzzaman M, Sharoni AM, Kikuchi S (2013) Roles of NAC transcription factors in the regulation of biotic and abiotic stress responses in plants. Front Microbiol 4:248. doi: 10.3389/fmicb.2013.00248

80. Agrawal GK, Iwahashi H, Rakwal R (2003) Small GTPase “Rop”: molecular switch for plant defense responses. FEBS Lett 546:173–180. doi: 10.1016/S0014-5793(03)00646-X

81. Beffa RS, Neuhaus J-M, Meins F (1993) Physiological compensation in antisense transformants: specific induction of an“ ersatz” glucan endo-1, 3-beta-glucosidase in plants infected with necrotizing viruses. Proc Natl Acad Sci 90:8792–8796.

82. Ranocha P, Dima O, Nagy R, et al (2013) Arabidopsis WAT1 is a vacuolar auxin transport facilitator required for auxin homoeostasis. Nat Commun 4:2625. doi: 10.1038/ncomms3625

83. Ranocha P, Denancé N, Vanholme R, et al (2010) Walls are thin 1 (WAT1), an Arabidopsis homolog of Medicago truncatula NODULIN21, is a tonoplast-localized protein required for secondary wall formation in fibers: Tonoplastic WAT1 and secondary wall formation. Plant J 63:469–483. doi: 10.1111/j.1365-313X.2010.04256.x

84. AbuQamar S (2014) Expansins: Cell Wall Remodeling Proteins with a Potential Function in Plant Defense. J Plant Biochem Physiol 2:1. doi: 10.4172/2329-9029.1000e118

85. Wu Y, Meeley RB, Cosgrove DJ (2001) Analysis and expression of the α-expansin and β-expansin gene families in maize. Plant Physiol 126:222–232. doi: 10.1104/pp.126.1.222

86. Marowa P, Ding A, Kong Y (2016) Expansins: roles in plant growth and potential applications in crop improvement. Plant Cell Rep 35:949–965. doi: 10.1007/s00299-016-1948-4

87. Wang X, Tang C, Deng L, et al (2010) Characterization of a pathogenesis-related thaumatin-like protein gene *TaPR5* from wheat induced by stripe rust fungus. Physiol Plant 139:27–38. doi: 10.1111/j.1399-3054.2009.01338.x

88. Liu J-J, Sturrock R, Ekramoddoullah AKM (2010) The superfamily of thaumatin-like proteins: its origin, evolution, and expression towards biological function. Plant Cell Rep 29:419–436. doi: 10.1007/s00299-010-0826-8

89. Girard C, Chelysheva L, Choinard S, et al (2015) AAA-ATPase FIDGETIN-LIKE 1 and Helicase FANCM Antagonize Meiotic Crossovers by Distinct Mechanisms. PLOS Genet 11:e1005369. doi: 10.1371/journal.pgen.1005369

90. Lascombe M-B, Bakan B, Buhot N, et al (2008) The structure of "defective in induced resistance" protein of *Arabidopsis thaliana* , DIR1, reveals a new type of lipid transfer protein. Protein Sci 17:1522–1530. doi: 10.1110/ps.035972.108

91. Maldonado AM, Doerner P, Dixon RA, et al (2002) A putative lipid transfer protein involved in systemic resistance signalling in Arabidopsis. Nature 419:399–403. doi: 10.1038/nature00962

92. Srivastava PK, Moturu TR, Pandey P, et al (2014) A comparison of performance of plant miRNA target prediction tools and the characterization of features for genome-wide target prediction. BMC Genomics 15:348. doi: 10.1186/1471-2164-15-348

93. Ding J, Zhou S, Guan J (2012) Finding MicroRNA Targets in Plants: Current Status and Perspectives. Genomics Proteomics Bioinformatics 10:264–275. doi: 10.1016/j.gpb.2012.09.003

94. Meng J, Shi L, Luan Y (2014) Plant microRNA-Target Interaction Identification Model Based on the Integration of Prediction Tools and Support Vector Machine. PLoS ONE 9:e103181. doi: 10.1371/journal.pone.0103181

