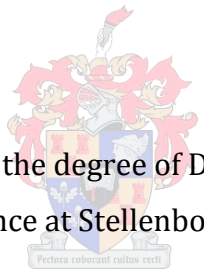
Small RNA profiling of virus-infected apple plants

by

Marike Visser



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

Supervisor: Prof. Johan T. Burger

Co-supervisors: Dr. Hans J. Maree and Dr. D. Jasper. G. Rees

December 2014

Declaration

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

December 2014

Abstract

Apple stem grooving virus (ASGV) is globally associated with latent infection in commercial apple trees. Little is known about this plant-virus interaction. This study made use of next-generation sequencing to investigate the effect of virus-infection on the expression of the different small RNA (sRNA) species namely, miRNAs, nat-siRNAs, phasiRNAs, rasiRNAs, tRNA-derived sRNAs and vsiRNAs. Broad and narrow size-range datasets were generated using sRNA libraries prepared from total and size-selected RNA, respectively. Through bioinformatic data analyses, 130 genomic loci were predicted to give rise to miRNAs, 85 of which were novel *MIR* genes. Targets were predicted for the majority of miRNAs, a few of which could be validated with a publicly available degradome dataset. Cis- and trans-natural antisense transcripts (NATs) were identified, of which only the latter were highly enriched for sRNAs in their overlapping regions. Transcript as well as genomic regions were also identified that can give rise to phasiRNAs. For 25 of these loci an in-phase miRNA target site was identified, half of which could be validated with the degradome dataset. Nearly all apple repeat sequences in Repbase were associated with sRNA synthesis. sRNAs derived from both ends of mature tRNAs were identified. These sRNAs corresponded to tRFs and tRNA-halves. Reads associated with tRNA-halves were prominent in the broad range datasets. sRNAs, originating from the central regions of tRNAs, were also observed. Analysis of the vsiRNAs suggested the presence of two ASGV genetic variants in two of the samples, while the third sample was infected with only one variant. Comparison of the vsiRNA profiles generated from the two datasets highlighted the influence of library preparation on the interpretation of results. Differential expression analysis of the identified apple sRNA species showed no variation between healthy and infected plants, except for the tRNA-derived sRNAs, which did show altered expression levels. Taken together, the various sRNA species characterised in this study significantly extended the existing knowledge of apple sRNAs and provide a broad platform for future functional studies in apple. This study also presents the first and most comprehensive report on sRNAs involved in ASGV infection in apple.

Opsomming

Appel gleufstam virus (ASGV) word wêreldwyd geassosieer met latente infeksie in kommersiële appelbome. Min inligting oor hierdie plant-virus interaksie is beskikbaar. Hierdie studie het van volgende-generasie volgordebepaling gebruik gemaak om die effek van virusinfeksie op die uitdrukking van verskillende klein RNA (sRNA) spesies, nl. miRNAs, nat-siRNAs, phasiRNAs, rasiRNAs, tRNA-afkomstige sRNAs en vsiRNAs, te ondersoek. Datastelle met breë en smal grootte-verspreiding is gegenereer m.b.v. sRNA biblioteke wat onderskeidelik voorberei is vanaf totale RNA en RNA van ‘n bepaalde grootte. Deur middel van bioinformatiese data-ontleding is 130 genomiese loci voorspel wat aanleiding kan gee tot miRNAs, waarvan 85 nuwe *MIR* gene is. Teikens is voorspel vir die meerderheid van die miRNAs en 'n aantal daarvan kon bevestig word m.b.v. 'n publiek-beskikbare degradoom datastel. Cis- en trans-natuurlike antisense transkripte (NATs) is geïdentifiseer, waarvan slegs die laasgenoemde verryk was vir sRNAs in hul oorvleuelende areas. Transkrip sowel as genomiese areas, wat aanleiding kan gee tot phasiRNAs, is ook geïdentifiseer. Vir 25 van hierdie loci is 'n in-fase miRNA teiken geïdentifiseer, waarvan die helfte bevestig kon word met die degradoom datastel. Byna al die appel herhalende volgordes in Repbase was geassosieer met sRNA sintese. sRNAs afkomstig van beide kante van volwasse tRNAs is geïdentifiseer. Hierdie sRNAs het ooreengestem met tRFs en tRNA-helftes. Volgordes geassosieer met tRNA-helftes was prominent in die breë grootte-verspreiding datastelle. sRNAs, afkomstig van die sentrale dele van tRNAs, is ook waargeneem. Ontleding van die vsiRNAs het die teenwoordigheid van twee ASGV genetiese variante in twee van die monsters aangetoon, terwyl die derde monster met slegs een variant geïnfekteer was. Die vergelyking van vsiRNA profiele, gegenereer vanaf die twee datasteltipes, beklemtoon die invloed van biblioteek voorbereiding op die interpretasie van resultate. Ontleding van die differensiële uitdrukking van die geïdentifiseerde appel sRNA spesies het geen verskil tussen gesonde en geïnfekteerde plante getoon nie, behalwe vir die tRNA-afkomstige sRNAs, wat wel verandering die vlak van uitdrukking getoon het. Die verskillende sRNA spesies wat in hierdie studie geïdentifiseer is, het die bestaande kennis van appel sRNAs aansienlik uitgebrei en bied 'n breë platform vir toekomstige funksionele studies in appel. Hierdie studie bied

ook die eerste, en mees omvattende verslag oor sRNAs betrokke in ASGV infeksie in appel.

List of abbreviations

2BWT	Two-way Burrows-Wheeler Transform
AFB	Auxin Signalling F-Box
AGO	Argonaut
ARF	Auxin Responsive Factor
ASGV	*Apple stem grooving virus*
BLAST	Basic Local Alignment Search Tool
BRL	Broad Range Library
BWA	Burrows-Wheeler Alignment Tool
BWT	Burrows-Wheeler Transform
CASHX	Cache-Assisted Hash Search with XOR Digital Logic
CP	Coat Protein
CPU	Central Processing Unit
cv.	Cultivar
DCL	Dicer-like
dsRNA	Double-stranded RNA
GDR	Genome Database for Rosaceae
GPU	Graphics Processing Unit
GUI	Graphical User Interface
Hel	Helicase
HEN1	HUA ENHANCER1
HST	HASTY
kb	Kilobases
kDa	Kilodalton
M. domestica	*Malus* x *domestica*
MAQ	Mapping and Assembly with Quality
mdm-miRNA	Apple MicroRNA
MIR	MicroRNA Gene
miRBase	MicroRNA Registry Database
miRNA	MicroRNA
MP	Movement Protein
Mt	Methyltransferase

MYB	Myeloblastosis
NAT	Natural Antisense Transcript
nat-siRNAs	Natural Antisense Transcript Small Interfering RNA
NB-LRR	Nucleotide Binding Site Leucine Rich Repeat
NGS	Next-Generation Sequencing
NRL	Narrow Range Library
nt	Nucleotide
ORF	Open Reading Frame
P-pro	Papain-like Protease
padj	Adjusted P-value
PHAS	Phased Small Interfering RNA Gene
phasiRNA	Phased Small Interfering RNA
Pol	DNA-dependent RNA Polymerase
PPR	Pentatricopeptide Repeat
pre-miRNA	MicroRNA Precursor
pre-tRNA	Transfer RNA Precursor
pri-miRNA	Primary MicroRNA Transcript
PTGS	Post-Transcriptional Gene Silencing
rasiRNAs	Repeat-associated Small Interfering RNA
RdDM	RNA-dependent DNA methylation
RDR	(Plant) RNA-dependent RNA polymerase
RdRp	(Virus) RNA-dependent RNA polymerase
RISC	RNA-Induced Silencing Complex
RITS	RNA-Induced Initiation of Transcriptional Gene Silencing
RNase	Ribonuclease
RT-PCR	Reverse Transcription PCR
RTE	Retrotransposon
sgRNA	Subgenomic RNA
siRNA	Small Interfering RNA
SOAP	Short Oligonucleotide Analysis Package
sRNA	Small RNA
sRNAome	Small RNA transcriptome
ssRNA	Single-stranded RNA

tasiRNA	Trans-acting Small Interfering RNA
TAS	Trans-acting Small Interfering RNA Gene
TBE	Tris/Boric acid/Ethylene-diamine-tetra-acetic acid
TE	Transposable Elements
TGS	Transcriptional Gene Silencing
tRF	tRNA-derived RNA Fragment
tRNA	Transfer RNA
UEA	University of East Anglia
V1	Variable Region 1
V2	Variable Region 2
vsiRNA	Virus-derived Small Interfering RNA
VSR	Viral Suppressor of RNA Silencing

Acknowledgements

I would like to express my gratitude to the following people and institutions:

Prof. Johan Burger, for support and encouragement as well as allowing my to be part of his research group.

Dr. Jasper Rees, for allowing me to be part of the Biotechnology Platform and PDP program of the ARC. For enabling this study and supporting me to attend conferences, locally and internationally.

Dr. Hano Maree, for support, guidance and encouragement.

Anelda van der Walt, for bioinformatic assistance and guidance, as well as encouraging and inspiring conversations.

Bernard Visser, for endless help, support and love.

Theo Pepler, for help with statistical analyses.

Dr. Lucienne Mansvelt, for support and assistance.

The international colleagues whom I never met, especially Michael Hackenberg, for their patience, willingness to help and advice.

Rosenhof nursery, for providing the plant material used in this study.

NRF/DST, for a personal bursary. Opinions expressed and conclusions arrived at, are those of the authors and are not necessarily to be attributed to the NRF.

Stellenbosch University and Whitehead Scientific, for a travel bursary to attend the 10th Congress of Plant Pathology in Beijing, China (2013).

My colleagues at the ARC Biotechnology Platform and Cultivar Development as well as fellow Vitis-lab members, for assistance, support, encouragement and a memorable time.

My family and friends, especially my parents, for love and support.

My Heavenly Father.

Contents

Chapter 1

Introduction

1.1 General introduction

South Africa has a large apple industry that produced approximately 880 kilotons of apples during the 2012/2013 season with a total value of production estimated at approximately R4.84 billion [1]. In 2013 the industry employed around 28,220 people with 112,882 dependants (http://www.hortgro.co.za/market-intelligence-statistics/key-deciduous-fruit-statistics/). Apples represent 29% (22,501 ha) of the total area under deciduous fruit cultivation in South Africa with the largest production areas the Ceres, Groenland, Langkloof East and Villiersdorp/Vyeboom districts. The most prevalent cultivars are Golden Delicious, Granny Smith and Royal Gala/Gala.

Due to the economical importance of apples, it is essential to study their pathogens. Apple trees are subjected to infections from a diversity of viruses from different genera. *Apple stem grooving virus* (ASGV) is one of the four viruses that are known to be present in South African apple orchards, the other three being *Apple chlorotic leaf spot virus*, *Apple stem pitting virus* and *Apple mosaic virus*. A recent study evaluated the effect of ASGV infection on the transcriptome of *in vitro* cultured apple shoots [2]. Despite the lack of symptom development the expression of more than 300 genes were altered as a result of ASGV infection. Besides the aforementioned study, limited information is available on plant defence responses associated with these viruses.

Variation in gene expression may result from RNA silencing, a regulatory process guided by small RNA (sRNA) molecules. Different sRNA species have evolved, with distinct synthesis and silencing pathways. The use of next-generation sequencing (NGS) coupled with bioinformatic analysis makes provision for the global

investigation of sRNA populations, allowing for the identification of sRNA species involved in specific biological processes such as virus infection.

1.2 Aims and objectives

This study aimed to characterise the sRNA population in apple, as well as to identify sRNAs involved in plant-viral interactions. The following objectives were set in order to achieve this goal:

- To identify healthy as well as ASGV singly infected apple plants that would serve as sample material for sRNA analysis
- To generate and sequence sRNA libraries on an NGS platform
- To implement bioinformatic software for the identification and characterisation of different apple sRNA species
- To predict potential targets for the predicted microRNAs
- To determine which sRNAs were differentially expressed between samples and therefore potentially involved in apple-virus interaction

1.3 Summary of dissertation chapters

This dissertation consists of five chapters, which are individually introduced, concluded and referenced. Each chapter is briefly outlined in the section below.

Chapter 1: *Introduction*

A general introduction is given along with the aims and objectives that were set for this study. An outline of the thesis is provided and the scientific outputs generated by this study are listed.

Chapter 2: *Literature review*

Along with an overview of ASGV, the biogenesis and function of different sRNA species involved in RNA silencing are discussed. Techniques for sRNA analyses are also briefly summarised, while the use of NGS and bioinformatics for sRNA analysis are discussed in more detail.

Chapter 3: *Extending the sRNAome of apple by next-generation sequencing*

This research chapter describes the use of NGS and bioinformatics to identify and characterise various sRNA species in apple. It discusses novel apple microRNAs and phased siRNAs that were added to characterised members of these sRNA species. Furthermore, apple natural antisense transcripts as well as the siRNA originating from their overlaps that were identified along with siRNAs associated with apple repeat sequences are described.

Chapter 4: *High-throughput sequencing reveals small RNAs involved in ASGV infection*

This research chapter focuses on the results of a comprehensive analysis of the apple sRNA-response to ASGV infection. Virus-derived siRNAs originating from ASGV are discussed. Newly characterised tRNA-derived sRNAs are described and their differential expression as a result of virus infection is shown. The importance of library preparation is also highlighted by the analysis of two differently prepared sRNA data sets.

Chapter 5: *Conclusion*

Concluding remarks are given and future prospects discussed.

1.4 Research outputs

1.4.1 Publications

1. **Visser M**, Van der Walt AP, Maree HJ, Rees DJG, Burger JT (2014) Extending the sRNAome of apple by next-generation sequencing. PLoS ONE 9: e95782. doi:10.1371/journal.pone.0095782

   This publication forms the basis for Chapter 3 and is almost entirely the work of Mrs. Visser. Mrs. Van der Walt provided assistance with the identification of natural antisense transcripts. Two non-authors, listed in the acknowledgment section of the paper, also provided bioinformatic assistance.

2. **Visser M**, Maree HJ, Rees DJG, Burger JT (2014) High-throughput sequencing reveals small RNAs involved in ASGV infection. BMC Genomics 15: 568. doi:10.1186/1471-2164-15-568.

   This research paper forms the basis for Chapter 4 and is almost entirely the work of Mrs. Visser. Two non-authors, listed in the acknowledgment section of the paper, provided bioinformatic assistance.

1.4.2 Conference proceeding

Visser M, Maree HJ, Rees DJG, Burger JT (2013) O23.016 Investigation of the role of small RNA in plant-virus interactions in apple trees. Acta Phytopathologica Sinica 43 (suppl.): 308. ISSN 0412-0914

This proceeding includes the work described in Chapter 3 and 4 and is almost entirely the work of Mrs. Visser.

1.4.3 Workshop presentation

Visser M, Van der Walt AP, Maree HJ, Rees DJG, Burger JT. Small RNA: How deep does the rabbit hole go? 11th Australasian Plant Virology Workshop, 13-15 August 2014, Brisbane, Australia.

This oral presentation, which includes the work from Chapter 3 and 4 of this thesis, was presented by H.J. Maree.

References

1. Directorate Statistics and Economic Analysis (2014) Abstract of agricultural statistics. p. 34. South Africa: Department of Agriculture, forestry and fisheries, Pretoria.
2. Chen S, Ye T, Hao L, Chen H, Wang S, *et al*. (2014) Infection of apple by *Apple stem grooving virus* leads to extensive alterations in gene expression patterns but no disease symptoms. PLoS ONE 9: e95239.

Chapter 2

Literature review

2.1 Introduction

Viral infections in agricultural crops can cause great economic losses and therefore many studies focus on these virulent viruses, investigating their genome structure, mode of transmission, as well as the molecular interaction with their hosts. The same holds true for pathogenic bacteria and fungi. However, viruses infecting commercial apples, such as *Apple stem grooving virus* (ASGV), are mostly latent and little is known about this apparent innocuous interaction. Investigating gene regulation, which include the analysis of small regulatory RNAs, on a genomic scale will provide a better understanding of the mechanisms involved in latent virus infection.

This chapter provides an overview of the literature related to the research undertaken in this study. The review is divided into three main sections. The first provides an outline on ASGV, while the second section deals with the different functional small RNA (sRNA) species analysed in the study. The last section focuses on the methods for sRNA analysis, specifically the use of next-generation sequencing (NGS) and bioinformatics.

2.2 *Apple stem grooving virus*

Apple stem grooving virus is generally considered to be latent, and hence less important in domesticated apple (*Malus* x *domestica*). This perceived diminished importance probably contributes to its global distribution. ASGV has been found in many countries, which include Brazil [1], Bosnia [2], China [3], Egypt [4], France and Germany [5], India [6], Japan [7], New Zealand [8], Turkey [9], United States of America [10] and South Africa [11].

2.2.1 Taxonomy and morphology

ASGV is the type member of the genus *Capillovirus* (family *Betaflexiviridae*, order *Tymovirales*) [12]. The only other official member of this genus is *Cherry virus A*. ASGV synonyms include *Citrus tatter leaf virus* [12–14] and *Pear black necrotic leaf spot virus* [15,16]. As a member of the *Betaflexiviridae* family the virions of ASGV are flexuous filamentous and 620-650 nm x 12 nm in size (Figure 2.1) [17].

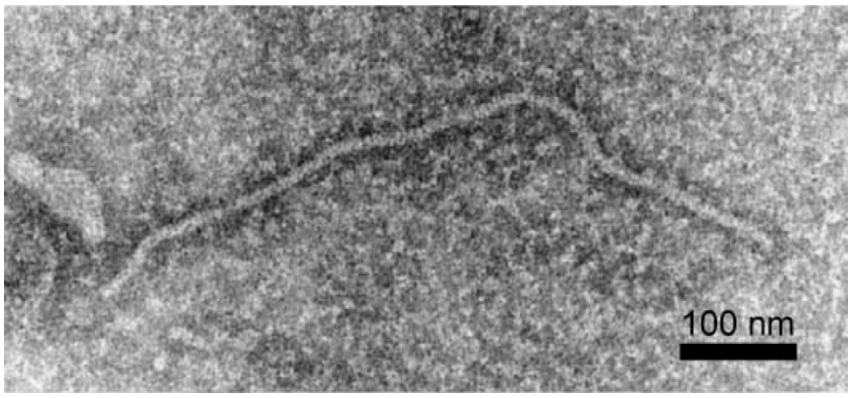


Figure 2.1. Electron micrograph of an *Apple stem grooving virus* virion isolated from kiwifruit. (Image taken from Clover *et al*. [18])

2.2.2 Genome organisation

ASGV is a positive-sense single-stranded RNA virus with a genome size of approximately 6.5 kilobases (kb), which is organised in two overlapping open reading frames (ORFs) (Figure 2.2). The monopartite genome contains a 3′ polyadenylated tail [19]. Yoshikawa *et al*. were the first group to sequence the genome of ASGV [20]. The first ORF (ORF1, bases 37 to 6345) encodes a 241 kilodalton (kDa) polyprotein, containing the helicase, methyltransferase, papain-like protease, and the RNA-dependent RNA polymerase (RdRp) domains, as well as the 27 kDa coat protein (CP) [19–21]. ORF2 (bases 4788 to 5750), which falls within ORF1 and has a different reading frame, encodes a putative 36 kDa movement protein (MP) [20,21]. Two highly variable regions have been characterised at amino acid level [14,22,23]. The two regions, V1 and V2, stretches from amino acids 532 to 570, and from 1583 to 1868, respectively. Although the genomic region of V2 covers the overlap between ORFs 1 and 2 (Figure 2.2), it has no influence on the amino acid sequence of the highly conserved ORF2.

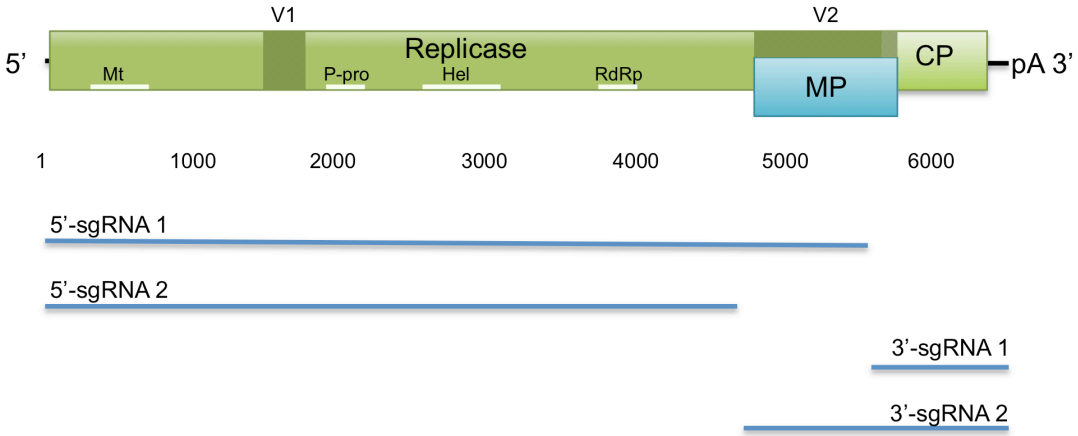


Figure 2.2. Diagram illustrating the genome organisation of *Apple stem grooving virus*. ORF1 and ORF2 are shown in green and blue respectively. The regions coding for the movement protein (MP) and coat protein (CP) are also indicated along with the methyltransferase (Mt), papain-like protease (P-pro), helicase (Hel) and RNA-dependent RNA polymerase (RdRp) protein domains. Shaded regions shown the two variable regions (V1 and V2). The regions for subgenomic RNA (sgRNA) expression are shown for the two 5′-terminal and two 3′-terminal sgRNAs.

Four subgenomic RNAs (sgRNAs), two 3′-terminal sgRNAs and two 5′-terminal sgRNAs, were identified for ASGV and are believed to facilitate the expression of the MP and CP [24,25]. Mutational analysis has indicated that expression of the CP as part of the ORF1 polyprotein is not essential for systemic infection [26]. Instead, sgRNA-expression of the CP is not only sufficient for systemic infection to occur, but is also compulsory [27]. The replication and pathogenicity of the virus were suggested to be dependent on the expression of the CP as part of the polyprotein i.e. the full-length expression of ORF1 [26].

2.2.3 Host range and symptomology

While ASGV-infected commercial apple cultivars are mostly symptomless, studies have reported a decrease in plant growth and trunk diameter, as well as chlorotic and deformed leaves and stem grooving [28,29]. Graft union necrosis has also been reported in certain *Malus* species [30]. Different isolates and scion-rootstock combinations seem to affect symptom development [30]. Symptoms can also result from mixed-infections, for example, in combination with 'Candidatus *Phytoplasma mali*' and *Apple chlorotic leafspot virus*, ASGV infection have displayed tree decline in apple plants grafted onto either *Malus sieboldii* or *Malus sargentii* hybrid rootstock [31].

Malus sylvestris cv. 'Virginia Crab' and *Pyronia veitchii* are used as woody indicator plants for ASGV, and develop symptoms such as a brown line at the graft union due to necrosis (Figure 2.3), leaf deformation and chlorotic leaf-spots symptoms upon graft-infection [32]. *Malus micromalus*, *Malus yunnanensis* and *Malus tschonoskii* display similar symptoms. *Malus micromalus* was suggested as a rapid woody plant indicator. In comparison with Virginia Crab, the use of *Malus micromalus* as indicator had the advantages of faster symptom development and more consistent symptom expression.