95. Qi X, Bao FS, Xie Z (2009) Small RNA Deep Sequencing Reveals Role for Arabidopsis thaliana RNA-Dependent RNA Polymerases in Viral siRNA Biogenesis. PLoS ONE 4:e4971. doi: 10.1371/journal.pone.0004971

96. Smith NA, Eamens AL, Wang M-B (2011) Viral Small Interfering RNAs Target Host Genes to Mediate Disease Symptoms in Plants. PLoS Pathog 7:e1002022. doi: 10.1371/journal.ppat.1002022

97. Shimura H, Pantaleo V, Ishihara T, et al (2011) A Viral Satellite RNA Induces Yellow Symptoms on Tobacco by Targeting a Gene Involved in Chlorophyll Biosynthesis using the RNA Silencing Machinery. PLoS Pathog 7:e1002021. doi: 10.1371/journal.ppat.1002021

98. Xia Z, Peng J, Li Y, et al (2014) Characterization of Small Interfering RNAs Derived from Sugarcane Mosaic Virus in Infected Maize Plants by Deep Sequencing. PLoS ONE 9:e97013. doi: 10.1371/journal.pone.0097013

99. Miozzi L, Gambino G, Burgyan J, Pantaleo V (2013) Genome-wide identification of viral and host transcripts targeted by viral siRNAs in Vitis vinifera. Mol Plant Pathol 14:30–43. doi: 10.1111/j.1364-3703.2012.00828.x

Chapter 6: Conclusion

Virus-associated diseases of grapevine are a major limiting factor to the sustainability of the viticulture industry. Grapevine leafroll disease (GLD) is present in all grapevine-growing countries in the world and has a detrimental effect on vine health, as well as crop yield and quality. Even though Koch's postulate has not yet been fulfilled for this disease, *Grapevine leafroll-associated virus 3* is considered the leading causative agent due to its consistent association with typical leafroll disease symptoms. To understand the disease and develop effective control strategies, it is necessary to characterise the molecular interaction between the virus and its host. Functional small RNA (sRNA) molecules have been shown to play an important role in gene regulation and specifically in RNA silencing, which provides a feasible hypothesis that sRNAs can be involved in the plant's defence response to biotic stress. Therefore, one of the main aims of this study was to characterise and compare the sRNA species in healthy and infected grapevines of different cultivars, thereby contributing to the establishment of a comprehensive database of sRNAs present in *Vitis vinifera*. In addition, the differential expression of these sRNAs, and that of their possible gene targets was evaluated to identify sRNAs associated with GLRaV-3 infection.

Next-generation sequencing (NGS) and bioinformatic analyses were used to identify sRNA species in grapevine phloem tissue. Initially, own-rooted Cabernet Sauvignon plants were studied using microarray and NGS technology to identify known microRNAs (miRNAs) modulated by GLRaV-3 infection. The anti-correlation of miRNA expression (up-regulation of vvi-miR398b-c and vvi-miR395a-m) and putative target expression (down-regulation of serine threonine-protein kinase and ATP sulfurylase), that was confirmed with microarray analysis, sRNA NGS, transcriptome NGS and qPCR assays, provided support for a possible regulation mechanism in response to GLRaV-3 infection. This preliminary study was followed up with the large-scale characterisation of sRNA species in healthy and infected plants of different cultivars.

Both known and novel miRNAs were identified in Chardonnay, Chenin blanc, Cabernet Sauvignon and own-rooted Cabernet Sauvignon. The novel miRNAs identified in the different cultivars contribute significantly to the grapevine miRNA knowledge base, since the majority of the miRNAs present in miRBase, was identified in Pinot noir. In some cases, the mature miRNA identified for a known miRNA precursor, had a higher read count than the known miRNA for that specific precursor. These isomiRs (sequence variants) of known vvi-miRNAs, as well as miRNAs identical to miRNAs from other plant species, suggest that different miRNAs from the same hairpin precursor can be expressed at different levels relative to each other as a result of cultivar, tissue type or environmental differences. The up-regulation of miR398b-c was detected in both infected Chardonnay and own-rooted Cabernet Sauvignon plants, implying a strong connection to host specificity as both Chardonnay and Cabernet Sauvignon are GLRaV-3 symptomatic cultivars, compared to Chenin blanc

that displays little or no visual symptoms. Conversely, a transcript that was identified as the potential target of miR398b-c, was significantly down-regulated in RNA-Seq experiments. This putative target had high sequence similarity to a serine threonine-protein kinase gene, which plays a key role in signalling during pathogen recognition and the activation of plant defence mechanisms, including the generation of nitric oxide and superoxide and programmed cell death [1]. This expression anti-correlation, coupled with the function of the predicted target, suggest a likely defence response.

Phased loci were predicted for all cultivars, with several of these being cultivar specific. The majority of the predicted phased loci overlapped with known *Vitis vinifera* transcripts and of these, 60 % had high sequence similarity to disease resistance proteins. In both Chardonnay and Chenin blanc, a phased locus which overlaps with a TAS4 locus was identified. The phased small interfering RNAs (phasiRNAs) generated from this locus are believed to target MYB transcription factors that in turn regulates anthocyanin biosynthesis [2, 3]. The lower phasing signature for this locus observed in Cabernet Sauvignon samples can be linked to the cultivar specificity of anthocyanin production seen in white versus red cultivars. Novel phased loci with high phasing signatures were also identified in all cultivars, implying a role in normal plant development. The function of the phasiRNAs generated from these loci, and how the phasing is initiated, remains to be determined. Only one miRNA cleavage site (miR3634-3p) fell into the phased register of one of the phased loci identified in Chardonnay, Cabernet Sauvignon and own-rooted Cabernet Sauvignon. This locus has a high sequence similarity to an Ankyrin repeat-containing gene, which has been linked with systemic resistance to pathogens in plants [4]. This miRNA was also up-regulated in infected own-rooted Cabernet Sauvignon plants, suggesting a possible defence mechanism. The absence of more in-phase miRNA cleavage sites may be due to initiation through a different sRNA species or cleavage by a second siRNA in the phasing registers. A number of phased loci were significantly differentially enriched for siRNAs, however only four of these enriched loci were shared between the different cultivars, with the remainder being cultivar specific. The same was observed for the read counts of individual phasiRNAs, suggesting a link between phasiRNA expression and a potential red-white, cultivar-specific defence response. However, this will need to be confirmed using additional cultivars and more biological replicates. A complete understanding of the phasing criteria and phasiRNAs' contribution to gene regulation can in future provide a potential system to manipulate for the modulation of specific genes.

Natural antisense transcript (NAT) pairs were identified by searching for transcript loci that overlap in an antiparallel manner (*cis*-NAT) or for unrelated transcripts with complementarity (*trans*-NAT). Twenty-five transcript pairs were identified that could form potential RNA-RNA duplexes. However, to confirm that these transcripts were expressed in the same cell and under the same conditions, sRNAs originating from the overlapping region had to be present in the samples. Even though sRNA reads with sequence identity to the overlap regions were identified, none of the 25 NAT overlap

regions had a significant enrichment for siRNAs when GLRaV-3 positive and negative samples were compared. This can be due to an underestimation of the number of NATs in *Vitis vinifera* since a very strict selection criterion was applied. The absence of natural antisense transcript siRNA (natsiRNA) differential expression can also imply that the natsiRNA mechanism in plants is not regulated by biotic stress.

Small RNAs associated with all the *Vitis vinifera* repeat sequences in Repbase were identified. The largest cluster of reads mapped to retrotransposons and a strand bias was observed for several repeat sequences in all cultivar groups. Significant differential read counts for siRNAs associated with the EnSpm-3 VV EnSpm/CACTA repeat were observed in infected samples of all the cultivars. The representative sRNA with the highest read count had sequence identity to a locus upstream of a transcript with high sequence similarity to a serine/threonine-protein kinase gene. This transcript was also found to be down-regulated in GLRaV-3 infected Chardonnay and own-rooted Cabernet Sauvignon in the transcriptome NGS data. Due to the read mapping strategy chosen, this siRNA was identified in the repeat-associated siRNA (rasiRNA) pool, however based on sequence identity, this siRNA was most likely virus-derived. Since this repeat region is a transposable element, it poses the question whether this sRNA is not an evolutionary remnant of the virus due to the co-evolution of grapevine and GLRaV-3. Whether this represents a possible interaction between the virus and the host genome or just a coincidence, remains to be determined.