ASGV-infection is not limited to *Malus* species (Figure 2.3). In pears ASGV infection can cause feeder-root die-back and trunk grooving, leading to plant decay [11], as well as visible symptoms ranging from black necrotic leaf spots to leaves becoming entirely black [15]. Leaf symptoms were also observed in kiwifruit (Chinese gooseberries) and include interveinal mottling, chlorotic mosaics and ringspots [18]. ASGV infection has been linked to fruit intumescence in Cleopatra mandarin [33]. Other fruit trees in which ASGV infection has been detected include apricots [34], kumquat (accession no. AY646511) and Meyer lemon trees [10].

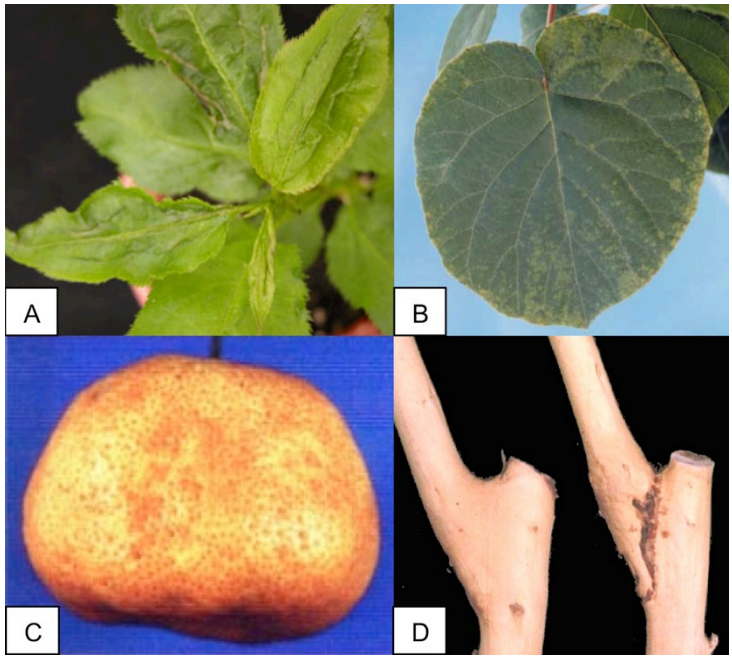


Figure 2.3. Photographs illustrating symptoms associated with ASGV infection in different plant species. A) Pear seedlings showing leaf deformation [35]. B) Kiwifruit leaf showing chlorotic mosaic symptoms [18]. C) Fruit intumescence in Cleopatra mandarin [33]. D) A healthy plant (left) as well as brown necrotic line developing at the graft union of infected (right) Virginia crab plants (http://www.dpvweb.net/dpv/showfig.php?dpvno=376&figno=01).

2.2.4 Transmission

ASGV transmission can occur by means of grafting [30,32,36,37], as well as mechanical [18] and sap inoculation [30,36,38]. Isolate specific seed transmission was suggested for *Lilium longiflorum* (lily) [39], as well as *Chenopodium quinoa* and *Malus platycarpa* [36]. The fungus *Talaromyces flavus* was shown to facilitate ASGV transmission [16]. The mechanism of this fungal-mediated transmission, however, remains to be elucidated.

2.3 RNA silencing

2.3.1 Background

Plants have developed the ability to also regulate gene expression through a process commonly known as RNA silencing. The production of small RNAs (sRNAs), approximately 17 to 26 nt in length, is required, which directs the regulation of gene expression. Based on their biogenesis and function, sRNAs have been classified into different sRNA species. sRNA synthesis can lead to transcriptional (TGS) or post-transcriptional gene silencing (PTGS), as well as translational repression [40–42]. Together these sRNA-directed pathways are involved in gene regulation of various aspects of plant development, as well as biotic and abiotic stress responses.

2.3.2 The central elements of sRNA biogenesis and function

The biogenesis of the different sRNA species each require the establishment of a double-stranded RNA (dsRNA) precursor, which can either be a dsRNA molecule or a single-stranded RNA (ssRNA) molecule with a partial double-stranded conformation. Three proteins central to sRNA pathways are; RNA-dependent RNA polymerase (RDR), Dicer-like (DCL) and Argonaute (AGO) proteins. The biogenesis and function of each sRNA species make use of a specific set of family members of these core enzymes.

Arabidopsis produces six different RDRs, three of which (RDR1, RDR2 and RDR6) are involved in RNA silencing pathways [43–51] (reviewed in Willmann *et al*. [52]). To

generate dsRNA precursors, RDRs can synthesise a complementary strand onto an ssRNA template.

Plant DCL proteins are ribonuclease III (RNase III) enzymes responsible for the cleavage of dsRNA to produce sRNA molecules [53]. In Arabidopsis DCL1 and DCL4 produce 21-nt sRNAs while DCL2 and DCL3 are associated with sRNAs 22 and 24 nt in size respectively [54–56]. The support of double-strand binding proteins (DRBs) is required for DCL-dsRNA association [57]. The double-stranded sRNA molecules processed by DCL contain 2-nt 3′ overhangs [53] and are subsequently 2′-*O* methylated by the HUA ENHANCER1 (HEN1) methyltransferase to prevent sRNA degradation [58–60].

Depending on the silencing approach, sRNAs are incorporated along with AGOs [61] into either an RNA-induced silencing complex (RISC) [62] or an RNA-induced initiation of transcriptional gene silencing (RITS) complex [63]. Directed by one strand of the sRNA duplex, known as the "guide-strand", AGOs can mediate gene regulation through chemical modification of a DNA target (TGS) [64,65], transcript cleaving (transcript degradation) (PTGS) [55,66] or translational repression [42,67,68]. AGO proteins associated with transcript cleaving are also referred to as "slicers" [66]. The three best-characterised AGO proteins in Arabidopsis are AGO1 and AGO2, predominantly associated with the cleaving of an RNA target [66,69], as well as AGO4, associated with transposon and heterochromatic DNA modification [64]. Generally, the sRNA-sequence specifies the target, through complementation, while the AGO proteins govern the mode of RNA silencing in an sRNA-target complementarity-dependent manner.

2.3.3 Endogenous sRNA species

2.3.3.1 MicroRNAs

Since their discovery in *Caenorhabditis elegans* in 1993 [70], microRNAs (miRNAs) have probably become the best-characterised sRNA species. The current miRNA registry, miRBase (release 21), contains miRNA entries for 73 different plant species, which include 207 mature apple miRNAs from 206 microRNA genes (*MIR*) [71–74].

Biogenesis of miRNAs

Plant miRNAs are predominantly expressed from non-coding *MIR* genes by means of RNA Polymerase II (Pol II) (Figure 2.4) [75,76]. The primary miRNA transcript (pri-miRNA) is 5′-capped as well as 3′-polyadenylated. The precursor molecule folds into a secondary hairpin structure and after at least two successive DCL1 cleavage reactions the mature miRNA, of ~21 nt, is liberated [58,77,78]. After methylation the mature miRNA is exported from the nucleus to the cytoplasm. While HASTY (HST) cannot be held fully accountable, this enzyme can facilitate miRNA export in plants [79]. Whether the miRNA is exported as a small dsRNA duplex or as ssRNAs also remains to be determined.

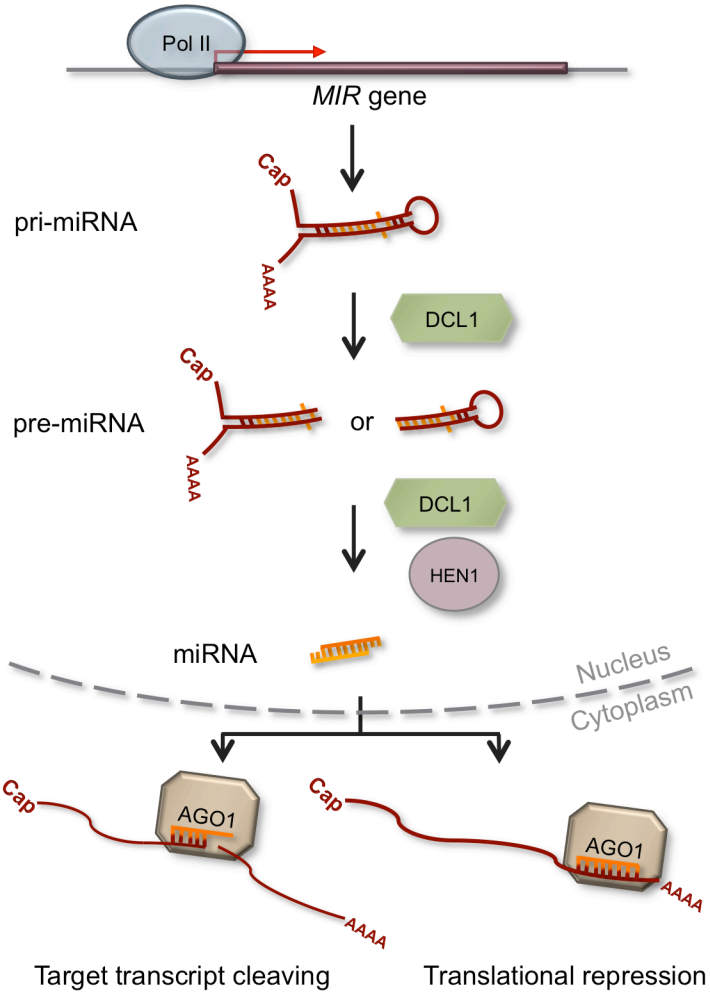


Figure 2.4. Diagram illustrating plant miRNA biogenesis and function.

Plant miRNAs mainly originate from non-coding genes. Recently, miRNAs have also been identified, which are synthesised from the exons [80] as well as introns of protein coding genes [80–84]. The latter molecules are known as mirtrons.

Function of miRNAs

In the cytoplasm the mature miRNA is associated mainly with AGO1 as part of the RISC [55,66]. miRNAs are therefore predominantly involved in PTGS by directing the cleaving of their complementary targets, leading to their subsequent degradation. Targets are cleaved between position 10 and 11 of the complementary miRNA [85,86]. A subset of miRNA targets undergoes advanced systematic degradation after their initial miRNA-guided cleavage. This phenomenon, known as phasing, is discussed under phased siRNAs below. The miRNA-AGO complex can also control the translational repression of their targets [42,67,68,87–89] and regulate target transcription through DNA methylation [65,90,91].

miRNAs play an essential regulatory role during plant development and morphogenesis [92–95], as well as in plant defence response to biotic and abiotic stresses. Differential miRNA expression levels have been reported for abiotic stresses such as; extreme heat [96,97], extreme cold [95,98,99], water deficiency [100–103], hypoxia [104], nutrient starvation [105,106] salinity [102,107,108] as well as biotic stresses such as; fungal [96,109,110] or bacterial [111–113]. Numerous reports have suggested the involvement of miRNAs in plant-virus interaction. These include studies on virus-infection in *Nicotiana benthamiana* [114], tobacco [115,116], Brassica [117], Arabidopsis [118], cotton [119], rice [120], tomato [121–125], kenaf [126], papaya [127], grapevine [128,129] and citrus [130]. Our current understanding of many miRNA-related plant stress responses is, however, mostly based on altered miRNA levels and remains to be functionally analysed. Although computational analyses have examined the possibility that plant-encoded miRNAs can play a direct role in plant defence against pathogen infection [131–133], no direct evidence exists for the natural efficacy of endogenous sRNAs against viral nucleic acids.

2.3.3.2 Phased siRNAs

Phased siRNAs (phasiRNAs) refer to the secondary sRNAs that are synthesised in precise increments from an RNA molecule that was initially cleaved by a primary sRNA [48,50].

Biogenesis of phased siRNAs

phasiRNA biogenesis is initiated by sRNA-guided cleaving of a target transcript (Figure 2.5) [50]. A complementary RNA strand is synthesised onto one of the remaining RNA fractions by RDR6, generating a long dsRNA precursor [48]. The dsRNA molecule is subsequently processed by DCL4 [134–136], in a phased manner, starting from the initial cleaved site [50]. The resulting phasiRNAs are equal in size and generally 21 nt in length, however, clusters of 24 nt long phasiRNAs have also been identified [136–138]. These 24-nt phasiRNAs are processed by DCL5 (formerly termed DCL3b) in rice [136]. Interestingly, DCL4-dependent phased sRNA synthesis from non-miRNA hairpin structured precursors has also been proposed [120].

Coding as well as non-protein coding genes can give rise to phasiRNAs. The sub-group of phasiRNAs that was first characterised is known as trans-acting siRNAs (tasiRNAs) and are synthesised from non-coding gene transcripts called trans-acting siRNA genes (*TAS*) [50]. As the name implies tasiRNAs regulate targets other than their transcript of origin. Cis-acting siRNAs have also been identified but are not well characterised [50,139–143]. Unless their cis or trans actions are proven, sRNAs generated by phased processing from a non-*TAS* homologue are referred to as phasiRNAs from *PHAS* genes [140]. Protein-coding genes that can also give rise to phasiRNAs include Auxin Signalling F-Box (*AFB*) [144,145], myeloblastosis (*MYB*) [144,146,147], nucleotide binding site Leucine rich repeat (NB-LRR) [140,145,148–150] and pentatricopeptide repeat (PPR) [145,148,151,152] protein family members.

The sRNAs, which guides the initial cleaving of the transcript, can either be a miRNA or a different sRNA species, which include phasiRNAs itself [50,139,145,152,153]. Initiator-sRNAs of specific lengths are believed to be associated with diverse phasiRNA biogenesis pathways. miRNAs and siRNAs, 22 nt in length, generally conform to a “1-hit” model where a single sRNA recognition site is present and

cleaved to initiate the phasing (also known as a "1_{22}" model, Figure 2.5) [140,153]. *TAS1*, *TAS2* and *TAS4* gene transcripts each contain a single miRNA recognition site, which is cleaved by AGO1 during phasing-initiation [154]. *TAS1* and *TAS2* transcripts contain a single miR173 recognition site [49,50], while TAS4 contains a single miR828 recognition site [155]. In all these cases the tasiRNAs are synthesised from the 3′ cleavage fragment [50]. Recent studies also illustrated the additional cleaving of *TAS1* and *TAS2* transcripts guided by a 21-nt *TAS1*-derived tasiRNA which restricts the phasing to the region between the miR173 and tasiRNA cleaved sites [141].

In contrast to the "1-hit" model a "2-hit" model exist where two sRNA recognition sites are present on the transcript (Figure 2.5) [139]. These initiator-sRNAs are 21 nt in length. *TAS3* processing is an example of a "2-hit" model (also known as a "2_{21}" model) [139,140]. The *TAS3* transcript contains two miR390 recognition sites [139]. The transcript is, however, only cleaved at the 3′ target site and the resulting tasiRNAs are synthesised from the 5′ cleavage fragment between the two miRNA recognition sites [50,139]. In contradiction to *TAS1*, *TAS2* and *TAS4* processing, *TAS3* tasiRNA biogenesis occurs in a AGO7-dependent manner [156]. Related "2-hit" models were also proposed for miRNA or tasiRNA initiated phasing in other genes, which include Auxin responsive factors (*ARF*s) and *PPR*s [139]. These proposed models included the possibility of cleavage at both sRNA recognition sites.

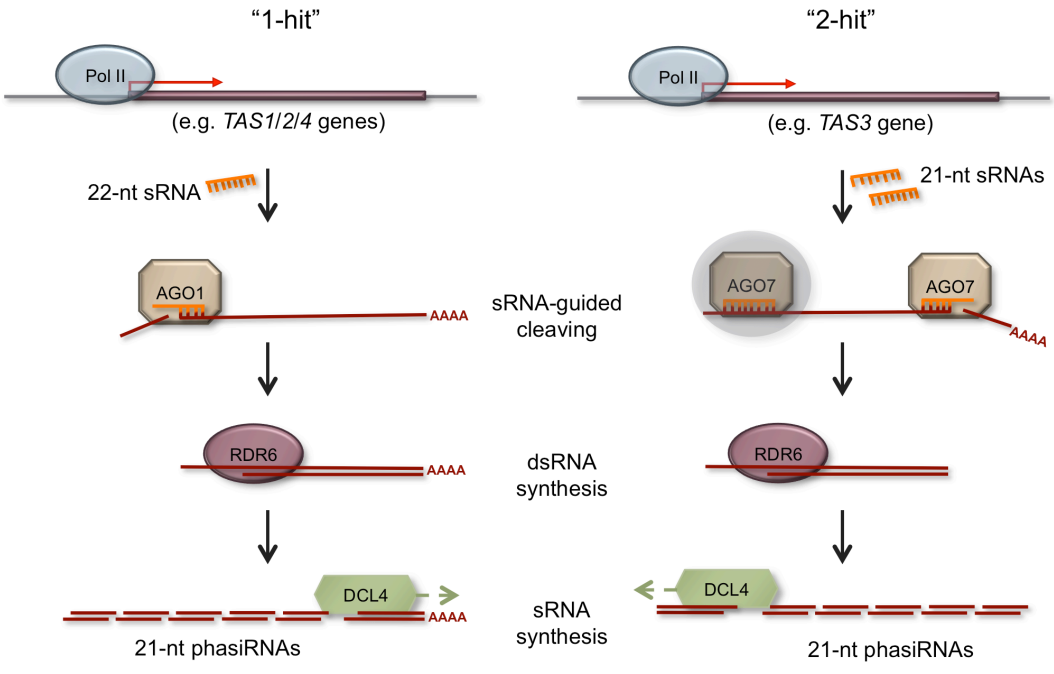


Figure 2.5. Diagram illustrating the "1-hit" and "2-hit" models for phasiRNA biogenesis.

Exceptions to the above two systems include two proposed "2_{22}" models. The first entails the 22-nt miRNA guided cleavage at only the 5′ target site of a "2-hit" system [151,157]. In the second "2_{22}" model miRNA directed cleaving occurred at both target sites with each side of the central region being in phase with the nearest cleaved site [140].

Various 22-nt miRNAs have been identified that can initiate phasing from protein and non-protein coding transcripts, as part of a "1-hit" or a "2-hit" system [137,140,149–151,153,158]. Not all 22-nt miRNAs can, however, initiate phasing, neither does 22-nt with a known phase-initiating function cause phasing at all its target sites. A recent study investigated the properties of target sites and although they were not able to completely explain phasing-site discrimination, they did highlight the importance of perfect base-pairing at the 3′ end of the miRNA [159].

Function of phased siRNAs

PhasiRNAs regulate their targets by means of cleaving [48] but can also initiate phasing in their targets, resulting in a cleavage cascade strengthening PTGS [142,143,152].

TAS1- and *TAS2*-derived tasiRNAs can target the *PPR* genes [48–50]. *TAS1*-derived tasiRNAs are involved in the regulation of genes which respond to temperature fluctuation [160]. *TAS3*-derived tasiRNAs down-regulate *ARF*s [50,161] and are therefore implicated in various aspects of plant development such as root development [162,163], leaf polarity [164,165] and lateral root growth [162,163]. The involvement of phasiRNAs in emerging inflorescences has also been demonstrated [137]. *TAS4*-derived tasiRNAs regulated the expression of MYB transcription factors [105,144,155,166].

In addition to their role in development and morphogenesis, phasiRNAs are also involved in plant stress responses. Phosphate deficiency increases the expression of miR828, as well as the miR828-regulated *TAS4*-derived tasiRNAs [105]. This process forms part of the auto-regulation of MYB transcription factors to regulate anthocyanin levels. *TAS1*, *-2* and *-3* are also influenced by hypoxia [104]. Studies have illustrated an increased synthesis or repression of phasiRNAs during bacterial and

viral challenges, which include phasiRNAs associated with pathogen response *NB-LRR* genes [114,120,149,157,167].

2.3.3.3 Natural antisense transcript siRNAs

Two complementary transcripts that can form dsRNA duplexes are considered to be a natural antisense transcript (NAT) pair. These transcripts are independently transcribed either from opposite strands of the same genomic locus or from two non-overlapping loci, known as cis- and trans-NATs respectively [168–171]. In 2003 Osato *et al*. suggested that bidirectional transcripts may also function as part of a plant's PTGS system [169]. In search of their role in PTGS, sRNAs originating from NATs termed natural antisense transcript siRNA (nat-siRNA), were identified [172].

Biogenesis of nat-siRNAs

Various pathways have been suggested for the biogenesis of plant nat-siRNAs. The first model, proposed by Borsani *et al*., involves the DCL2-dependent cleavage of a 24-nt nat-siRNA from the double stranded duplex region between a constitutively expressed transcript and an antisense transcript expressed under abiotic stress (Figure 2.6) [172]. It is anticipated that an increase in the 24-nt nat-siRNA levels involves DNA-dependent RNA polymerase IV (Pol IV) and RDR6 activity. The 24-nt nat-siRNA guides the cleaving of the constitutively expressed transcript, which sets the phase for the DCL1-dependent production of 21-nt nat-siRNAs, similar to phasiRNAs. RDR6 is required for the synthesis of both types of nat-siRNAs. The cleaving of the constitutively expressed transcript by the 24-nt nat-siRNA leads to the down-regulation thereof. Down-regulation is in turn amplified due to its processing into 21-nt nat-siRNAs, which are not only degradation products, but can also guide the cleaving of other copies of the same transcript. This proposed model was based on a cis-NAT pair.

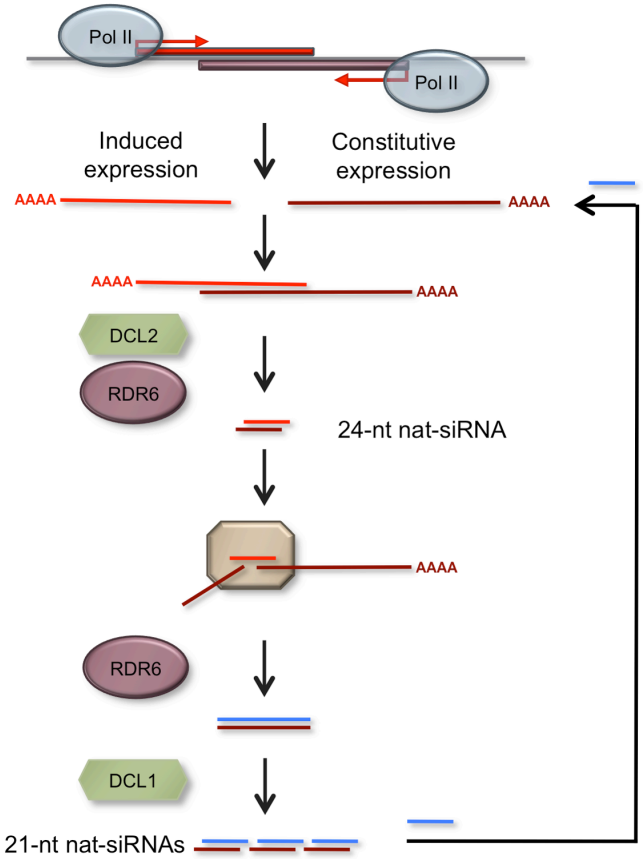


Figure 2.6. Diagram illustrating the model for nat-siRNA biogenesis and function as described by Borsani *et al.* [172].

Katiyar-Agarwal *et al*. also proposed a model in which a 22-nt nat-siRNA, generated from the overlapping region of a cis-NAT pair in a DCL1-RDR6-Pol IV dependent manner, down-regulates one of the transcripts [173]. Subsequently, the same research group discovered nat-siRNAs of 30 to 40 nt in length being synthesised from cis-NATs [174]. In addition to being DCL1, RDR6 and Pol IV dependent; their biogenesis also relied on DCL4 and AGO7. Contradicting previous predictions, Ron *et al*. demonstrated that 21-nt cis-nat-siRNA biogenesis was, along with DCL1 and Pol IV, also depended on RDR2 instead of RDR6 [175].