Almost all mature tRNAs were found to give rise to sRNAs and more than 98 % of the tRNA-derived sRNAs identified were present in all cultivar groups. This would suggest that these siRNA species have a specific biogenesis and are not random degradation products. The number of tRNA-halves identified in Chenin blanc was lower compared to the other cultivars, suggesting a possible difference in how the defence response of Chenin blanc can be adapted, since tRNA-halves have been linked to stress responses. A significant number of sRNAs were associated with specific tRNAs in GLRaV-3 infected, compared to healthy plants, in all cultivars and a significant variation in read counts for individual tRNA-derived siRNAs were observed. The majority of these differentially expressed siRNAs were unique for each cultivar, suggesting a cultivar-specific tRNA-mediated regulation mechanism in response to GLRaV-3 stress. The specific function of the tRNA-derived sRNAs remains to be elucidated.

The production of virus-derived siRNAs (vsiRNAs) was also investigated and a complete genome sequence for GLRaV-3 could be assembled from the sRNA sequences. The analysis of vsiRNAs confirmed the presence of only GLRaV-3 variant group II in the infected samples, and no GLRaV-3 infection was detected in the healthy samples. The GLRaV-3-specific read counts was also compared to the virus concentration ratio (VCR) measured with RT-qPCR. Even though a significant higher VCR was detected in Chenin blanc compared to the other cultivars, no positive correlation was

observed between VCR and read counts. This could suggest that vsiRNAs will accumulate to a specific level, irrespective of VCR. Both the cultivar-associated VCR differences and the lack of a correlation between VCR and read counts warrant further investigation. In all cultivars, a high number of reads mapped to open reading frame 10 (ORF10). This ORF is predicted to be involved in the suppression of RNA silencing [5–7], therefore this high-density read mapping to ORF10 can either be due to a host-pathogen interaction to suppress the plant's antiviral response, or the presence of an elevated number of templates for degradation, as a result of the transcription of viral sgRNAs [8].

The absence of differentially expressed sRNAs in the Cabernet Sauvignon samples was unexpected; however, the extended time it took for the canes of this cultivar to lignify compared to the other cultivars, may imply a developmental-stage difference between the cultivars. Also, the own-rooted Cabernet Sauvignon plants were established from cuttings made from the original virus source. These plants were used to inoculate the other cultivars, which did display differential sRNA expression, suggesting that GLRaV-3 infection in Cabernet Sauvignon takes longer to establish to the same degree as in the other cultivars. The inclusion of the own-rooted Cabernet Sauvignon plants in the experimental design was to investigate whether grafting a scion onto a rootstock influences the molecular response to GLRaV-3 infection. Since the own-rooted Cabernet Sauvignon responded differently than the Cabernet Sauvignon plants, sRNA expression during GLRaV-3 infection for different scion-rootstock combinations remain an avenue to explore. A third level of comparison was planned for this experiment, which involved establishing GLRaV-3 infections in rootstock material. However, the effect of the greenhouse environment or the grafting of the virus-infected buds caused the plants to remain stunted and did not yield adequate new growth for analyses. The potential regulation mechanism involved in the scion, rootstock and virus interaction remains to be investigated.

Transcripts with high sequence similarity to serine threonine-protein kinase genes were identified as potential targets in both the miRNA and the rasiRNA/vsiRNA analyses, suggesting that this gene family can play a significant role in the host-virus interaction. The association of this gene family to pathogen recognition and the activation of the plant defence response, indicates an ideal target for modulation to enhance tolerance in plants. The identification of a high number of rasiRNAs and tRNA-derived siRNAs extends the sRNA profile of grapevine and demonstrates the diversity of potential genome regions and RNA molecules that can produce sRNAs. Information regarding the function of tRNA-derived sRNAs in plants is still limited, however the altered levels of these sRNAs in infected plants, suggest that they play a role during virus infection. Next-generation sequencing allowed for the identification of a large number of sRNAs, however most of these reads remain unidentified. Although not differentially expressed in infected samples, these reads may represent as yet uncharacterised sRNA species.