Another model was suggested by Zubko and Meyer [176]. In this model a group of sRNAs are produce by [DICER] from the dsRNA region of a cis-NAT duplex. One of the transcripts (the sense transcript) acts as template for the synthesis of 35-nt sRNAs by RDRs guided by the initial sRNAs, without them serving as a primer. Primed by the 35-nt sRNAs a second strand is synthesised onto the sense transcript, starting at different regions depending on the 35-nt primer. Finally, [DICER] cleaves these various dsRNA molecules into nat-siRNAs roughly 24 nt in length.

Most predicted models agree with some aspects of nat-siRNA biogenesis and function, however, the lack of a distinct model that has been agreed upon highlights the complexity of the system. A recent study by Zhang *et al*. added another level of complexity, suggesting the RDR and Pol IV independency during the biogenesis of some nat-siRNAs, as well as the DCL3-RDR2-Pol IV dependency of 24-nt nat-siRNA synthesis [177].

Function of nat-siRNAs

Our understanding of the biological role of nat-siRNAs is incomplete. While nat-siRNAs are believed to target one member of the NAT pair for degradation and therefore down-regulation, many studies have disagreed on the importance of specifically cis-nat-siRNAs. Some studies have shown that sRNA synthesis from cis-NATs is limited, and thus comparable with those produced from non-NATs [178–181]. These studies were based on genome-wide cis-NAT analyses and challenged the role of cis-nat-siRNAs as regulators of NATs even though a degree of enrichment in overlapping regions was observed when compared to the non-overlapping regions of the NAT pairs. The potential role of cis-nat-sRNAs in PTGS has been anticipated due to their enrichment in the overlapping regions, the observed bias towards one of the strands, and/or the anti-correlation of the transcript levels of the genes in the cis-NAT pairs [167,172–174,177,182,183]. Despite the lack of functional studies on trans-nat-siRNAs global analysis of their expression revealed the enrichment of their synthesis in trans-NAT and specifically to the overlapping regions op the transcript pair [179,180,183]. Along with the anti-correlation of the trans-NAT pair, bias towards one of the transcripts was also observed.

The functional role of a number of nat-siRNAs has been recognised. They can act upon biotic and abiotic stress, as well as play a role during development. Studies have shown that cis-nat-siRNAs are important during plant reproduction [175], cytokinin regulation [176] and are also projected as regulators of cell-wall synthesis [184]. Stress conditions in which the role nat-siRNAs have been anticipated include salinity [172], hypoxia [104], extreme heat conditions [185] as well as bacterial infection [167,173,174].

2.3.3.4 Repeat-associated siRNAs

The large subset of sRNAs that is synthesised from repetitive DNA sequences, such as satellite DNA and transposable elements (TE), in plant genomes is called repeat-associated siRNAs (rasiRNAs), also known as heterochromatin-associated siRNAs (hc-siRNAs). The current model for rasiRNAs biogenesis and function in RNA-dependent DNA methylation (RdDM) is described below.

Biogenesis of rasiRNAs

During the first step of rasiRNA biogenesis, Pol IV transcribes an ssRNA molecule (Figure 2.7) [186–188]. RDR2 synthesize a complementary RNA strand onto the ssRNA molecule to produce a dsRNA precursor [54,186]. This dsRNA molecule is subsequently cleaved by DCL3 to produce rasiRNAs, which are predominantly 24 nt in length.

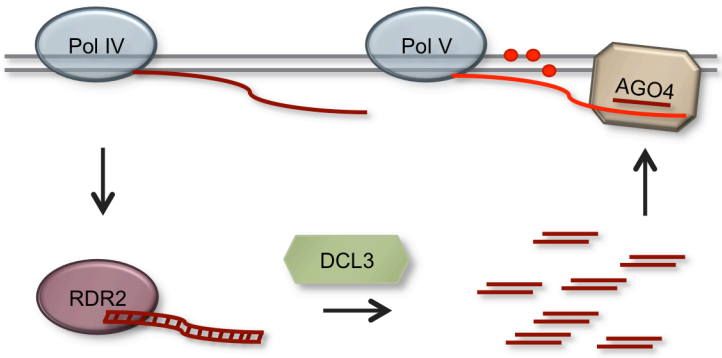


Figure 2.7. Diagram illustrating the biogenesis and function of rasiRNAs.

Function of rasiRNAs

rasiRNAs function through de novo RdDM, a concept introduced by Wassenegger and colleagues in 1994 [189]. A transcript scaffold is required to enable the recognition of a DNA target by the RITS complex, which consists of a rasiRNA-guided AGO4 [54,64,190]. DNA-dependent RNA polymerase V (Pol V) [188] produces a transcript at the target DNA loci that forms a scaffold that together with a rasiRNA-AGO4 complex directs cytosine methylation by Domains Rearranged 2 (DRM2) at the target DNA [191–193]. Consequently the Pol II or Pol III transcribed target genes are

silenced. RdDM can result in heterochromatin formation [186,194] and transposition control [195]. rasiRNAs can guide histone as well as DNA methylation [64].

Through their predicted regulation of gibberellin and brassinosteroid homeostasis, the importance of rasiRNA synthesis for optimal plant growth, leaf angle and secondary bud formation was proposed [194]. Their significance in seed development has also been suggested [196]. Stress responses where rasiRNAs are believed to play a role include heat stress [195], phosphate deficiency [105,197] and virus infection [119].

2.3.3.5 tRNA-derived sRNAs

tRNAs have long been known for their fundamental role during translation. During the past decade, however, reports have shown that tRNAs can give rise to functional sRNAs. This relative newly classified sRNA species has been well studied in mammalian systems, but its biogenesis and function in plants remain to be elucidated.

Biogenesis of tRNA-derived sRNAs

Based on their length, two main classes of tRNA-derived sRNAs were defined namely tRNA-halves and tRNA-derived RNA fragments (tRFs) (Figure 2.8). Mature tRNAs are cleaved in the anticodon region by ribonuclease enzymes and give rise to tRNA-halves (~28-36 nt) [198]. Both halves, however, are not always detected after cleaving [199]. Bacterial tRNAs are cleaved by anticodon-targeting nucleases [200], while in fungi, the RNase in Yeast 1 (Rny1) protein, a member of the RNase T2 family, cleaves the tRNA [201,202]. In humans, this function is performed by Angiogenin, which belongs to the RNase A family [203,204]. Both mature and immature tRNAs (pre-tRNA) can be processed to produce tRFs. Mature tRNAs cleaved in either the T- or the D-loop or stems respectively give rise to tRFs, which are roughly 13-26 nt in length [205]. The 3′ cleavage fragment (3′-trailer) of a pre-tRNA, released during tRNA maturation, produces an additional tRF. Evidence exists for Dicer-dependent tRF synthesis in humans [206–210]. tRNA-derived sRNAs longer than the typical tRNA-halves have additionally been described, but their synthesis remains uncharacterised [211].

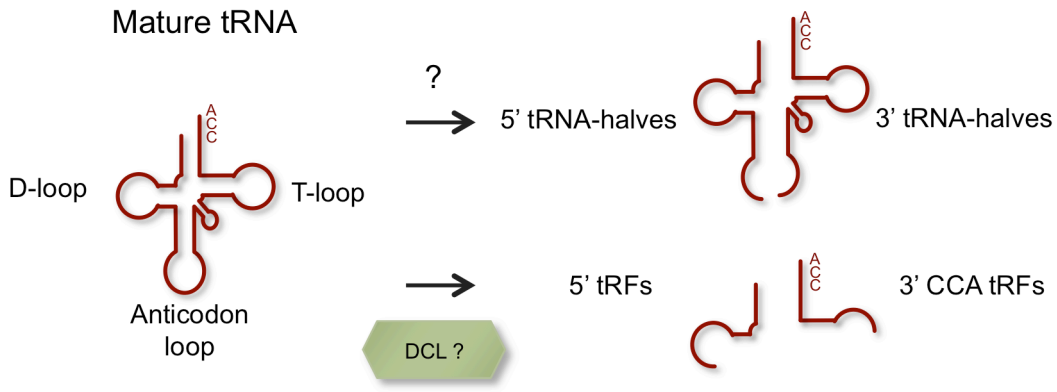


Figure 2.8. Diagram illustrating the functional sRNA species that can originate from mature tRNAs.

Function of tRNA-derived sRNAs

Whilst in bacteria more tRNA-derived sRNAs were found associated with the most frequently use tRNAs [212], Hsieh *et al*. found no such relationship between tRNA-derived sRNAs and the bias towards specific codons in Arabidopsis [105]. This corresponds to the lack of correlation seen between tRNA-derived sRNAs and codon usage in protozoa [198]. Furthermore, since there is generally no correlation between tRNA-derived sRNAs synthesis and significant tRNA reduction, it is considered that the purpose of tRNA-derived sRNAs biogenesis is not for translation inhibition by means of depletion of the tRNA pool [198,203,204,211–213].

Suggestions have been made for the possible function of tRNA-derived sRNAs. In addition to studies performed in human cells [204,214,215], two plant-based studies have presented results indicating that tRNA-derived sRNAs are involved in the inhibition of translation [199,211]. This translational repression can occur in a non sequence-specific manner. In Archaea a stress-induced 5' tRF could repress global protein synthesis by binding specifically to the small ribosomal subunit [216]. Arabidopsis tRFs were found in association with various AGO proteins, and indicate that they can potentially regulate gene expression in a way similar to other sRNAs such as miRNAs or siRNAs [217]. Similar analyses on human tRFs support the view that these sRNAs can be functionally active in an RNA silencing system [207,209,210]. Another proposed method, which remains to be investigated for plants, suggested that tRNA-derived sRNAs could direct RNase proteins, involved in tRNA processing, in order to cleave defined transcripts [218,219].

In plants, tRNA-derived sRNAs appear to play an important role especially during stress response. While the specific function of the sRNAs remains to be elucidated, the differential expression of tRNA-derived sRNAs have been shown in association with phosphate deficiency [105,197], drought [217] and oxidative stress [213], as well as bacterial challenges [217].

2.3.4 Virus-derived siRNAs

Plants have developed RNA silencing systems to protect themselves against exogenous nucleic acids such as transgenes and virus genomes. This host defence system involves the production of virus-derived siRNAs (vsiRNAs) through a process similar to the synthesis of endogenous siRNAs.

Biogenesis of vsiRNAs

There are thee types of vsiRNAs: primary, secondary and structure-related vsiRNAs (Figure 2.9). Primary vsiRNAs are processed from dsRNA precursors that resulted from viral replication or transcription. During the replication of ssRNA viruses the viral RNA-dependent RNA polymerase (RdRp) synthesises a complementary RNA strand onto the virus genome. This dsRNA intermediate molecule [220,221] is processed by DCL4 to produce 21-nt vsiRNAs [56,222,223]. In the absence of DCL4, DCL2 can act as a substitute for the processing of vsiRNAs from ssRNA viruses [45,56,222,223]. These vsiRNAs are 22 nt in length.

In the case of DNA viruses, the bidirectional transcription of their genes resulted in dsRNAs, which also can serve as precursors for primary vsiRNA synthesis [223,224]. These viral transcript-duplexes can be processed by any one of the DCL proteins [223,225]. To strengthen virus-directed RNA silencing, plant hosts could apply their own RNA-dependent RNA polymerases to produce dsRNA silencing precursors from viral ssRNAs for secondary vsiRNAs synthesis. RDR1 and RDR6 can fulfil this function [44,45,226]. The potential role of these two RDRs has also been implicated in systemic silencing [45,51,226,227], along with the importance of DCL2 and DCL4 [45,223]. Lastly, vsiRNAs can also be processed from double-stranded regions of viral genomic ssRNA secondary structures [228–230].

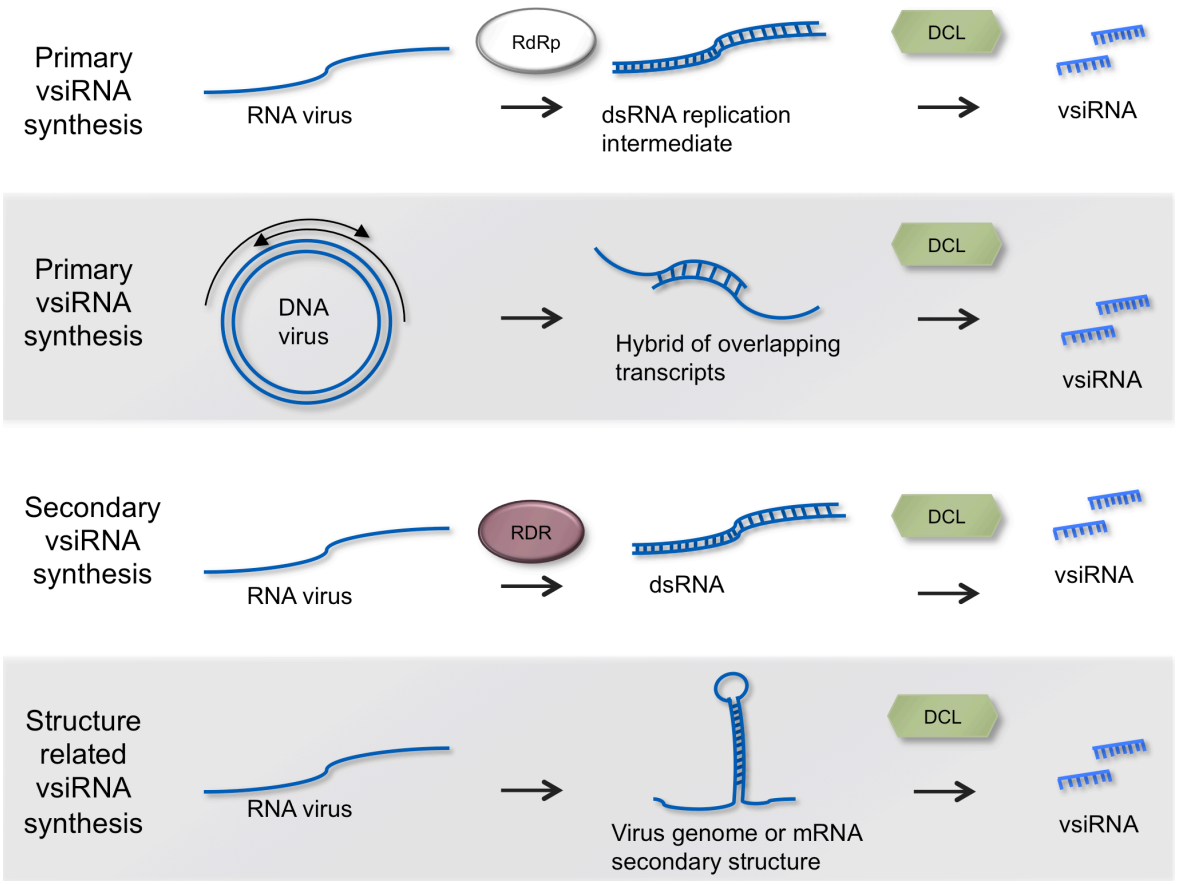


Figure 2.9. Diagram illustrating the biogenesis of primary, secondary and structure related vsiRNAs.

Function of vsiRNAs

During virus-host interaction the cleaving of viral RNAs during primary and secondary vsiRNA synthesis forms the first line of defence. This initial response is not always sufficient to provide complete resistance against the virus; additional cleaving of the viral RNA may be needed [231]. Although a number of AGO proteins might be involved in host-defence, substantial evidence exists supporting the need for AGO1 and AGO2 to mediate resistance [69,231–238]. As part of a RISC these two AGOs can target viral RNA through vsiRNA guided cleaving [231,236,237,69]. AGO1 and AGO2, guided specifically by secondary vsiRNAs, can jointly regulate antiviral resistance [231]. The inhibition of AGO1 by viral proteins can furthermore lift the miRNA-mediated suppression of AGO2, which can subsequently lead the antiviral defence [235].

Initial studies, based on *in silico* predictions and strengthened by experimental analysis, suggested that vsiRNAs from *Cauliflower mosaic virus* [225] and *Tobacco mosaic virus* [239] could target host genes for down-regulation. Subsequently, vsiRNAs derived from the Y-satellite RNA of *Cucumber mosaic virus* were shown to cleave two host genes, chlorophyll biosynthetic gene (*CHL1*) [240] and the magnesium protoporphyrin chelatase subunit I (*ChlI*) [241], leading to symptom development. *Grapevine fleck virus* and *Grapevine rupestris stem pitting-associated virus* vsiRNAs were additionally shown to potentially be responsible for cleaving numerous host transcripts [242]. The role of vsiRNAs derived from *Sugarcane mosaic virus* in translational inhibition of host genes was also suggested [238].

2.3.5 Viral suppressor of RNA silencing

Viruses have evolved counter-responses against host defence mechanisms. For a number of plant viruses these responses come in the form of proteins known as viral suppressors of RNA silencing (VSRs). These viral proteins can prevent sRNA synthesis, accumulation or activity [243]. There is no evidence to suggest the existence of a VSR produced by ASGV. For a comprehensive review on the various VSRs refer to Guo *et al*. [243].

2.4 Small RNA analysis

Methods to analyse sRNA expression were developed based on the availability (or not) of prior sequence information. Probe-based methods, which depend on prior sequence information, include Northern- [77,120,244] and *in situ* hybridisation [165,245,246]. These techniques are only sensible for use on a relatively small scale. Microarrays, on the other hand, are high-throughput hybridisation-based assays, which have been applied for miRNA expression analysis in a number of organisms [98,247,248]. In spite of their high-throughput advantage, all hybridisation-based systems are limited to sRNAs with a known sequence, and the inability to clearly discriminate between sRNAs differing from each other only at their termini [249]. Another approach to confirm and/or quantify sRNA expression is by means of reverse transcription quantitative PCR (RT-qPCR). One of the most widely used RT-PCR based strategies applies hairpin primers for reverse transcription [250–253].

These assays can be used in a multiplex reaction to increase throughput, though not always for quantitation [251,252].

Earlier studies used cloning and Sanger sequencing to identify and analyse the expression of known and novel sRNAs without any prior sequence knowledge [77,244]. Recent developments in next-generation sequencing (NGS) technologies now enable sRNA analysis to be performed on a genomic scale [254].

2.4.1 Next-generation sequencing

NGS technologies allow high-throughput sequencing of DNA in an unbiased manner and without any prior sequence information. The increased depth of coverage afforded by the multiple sequencing of millions of DNA fragments led to NGS also being referred to as deep sequencing or high-throughput sequencing. Several million sequence reads can be generated from a single run on an NGS platform. Platforms include the Illumina [255], Roche-454 [256] and ABI SOLiD [257] systems. The number of studies that apply NGS to identify and characterise sRNAs in plants is increasing significantly.

2.4.2 The bioinformatics of sRNA NGS data

Bioinformatic tools were developed to process the large amount of data generated by NGS. NGS sRNA data analysis requires an approach that first removes artefacts (adapter sequences) from library preparation, which are still present after sequencing, then filters the reads based on sequencing quality. The high-quality reads are used to identify sRNA precursors, which allow a short read to be classified as a specific sRNA species. This primary analysis is followed by sRNA target prediction and differential expression analysis between sample groups.

2.4.2.1 The basics of miRNA analysis

Since many miRNAs are widely conserved and well characterised, numerous algorithms and software have been developed for their analyses. These tools enable miRNA identification, prediction and visualisation, the comparison of miRNA expression, as well as the prediction and validation of targets. Due to the differences

in plant and animal miRNAs, algorithms were developed specifically for plant miRNA analysis. The process of identifying novel miRNAs from NGS sRNA datasets, generally comprise of at least three steps. First, reads are mapped onto a reference genome to identify sRNA reads that cluster together. Second, loci associated with these read clusters are extended to include defined flanking regions. Sequences of these extended genomic regions are subsequently extracted. Lastly, the ability of transcripts, which can potentially be expressed from these genomic loci, to fold into secondary hairpin structures, is evaluated. Such hairpin structures should have the same thermodynamic and structural properties as known pre-miRNAs. In order to strengthen *MIR* prediction, some software additionally analyse the sRNA distribution along a potential pre-miRNA [258]. Meyers *et al*. presented a set of criteria for the annotation of plant miRNAs, which is often applied by prediction tools [259].

2.4.2.2 Short-read alignment tools

Early tools developed for mapping large numbers of short reads onto reference genomes were based on storing sequence information into hash tables to facilitate sequence searches. Such tools include Cache-Assisted Hash Search with XOR digital logic (CASHX) [260], Mapping and Assembly with Quality (MAQ) [261] and Short Oligonucleotide Analysis Package (SOAP) [262]. MAQ stores the reads in a hash table while SOAPv1 and CASHX hashes the reference genome. CHASHX allows only precise alignments of the reads with the genome.

More recently the use of a hash table was replaced by the implementation of Burrows-Wheeler transform (BWT) based algorithms [263] for reference genome indexing. Bowtie [264] and a descendant of MAQ, Burrows-Wheeler Alignment tool (BWA) [265], are two of the most popular alignment tools that make use BWT. While BWA can account for indels, Bowtie is unable to do so. Improving on its predecessor, SOAPv1, another short read aligner SOAP2 applies a two-way-BWT (2BWT) algorithm for indexing [266]. In comparison to their predecessors, BWT-based alignment tools are optimised for high-speed alignment of large numbers of short sequence reads while utilising very little memory. The performance of the different alignment tools varies depending on the read length as well as the number of mismatches allowed [265–267].

SOAP3 is the latest addition to the SOAP alignment family [267]. In addition to a multi-core central processing unit (CPU), this program requires a graphics processing unit (GPU) and applies an adapted indexing algorithm namely GPU-2BWT. The application of a GPU allows for major parallelism, leading to a drastic decrease the running time.

2.4.2.3 Tools for plant miRNA prediction

A number of bioinformatic tools have been developed for the identification of known and novel miRNAs. Popular prediction tools for analysing NGS sRNA data include mirDeep-P, ShortStack and miREAP. miRDeep-P is the plant-specific version of miRDeep [268,269]. It consists of a series of Pearl scripts that orchestrate read-mapping, secondary structure prediction and miRNA characteristic-specific filtering. Bowtie performs reference alignment, while secondary structure prediction of a potential pre-miRNA is performed using the Vienna RNA package [270]. To retain plant miRNA-specific predictions, the plant-specific criteria as described by Meyers *et al*. [259], and criteria for miRNA-specific read distribution (from the core miRDeep algorithm) are used. ShortStack [258] applies similar steps for miRNA predictions. It offers the user more control over the choice of reference-alignment tool and allows users to set more parameters manually. The criteria set by ShortStack for miRNA prediction are focussed on specificity, which, to some degree, come at the cost of sensitivity. In spite of requiring substantial pre- and post-processing, miREAP (http://sourceforge.net/projects/mireap/), developed by the Beijing Genomic Institute, is widely cited in literature. Users have to apply their own filtering steps in order to select miRNAs with the desired read-distribution and valid stem-loop structures, from the list of predicted structures.

Some miRNA prediction tools make use of a machine-learning approach. These predictions are strongly dependent on the training data used. Machine-learning algorithms include support vector machine [271], hidden Markov model [272] or naïve Bayes classifier [273]. miRanalyzer, for example, follows a machine-learning approach that can be used for plant miRNA predictions [274,275]. It applies five random forest models [276] that were built on pre-miRNA information from four plant species (*Arabidopsis thaliana*, *Vitis vinifera*, *Oryza sativa* and *Zea mays*).

2.4.2.4 miRNA target validation

Degradome sequencing is a high-throughput method for miRNA target validation. Short-read libraries, which represent the 5′ ends of the 3′ cleaved fragments of miRNA targets (known as a degradome) can be generated [277–279]. This sequence data can be used to validate *in silico* predicted miRNA targets. CleaveLand [280], SeqTar [281] and PAREsnip [282] are examples of bioinformatic tools for the analysis of user-defined degradome sequencing data. All three these software tools make use of a complementarity-based scoring system. While CleaveLand is strict at the expense of potentially generating more false-negatives, SeqTar allows more miss-matches but apply two statistical analyses in scoring the strength of the prediction [281].

2.4.2.5 Tools for non-miRNA sRNA species analysis

Despite the fact that miRNAs are the best-studied sRNA species, other species such as phasiRNAs, nat-siRNA, rasiRNAs and vsiRNAs have also been extensively studied. To assist in non-miRNA sRNA analysis, tools such as SearchSmallRNA [283] and the R package, viRome [284] for vsiRNA analysis, as well as, the UEA sRNA Workbench [285], pssRNAMiner [286], TasExpAnalysis [287] and ShortStack [258] for phasiRNA prediction were developed. Software for the focused analysis of more recently described sRNA groups, such as tRNA-derived RNA fragments remain to be developed. These sRNAs can, nonetheless, often be analysed by implementing short-read alignment tools and additional in-house scripts.

2.4.2.6 Differential expression analysis using NGS data

One of the ultimate goals of sRNA analysis is to determine whether the identified sRNAs have biological function. Investigating the variation in the expression patterns of specific sample groups may provide evidence for biological function. Software were developed, which apply different statistical packages to determine the significance of variation in sRNA expression, based on the NGS data. The Bioconductor package (of the R software) provides a number of individual packages applicable to NGS differential expression analysis. The two most frequently used Bioconductor packages are DESeq [288] and edgeR [289], and were designed to make used of count

data generated through NGS, such as RNA-Seq, or ChipSeq directed experiments. Both these algorithms are based on negative binomial distribution and assume that the majority of the genes show constant expression but differ in the way they normalise read counts. To perform normalisation DESeq makes used of an estimate size factor generated for each sample [288]. This size factor is the median of all the gene-specific ratios for that sample. A gene-specific ratio for a sample is calculated by dividing its read count by the geometric mean of that gene's read counts across all samples. Normalisation in the edgeR package can be performed in three ways, one similar as within DESeq and two additional ways making use of the weighted trimmed mean of M-values method [290] or an upper-quartile normalisation method [291]. edgeR also gives the option of no-normalisation (a normalisation factor of one). Since the last DESeq versions were considered to be to overly strict in their analysis a new advanced version, DESeq2, was recently developed that provide a better balance between being sensitive enough and preventing false positive [292]. In addition to using empirical Bayes shrinkage for estimation of dispersions, DESeq2 also applies it for logarithmic fold change estimation. This is done partly to compensate for the higher degree a variance seen in cases where genes have a low number of reads associated with it.

2.4.2.7 Genome assembly

An additional application for NGS sRNA data in plant analysis is the identification of infectious agents. Virus genomes can be reconstructed by using software for de novo or reference-based assembly. Velvet is probably the most widely used de novo short read assembler [293]. It makes use of de Bruijn graphs [294,295] to represent the overlaps between short reads. Other de novo short read assemblers that also make use of de Bruijn graphs include SOAPdenovo(2) [296,297], SSAKE [298] and VCAKE [299]. The recently described tool, SearchSmallRNA, was developed for the specific purpose of assembling virus genomes from sRNA sequence data with the help of a reference genome [283]. The software provides general information about the vsiRNAs and additionally provides a graphical representation of the genomic distribution of the vsiRNAs.

2.4.2.8 Final comments on sRNA NGS data analysis

NGS sRNA data analysis can be complex and computationally demanding. User-friendly packages such as the UEA sRNA workbench [285] have been developed to provide a comprehensive set of tools for various aspects of sRNA analysis. Most of the commonly used bioinformatics software are available as a standalone version for different platforms (e.g. Linux, OS X or Windows). These packages are generally command-line driven though some do have a graphical user interface (GUI). Web-based tools provide a platform to run processing and memory intensive software in the absence of a local server. They are, however, sometimes restricted to specific preloaded plant species, databases, datasets or pre-set parameters. They also offer the end-user no control over the version of backend software, for example the version of the R package used for statistical analysis.

2.5 Conclusion

Plants have developed various mechanisms to control gene expression, one of which is regulation by means of functional small RNAs. Investigation of the different sRNA species can therefore provide valuable insight into genes involved in specific developmental stages or during stress-response. This study focussed on applying bioinformatics tools to analyse NGS small RNA data in order to gain a better understanding of the processes involved in virus infection in apple.

References

1. Nickel O, Fajardo TV, Jelkmann W, Kuhn GB (2001) Sequence analysis of the capsid protein gene of an isolate of *Apple stem grooving virus*, and its survey in Southern Brazil. Fitopatol Bras 26: 655–659.
2. Lolić B, Matić S, ĐJurić G, Hassan M, Di Serio F, *et al*. (2010) Pome fruit viruses in Bosnia and Herzegovina. Julius-Kühn-Arch 427: 245–247.
3. Liu P, Zhang L, Zhang H, Jiao H, Wu Y (2013) Detection and molecular variability of *Apple stem grooving virus* in Shaanxi, China. J Phytopathol 161: 445–449.
4. Youssef SA, Moawad SM, Nosseir FM, Shalaby AA (2010) Detection and identification of *Apple stem pitting virus* and *Apple stem grooving virus* affecting apple and pear trees in Egypt. Julius-Kühn-Arch 427: 248–2452.
5. Menzel W, Jelkmann W, Maiss E (2002) Detection of four apple viruses by multiplex RT-PCR assays with coamplification of plant mRNA as internal control. J Virol Methods 99: 81–92.
6. Kumar S, Singh RM, Ram R, Badyal J, Hallan V, *et al*. (2012) Determination of major viral and sub viral pathogens incidence in apple orchards in Himachal Pradesh. Indian J Virol 23: 75–79.

7. Ito T, Namba N, Ito T (2003) Distribution of citrus viroids and *Apple stem grooving virus* on citrus trees in Japan using multiplex reverse transcription polymerase chain reaction. J Gen Plant Pathol 69: 205–207.

8. Pearson MN, Clover GRG, Guy PL, Fletcher JD, Beever RE (2006) A review of the plant virus, viroid and mollicute records for New Zealand. Australas Plant Pathol 35: 217–252.

9. Birişik N, Hassan M, Baloğlu S (2008) A preliminary account on apple viruses in Mediterranean region of Turkey. Acta Hortic 781: 125–130.

10. Garnsey S (1964) Detection of tatter leaf virus of citrus in Florida. Proc Fla State Hortic Soc 77: 106–109.

11. Siebert VZ, Engelbrecht DJ (1982) Association of *Apple stem grooving virus* with a delcine of Packham's Triumph pear on seedling rootstock. Acta Hort 130: 47–52.

12. Adams, MJ, Candresse T, Hammond J, Kreuze JF, Martelli, et al. (2011) *Betaflexiviridae*. In: King AMQ, Adams MJ, Carstens EB, Lefkowitz EJ, editors. Virus Taxonomy: Ninth Report of the International Committee on Taxonomy of Viruses. Elsevier. pp. 920–941.

13. Yoshikawa N, Imaizumi M, Takahashi T, Inouye N (1993) Striking similarities between the nucleotide sequence and genome organization of *Citrus tatter leaf* and *Apple stem grooving capilloviruses*. J Gen Virol 74 (Pt 12): 2743–2747.

14. Tatineni S, Afunian MR, Hilf ME, Gowda S, Dawson WO, *et al*. (2009) Molecular characterization of *Citrus tatter leaf virus* historically associated with Meyer Lemon trees: Complete genome sequence and development of biologically active *in vitro* transcripts. Phytopathology 99: 423–431.

15. Shim H, Min Y, Hong S, Kwon M, Kim D, *et al*. (2004) Nucleotide sequences of a Korean isolate of *Apple stem grooving virus* associated with black necrotic leaf spot disease on pear (Pyrus pyrifolia). Mol Cells 18: 192–199.

16. Shim HK, Hwang KH, Shim CK, Son SW, Kim DG, *et al*. (2006) The pear black necrotic leaf spot disease transmitted by *Talaromyces flavus* displays pathogenicity similar to *Apple stem grooving virus* strains. Plant Pathol J 22: 255–259.

17. Yoshikawa N (2000) *Apple stem grooving virus*. CMI/AAB Description of plant viruses No. 376.

18. Clover GRG, Pearson MN, Elliott DR, Tang Z, Smales TE, *et al*. (2003) Characterization of a strain of *Apple stem grooving virus* in *Actinidia chinensis* from China. Plant Pathol 52: 371–378.

19. Yoshikawa N, Takahashi T (1988) Properties of RNAs and proteins of *Apple stem grooving* and *Apple chlorotic leaf spot viruses*. J Gen Virol - J Gen Virol 69: 241–245.

20. Yoshikawa N, Sasaki E, Kato M, Takahashi T (1992) The nucleotide sequence of *Apple stem grooving capillovirus* genome. Virology 191: 98–105.

21. Ohira K, Namba S, Rozanov M, Kusumi T, Tsuchizaki T (1995) Complete sequence of an infectious full-length cDNA clone of *Citrus tatter leaf capillovirus*: comparative sequence analysis of capillovirus genomes. J Gen Virol 76 (Pt 9): 2305–2309.

22. Magome H, Yoshikawa N, Takahashi T, Ito T, Miyakawa T (1997) Molecular variability of the genomes of capilloviruses from apple, Japanese pear, European pear, and citrus trees. Phytopathology 87: 389–396.

23. Liebenberg A, Moury B, Sabath N, Hell R, Kappis A, *et al*. (2012) Molecular evolution of the genomic RNA of *Apple stem grooving capillovirus*. J Mol Evol 75: 92–101.

24. Hirata H, Lu X, Yamaji Y, Kagiwada S, Ugaki M, *et al*. (2003) A single silent substitution in the genome of *Apple stem grooving virus* causes symptom attenuation. J Gen Virol 84: 2579–2583.

25. Tatineni S, Afunian MR, Gowda S, Hilf ME, Bar-Joseph M, *et al*. (2009) Characterization of the 5′- and 3′-terminal subgenomic RNAs produced by a capillovirus: Evidence for a CP subgenomic RNA. Virology 385: 521–528.

26. Hirata H, Yamaji Y, Komatsu K, Kagiwada S, Oshima K, *et al*. (2010) Pseudo-polyprotein translated from the full-length ORF1 of capillovirus is important for pathogenicity, but a

truncated ORF1 protein without variable and CP regions is sufficient for replication. Virus Res 152: 1–9.

27. Komatsu K, Hirata H, Fukagawa T, Yamaji Y, Okano Y, *et al*. (2012) Infection of capilloviruses requires subgenomic RNAs whose transcription is controlled by promoter-like sequences conserved among flexiviruses. Virus Res 167: 8–15.

28. Maxim A, Zagrai L, Zagrai I, Isac M (2004) Studies on the influence of *Apple stem grooving virus* on tree growth of various apple cultivars in the nursery. Acta Hort 657: 41–44.

29. Birişik N, Baloğlu S (2010) Determination of the effects of *Apple stem grooving virus* on some commercial apple cultivars. Julius-Kühn-Arch 427: 240–244.

30. Yanase H (1982) Back transmission of *Apple stem grooving virus* to apple seedlings and induction of symptoms of apple topworking disease in Mitsuba Kaido (*Malus sieboldii*) and Kobano Zumi (*Malus sieboldii* var. arborescens) rootstocks. Acta Hortic 130: 117–122.

31. Seemüller E, Moll E, Schneider B (2008) Apple proliferation resistance of *Malus sieboldii*-based rootstocks in comparison to rootstocks derived from other *Malus* species. Eur J Plant Pathol 121: 109–119.

32. Howell WE, Mink GI, Hurtt SS, Foster JA, Postman JD (1996) Select *Malus* clones for rapid detection of *Apple stem grooving virus*. Plant Dis 80: 1200–1202.

33. Lovisolo O, Accotto GP, Masenga V, Colariccio A (2003) An isolate of *Apple stem grooving virus* associated with Cleopatra mandarin fruit intumescence. Fitopatol Bras 28: 54–58.

34. Takahashi T, Saito N, Goto M, Kawai A, Namba S, *et al*. (1990) *Apple stem grooving virus* isolated from Japanese apricot (*Prunus mume*) imported from China. Res Bull Plant Prot Serv Jpn: 15–21.

35. Wu Z-B, Ku H-M, Su C-C, Chen I-Z, Jan F-J (2010) Molecular and biological characterization of an isolate of *Apple stem pitting virus* causing pear vein yellows disease in Taiwan. J Plant Pathol 92: 721–728.

36. Van der Meer FA (1976) Observations on *Apple stem grooving virus*. Acta Hortic 67: 293–304.

37. Kirby MJ, Guise CM, Adams AN (2001) Comparison of bioassays and laboratory assays for *Apple stem grooving virus*. J Virol Methods 93: 167–173.

38. James D (1999) A simple and reliable protocol for the detection of *Apple stem grooving* virus by RT–PCR and in a multiplex PCR assay. J Virol Methods 83: 1–9.

39. Inouye N, Maeda T, Mitsuhata K (1979) *Citrus tatter leaf virus* isolated from lily. Ann Phytopathol Soc Jpn 45: 712–720.

40. Hamilton AJ, Baulcombe DC (1999) A species of small antisense RNA in posttranscriptional gene silencing in plants. Science 286: 950–952.

41. Mette MF, Aufsatz W, van der Winden J, Matzke MA, Matzke AJ (2000) Transcriptional silencing and promoter methylation triggered by double-stranded RNA. EMBO J 19: 5194–5201.

42. Brodersen P, Sakvarelidze-Achard L, Bruun-Rasmussen M, Dunoyer P, Yamamoto YY, *et al*. (2008) Widespread translational inhibition by plant miRNAs and siRNAs. Science 320: 1185–1190.

43. Donaire L, Barajas D, Martínez-García B, Martínez-Priego L, Pagán I, *et al*. (2008) Structural and genetic requirements for the biogenesis of *Tobacco rattle virus*-derived small interfering RNAs. J Virol 82: 5167–5177.

44. Wang X-B, Wu Q, Ito T, Cillo F, Li W-X, *et al*. (2010) RNAi-mediated viral immunity requires amplification of virus-derived siRNAs in *Arabidopsis thaliana*. Proc Natl Acad Sci U S A 107: 484–489.

45. Garcia-Ruiz H, Takeda A, Chapman EJ, Sullivan CM, Fahlgren N, *et al*. (2010) Arabidopsis RNA-dependent RNA polymerases and Dicer-like proteins in antiviral defense and small interfering RNA biogenesis during *Turnip mosaic virus* infection. Plant Cell Online 22: 481–496.

46. Kasschau KD, Fahlgren N, Chapman EJ, Sullivan CM, Cumbie JS, *et al*. (2007) Genome-wide profiling and analysis of Arabidopsis siRNAs. PLoS Biol 5: e57.

47. Peragine A, Yoshikawa M, Wu G, Albrecht HL, Poethig RS (2004) SGS3 and SGS2/SDE1/RDR6 are required for juvenile development and the production of trans-acting siRNAs in Arabidopsis. Genes Dev 18: 2368–2379.

48. Vazquez F, Vaucheret H, Rajagopalan R, Lepers C, Gasciolli V, *et al*. (2004) Endogenous trans-acting siRNAs regulate the accumulation of Arabidopsis mRNAs. Mol Cell 16: 69–79.

49. Yoshikawa M, Peragine A, Park MY, Poethig RS (2005) A pathway for the biogenesis of trans-acting siRNAs in Arabidopsis. Genes Dev 19: 2164–2175.

50. Allen E, Xie Z, Gustafson AM, Carrington JC (2005) microRNA-directed phasing during trans-acting siRNA biogenesis in plants. Cell 121: 207–221.

51. Qu F, Ye X, Hou G, Sato S, Clemente TE, *et al*. (2005) RDR6 has a broad-spectrum but temperature-dependent antiviral defense role in *Nicotiana benthamiana*. J Virol 79: 15209–15217.

52. Willmann MR, Endres MW, Cook RT, Gregory BD (2011) The Functions of RNA-dependent RNA polymerases in Arabidopsis. Arab Book 9: e0146.

53. Tang G, Reinhart BJ, Bartel DP, Zamore PD (2003) A biochemical framework for RNA silencing in plants. Genes Dev 17: 49–63.

54. Xie Z, Johansen LK, Gustafson AM, Kasschau KD, Lellis AD, *et al*. (2004) Genetic and functional diversification of small RNA pathways in plants. PLoS Biol 2: e104.

55. Qi Y, Denli AM, Hannon GJ (2005) Biochemical specialization within Arabidopsis RNA silencing pathways. Mol Cell 19: 421–428.

56. Deleris A, Gallego-Bartolome J, Bao J, Kasschau KD, Carrington JC, *et al*. (2006) Hierarchical action and inhibition of plant Dicer-like proteins in antiviral defense. Science 313: 68–71.

57. Hiraguri A, Itoh R, Kondo N, Nomura Y, Aizawa D, *et al*. (2005) Specific interactions between Dicer-like proteins and HYL1/DRB-family dsRNA-binding proteins in *Arabidopsis thaliana*. Plant Mol Biol 57: 173–188.

58. Park W, Li J, Song R, Messing J, Chen X (2002) CARPEL FACTORY, a Dicer homolog, and HEN1, a novel protein, act in microRNA metabolism in *Arabidopsis thaliana*. Curr Biol CB 12: 1484–1495.

59. Yu B, Yang Z, Li J, Minakhina S, Yang M, *et al*. (2005) Methylation as a crucial step in plant microRNA biogenesis. Science 307: 932–935.

60. Li J, Yang Z, Yu B, Liu J, Chen X (2005) Methylation protects miRNAs and siRNAs from a 3'-end uridylation activity in Arabidopsis. Curr Biol CB 15: 1501–1507.

61. Hammond SM, Boettcher S, Caudy AA, Kobayashi R, Hannon GJ (2001) Argonaute2, a link between genetic and biochemical analyses of RNAi. Science 293: 1146–1150.

62. Hammond SM, Bernstein E, Beach D, Hannon GJ (2000) An RNA-directed nuclease mediates post-transcriptional gene silencing in Drosophila cells. Nature 404: 293–296.

63. Verdel A, Jia S, Gerber S, Sugiyama T, Gygi S, *et al*. (2004) RNAi-mediated targeting of heterochromatin by the RITS complex. Science 303: 672–676.

64. Zilberman D, Cao X, Jacobsen SE (2003) ARGONAUTE4 Control of locus-specific siRNA accumulation and DNA and histone methylation. Science 299: 716–719.

65. Wu L, Zhou H, Zhang Q, Zhang J, Ni F, *et al*. (2010) DNA methylation mediated by a microRNA pathway. Mol Cell 38: 465–475.

66. Baumberger N, Baulcombe DC (2005) Arabidopsis ARGONAUTE1 is an RNA Slicer that selectively recruits microRNAs and short interfering RNAs. Proc Natl Acad Sci 102: 11928–11933.

67. Lanet E, Delannoy E, Sormani R, Floris M, Brodersen P, *et al*. (2009) Biochemical evidence for translational repression by Arabidopsis microRNAs. PLANT CELL ONLINE 21: 1762–1768.

68. Li S, Liu L, Zhuang X, Yu Y, Liu X, *et al*. (2013) MicroRNAs inhibit the translation of target mRNAs on the endoplasmic reticulum in Arabidopsis. Cell 153: 562–574.

69. Schuck J, Gursinsky T, Pantaleo V, Burgyán J, Behrens S-E (2013) AGO/RISC-mediated antiviral RNA silencing in a plant in vitro system. Nucleic Acids Res 41: 5090–5103.

70. Lee RC, Feinbaum RL, Ambros V (1993) The *C. elegans* heterochronic gene lin-4 encodes small RNAs with antisense complementarity to lin-14. Cell 75: 843–854.

71. Griffiths-Jones S, Grocock RJ, Dongen S van, Bateman A, Enright AJ (2006) miRBase: microRNA sequences, targets and gene nomenclature. Nucleic Acids Res 34: D140–D144.

72. Griffiths-Jones S, Saini HK, Dongen S van, Enright AJ (2008) miRBase: tools for microRNA genomics. Nucleic Acids Res 36: D154–D158.

73. Kozomara A, Griffiths-Jones S (2011) miRBase: integrating microRNA annotation and deep-sequencing data. Nucleic Acids Res 39: D152–D157.

74. Kozomara A, Griffiths-Jones S (2013) miRBase: annotating high confidence microRNAs using deep sequencing data. Nucleic Acids Res 42: D68–D73.

75. Lee Y, Kim M, Han J, Yeom K-H, Lee S, *et al*. (2004) MicroRNA genes are transcribed by RNA polymerase II. EMBO J 23: 4051–4060.

76. Xie Z, Allen E, Fahlgren N, Calamar A, Givan SA, *et al*. (2005) Expression of Arabidopsis *MIRNA* Genes. PLANT Physiol 138: 2145–2154.

77. Reinhart BJ, Weinstein EG, Rhoades MW, Bartel B, Bartel DP (2002) MicroRNAs in plants. Genes Dev 16: 1616–1626.

78. 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.

79. Park MY, Wu G, Gonzalez-Sulser A, Vaucheret H, Poethig RS (2005) Nuclear processing and export of microRNAs in Arabidopsis. Proc Natl Acad Sci U S A 102: 3691–3696.

80. 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.

81. Zhu Q-H, Spriggs A, Matthew L, Fan L, Kennedy G, *et al*. (2008) A diverse set of microRNAs and microRNA-like small RNAs in developing rice grains. Genome Res 18: 1456–1465.

82. Joshi PK, Gupta D, Nandal UK, Khan Y, Mukherjee SK, *et al*. (2012) Identification of mirtrons in rice using MirtronPred: A tool for predicting plant mirtrons. Genomics 99: 370–375.

83. Yang GD, Yan K, Wu BJ, Wang YH, Gao YX, *et al*. (2012) Genomewide analysis of intronic microRNAs in rice and Arabidopsis. J Genet 91: 313–324.

84. Meng Y, Shao C (2012) Large-scale identification of mirtrons in Arabidopsis and rice. PLoS ONE 7: e31163.

85. Elbashir SM, Martinez J, Patkaniowska A, Lendeckel W, Tuschl T (2001) Functional anatomy of siRNAs for mediating efficient RNAi in *Drosophila melanogaster* embryo lysate. EMBO J 20: 6877–6888.

86. Llave C, Xie Z, Kasschau KD, Carrington JC (2002) Cleavage of scarecrow-like mRNA targets directed by a class of Arabidopsis miRNA. Science 297: 2053–2056.

87. Aukerman MJ, Sakai H (2003) Regulation of flowering time and floral organ identity by a microRNA and its APETALA2-like target genes. PLANT CELL ONLINE 15: 2730–2741.

88. Chen X (2004) A microRNA as a translational repressor of APETALA2 in Arabidopsis flower development. Science 303: 2022–2025.

89. Gandikota M, Birkenbihl RP, Höhmann S, Cardon GH, Saedler H, *et al*. (2007) The miRNA156/157 recognition element in the 3' UTR of the Arabidopsis SBP box gene *SPL3* prevents early flowering by translational inhibition in seedlings. Plant J Cell Mol Biol 49: 683–693.

90. Bao N, Lye K-W, Barton MK (2004) MicroRNA binding sites in Arabidopsis class III HD-ZIP mRNAs are required for methylation of the template chromosome. Dev Cell 7: 653–662.

91. Khraiwesh B, Arif MA, Seumel GI, Ossowski S, Weigel D, *et al*. (2010) Transcriptional control of gene expression by microRNAs. Cell 140: 111–122.