Since sRNAs are not necessarily the only regulators of GLRaV-3 responses in grapevine, differential

gene expression was also investigated independently from sRNAs. Seventeen genes were differentially expressed exclusively in the two symptomatic cultivars (Chardonnay and both the Cabernet Sauvignon groups), providing possible targets for investigating symptom expression. Twelve genes were identified as differentially expressed in all GLRaV-3 infected samples, all with a possible link to plant stress responses. The direction of regulation of these 12 genes was confirmed with RT-qPCR assays, except for one gene in Chenin blanc and two genes in own-rooted Cabernet Sauvignon. Both the NGS and RT-qPCR data identified one gene that was consistently and significantly up-regulated in all GLRaV-3 infected samples of all the cultivars. This gene has high sequence identity to a gene in the expansin family, which was shown to play a role in cell structure and in plant-microbe interactions [9–13]. Studies showed that the suppression of expansin genes can enhance the protection of plants against pathogens through preventing plant cell walls from loosening [13]. Assuming that the opposite can also be true, the up-regulation of these genes in infected plants could enhance viral movement. This finding identified a possible mechanism to exploit for engineering viral tolerance in grapevine.

Future prospects include functional studies to elucidate the role of the sRNA species with unknown function, identified in this study. The development of sRNA target prediction tools, customised for the different sRNA species, other than miRNAs, will contribute significantly to the functional annotations of these sRNAs. Such tools will however be contingent on the elucidation of these sRNA-target recognition mechanisms. It will also be valuable to use functional studies to test the anti-correlation between miRNAs and their predicted targets. Although the anti-correlation was validated with transcriptome NGS data and RT-qPCR, correlation does not neccesarily imply causation. Large-scale validation of predicted targets, using degradome sequencing will also add significant value to future studies.

The number of biological replicates utilised in the experimental design was a potential limitation of this study. The natural variation between plants may be too large to be compensated for by only three replicates, thus the addition of more plants per group could potentially lead to a clearer picture of grapevine responses to GLRaV-3 infection. The selection of three plants per group was largely due to resource constraints, since the inclusion of different cultivars to the experimental design was perceived more important than more biological replicates. The cultivar specificity observed with regards to the differentially expressed sRNAs and genes remains to be confirmed by the comparison of additional cultivars. These differences observed between Chenin blanc and the other cultivars provide valuable leads for functional studies to potentially identify a natural resistance mechanism. Through the comparison of symptomatic and asymptomatic GLRaV-3-infected plants, sRNAs and genes that are potentially responsible for inhibiting symptom development can be identified.

Little is known about the biological properties of the different GLRaV-3 genetic variant groups and specifically whether the level of pathogenicity differs between these. Based on data collected from a

survey of Western Cape vineyards, variant group II and VI infections were the most abundant [14]. These variant groups are also found worldwide [15–17] and for this reason, GLRaV-3 variant group II was chosen for the current study. It will be of interest to investigate the influence of the other variant groups, in either single or mix infections, on the sRNA profiles and gene expression patterns of infected plants. The development of a GLRaV-3 infectious clone will allow for the identification of specific elements that are essential for GLRaV-3 infection. Such an infectious clone will also be valuable to study the influence of the variant group-specific genome regions on the host response.

The current study was performed in a greenhouse to eliminate the influence of environmental factors that could interfere with the defence responses of the plant against the virus. Even though this provides an ideal environment to answer specific biological questions, the real-life scenario involves influences from climate, soil, different irrigation regimes and co-infection with other pathogens and viruses. The leads generated in the greenhouse trials can be used to extent the survey to field conditions by including plants infected with multiple GLRaV-3 variant groups and other grapevine viruses. This will determine if the responses identified in this study is GLRaV-3 specific, or a universal stress response.

This study have contributed significantly to the knowledge of sRNAs produced in grapevine and provides a number of sRNAs and target genes that can be utilised as potential targets of grapevine functional studies. The negative effect of GLRaV-3 on plant growth could be linked to the differentially regulated gene targets identified in the study, since the modulation of these targets can result in reduced plant growth and a lower resistance to biotic stresses. The knowledge generated in the study has contributed to the characterisation of the host defence and viral counter-defence strategies and has the potential to be utilised for the impairment of the virus's ability to induce pathogenesis in plants.