92. Palatnik JF, Allen E, Wu X, Schommer C, Schwab R, *et al*. (2003) Control of leaf morphogenesis by microRNAs. Nature 425: 257–263.

93. Baker CC, Sieber P, Wellmer F, Meyerowitz EM (2005) The early extra petals1 mutant uncovers a role for microRNA miR164c in regulating petal number in Arabidopsis. Curr Biol CB 15: 303–315.

94. Mallory AC, Bartel DP, Bartel B (2005) MicroRNA-directed regulation of Arabidopsis AUXIN RESPONSE FACTOR17 is essential for proper development and modulates expression of early auxin response genes. PLANT CELL ONLINE 17: 1360–1375.

95. Yang C, Li D, Mao D, Liu X, Ji C, *et al*. (2013) Overexpression of microRNA319 impacts leaf morphogenesis and leads to enhanced cold tolerance in rice (Oryza sativa L.). Plant Cell Environ 36: 2207–2218.

96. Xin M, Wang Y, Yao Y, Xie C, Peng H, *et al*. (2010) Diverse set of microRNAs are responsive to powdery mildew infection and heat stress in wheat (*Triticum aestivum* L.). BMC Plant Biol 10: 123.

97. Guan Q, Lu X, Zeng H, Zhang Y, Zhu J (2013) Heat stress induction of miR398 triggers a regulatory loop that is critical for thermotolerance in Arabidopsis. Plant J Cell Mol Biol 74: 840–851.

98. Lv D-K, Bai X, Li Y, Ding X-D, Ge Y, *et al*. (2010) Profiling of cold-stress-responsive miRNAs in rice by microarrays. Gene 459: 39–47.

99. Zhang X-N, Li X, Liu J-H (2014) Identification of conserved and novel cold-responsive microRNAs in trifoliate orange (*Poncirus trifoliata* (L.) Raf.) using high-throughput sequencing. Plant Mol Biol Report 32: 328–341.

100. Li W-X, Oono Y, Zhu J, He X-J, Wu J-M, *et al*. (2008) The Arabidopsis NFYA5 transcription factor is regulated transcriptionally and posttranscriptionally to promote drought resistance. Plant Cell Online 20: 2238–2251.

101. Bertolini E, Verelst W, Horner DS, Gianfranceschi L, Piccolo V, *et al*. (2013) Addressing the role of microRNAs in reprogramming leaf growth during drought stress in *Brachypodium distachyon*. Mol Plant 6: 423–443.

102. Zhou M, Li D, Li Z, Hu Q, Yang C, *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.

103. Ghorecha V, Patel K, Ingle S, Sunkar R, Krishnayya NSR (2014) Analysis of biochemical variations and microRNA expression in wild (*Ipomoea campanulata*) and cultivated (*Jacquemontia pentantha*) species exposed to *in vivo* water stress. Physiol Mol Biol Plants 20: 57–67.

104. Moldovan D, Spriggs A, Yang J, Pogson BJ, Dennis ES, *et al*. (2010) Hypoxia-responsive microRNAs and trans-acting small interfering RNAs in Arabidopsis. J Exp Bot 61: 165–177.

105. Hsieh L-C, Lin S-I, Shih AC-C, Chen J-W, Lin W-Y, *et al*. (2009) Uncovering small RNA-mediated responses to phosphate deficiency in Arabidopsis by deep sequencing. PLANT Physiol 151: 2120–2132.

106. Zhao M, Ding H, Zhu J-K, Zhang F, Li W-X (2011) Involvement of miR169 in the nitrogen-starvation responses in Arabidopsis. New Phytol 190: 906–915.

107. 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.

108. 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.

109. Chen L, Ren Y, Zhang Y, Xu J, Zhang Z, *et al*. (2012) Genome-wide profiling of novel and conserved Populus microRNAs involved in pathogen stress response by deep sequencing. Planta 235: 873–883.

110. Li Y, Lu Y-G, Shi Y, Wu L, Xu Y-J, *et al*. (2014) Multiple rice microRNAs are involved in immunity against the blast fungus *Magnaporthe oryzae*. Plant Physiol 164: 1077–1092.

111. Navarro L, Dunoyer P, Jay F, Arnold B, Dharmasiri N, *et al*. (2006) A plant miRNA contributes to antibacterial resistance by repressing auxin signaling. Science 312: 436–439.

112. Jagadeeswaran G, Saini A, Sunkar R (2009) Biotic and abiotic stress down-regulate miR398 expression in Arabidopsis. Planta 229: 1009–1014.

113. Zhang W, Gao S, Zhou X, Chellappan P, Chen Z, *et al*. (2011) Bacteria-responsive microRNAs regulate plant innate immunity by modulating plant hormone networks. Plant Mol Biol 75: 93–105.

114. Xiao B, Yang X, Ye C-Y, Liu Y, Yan C, *et al*. (2014) A diverse set of miRNAs responsive to begomovirus-associated betasatellite in *Nicotiana benthamiana*. BMC Plant Biol 14: 60.

115. 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.

116. Lang Q, Jin C, Lai L, Feng J, Chen S, *et al*. (2011) Tobacco microRNAs prediction and their expression infected with *Cucumber mosaic virus* and *Potato virus X*. Mol Biol Rep 38: 1523–1531.

117. He X-F, Fang Y-Y, Feng L, Guo H-S (2008) Characterization of conserved and novel microRNAs and their targets, including a TuMV-induced TIR–NBS–LRR class R gene-derived novel miRNA in Brassica. FEBS Lett 582: 2445–2452.

118. Tagami Y, Inaba N, Kutsuna N, Kurihara Y, Watanabe Y (2007) Specific enrichment of miRNAs in *Arabidopsis thaliana* infected with *Tobacco mosaic virus*. DNA Res 14: 227–233.

119. Romanel E, Silva TF, Corrêa RL, Farinelli L, Hawkins JS, *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.

120. Du P, Wu J, Zhang J, Zhao S, Zheng H, *et al*. (2011) Viral infection induces expression of novel phased microRNAs from conserved cellular microRNA precursors. PLoS Pathog 7: e1002176.

121. Naqvi AR, Haq QM, Mukherjee SK (2010) MicroRNA profiling of *Tomato leaf curl new delhi virus* (tolcndv) infected tomato leaves indicates that deregulation of mir159/319 and mir172 might be linked with leaf curl disease. Virol J 7: 281.

122. Lang Q, Zhou X, Zhang X, Drabek R, Zuo Z, *et al*. (2011) Microarray-based identification of tomato microRNAs and time course analysis of their response to *Cucumber mosaic virus* infection. J Zhejiang Univ Sci B 12: 116–125.

123. Feng J, Lai L, Lin R, Jin C, Chen J (2012) Differential effects of *Cucumber mosaic virus* satellite RNAs in the perturbation of microRNA-regulated gene expression in tomato. Mol Biol Rep 39: 775–784.

124. Feng J, Wang Y, Lin R, Chen J (2013) Altered expression of microRNAs and target mRNAs in tomato root and stem tissues upon different viral infection. J Phytopathol 161: 107–119.

125. 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.

126. 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.

127. Aryal R, Yang X, Yu Q, Sunkar R, Li L, *et al*. (2012) Asymmetric purine-pyrimidine distribution in cellular small RNA population of papaya. BMC Genomics 13: 682.

128. 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.

129. Alabi OJ, Zheng Y, Jagadeeswaran G, Sunkar R, Naidu RA (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.

130. Ruiz-Ruiz S, Navarro B, Gisel A, Peña L, Navarro L, *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.

131. Pérez-Quintero ÁL, Neme R, Zapata A, López C (2010) Plant microRNAs and their role in defense against viruses: a bioinformatics approach. BMC Plant Biol 10: 138.

132. Naqvi AR, Choudhury NR, Mukherjee SK, Haq QMR (2011) *In silico* analysis reveals that several tomato microRNA/microRNA* sequences exhibit propensity to bind to *Tomato leaf curl virus* (ToLCV) associated genomes and most of their encoded open reading frames (ORFs). Plant Physiol Biochem PPB Société Fr Physiol Végétale 49: 13–17.

133. Feng J, Chen J (2013) In silico analysis the complementarity of tomato microRNA/microRNA* sequences with *Cucumber mosaic virus* (CMV) genomic RNAs. J Nanosci Nanotechnol 13: 4421–4426.

134. Gasciolli V, Mallory AC, Bartel DP, Vaucheret H (2005) Partially redundant functions of Arabidopsis DICER-like enzymes and a role for DCL4 in producing trans-acting siRNAs. Curr Biol CB 15: 1494–1500.

135. Xie Z, Allen E, Wilken A, Carrington JC (2005) DICER-LIKE 4 functions in trans-acting small interfering RNA biogenesis and vegetative phase change in *Arabidopsis thaliana*. Proc Natl Acad Sci U S A 102: 12984–12989.

136. Song X, Li P, Zhai J, Zhou M, Ma L, *et al*. (2012) Roles of DCL4 and DCL3b in rice phased small RNA biogenesis: Phased small RNA clusters biogenesis in rice. Plant J 69: 462–474.

137. Johnson C, Kasprzewska A, Tennessen K, Fernandes J, Nan G-L, *et al*. (2009) Clusters and superclusters of phased small RNAs in the developing inflorescence of rice. Genome Res 19: 1429–1440.

138. Tang S, Wang Y, Li Z, Gui Y, Xiao B, *et al*. (2012) Identification of wounding and topping responsive small RNAs in tobacco (*Nicotiana tabacum*). BMC Plant Biol 12: 28.

139. Axtell MJ, Jan C, Rajagopalan R, Bartel DP (2006) A two-hit trigger for siRNA biogenesis in plants. Cell 127: 565–577.

140. Zhai J, Jeong D-H, De Paoli E, Park S, Rosen BD, *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.

141. Rajeswaran R, Aregger M, Zvereva AS, Borah BK, Gubaeva EG, *et al*. (2012) Sequencing of RDR6-dependent double-stranded RNAs reveals novel features of plant siRNA biogenesis. Nucleic Acids Res 40: 6241–6254.

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

143. Zhang C, Li G, Wang J, Zhu S, Li H (2013) Cascading cis-cleavage on transcript from trans-acting siRNA-producing locus 3. Int J Mol Sci 14: 14689–14699.

144. Xia R, Zhu H, An Y, Beers EP, Liu Z (2012) Apple miRNAs and tasiRNAs with novel regulatory networks. Genome Biol 13: R47.

145. Howell MD, Fahlgren N, Chapman EJ, Cumbie JS, Sullivan CM, *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.

146. Zhu H, Xia R, Zhao B, An Y, Dardick CD, *et al*. (2012) Unique expression, processing regulation, and regulatory network of peach (*Prunus persica*) miRNAs. BMC Plant Biol 12: 149.

147. Rock CD (2013) Trans-acting small interfering RNA4: key to nutraceutical synthesis in grape development? Trends Plant Sci 18: 601–610.

148. Klevebring D, Street NR, Fahlgren N, Kasschau KD, Carrington JC, *et al*. (2009) Genome-wide profiling of Populus small RNAs. BMC Genomics 10: 620.

149. Li F, Pignatta D, Bendix C, Brunkard JO, Cohn MM, *et al*. (2012) MicroRNA regulation of plant innate immune receptors. Proc Natl Acad Sci 109: 1790–1795.

150. Kallman T, Chen J, Gyllenstrand N, Lagercrantz U (2013) A significant fraction of 21-nucleotide small RNA originates from phased degradation of resistance genes in several perennial species. PLANT Physiol 162: 741–754.

151. Xia R, Meyers BC, Liu Z, Beers EP, Ye S, *et al*. (2013) MicroRNA superfamilies descended from miR390 and their roles in secondary small interfering RNA biogenesis in eudicots. Plant Cell 25: 1555–1572.

152. Chen H-M, Li Y-H, Wu S-H (2007) Bioinformatic prediction and experimental validation of a microRNA-directed tandem trans-acting siRNA cascade in Arabidopsis. Proc Natl Acad Sci 104: 3318–3323.

153. Chen H-M, Chen L-T, Patel K, Li Y-H, Baulcombe DC, *et al*. (2010) 22-nucleotide RNAs trigger secondary siRNA biogenesis in plants. Proc Natl Acad Sci 107: 15269–15274.

154. Montgomery TA, Yoo SJ, Fahlgren N, Gilbert SD, Howell MD, *et al*. (2008) AGO1-miR173 complex initiates phased siRNA formation in plants. Proc Natl Acad Sci U S A 105: 20055–20062.

155. 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.

156. Montgomery TA, Howell MD, Cuperus JT, Li D, Hansen JE, *et al*. (2008) Specificity of ARGONAUTE7-miR390 interaction and dual functionality in TAS3 trans-acting siRNA formation. Cell 133: 128–141.

157. Shivaprasad PV, Chen H-M, Patel K, Bond DM, Santos BACM, *et al*. (2012) A microRNA superfamily regulates nucleotide binding site-leucine-rich repeats and other mRNAs. Plant Cell 24: 859–874.

158. Wang Y, Itaya A, Zhong X, Wu Y, Zhang J, *et al*. (2011) Function and evolution of a microRNA that regulates a Ca2+-ATPase and triggers the formation of phased small interfering RNAs in tomato reproductive growth. PLANT CELL ONLINE 23: 3185–3203.

159. Zhang C, Ng DW-K, Lu J, Chen ZJ (2012) Roles of target site location and sequence complementarity in trans-acting siRNA formation in Arabidopsis. Plant J Cell Mol Biol 69: 217–226.

160. 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.

161. Williams L, Carles CC, Osmont KS, Fletcher JC (2005) A database analysis method identifies an endogenous trans-acting short-interfering RNA that targets the Arabidopsis *ARF2*, *ARF3*, and *ARF4* genes. Proc Natl Acad Sci U S A 102: 9703–9708.

162. Marin E, Jouannet V, Herz A, Lokerse AS, Weijers D, *et al*. (2010) miR390, Arabidopsis TAS3 tasiRNAs, and their AUXIN RESPONSE FACTOR targets define an autoregulatory network quantitatively regulating lateral root growth. Plant Cell 22: 1104–1117.

163. Yoon EK, Yang JH, Lim J, Kim SH, Kim S-K, *et al*. (2010) Auxin regulation of the microRNA390-dependent transacting small interfering RNA pathway in Arabidopsis lateral root development. Nucleic Acids Res 38: 1382–1391.

164. Nogueira FTS, Madi S, Chitwood DH, Juarez MT, Timmermans MCP (2007) Two small regulatory RNAs establish opposing fates of a developmental axis. Genes Dev 21: 750–755.

165. Nogueira FTS, Chitwood DH, Madi S, Ohtsu K, Schnable PS, *et al*. (2009) Regulation of small RNA accumulation in the maize shoot apex. PLoS Genet 5: e1000320.

166. 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.

167. 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.

168. Yamada K, Lim J, Dale JM, Chen H, Shinn P, *et al*. (2003) Empirical analysis of transcriptional activity in the Arabidopsis genome. Science 302: 842–846.

169. Osato N, Yamada H, Satoh K, Ooka H, Yamamoto M, *et al*. (2003) Antisense transcripts with rice full-length cDNAs. Genome Biol 5: R5.

170. 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.

171. Wang H, Chua N-H, Wang X-J (2006) Prediction of trans-antisense transcripts in *Arabidopsis thaliana. Genome* Biol 7: R92.

172. Borsani O, Zhu J, Verslues PE, Sunkar R, Zhu J-K (2005) Endogenous siRNAs derived from a pair of natural cis-antisense transcripts regulate salt tolerance in Arabidopsis. Cell 123: 1279–1291.

173. Katiyar-Agarwal S, Morgan R, Dahlbeck D, Borsani O, Villegas A, *et al*. (2006) A pathogen-inducible endogenous siRNA in plant immunity. Proc Natl Acad Sci 103: 18002–18007.

174. Katiyar-Agarwal S, Gao S, Vivian-Smith A, Jin H (2007) A novel class of bacteria-induced small RNAs in Arabidopsis. Genes Dev 21: 3123–3134.

175. Ron M, Alandete Saez M, Eshed Williams L, Fletcher JC, McCormick S (2010) Proper regulation of a sperm-specific cis-nat-siRNA is essential for double fertilization in Arabidopsis. Genes Dev 24: 1010–1021.

176. 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.

177. Zhang X, Xia J, Lii YE, Barrera-Figueroa BE, Zhou X, *et al*. (2012) Genome-wide analysis of plant nat-siRNAs reveals insights into their distribution, biogenesis and function. Genome Biol 13: R20.

178. Henz SR, Cumbie JS, Kasschau KD, Lohmann JU, Carrington JC, *et al*. (2007) Distinct expression patterns of natural antisense transcripts in Arabidopsis. PLANT Physiol 144: 1247–1255.

179. Zhou X, Sunkar R, Jin H, Zhu J-K, Zhang W (2009) Genome-wide identification and analysis of small RNAs originated from natural antisense transcripts in *Oryza sativa*. Genome Res 19: 70–78.

180. Chen D, Meng Y, Ma X, Mao C, Bai Y, *et al*. (2010) Small RNAs in angiosperms: sequence characteristics, distribution and generation. Bioinformatics 26: 1391–1394.

181. Zhan S, Lukens L (2013) Protein-coding cis-natural antisense transcripts have high and broad expression in Arabidopsis. PLANT Physiol 161: 2171–2180.

182. Jin H, Vacic V, Girke T, Lonardi S, Zhu J-K (2008) Small RNAs and the regulation of cis-natural antisense transcripts in Arabidopsis. BMC Mol Biol 9: 6.

183. 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.

184. Held MA, Penning B, Brandt AS, Kessans SA, Yong W, *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.

185. Yu X, Yang J, Li X, Liu X, Sun C, *et al*. (2013) Global analysis of cis-natural antisense transcripts and their heat-responsive nat-siRNAs in *Brassica rapa*. BMC Plant Biol 13: 208.

186. Onodera Y, Haag JR, Ream T, Costa Nunes P, Pontes O, *et al*. (2005) Plant nuclear RNA polymerase IV mediates siRNA and DNA methylation-dependent heterochromatin formation. Cell 120: 613–622.

187. Herr AJ, Jensen MB, Dalmay T, Baulcombe DC (2005) RNA polymerase IV directs silencing of endogenous DNA. Science 308: 118–120.

188. Pontier D, Yahubyan G, Vega D, Bulski A, Saez-Vasquez J, *et al*. (2005) Reinforcement of silencing at transposons and highly repeated sequences requires the concerted action of two distinct RNA polymerases IV in Arabidopsis. Genes Dev 19: 2030–2040.

189. Wassenegger M, Heimes S, Riedel L, Sänger HL (1994) RNA-directed de novo methylation of genomic sequences in plants. Cell 76: 567–576.

190. Zilberman D, Cao X, Johansen LK, Xie Z, Carrington JC, *et al*. (2004) Role of Arabidopsis ARGONAUTE4 in RNA-directed DNA methylation triggered by inverted repeats. Curr Biol CB 14: 1214–1220.

191. Wierzbicki AT, Haag JR, Pikaard CS (2008) Noncoding transcription by RNA polymerase Pol IVb/Pol V mediates transcriptional silencing of overlapping and adjacent genes. Cell 135: 635–648.

192. Wierzbicki AT, Ream TS, Haag JR, Pikaard CS (2009) RNA polymerase V transcription guides ARGONAUTE4 to chromatin. Nat Genet 41: 630–634.

193. Wierzbicki AT, Cocklin R, Mayampurath A, Lister R, Rowley MJ, *et al*. (2012) Spatial and functional relationships among Pol V-associated loci, Pol IV-dependent siRNAs, and cytosine methylation in the Arabidopsis epigenome. Genes Dev 26: 1825–1836.

194. Wei L, Gu L, Song X, Cui X, Lu Z, *et al*. (2014) Dicer-like 3 produces transposable element-associated 24-nt siRNAs that control agricultural traits in rice. Proc Natl Acad Sci: 201318131.

195. Ito H, Gaubert H, Bucher E, Mirouze M, Vaillant I, *et al*. (2011) An siRNA pathway prevents transgenerational retrotransposition in plants subjected to stress. Nature 472: 115–119.

196. Sun F, Guo W, Du J, Ni Z, Sun Q, *et al*. (2013) Widespread, abundant, and diverse TE-associated siRNAs in developing wheat grain. Gene 522: 1–7.

197. Hackenberg M, Huang P-J, Huang C-Y, Shi B-J, Gustafson P, *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.

198. Lee SR, Collins K (2005) Starvation-induced cleavage of the tRNA anticodon loop in *Tetrahymena thermophila*. J Biol Chem 280: 42744–42749.

199. Nowacka M, Strozycki PM, Jackowiak P, Hojka-Osinska A, Szymanski M, *et al*. (2013) Identification of stable, high copy number, medium-sized RNA degradation intermediates that accumulate in plants under non-stress conditions. Plant Mol Biol 83: 191–204.

200. Ogawa T, Tomita K, Ueda T, Watanabe K, Uozumi T, *et al*. (1999) A cytotoxic ribonuclease targeting specific transfer RNA anticodons. Science 283: 2097–2100.

201. MacIntosh GC, Bariola PA, Newbigin E, Green PJ (2001) Characterization of Rny1, the *Saccharomyces cerevisiae* member of the T2 RNase family of RNases: unexpected functions for ancient enzymes? Proc Natl Acad Sci U S A 98: 1018–1023.

202. Thompson DM, Parker R (2009) The RNase Rny1p cleaves tRNAs and promotes cell death during oxidative stress in *Saccharomyces cerevisiae*. J Cell Biol 185: 43–50.

203. Fu H, Feng J, Liu Q, Sun F, Tie Y, *et al*. (2009) Stress induces tRNA cleavage by angiogenin in mammalian cells. FEBS Lett 583: 437–442.

204. Yamasaki S, Ivanov P, Hu G-F, Anderson P (2009) Angiogenin cleaves tRNA and promotes stress-induced translational repression. J Cell Biol 185: 35–42.

205. 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.

206. Cole C, Sobala A, Lu C, Thatcher SR, Bowman A, *et al*. (2009) Filtering of deep sequencing data reveals the existence of abundant Dicer-dependent small RNAs derived from tRNAs. RNA 15: 2147–2160.

207. Haussecker D, Huang Y, Lau A, Parameswaran P, Fire AZ, *et al*. (2010) Human tRNA-derived small RNAs in the global regulation of RNA silencing. RNA N Y N 16: 673–695.

208. Langenberger D, Çakir MV, Hoffmann S, Stadler PF (2013) Dicer-processed small RNAs: Rules and exceptions. J Exp Zoolog B Mol Dev Evol 320: 35–46.

209. Li Z, Ender C, Meister G, Moore PS, Chang Y, *et al*. (2012) Extensive terminal and asymmetric processing of small RNAs from rRNAs, snoRNAs, snRNAs, and tRNAs. Nucleic Acids Res 40: 6787–6799.

210. Maute RL, Schneider C, Sumazin P, Holmes A, Califano A, *et al*. (2013) tRNA-derived microRNA modulates proliferation and the DNA damage response and is down-regulated in B cell lymphoma. Proc Natl Acad Sci 110: 1404–1409.

211. 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.

212. Haiser HJ, Karginov FV, Hannon GJ, Elliot MA (2007) Developmentally regulated cleavage of tRNAs in the bacterium *Streptomyces coelicolor*. Nucleic Acids Res 36: 732–741.

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

214. Sobala A, Hutvagner G (2013) Small RNAs derived from the 5′ end of tRNA can inhibit protein translation in human cells. RNA Biol 10: 553–563.

215. Ivanov P, Emara MM, Villen J, Gygi SP, Anderson P (2011) Angiogenin-induced tRNA fragments inhibit translation initiation. Mol Cell 43: 613–623.

216. 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. Article ID 260909.

217. 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: 6.

218. Nashimoto M (2000) Anomalous RNA substrates for mammalian tRNA 3' processing endoribonuclease. FEBS Lett 472: 179–186.

219. Elbarbary RA, Takaku H, Uchiumi N, Tamiya H, Abe M, *et al*. (2009) Modulation of gene expression by human cytosolic tRNase ZL through 5′-half-tRNA. PLoS ONE 4: e5908.

220. Ho T, Wang H, Pallett D, Dalmay T (2007) Evidence for targeting common siRNA hotspots and GC preference by plant Dicer-like proteins. FEBS Lett 581: 3267–3272.

221. Ho T, Rusholme Pilcher RL, Edwards M-L, Cooper I, Dalmay T, *et al*. (2008) Evidence for GC preference by monocot Dicer-like proteins. Biochem Biophys Res Commun 368: 433–437.

222. 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.

223. Blevins T, Rajeswaran R, Shivaprasad PV, Beknazariants D, Si-Ammour A, *et al*. (2006) Four plant Dicers mediate viral small RNA biogenesis and DNA virus induced silencing. Nucleic Acids Res 34: 6233–6246.

224. Chellappan P, Vanitharani R, Pita J, 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.

225. Moissiard G, Voinnet O (2006) RNA silencing of host transcripts by *Cauliflower mosaic virus* requires coordinated action of the four Arabidopsis Dicer-like proteins. Proc Natl Acad Sci U S A 103: 19593–19598.

226. Xie Z, Fan B, Chen C, Chen Z (2001) An important role of an inducible RNA-dependent RNA polymerase in plant antiviral defense. Proc Natl Acad Sci U S A 98: 6516–6521.

227. Schwach F, Vaistij FE, Jones L, Baulcombe DC (2005) An RNA-dependent RNA polymerase prevents meristem invasion by *Potato virus X* and is required for the activity but not the production of a systemic silencing signal. Plant Physiol 138: 1842–1852.

228. Molnar A, Csorba T, Lakatos L, Varallyay E, Lacomme C, *et al*. (2005) Plant virus-derived small interfering RNAs originate predominantly from highly structured single-stranded viral RNAs. J Virol 79: 7812–7818.

229. Ho T, Pallett D, Rusholme R, Dalmay T, Wang H (2006) A simplified method for cloning of short interfering RNAs from *Brassica juncea* infected with *Turnip mosaic potyvirus* and *Turnip crinkle carmovirus*. J Virol Methods 136: 217–223.

230. Du Q-S, Duan C-G, Zhang Z-H, Fang Y-Y, Fang R-X, *et al*. (2007) DCL4 targets *Cucumber mosaic virus* Satellite RNA at novel secondary structures. J Virol 81: 9142–9151.

231. Wang X-B, Jovel J, Udomporn P, Wang Y, Wu Q, *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.

232. Morel J-B, Godon C, Mourrain P, Béclin C, Boutet S, *et al*. (2002) Fertile hypomorphic ARGONAUTE (ago1) mutants impaired in post-transcriptional gene silencing and virus resistance. Plant Cell 14: 629–639.

233. Scholthof HB, Alvarado VY, Vega-Arreguin JC, Ciomperlik J, Odokonyero D, *et al*. (2011) Identification of an ARGONAUTE for antiviral RNA silencing in *Nicotiana benthamiana*. Plant Physiol 156: 1548–1555.

234. Jaubert M, Bhattacharjee S, Mello AFS, Perry KL, Moffett P (2011) ARGONAUTE2 mediates RNA-silencing antiviral defenses against *Potato virus X* in Arabidopsis. Plant Physiol 156: 1556–1564.

235. Harvey JJW, Lewsey MG, Patel K, Westwood J, Heimstädt S, *et al*. (2011) An antiviral defense role of AGO2 in plants. PloS ONE 6: e14639.

236. Zhang X, Yuan Y-R, Pei Y, Lin S-S, Tuschl T, *et al*. (2006) *Cucumber mosaic virus*-encoded 2b suppressor inhibits Arabidopsis Argonaute1 cleavage activity to counter plant defense. Genes Dev 20: 3255–3268.

237. Pantaleo V, Szittya G, Burgyan J (2007) Molecular bases of viral RNA targeting by viral small interfering RNA-programmed RISC. J Virol 81: 3797–3806.

238. Xia Z, Peng J, Li Y, Chen L, Li S, *et al*. (2014) Characterization of small interfering RNAs derived from *Sugarcane mosaic virus* in infected maize plants by deep sequencing. PloS ONE 9: e97013.

239. 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.

240. 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.

241. Shimura H, Pantaleo V, Ishihara T, Myojo N, Inaba J, *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.

242. 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.

243. Guo W, Liew JY, Yuan YA (2014) Structural insights into the arms race between host and virus along RNA silencing pathways in *Arabidopsis thaliana*: Structural insights into plant-virus interactions. Biol Rev 89: 337–355.

244. Lau NC, Lim LP, Weinstein EG, Bartel DP (2001) An abundant class of tiny RNAs with probable regulatory roles in *Caenorhabditis elegans*. Science 294: 858–862.

245. Válóczi A, Várallyay É, Kauppinen S, Burgyán J, Havelda Z (2006) Spatio-temporal accumulation of microRNAs is highly coordinated in developing plant tissues. Plant J 47: 140–151.

246. Chitwood DH, Nogueira FTS, Howell MD, Montgomery TA, Carrington JC, *et al*. (2009) Pattern formation via small RNA mobility. Genes Dev 23: 549–554.

247. Jin W, Wu F, Xiao L, Liang G, Zhen Y, *et al*. (2012) Microarray-based analysis of tomato miRNA regulated by *Botrytis cinerea*. J Plant Growth Regul 31: 38–46.

248. Liang R-Q, Li W, Li Y, Tan C, Li J-X, *et al*. (2005) An oligonucleotide microarray for microRNA expression analysis based on labeling RNA with quantum dot and nanogold probe. Nucleic Acids Res 33: e17.

249. Aravin A, Tuschl T (2005) Identification and characterization of small RNAs involved in RNA silencing. FEBS Lett 579: 5830–5840.

250. Chen C, Ridzon DA, Broomer AJ, Zhou Z, Lee DH, *et al*. (2005) Real-time quantification of microRNAs by stem-loop RT-PCR. Nucleic Acids Res 33: e179.

251. Lao K, Xu NL, Yeung V, Chen C, Livak KJ, *et al*. (2006) Multiplexing RT-PCR for the detection of multiple miRNA species in small samples. Biochem Biophys Res Commun 343: 85–89.

252. Tang F, Hajkova P, Barton SC, Lao K, Surani MA (2006) MicroRNA expression profiling of single whole embryonic stem cells. Nucleic Acids Res 34: e9.

253. Varkonyi-Gasic E, Wu R, Wood M, Walton EF, Hellens RP (2007) Protocol: a highly sensitive RT-PCR method for detection and quantification of microRNAs. Plant Methods 3: 12.

254. Lu C, Tej SS, Luo S, Haudenschild CD, Meyers BC, *et al*. (2005) Elucidation of the small RNA component of the transcriptome. Science 309: 1567–1569.

255. Bentley DR, Balasubramanian S, Swerdlow HP, Smith GP, Milton J, *et al*. (2008) Accurate whole human genome sequencing using reversible terminator chemistry. Nature 456: 53–59.

256. Margulies M, Egholm M, Altman WE, Attiya S, Bader JS, *et al*. (2005) Genome sequencing in microfabricated high-density picolitre reactors. Nature 437: 376–380.

257. Valouev A, Ichikawa J, Tonthat T, Stuart J, Ranade S, *et al*. (2008) A high-resolution, nucleosome position map of *C. elegans* reveals a lack of universal sequence-dictated positioning. Genome Res 18: 1051–1063.

258. Axtell MJ (2013) ShortStack: Comprehensive annotation and quantification of small RNA genes. RNA 19: 740–751.

259. Meyers BC, Axtell MJ, Bartel B, Bartel DP, Baulcombe D, *et al*. (2008) Criteria for annotation of plant microRNAs. Plant Cell Online 20: 3186–3190.

260. Fahlgren N, Sullivan CM, Kasschau KD, Chapman EJ, Cumbie JS, *et al*. (2009) Computational and analytical framework for small RNA profiling by high-throughput sequencing. RNA N Y N 15: 992–1002.

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

262. Li R, Li Y, Kristiansen K, Wang J (2008) SOAP: short oligonucleotide alignment program. Bioinforma Oxf Engl 24: 713–714.

263. Burrows M, Wheeler DJ (1994) A block-sorting lossless data compression algorithm. Technical report 124. Palo Alto, CA: Digital Equipment Corporation.

264. 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.

265. Li H, Durbin R (2009) Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinforma Oxf Engl 25: 1754–1760.

266. Li R, Yu C, Li Y, Lam T-W, Yiu S-M, *et al*. (2009) SOAP2: an improved ultrafast tool for short read alignment. Bioinformatics 25: 1966–1967.

267. Liu C-M, Wong T, Wu E, Luo R, Yiu S-M, *et al*. (2012) SOAP3: ultra-fast GPU-based parallel alignment tool for short reads. Bioinforma Oxf Engl 28: 878–879.

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

269. Friedländer MR, Chen W, Adamidi C, Maaskola J, Einspanier R, *et al*. (2008) Discovering microRNAs from deep sequencing data using miRDeep. Nat Biotechnol 26: 407–415.

270. Hofacker IL (2003) Vienna RNA secondary structure server. Nucleic Acids Res 31: 3429–3431.

271. Cortes C, Vapnik V (1995) Support-vector networks. Mach Learn 20: 273–297.

272. Fine S, Singer Y, Tishby N (1998) The hierarchical hidden Markov model: Analysis and applications. Mach Learn 32: 41–62.

273. Yousef M, Nebozhyn M, Shatkay H, Kanterakis S, Showe LC, *et al*. (2006) Combining multi-species genomic data for microRNA identification using a Naive Bayes classifier. Bioinforma Oxf Engl 22: 1325–1334.

274. Hackenberg M, Sturm M, Langenberger D, Falcón-Pérez JM, Aransay AM (2009) miRanalyzer: a microRNA detection and analysis tool for next-generation sequencing experiments. Nucleic Acids Res 37: W68–W76.

275. Hackenberg M, Rodríguez-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.

276. Breiman L (2001) Random Forests. Mach Learn 45: 5–32.

277. Addo-Quaye C, Eshoo TW, Bartel DP, Axtell MJ (2008) Endogenous siRNA and miRNA targets identified by sequencing of the Arabidopsis degradome. Curr Biol 18: 758–762.

278. German MA, Pillay M, Jeong D-H, Hetawal A, Luo S, *et al*. (2008) Global identification of microRNA–target RNA pairs by parallel analysis of RNA ends. Nat Biotechnol 26: 941–946.

279. Gregory BD, O'Malley RC, Lister R, Urich MA, Tonti-Filippini J, *et al*. (2008) A link between RNA metabolism and silencing affecting Arabidopsis development. Dev Cell 14: 854–866.

280. Addo-Quaye C, Miller W, Axtell MJ (2009) CleaveLand: a pipeline for using degradome data to find cleaved small RNA targets. Bioinformatics 25: 130–131.

281. 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.

282. Folkes L, Moxon S, Woolfenden HC, Stocks MB, Szittya G, *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.

283. De Andrade RR, Vaslin MF (2014) SearchSmallRNA: a graphical interface tool for the assemblage of viral genomes using small RNA libraries data. Virol J 11: 45.

284. Watson M, Schnettler E, Kohl A (2013) viRome: an R package for the visualization and analysis of viral small RNA sequence datasets. Bioinformatics 29: 1902–1903.

285. Stocks MB, Moxon S, Mapleson D, Woolfenden HC, Mohorianu I, *et al*. (2012) The UEA sRNA Workbench: A suite of tools for analysing and visualising next generation sequencing microRNA and small RNA datasets. Bioinformatics 28: 2059–2061.

286. Dai X, Zhao PX (2008) pssRNAMiner: a plant short small RNA regulatory cascade analysis server. Nucleic Acids Res 36: W114–W118.

287. Zhang C, Li G, Zhu S, Zhang S, Fang J (2014) tasiRNAdb: a database of ta-siRNA regulatory pathways. Bioinforma Oxf Engl 30: 1045–1046.

288. Anders S, Huber W (2010) Differential expression analysis for sequence count data. Genome Biol 11: R106.

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

290. Robinson MD, Oshlack A (2010) A scaling normalization method for differential expression analysis of RNA-seq data. Genome Biol 11: R25.

291. Bullard JH, Purdom E, Hansen KD, Dudoit S (2010) Evaluation of statistical methods for normalization and differential expression in mRNA-Seq experiments. BMC Bioinformatics 11: 94.

292. Love MI, Huber W, Anders S (2014) Moderated estimation of fold change and dispersion for RNA-Seq data with DESeq2. bioRxiv: http://dx.doi.org/10.1101/002832.

293. Zerbino DR, Birney E (2008) Velvet: Algorithms for de novo short read assembly using de Bruijn graphs. Genome Res 18: 821–829.

294. De Bruijn NG (1946) A combinatorial problem. Proc K Ned Akad V Wet 49: 758–764.

295. Good IJ (1946) Normal recurring decimals. J Lond Math Soc s1-21: 167–169.

296. Li R, Zhu H, Ruan J, Qian W, Fang X, *et al*. (2010) De novo assembly of human genomes with massively parallel short read sequencing. Genome Res 20: 265–272.

297. Luo R, Liu B, Xie Y, Li Z, Huang W, *et al*. (2012) SOAPdenovo2: an empirically improved memory-efficient short-read de novo assembler. Gigascience 1: 18.

298. Warren RL, Sutton GG, Jones SJM, Holt RA (2007) Assembling millions of short DNA sequences using SSAKE. Bioinforma Oxf Engl 23: 500–501.

299. Jeck WR, Reinhardt JA, Baltrus DA, Hickenbotham MT, Magrini V, *et al*. (2007) Extending assembly of short DNA sequences to handle error. Bioinformatics 23: 2942–2944.

Chapter 3

Extending the sRNAome of apple by next-generation sequencing

3.1 Abstract

The global importance of apple as a fruit crop necessitates investigations into molecular aspects of the processes that influence fruit quality and yield, including plant development, fruit ripening and disease resistance. In order to study and understand biological processes it is essential to recognise the range of molecules, which influence these processes. Small non-coding RNAs are regulatory agents involved in diverse plant activities, ranging from development to stress response. The occurrence of these molecules in apple leaves was studied by means of next-generation sequencing. 85 novel microRNA (miRNA) gene loci were predicted and characterized along with known miRNA loci. Both cis- and trans-natural antisense transcript pairs were identified. Although the trans-overlapping regions were enriched in small RNA (sRNA) production, cis-overlaps did not seem to agree. More than 150 phased regions were also identified, and for a small subset of these, potential miRNAs that could initiate phasing, were revealed. Repeat-associated siRNAs, which are generated from repetitive genomic regions such as transposons, were also analysed. For this group almost all available repeat sequences, associated with the apple genome and present in Repbase, were found to produce siRNAs. Results from this study extend our current knowledge on apple sRNAs and their precursors significantly. A rich molecular resource has been created and is available to the research community to serve as a baseline for future studies.

3.2 Introduction

Apple (*Malus* x *domestica*) is one of the world's most important fruit crops. Due to the efforts of several individuals and working groups, a number of genomic resources have become available to form the basis for studying various biological processes in apple. These include the draft genome sequence (~742.3 Mb), genome annotation, and various transcriptome datasets available in public databases, including datasets describing small RNAs, degradome, and expressed transcripts [1–6].

Plants acquired a variety of systems to regulate gene expression, including transcriptional (TGS) and post-transcriptional gene silencing (PTGS) [7,8]. These regulatory processes can be triggered by double-stranded RNA (dsRNA) precursors that lead to the generation of small RNA (sRNA) molecules (~17-26 nt), which target specific RNA molecules. One of the sRNA strands, known as the "guide strand", associates with enzymes called Argonautes (AGOs) [9], as well as other proteins in either the RNA-induced silencing complex (RISC) [10] or the RNA-induced initiation of transcriptional gene silencing (RITS) complex [11]. Base-pairing to target nucleic acids complementary to the sRNA subsequently triggers silencing. Although silencing can result from DNA methylation and histone modifications, in plants it is most often the result of cleaving and degradation of the target RNA [12,13].

The two dominant types of sRNAs in plants are microRNAs (miRNAs) and small interfering RNAs (siRNAs). miRNAs, despite not being the most abundant sRNA type, are the best-studied group. Primary miRNAs (pri-miRNA) are transcribed by RNA polymerase II from endogenous genes [14,15]. These transcripts have a 5'-cap and a 3'-polyadenylated tail and fold into hairpin structures. Two successive cleavage reactions by Dicer-like (DCL) type III RNases result in the mature miRNA – a short (~21 bp) double-stranded molecule containing a small number of mismatches between the miRNA and its antisense strand (previously known as miRNA*) [16,17].

Small interfering RNAs are processed by DCL enzymes from long dsRNAs that are perfectly base-paired. This group can be divided into several sub-groups. See the review article by Axtell, for a comprehensive discussion on plant siRNAs [18]. The various siRNA species are produced via diverse biosynthetic pathways and affect

gene regulation through different modes of action. One siRNA species, known as natural antisense transcript siRNA (nat-siRNA), originate from the overlapping regions of complementary transcripts, which form dsRNA duplexes [19]. In contrast, the dsRNA precursors for phased siRNA (phasiRNA) are generated by RNA-dependent RNA polymerase (RDR) activity [20]. This group includes the well-characterized trans-acting siRNAs (tasiRNAs) [21]. The siRNAs are spawned in a phased manner starting from the cleaved site. Repetitive genetic elements such as transposons and satellite DNA can also give rise to a group known as repeat-associated siRNAs (rasiRNAs) [22,23]. Other more recently identified and less characterised functional sRNAs include species derived from small nucleolar RNA (snoRNA), ribosomal RNA (rRNA) and transfer RNA (tRNA) [24–28].

The role of sRNAs in the regulation of important biological processes is well documented. The present study improves the current apple sRNA species database, by adding and categorising novel and known sRNAs. A next-generation sequencing (NGS) approach was followed to sequence the sRNA transcriptome (sRNAome) of apple leaves. Computational analysis of the sRNA data provides a comprehensive resource to support future studies in order to investigate the role of specific miRNAs, phasiRNAs, nat-siRNAs, as well as rasiRNAs in various biological processes of apple.

3.3 Results and Discussion

3.3.1 Apple sRNA next-generation sequencing data

The ability of NGS to detect low titres of sRNA species in plant cells was exploited to determine the sRNAome of apple leaves. Of the 71,273,331 high quality sequence reads, 96.58% were 17 to 26 nt in length (Table 3.1). The majority of functional sRNA species involved in TGS and PTGS is considered to fall within this range. The library was dominated by reads 24 nt in length (37%) followed by 21 nt long reads (31%), which also displayed the greatest redundancy (95%) (Figure 3.1). This high level of redundancy may be attributed to a small group of 21 nt long sRNAs with an elevated demand by the cell. sRNAs that often fall into this size group are miRNAs and phasiRNAs. Analysis of the miRNAs (Table 3.1) showed a high level of redundancy, although the miRNA group alone cannot fully explain the redundancy of the 21-nt size

group. The dominance of the two size groups highlighted their probable significance in regulating biological processes. sRNAs from these size groups include heterochromatic siRNAs, which are 24 nt in length and function by means of RNA-mediated methylation of DNA targets [13], as well as phasiRNAs and miRNAs, which can be of either length. It is important to note that miRNAs are not necessarily restricted to a length of 21 or 24 nt, but have been found to range between 20 and 24 nt in length.

Table 3.1. Summary of the sequenced reads.

Small RNA	Unique	Total
Adapter trimmed	14,027,369	77,651,426
High quality reads	12,969,231	71,273,331
17-26 nt reads	12,422,959	68,837,477
miRNA reads[a]	249	12,119,076
nat-siRNA reads	108,657	813,241
phasiRNA reads[b]	363	30,500
rasiRNA reads	1,139,528	5,526,689

[a] Reads with perfect matches to known and novel mdm-miRNAs.

[b] Reads with perfect matches to phasiRNA which are in phase with miRNA cleave-sites.

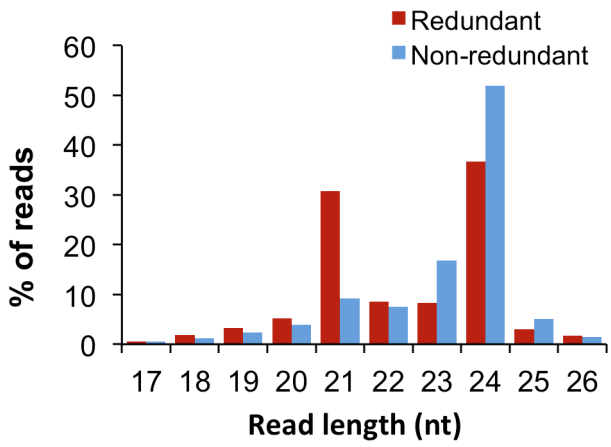


Figure 3.1. Sequencing library size distribution. Number of reads, 17 to 26 nt in length, as a percentage of either the total redundant or non-redundant reads in this size range.

3.3.2 Detection of known miRNAs

To identify known miRNAs present in the dataset, the reads were compared to the publicly available miRNA Registry Database, miRBase (version 20) [3]. When only allowing perfect matches, 11,847,841 reads mapped to apple miRNAs (mdm-miRNAs) recorded in miRBase. The database contains 207 mature mdm-miRNAs (from 43 families), which were predicted from Golden Delicious plants. During this study, 195 of the listed mdm-miRNAs, belonging to 40 families, could be detected in leaf material (Table S1A). No members were detected for the miR828, miR2111 and miR7128 families. The abundance of individual known miRNAs ranged from single reads to a few million, with most (75.9%) of the mature mdm-miRNAs having a read count of greater than 100. The miRNA cluster with the highest read count, mdm-miR166a-i, accounted for 91.51% of all the reads mapped to known apple miRNAs. In total, miR166 was the largest represented miRNA family followed by miR396 and miR398 (Table S1B). This result differs from a study by Xia *et al*. [5], who found miR167 to be the most abundant in leaf sequencing data, closely followed by miR165/166. However, differences in expression levels of miRNA families between studies can be expected and may be attributed to differences in developmental stages, on-going physiological processes, and environmental conditions. mdm-miR166 was previously proven to target apple homeobox-leucine zipper proteins [5]. This protein family is involved in a range of plant processes including growth and morphogenesis [29]. The vast number of miR166 reads, mirroring their high expression level, is a clear indication of the central role of this miRNA species in regulating apple processes.

Besides known mdm-miRNAs, 77 unique reads with 100% homology to miRNAs from other plant species, not yet identified in apple, were also detected (Table S1C). These reads numbered 198,840, with the highest represented sequence having an individual read count of 180,963. Twenty of the reads homologous to non-apple Viridiplantae species had a sum of more than 100. However, the presence of homologous sequences in the apple sRNA dataset is not sufficient evidence for these to be considered apple miRNAs. This matter can possibly be resolved by analysing their region of origin on the apple genome during novel miRNA prediction.