References

1. Afzal AJ, Wood AJ, Lightfoot DA (2008) Plant receptor-like serine threonine kinases: roles in signaling and plant defense. Mol Plant Microbe Interact 21:507–517. doi: 10.1094 / MPMI -21-5-0507
2. Zheng Y, Wang Y, Wu J, et al (2015) A dynamic evolutionary and functional landscape of plant phased small interfering RNAs. BMC Biol 13:32. doi: 10.1186/s12915-015-0142-4
3. Rock CD (2013) Trans-acting small interfering RNA4: key to nutraceutical synthesis in grape development? Trends Plant Sci 18:601–610. doi: 10.1016/j.tplants.2013.07.006
4. Voronin DA, Kiseleva EV (2008) Functional role of proteins containing ankyrin repeats. Cell Tissue Biol 2:1–12. doi: 10.1134/S1990519X0801001X
5. Reed JC, Kasschau KD, Prokhnevsky AI, et al (2003) Suppressor of RNA silencing encoded by Beet yellows virus. Virology 306:203–209. doi: 10.1006/S0042-6822(02)00051-X
6. Lu R, Folimonov A, Shintaku M, et al (2004) Three distinct suppressors of RNA silencing encoded by a 20-kb viral RNA genome. Proc Natl Acad Sci U S A 101:15742–15747. doi: 10.1073/pnas.0404940101
7. Chiba M, Reed JC, Prokhnevsky AI, et al (2006) Diverse suppressors of RNA silencing enhance agroinfection by a viral replicon. Virology 346:7–14. doi: 10.1016/j.virol.2005.09.068

8. Jarugula S, Gowda S, Dawson WO, Naidu RA (2010) 3'-coterminal subgenomic RNAs and putative cis-acting elements of Grapevine leafroll-associated virus 3 reveals "unique" features of gene expression strategy in the genus Ampelovirus. Virol J 7:180. doi: 10.1186/1743-422X-7-180

9. Cho H-T, Kende H (1997) Expression of expansin genes is correlated with growth in deepwater rice. Plant Cell 9:1661–1671. doi: 10.1105/tpc.9.9.1661

10. AbuQamar S (2014) Expansins: Cell Wall Remodeling Proteins with a Potential Function in Plant Defense. J Plant Biochem Physiol 2:1. doi: 10.4172/2329-9029.1000e118

11. Sampedro J, Cosgrove DJ (2005) The expansin superfamily. Genome Biol 6:242. doi: 10.1186/gb-2005-6-12-242

12. Wu Y, Meeley RB, Cosgrove DJ (2001) Analysis and expression of the α-expansin and β-expansin gene families in maize. Plant Physiol 126:222–232. doi: 10.1104/pp.126.1.222

13. Marowa P, Ding A, Kong Y (2016) Expansins: roles in plant growth and potential applications in crop improvement. Plant Cell Rep 35:949–965. doi: 10.1007/s00299-016-1948-4

14. Jooste AEC, Molenaar N, Maree HJ, et al (2015) Identification and distribution of multiple virus infections in Grapevine leafroll diseased vineyards. Eur J Plant Pathol 142:363–375. doi: 10.1007/s10658-015-0620-0

15. Sharma AM, Wang J, Duffy S, et al (2011) Occurrence of Grapevine Leafroll-Associated Virus Complex in Napa Valley. PLoS ONE 6:e26227. doi: 10.1371/journal.pone.0026227

16. Gouveia P, Santos MT, Eiras-Dias JE, Nolasco G (2011) Five phylogenetic groups identified in the coat protein gene of grapevine leafroll-associated virus 3 obtained from Portuguese grapevine varieties. Arch Virol 156:413–420. doi: 10.1007/s00705-010-0878-7

17. Chooi KM, Cohen D, Pearson MN (2013) Generic and sequence-variant specific molecular assays for the detection of the highly variable Grapevine leafroll-associated virus 3. J Virol Methods 189:20–29. doi: 10.1016/j.jviromet.2012.12.018