3.3.3 Novel miRNA and target prediction

Due to the essential regulatory role that miRNAs play in many biological processes it is important to expand the available miRNA knowledge base. To identify novel apple miRNAs we performed a miRBase-independent, computational miRNA prediction analysis, based on the sRNA sequencing data. 130 genomic loci were predicted to be miRNA genes, each having a mature miRNA represented by at least 10 reads (Figure 3.2 and Table S1D). Nine of these miRNA genes have more than one potential mature miRNA pair. The predicted genomic regions of 45 of the miRNA precursors overlapped with the loci of known mdm-miRNAs. For the majority of these precursors, at least one of the predicted mature miRNAs was a known mdm-miRNA. For two of the precursors, mdm-MIR399e and mdm-MIR5225c, the current analysis predicted the complement of the mature miRBase entry sequence as the novel miRNA. Some of the predicted novel mature miRNAs were isomiRs (sequence variants) of existing miRBase mature entries, some of which were homologous to miRNAs from other plant species. At four of these known miRNA gene loci, the newly predicted miRNAs had read counts that were higher than those of the current miRBase entry. Three of these miRNAs were homologous to miRNAs from other plant species. Figure 3.3 illustrates two cases where the mature sequence differed from the miRBase entry or where the isomiR was predicted as mature along with the miRBase entry. The fact that the mature sequence, as registered in miRBase, does not correspond to the dominant miRNA for the precursor in this dataset does not necessarily imply that the registry entry is a miss-annotation. As can be seen from the data by Xia *et al*. [5], different isomiRs can be expressed at different levels relative to each other depending on the tissue type. Similar variation in expression levels can probably also be ascribed to differences in environmental conditions.

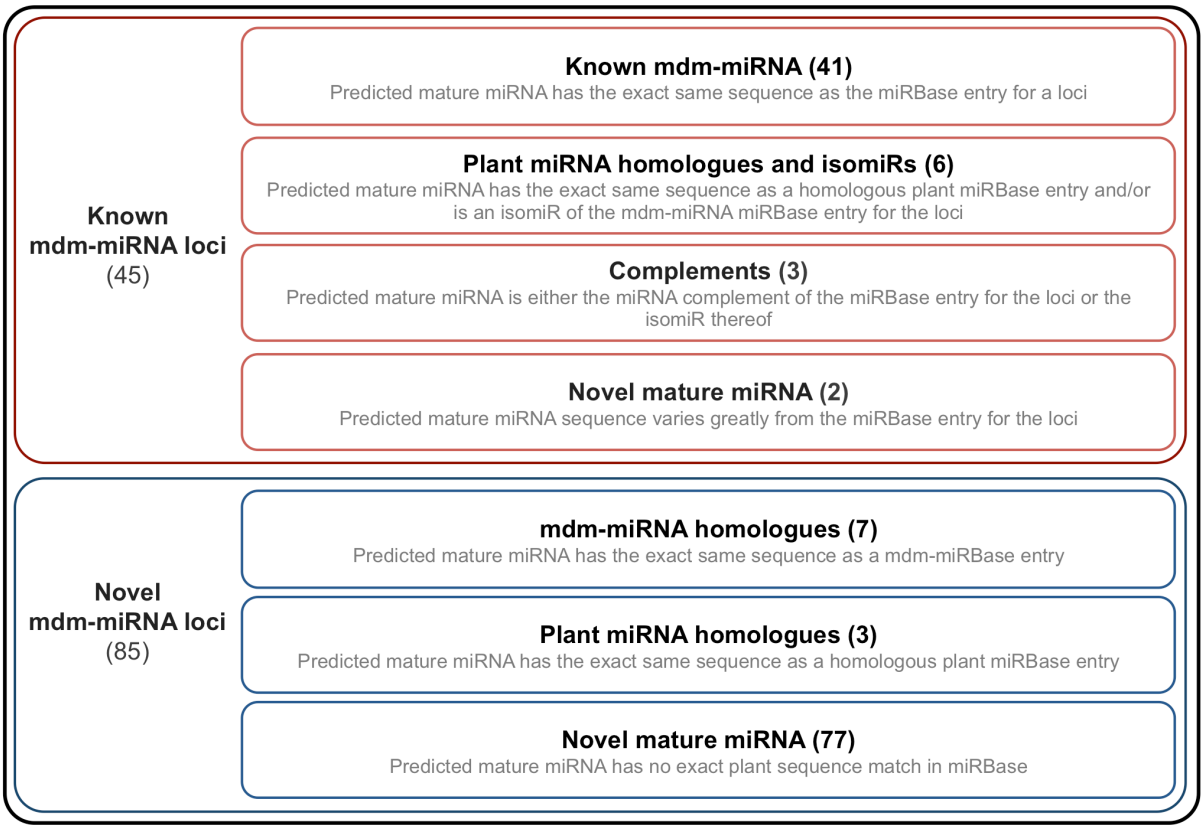


Figure 3.2. Known and novel miRNA predictions. Diagram defining the different classifications used for known and novel mdm-miRNA loci predictions. The predicted mature miRNA at known apple miRNA loci belonged to one of four classes: it could have the same sequence as the mature apple miRBase entry; it could have the same sequence as another plant homologue which can also be an isomiR of the apple miRBase entry; it could be the antisense-complement of the miRBase entry or an isomiR thereof; or it could have a sequence that varies significantly from the miRBase entry and therefore is classified as a novel miRNA. Novel miRNA loci had mature miRNAs, which belonged to one of three classes: it could have a sequence which is the same as another apple miRNA already present in miRBase and may therefore fall into the same family, it could have a sequence which is the same as a homologous plant miRBase entry; or it could have a sequence for which there is no exact plant sequence entry in miRBase.

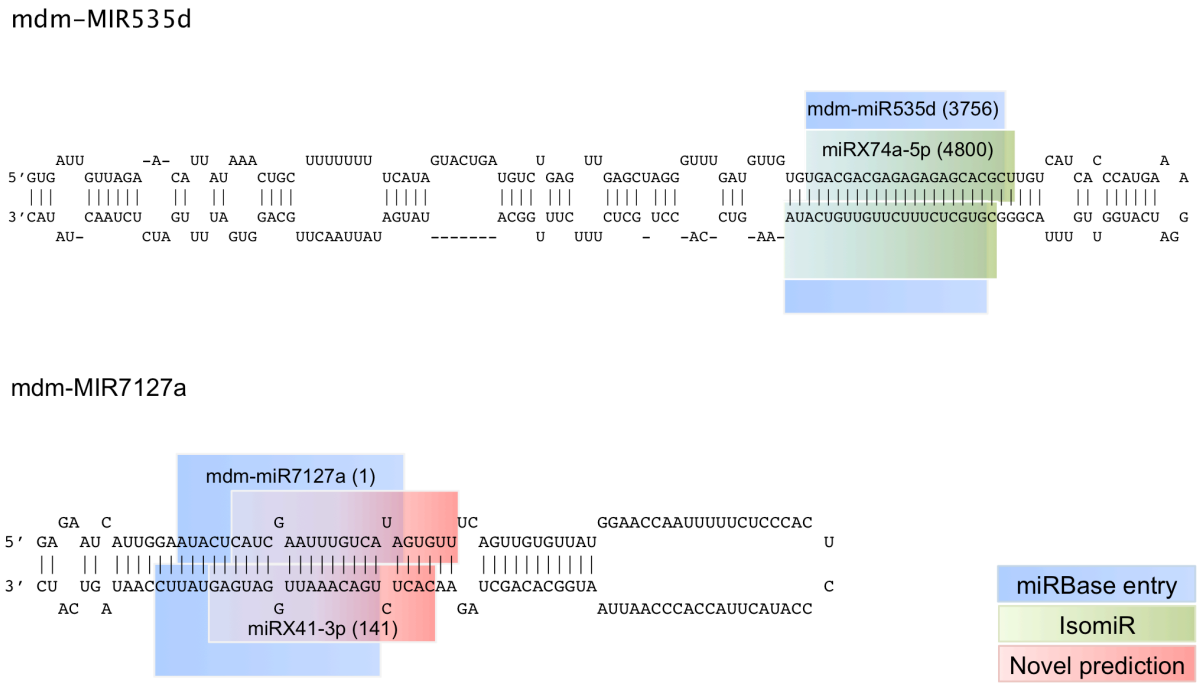


Figure 3.3. miRNA examples. (A) Example of a miRNA precursor for which the miRBase mature entry, as well as its isomiR was predicted as mature miRNAs. (B) Example of a miRNA precursor for which the miRBase mature entry was not predicted as a mature miRNA, but rather a mature miRNA varying significantly from the miRBase entry and was therefore classified as a novel miRNA. Read counts are given in brackets.

In addition to the known mdm-miRNA gene loci, 85 putatively novel precursor miRNA loci were identified. The mature miRNAs predicted for a few of these novel loci were the same as known mdm-miRNAs and can therefore be considered new members of the already known mdm-miRNA gene family. Along with novel miRNA loci also having a novel mature sequence, additional precursors were identified with predicted mature sequences homologous to miRNAs from other plant species.

Of the predicted miRNA loci, 33 overlapped with predicted apple transcripts. Although plant miRNAs predominantly originate from intergenic regions, it was demonstrated earlier that they can also be derived from gene introns, known as mirtrons [30], and exons [31].

In silico analysis with psRNATarget predicted targets for 217 of the novel miRNAs. 81.9% of all targets were predicted to be down-regulated through cleavage (Table S1E). Additional analysis with TargetFinder and CleaveLand, applying a Golden Delicious degradome sequencing dataset, resulted in the successful validation of 26

cleaved mRNA targets (Table S1F). A dataset generated from the cleaved RNAs of different Golden Delicious tissue types was used to validate miRNA targets (accession no. SRR413929) [5]. Despite the fact that the publicly available degradome dataset was generated from the same apple variety as the miRNA dataset in the present study, differences in environmental conditions may have prevented the validation of a larger number of computationally-predicted miRNA targets. Another limitation of the degradome dataset is the fact that it was generated from a range of different plant tissues, which may have caused the under-representation of leaf material in the sample.

3.3.4 NAT and nat-siRNA identification

Earlier studies have expounded the contribution of natural antisense transcript (NAT) siRNAs (nat-siRNAs) in plant development [32,33], disease resistance [34,35] and stress responses [19]. nat-siRNAs are processed from the overlapping region of transcript hybrids and in general down-regulate the expression of one of the transcripts involved in the duplex [36]. We have identified 1423 cis-NAT and 2198 trans-NAT pairs, of which 19 and 3 pairs, respectively, contained more than 1 overlapping region (Table S2A and S2B). The Genome Database for Rosaceae (GDR) [1,2] contains a total of 63541 predicted apple transcripts, of which 3752 were predicted to be part of various combinations of NAT pairs. Of all transcripts, 4.4% were involved in cis-NATs and 1.5% in trans-NATs. A small subset of transcripts (5% of all NATs) could form both kinds of NAT pairs (Figure 3.4), similar to what was found in other studies [37–39].

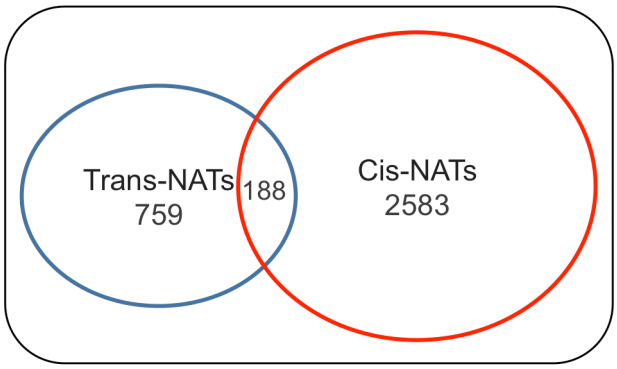


Figure 3.4. Transcripts forming cis- and trans-NAT pairs. Diagram illustrating the number of apple transcripts involved in either a cis- or trans-NAT relationship as well as the number of transcripts which are shared by the two groups of NATs.

A single transcript can be part of a duplex in a one-to-one (i.e. can form a duplex with only one other transcript), one-to-many or many-to-many relationship [37–40]. Figure 3.5 illustrates these criteria with reference to results from the present study. In our analysis, 81.5% of the NATs were involved in one-to-one, 4.6% in one-to-many and 13.9% in many-to-many bonds. These figures were 41.6%, 13.7% and 44.7% for trans-NATs, and 90.7%, 5.2% and 4.1% for cis-NATs, respectively. This indicates that NATs are part of a complex gene regulatory network in apple, similar to what has been observed in other plants [38]. Although these computationally-predicted NATs have the potential to hybridise *in planta,* concurrent expression in the same cellular location must occur for these duplexes to form.

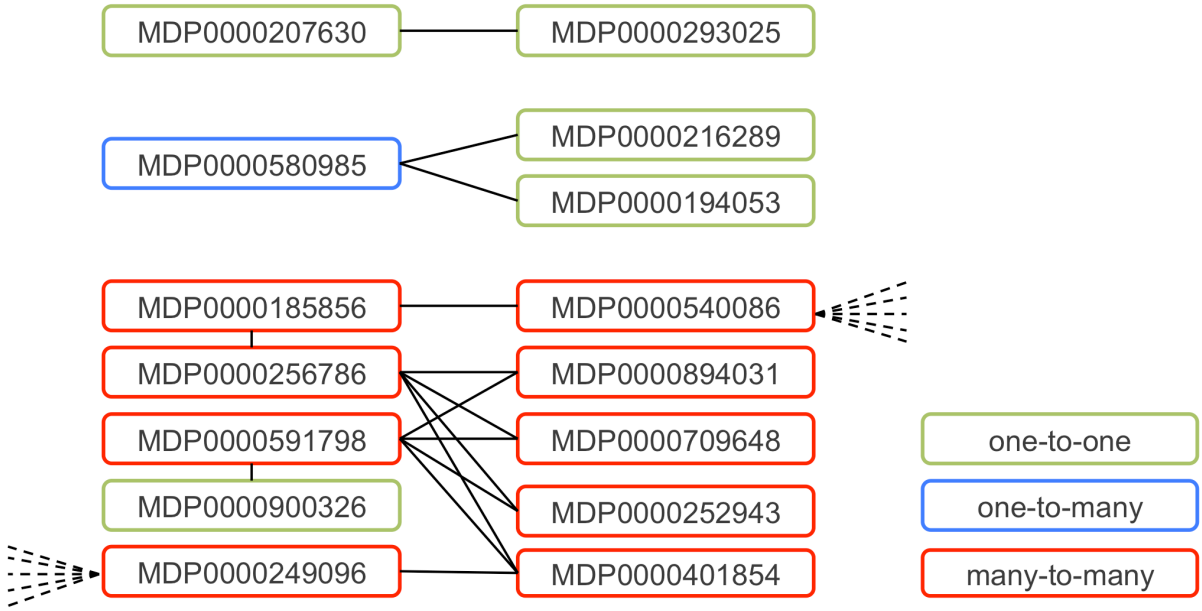


Figure 3.5. Natural antisense transcript networks. Diagram illustrating the three different relationships a NAT can be involved in i.e. a one-to-one (green), one-to-many (blue) or many-to-may (red) relationship. Solid lines indicate NAT pairs while dashed lines indicate a NAT relationship with a transcript not shown in the diagram.

In order to determine whether the siRNA spawned from overlapping regions was not purely by chance, the sRNA density (number of reads per kb of transcript) of the overlaps was compared to that of the rest of the NATs. The median of the siRNA densities of the overlapping regions of cis-NATs was 6.7 reads/kb while the reads on the overlapping regions of the trans-NATs had a density of 299,700 reads/kb (Table 3.2). Previous studies demonstrated that cis-NATs of protein-coding genes generally yield low levels of sRNAs when compared to non-cis-NATs [41,42]. Mapping analysis

indicated that trans-NATs of apple are significantly enriched for sRNAs in their overlap regions when compared to the rest of the transcripts ($p < 2.2 \times 10^{-16}$). Two examples of this enrichment are given in Figure 3.6. Conversely, the cis-NATs did not produce more sRNAs from their overlaps, it rather showed a reduction in sRNA generation ($p < 2.2 \times 10^{-16}$). This is not in accordance with what has been found for other plant species by Zhou *et al*. and Henz *et al*. [38,41]. It is important to note that while these studies found an increase in sRNA production from the cis-NAT overlapping regions when compared to the non-overlapping regions of the NATs, they found no significant difference between siRNA production from cis-NATs and transcripts uninvolved in NAT formation. The latter study could not find any evidence to support the regulation of cis-NAT by siRNA more than was the case for any non-overlapping transcript.

Table 3.2. Natural antisense transcript summary.

	Cis-NAT	Trans-NAT	Unclassified
Pairs	1423	2198	812
Portion of total transcripts (%)	4.4	1.5	0.35
Overlap length (nt)[a]	367	124	116
One-to-one (%)	90.7	41.6	68.3
One-to-many (%)	5.2	13.7	2.7
Many-to-many (%)	4.1	44.7	29
Density[b] in overlap/ transcripts	6.7/23.7	299,700/1,172	305,300/1,600
Overlap enrichment (p-value)	No ($< 2.2 \times 10^{-16}$)	Yes ($< 2.2 \times 10^{-16}$)	Yes ($< 2.2 \times 10^{-16}$)
> 2-fold stand bias (%)	75.5	42.7	47.1

[a] Values indicate the median

[b] Reads/kb

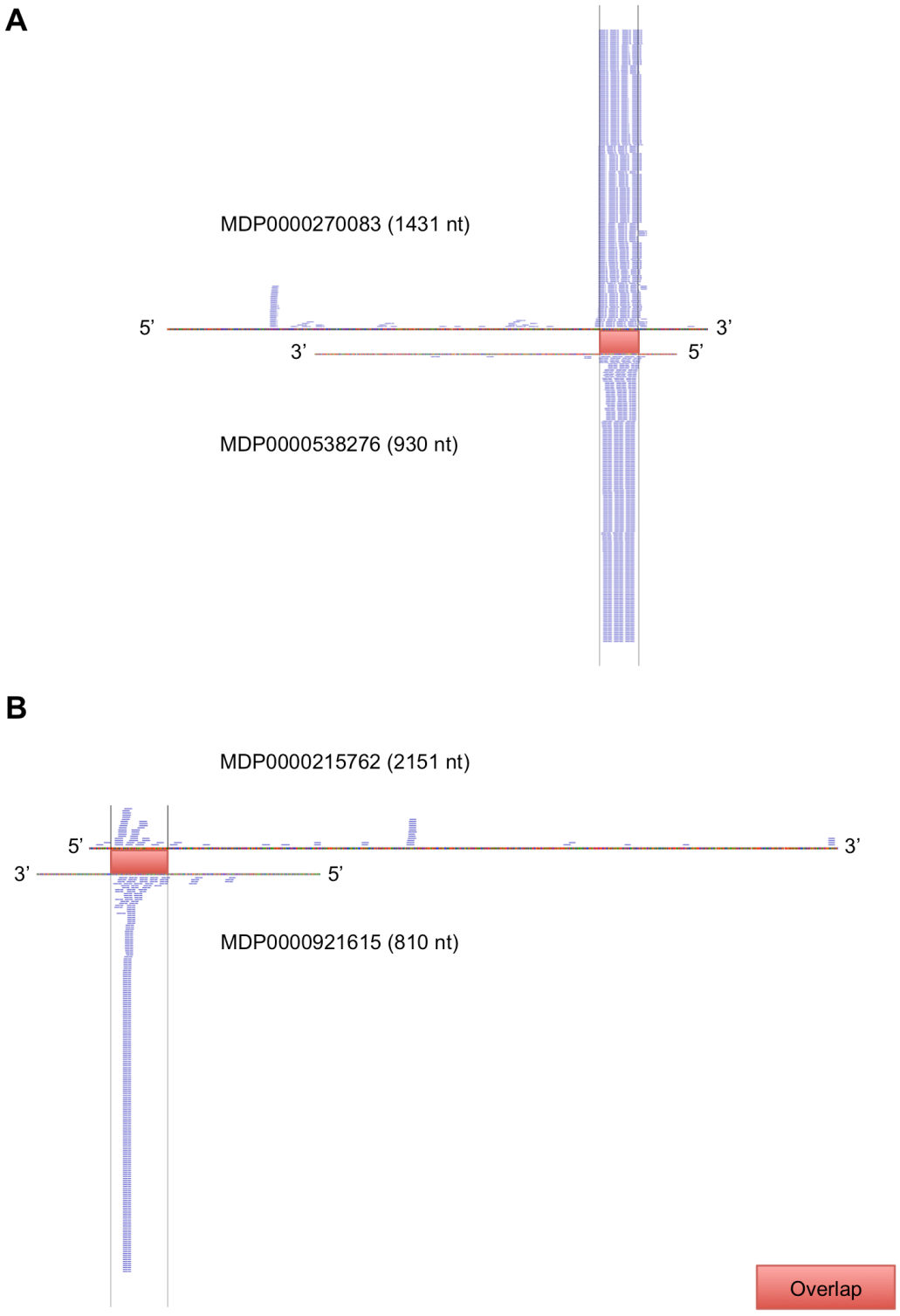


Figure 3.6. siRNA enrichment of trans-NAT overlaps. Illustration showing two trans-NAT pairs predominantly producing siRNAs from their overlapping regions.

To allow for the down-regulation of a transcript, the expression of the complementary transcript is first increased to stimulate nat-siRNA formation, which will target the constitutively expressed transcript [19,34,36,43]. This anti-correlation of transcripts [33,35,41] is not always observed [41]. Due to the absence of quantitative transcriptome data in the current study, it was not possible to analyse correlation of expression for paired transcripts. However, we did observe a strand-bias in the nat-siRNAs, of at least two-fold, for more than 53% of the NATs. Our data indicate that, for a significant number NAT pairs, the siRNAs were derived predominantly from one of the NATs, and thus suggests the preferred down-regulation thereof. These results support those of similar studies [35,38,39].

Although an earlier study has reported on apple trans-NATs [44], to our knowledge, this study is the first to report on apple cis-NATs and also the first to use annotated transcripts to investigate the production of NATs and nat-siRNAs in apple. The overlapping regions and nat-siRNAs identified here can be combined with transcriptome data in future studies to investigate gene regulation in which transcript hybridisation plays a central role. Besides cis- and trans-NATs a number of NAT pairs were also identified for which the chromosomal coordinates of at least one of the transcripts are unknown (Table S2C). These transcript pairs were therefore grouped into a third unclassified group, which followed the same trend as the trans-NAT group (Table 3.2).

3.3.5 *PHAS* identification and phasiRNA analysis

sRNAs produced from trans-acting siRNA genes (*TAS*), were first considered to only work in trans (hence the name tasiRNA). Subsequently, their cis-action was also suggested [45,46]. For this reason Zhai *et al*. introduce the term phasiRNA for all phased siRNAs, irrespective of whether they target other transcripts in trans or their own source [20]. The genes were called *PHAS* genes and included protein-coding as well as non-coding genes.

Trans-acting siRNA genes are involved in plant development [47,48], biotic stress [35] and abiotic stress [28,36,49]. Several *TAS* gene families have been recognized in diverse plant species, some conserved and some species-specific [45,50,51]. A recent

study by Xia *et al*. have identified and characterised the *TAS3* and *TAS4* families in apple [5]. They also discovered myeloblastosis (*MYB*) genes from which phasiRNAs were generated after cleavage by miR828. The same research group later characterised an additional novel *TAS* gene in apple, which they called Md-*TASL1*[52].

In the current study two approaches were followed to identify phased regions (clusters) in apple, by firstly implementing transcript data and secondly the genome. In total 157, 21 nt phased clusters were predicted to be statistically significant (Table S3A and S3B). Four of the transcripts were reported before to produce phasiRNAs, namely MDP0000578193 and MDP0000124555 [5], as well as MDP0000179176 and MDP0000302095 [52]. When the phased regions were examined using the NCBI BLAST database the majority of these aligned against disease-responsive genes, particularly genes belonging to protein families with nucleotide binding site leucine-rich repeats (NB-LRR) domains. The production of phasiRNAs was demonstrated earlier for NB-LRR protein families [20,50,53–55], as well as for pentatricopeptide repeat (PPR) [50,52,53,56], MYB [5,57,58] and Auxin Signalling F-Box (AFB) protein families [5,50]. All of these protein families also displayed phasiRNA generation in this study. Furthermore, besides being a source of phasiRNAs, pathogen resistance genes are also known to be targeted by this sRNA species [35,50,53,56,59], indicating the importance of phasiRNAs in integrated networks for gene regulation. BLAST hits included not only the above-mentioned protein-coding genes, but also *TAS3* gene homologues as expected.

When comparing the results for the two phasing analysis approaches, the genomic coordinates for a large number of phased transcript regions overlapped with that of phased genomic regions. This can be expected for phased regions, which do not span an intron. Not all transcripts are anchored onto the genome assembly used, which can additionally cause phased transcript regions to appear absent from the genomic results.

miRNAs can act as phase-initiators. After target cleavage dsRNA is formed through RDR activity followed by phasiRNA generation from the cleaved site [45]. To identify potential phase-initiators, miRNA targeting phased clusters were identified. Despite a number of miRNAs potentially targeting the phased clusters, only 26 had a miRNA

target cleavage site that fell into the dominant phasing register (Table S3C and S3D). Previous studies have also shown phased regions with a phase-initiating cleavage site being out-of-phase [5,50,52,57,60,61]. These instances, known as phase-drift [50,62], can be ascribed to DCL slippage leading to a slight shift in the phasing with regards to the cleaved site [50], or to the presence of an additional cleavage of the phased region by a phasiRNA [60,61]. In this study, the cleavage start sites for 10 clusters were validated with the apple degradome dataset (Table S3C and S3D). As mentioned previously, the differences in experimental conditions between the degradome sequencing and the sRNA-sequencing in this study may explain the lack of miRNA cleavage validation. Almost all (24 out of 29) initiator-miRNAs were 22 nt in length and had a uracil at the 5' end. The length of the sRNA initiator play a role in the phasing model. The phasing model can be based on either single (one-hit) or double (two-hit) miRNA target sites [60]. Although 22 nt miRNAs are mostly considered to be involved in phasing triggered by a single miRNA site [63], their association in the two-hit phase model was also demonstrated [20,52]. All the 22 nt miRNAs in this study complied to the one-hit phase model.

3.3.6 rasiRNA identification

This study demonstrated that 517 of the 524 repetitive sequence entries (Repbase *M. x domestica*) spawn sRNAs (Table S4). These entries included satellite DNA and integrated virus sequences, as well as retro- and DNA transposable element (TE) sequences. The largest cluster of reads (550,108 reads) mapped to retrotransposon-1 (RTE-1) followed by RTE-1B and DNA-transposon9-10. Generally, no particular strand-bias (more than 2-fold difference) was observed when investigating rasiRNAs mapping to the repeats. Some siRNA clusters, e.g. the long terminal repeat Copia-23, had a strong bias towards one of the repeat strands (more than 200-fold difference). The bulk (>50%) of rasiRNAs were 24 nt in length, a size-group frequently linked to heterochromatin-associated siRNAs [64]. Satellite 1 DNA is known to be associated with heterochromatin in the centromeres and other chromosomal regions [65]. Therefore it can be suggested that the siRNAs, which mapped to SAT1, can be classified as heterochromatic siRNAs. This group formed the sixth largest cluster of rasiRNAs (180,856). In our analysis, 9.2% and 8.5% of the rasiRNAs were 21 and 22

nt in length respectively and have also been implicated in TE silencing before [53,66,67].

3.4 Conclusion

The roles of small RNAs in gene regulation, and therefore in biological processes are being investigated for most important agricultural crops, including woody fruit crops. This study provides the most comprehensive single report on the sRNAs of apple. The apple miRNA database was extended significantly through the prediction of 85 novel miRNA precursor loci. Characterisation of these novel, and known loci, revealed mature miRNAs that were either known (mdm-miRNAs, miRBase 20), known plant homologues, or novel. Cis- and trans-NAT pairs, and the associated nat-siRNAs produced from their overlapping regions, were identified. Phased regions were identified at both the genome and transcriptome level. Besides non-coding loci, a number of protein-coding genes were shown to produce phasiRNAs. Finally, rasiRNAs for nearly all the apple repeat sequences in Repbase were identified.

This study, through NGS and computational analysis, identified a range of novel and known sRNA species in apple. Collectively they significantly add to the existing databases and will provide a platform for future functional studies in this important fruit crop.

3.5 Methods

3.5.1 NGS and sRNA dataset preparation

Sample material was collected from six, greenhouse grown, *M. x domestica* cv. 'Golden Delicious' (NIVV) seedlings, grafted onto MM.109 rootstocks. Total RNA was extracted from leaf material using the Plant RNA Reagent Kit (Invitrogen) and the small RNA fraction (17-29 nt) was purified from total RNA using a 15% TBE-urea polyacrylamide gel. Library preparation was performed by means of the TruSeq Small RNA library preparation kit from Illumina, and sequenced on an Illumina HiScan SQ instrument. The sequence data from the six libraries were pooled. The software cutadapt (V 1.0) [68] was applied to remove adapter sequences and the reads were

filtered for quality (phred score ≥ 20) using FASTX-toolkit (V 0.0.13, *http://hannonlab.cshl.edu/fastx_toolkit/index.html*). Only reads 17-26 nt in length were used for sRNA analysis. The adapter-trimmed libraries (unfiltered), of the separate six samples, were submitted to the NCBI-SRA database (accession no. SRR1136652 to SRR1136657).

3.5.2 miRNA analysis and target prediction

Known apple miRNAs as well as sRNAs homologous to known miRNAs of other species were identified using miRanalyzer (V 03/2012) [69,70]. To get an indication of the sRNA reads which represent an exact miRNA in the registry, no mismatches were allowed to miRBase entries, thus excluding any isomiRs. The "Plant mode" of ShortStack (V 0.4.1) [71] was used to perform novel miRNA prediction from sRNAs that were read-mapped, with a maximum of one mismatch to the *M. x domestica* genome primary pseudo-haplotype assembly (*M*. x *domestica* Whole Genome v1.0p) [1,2]. ShortStack filters predicted hairpin structures to identify miRNA precursors following the plant miRNA criteria as set by Meyers *et al*. [72]. It allows a maximum of 150 base pairs in a miRNA hairpin, a maximum of five unpaired nt in a mature miRNA duplex, unlimited loop length and a minimum fraction of 0.8 mappings within Dicer size range to annotate a locus as Dicer-derived.

Targets for the newly predicted miRNAs were first accessed using the web-based tool psRNATarget (http://plantgrn.noble.org/psRNATarget/) [73], applying default parameters. In an attempt to identify mRNA cleaved by the novel miRNA TargetFinder (V 1.6, http://carringtonlab.org/resources/targetfinder) along with CleaveLand (V 3.0.1) [74] was utilized to predict and validate miRNA cleavage sites.

To perform apple degradome analysis (also known as PARE: Parallel analysis of cDNA ends) the library preparation protocol by Jeong and Green (2012) [75] was followed using two different amounts of total RNA, 100 µg and 200 µg. The final library had an average size 110 nt in length as determined by polyacrylamide gel electrophoresis, which is ~20 nt shorter than expected. The integrity of the library was evaluated by Sanger sequencing of six samples. Five of these samples were individual clones, while the sixth represented a pooled subset of clones. Sequence analysis revealed the

complete expected library (adapters with flanking regions added during PCR amplification), without any inserts. Further attempts to generate a degradome library included increasing of adapter concentrations and replacing the phenol-chloroform clean-up and subsequent precipitation steps (after the short PCR amplification step) with a MinElute (Qiagen) clean-up step. Libraries generated after these modifications, still lacked any degradome inserts. The apple degradome library used for validation was, therefore, obtained from the NCBI-SRA database (accession no. SRR413929).

3.5.3 nat-siRNA identification

Cis- and trans-natural antisense transcripts were identified following a similar workflow to Zhou *et al*. [38]. Apple transcript sequences (*M*. x *domestica* v1.0 consensus CDS 300 flanking) were obtained from the GDR (http://www.rosaceae.org/species/malus/malus_x_domestica) [1,2]. Transcript sequences included coding regions as well as up to 300 nt up- and downstream. Duplex formation of overlapping genomic regions (>50 nt) formed by transcripts originating from opposite strands was validated using UNAfold (V 3.8) [76]. These hybridising molecules were considered to be cis-NAT pairs. Trans-natural antisense transcripts were identified by aligning the transcripts to each other using standalone BLAST (V 2.2.27+)[77]. The trans-NAT pairs were derived from diverse genomic regions, with an overlapping region of more than 100 nt having 100% identity. The same analysis, as for trans-NAT, was performed on transcripts for which the genomic region was unknown. UNAfold was again used to validate duplex formation.

The density of the sRNAs on the overlapping and non-overlapping regions of the NATs was compared to determine whether the overlapping regions of the NATs were significantly enriched with sRNAs. The density was determined by calculating the number of reads per kilobase of overlapping or non-overlapping NAT regions while the significance of the difference in densities was determined by mean of a Wilcoxon rank sum test.

3.5.4 Phased cluster and siRNA identification

Phased regions were identified using ShortStack [71], allowing a single mismatch of the sRNA to either the apple genome or computationally-predicted transcriptome. P-values were corrected for multiple testing and a Bonferroni adjusted significance level of 0.0034 or 0.001 was used for transcript and genomic analysis, respectively. Potential phase-initiating miRNAs were identified using psRNATarget [73] and cleavage at the target site was validated with an apple degradome sequencing dataset from the NCBI-SRA database (accession no. SRR413929) as described by Zhang *et al*. [51].

3.5.5 rasiRNA identification

miRanalyzer was used to identify rasiRNAs based on *M*. *x domestica* repeat sequences present in Repbase 17.12 [78,79]. After removing sequences that matched known mdm-miRNAs, a single mismatch was allowed between the sRNA read and the repeat sequence.

3.6 Supporting information

Table S1. miRNA results. The number of sRNA reads associated with apple miRBase entries and families, as well as other plant homologues. Predicted miRNA loci (with their properties) along with novel miRNA target prediction and degradome validation results are also given.
This document can be found online at:
http://www.plosone.org/article/fetchSingleRepresentation.action?uri=info:doi/10.1371/journal.pone.0095782.s001

Table S2. NAT results. Cis-, trans- and unclassified apple natural antisense transcript pairs with the sequence and coordinates of the overlapping regions.
This document can be found online at:
http://www.plosone.org/article/fetchSingleRepresentation.action?uri=info:doi/10.1371/journal.pone.0095782.s002

Table S3. phasiRNA results. Phased genomic and transcript regions with their properties such as phase-initiating miRNA, strandedness, phase-offset and alignment results for the region.

This document can be found online at:

http://www.plosone.org/article/fetchSingleRepresentation.action?uri=info:doi/10.1371/journal.pone.0095782.s003

Table S4. rasiRNAs results**.** The number of sRNA reads associated with both strands of apple repeat sequences in Repbase.

This document can be found online at:

http://www.plosone.org/article/fetchSingleRepresentation.action?uri=info:doi/10.1371/journal.pone.0095782.s004

3.7 Acknowledgments

The authors would like to thank Michael Hackenberg and Bernard Visser for bioinformatic support.

References

1. Jung S, Staton M, Lee T, Blenda A, Svancara R, *et al*. (2008) GDR (Genome Database for Rosaceae): integrated web-database for Rosaceae genomics and genetics data. Nucleic Acids Res 36: D1034–1040.
2. Velasco R, Zharkikh A, Affourtit J, Dhingra A, Cestaro A, *et al*. (2010) The genome of the domesticated apple (*Malus* × *domestica* Borkh.). Nat Genet 42: 833–839.
3. Kozomara A, Griffiths-Jones S (2011) miRBase: integrating microRNA annotation and deep-sequencing data. Nucleic Acids Res 39: D152–D157.
4. Krost C, Petersen R, Schmidt ER (2012) The transcriptomes of columnar and standard type apple trees (*Malus* x *domestica*) - a comparative study. Gene 498: 223–230.
5. Xia R, Zhu H, An Y, Beers EP, Liu Z (2012) Apple miRNAs and tasiRNAs with novel regulatory networks. Genome Biol 13: R47.
6. Krost C, Petersen R, Lokan S, Brauksiepe B, Braun P, *et al*. (2013) Evaluation of the hormonal state of columnar apple trees (*Malus x domestica*) based on high throughput gene expression studies. Plant Mol Biol 81: 211–220.
7. Mette MF, Aufsatz W, van der Winden J, Matzke MA, Matzke AJ (2000) Transcriptional silencing and promoter methylation triggered by double-stranded RNA. EMBO J 19: 5194–5201.
8. Hamilton AJ, Baulcombe DC (1999) A species of small antisense RNA in posttranscriptional gene silencing in plants. Science 286: 950–952.
9. Hammond SM, Boettcher S, Caudy AA, Kobayashi R, Hannon GJ (2001) Argonaute2, a link between genetic and biochemical analyses of RNAi. Science 293: 1146–1150.

10. Hammond SM, Bernstein E, Beach D, Hannon GJ (2000) An RNA-directed nuclease mediates post-transcriptional gene silencing in Drosophila cells. Nature 404: 293–296.

11. Verdel A, Jia S, Gerber S, Sugiyama T, Gygi S, *et al*. (2004) RNAi-mediated targeting of heterochromatin by the RITS complex. Science 303: 672–676.

12. Wu L, Zhou H, Zhang Q, Zhang J, Ni F, *et al*. (2010) DNA methylation mediated by a microRNA pathway. Mol Cell 38: 465–475.

13. Zakrzewski F, Weisshaar B, Fuchs J, Bannack E, Minoche AE, *et al*. (2011) Epigenetic profiling of heterochromatic satellite DNA. Chromosoma 120: 409–422.

14. Lee Y, Kim M, Han J, Yeom K-H, Lee S, *et al*. (2004) MicroRNA genes are transcribed by RNA polymerase II. EMBO J 23: 4051–4060.

15. Xie Z , Allen E, Fahlgren N, Calamar A, Givan SA, *et al*. (2005) Expression of Arabidopsis miRNA genes. PLANT Physiol 138: 2145–2154.

16. Reinhart BJ, Weinstein EG, Rhoades MW, Bartel B, Bartel DP (2002) MicroRNAs in plants. Genes Dev 16: 1616–1626.

17. 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.

18. Axtell MJ (2013) Classification and comparison of small RNAs from plants. Annu Rev Plant Biol 64: 137–159.

19. Borsani O, Zhu J, Verslues PE, Sunkar R, Zhu J-K (2005) Endogenous siRNAs derived from a pair of natural cis-antisense transcripts regulate salt tolerance in Arabidopsis. Cell 123: 1279–1291.

20. Zhai J, Jeong D-H, De Paoli E, Park S, Rosen BD, *et al*. (2011) MicroRNAs as master regulators of the plant NB-LRR defence gene family via the production of phased, trans-acting siRNAs. Genes Dev 25: 2540–2553.

21. Vazquez F, Vaucheret H, Rajagopalan R, Lepers C, Gasciolli V, *et al*. (2004) Endogenous trans-acting siRNAs regulate the accumulation of Arabidopsis mRNAs. Mol Cell 16: 69–79.

22. Hamilton A, Voinnet O, Chappell L, Baulcombe D (2002) Two classes of short interfering RNA in RNA silencing. EMBO J 21: 4671–4679.

23. Aravin AA, Lagos-Quintana M, Yalcin A, Zavolan M, Marks D, *et al*. (2003) The small RNA profile during *Drosophila melanogaster* development. Dev Cell 5: 337–350.

24. Taft RJ, Glazov EA, Lassmann T, Hayashizaki Y, Carninci P, *et al*. (2009) Small RNAs derived from snoRNAs. RNA 15: 1233–1240.

25. Li Y-F, Zheng Y, Jagadeeswaran G, Sunkar R (2013) Characterization of small RNAs and their target genes in wheat seedlings using sequencing-based approaches. Plant Sci 203–204: 17–24.

26. Wei H, Zhou B, Zhang F, Tu Y, Hu Y, *et al*. (2013) Profiling and identification of small rDNA-derived RNAs and their potential biological functions. PLoS ONE 8: e56842.

27. Li Y, Luo J, Zhou H, Liao J-Y, Ma L-M, *et al*. (2008) Stress-induced tRNA-derived RNAs: a novel class of small RNAs in the primitive eukaryote *Giardia lamblia*. Nucleic Acids Res 36: 6048–6055.

28. Hsieh L-C, Lin S-I, Shih AC-C, Chen J-W, Lin W-Y, *et al*. (2009) Uncovering small RNA-mediated responses to phosphate deficiency in Arabidopsis by deep sequencing. PLANT Physiol 151: 2120–2132.

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

30. Zhu Q-H, Spriggs A, Matthew L, Fan L, Kennedy G, *et al*. (2008) A diverse set of microRNAs and microRNA-like small RNAs in developing rice grains. Genome Res 18: 1456–1465.

31. 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.

32. 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.

33. Ron M, Alandete Saez M, Eshed Williams L, Fletcher JC, McCormick S (2010) Proper regulation of a sperm-specific cis-nat-siRNA is essential for double fertilization in Arabidopsis. Genes Dev 24: 1010–1021.

34. Katiyar-Agarwal S, Morgan R, Dahlbeck D, Borsani O, Villegas A, *et al*. (2006) A pathogen-inducible endogenous siRNA in plant immunity. Proc Natl Acad Sci 103: 18002–18007.

35. 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.

36. Moldovan D, Spriggs A, Yang J, Pogson BJ, Dennis ES, *et al*. (2010) Hypoxia-responsive microRNAs and trans-acting small interfering RNAs in Arabidopsis. J Exp Bot 61: 165–177.

37. Wang H, Chua N-H, Wang X-J (2006) Prediction of trans-antisense transcripts in *Arabidopsis thaliana*. Genome Biol 7: R92.

38. Zhou X, Sunkar R, Jin H, Zhu J-K, Zhang W (2009) Genome-wide identification and analysis of small RNAs originated from natural antisense transcripts in *Oryza sativa*. Genome Res 19: 70–78.

39. 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.

40. Lu T, Zhu C, Lu G, Guo Y, Zhou Y, *et al*. (2012) Strand-specific RNA-seq reveals widespread occurrence of novel cis-natural antisense transcripts in rice. BMC Genomics 13: 721.

41. Henz SR, Cumbie JS, Kasschau KD, Lohmann JU, Carrington JC, *et al*. (2007) Distinct expression patterns of natural antisense transcripts in Arabidopsis. PLANT Physiol 144: 1247–1255.

42. Zhan S, Lukens L (2013) Protein-coding cis-natural antisense transcripts have high and broad expression in Arabidopsis. PLANT Physiol 161: 2171–2180.

43. 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.

44. Chen D, Yuan C, Zhang J, Zhang Z, Bai L, *et al*. (2012) PlantNATsDB: a comprehensive database of plant natural antisense transcripts. Nucleic Acids Res 40: D1187–D1193.

45. Allen E, Xie Z, Gustafson AM, Carrington JC (2005) MicroRNA-directed phasing during trans-acting siRNA biogenesis in plants. Cell 121: 207–221.

46. Jagadeeswaran G, Zheng Y, Li Y-F, Shukla LI, Matts J, *et al*. (2009) Cloning and characterization of small RNAs from *Medicago truncatula* reveals four novel legume-specific microRNA families. New Phytol 184: 85–98.

47. Cho SH, Coruh C, Axtell MJ (2012) miR156 and miR390 regulate tasiRNA accumulation and developmental timing in *Physcomitrella patens*. Plant Cell 24: 4837–4849.

48. Johnson C, Kasprzewska A, Tennessen K, Fernandes J, Nan G-L, *et al*. (2009) Clusters and superclusters of phased small RNAs in the developing inflorescence of rice. Genome Res 19: 1429–1440.

49. 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 At1g51679 expression. Biosci Biotechnol Biochem 74: 1435–1440.

50. Howell MD, Fahlgren N, Chapman EJ, Cumbie JS, Sullivan CM, *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.

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

52. Xia R, Meyers BC, Liu Z, Beers EP, Ye S, *et al*. (2013) MicroRNA superfamilies descended from miR390 and their roles in secondary small interfering RNA biogenesis in eudicots. Plant Cell 25: 1555–1572.

53. Klevebring D, Street NR, Fahlgren N, Kasschau KD, Carrington JC, *et al*. (2009) Genome-wide profiling of Populus small RNAs. BMC Genomics 10: 620.

54. Li F, Pignatta D, Bendix C, Brunkard JO, Cohn MM, *et al*. (2012) MicroRNA regulation of plant innate immune receptors. Proc Natl Acad Sci 109: 1790–1795.

55. Kallman T, Chen J, Gyllenstrand N, Lagercrantz U (2013) A significant fraction of 21-nucleotide small RNA originates from phased degradation of resistance genes in several perennial species. PLANT Physiol 162: 741–754.

56. Chen H-M, Li Y-H, Wu S-H (2007) Bioinformatic prediction and experimental validation of a microRNA-directed tandem trans-acting siRNA cascade in Arabidopsis. Proc Natl Acad Sci 104: 3318–3323.

57. Zhu H, Xia R, Zhao B, An Y, Dardick CD, *et al*. (2012) Unique expression, processing regulation, and regulatory network of peach (*Prunus persica*) miRNAs. BMC Plant Biol 12: 149.

58. Rock CD (2013) Trans-acting small interfering RNA4: key to nutraceutical synthesis in grape development? Trends Plant Sci 18: 601–610.

59. Yoshikawa M, Peragine A, Park MY, Poethig RS (2005) A pathway for the biogenesis of trans-acting siRNAs in Arabidopsis. Genes Dev 19: 2164–2175.

60. Axtell MJ, Jan C, Rajagopalan R, Bartel DP (2006) A two-hit trigger for siRNA biogenesis in plants. Cell 127: 565–577.

61. Rajeswaran R, Aregger M, Zvereva AS, Borah BK, Gubaeva EG, *et al*. (2012) Sequencing of RDR6-dependent double-stranded RNAs reveals novel features of plant siRNA biogenesis. Nucleic Acids Res 40: 6241–6254.

62. De Paoli E, Dorantes-Acosta A, Zhai J, Accerbi M, Jeong D-H, *et al*. (2009) Distinct extremely abundant siRNAs associated with cosuppression in Petunia. RNA 15: 1965–1970.

63. Chen H-M, Chen L-T, Patel K, Li Y-H, Baulcombe DC, *et al*. (2010) 22-nucleotide RNAs trigger secondary siRNA biogenesis in plants. Proc Natl Acad Sci 107: 15269–15274.

64. Romanel E, Silva TF, Corrêa RL, Farinelli L, Hawkins JS, *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.

65. Pedrosa A, Jantsch MF, Moscone EA, Ambros PF, Schweizer D (2001) Characterisation of pericentrometric and sticky intercalary heterochromatin in *Ornithogalum longibracteatum* (Hyacinthaceae). Chromosoma 110: 203–213.

66. Slotkin RK, Vaughn M, Borges F, Tanurdžić M, Becker JD, *et al*. (2009) Epigenetic reprogramming and small RNA silencing of transposable elements in pollen. Cell 136: 461–472.

67. Wang X, Elling AA, Li X, Li N, Peng Z, *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.

68. Martin M (2011) Cutadapt removes adapter sequences from high-throughput sequencing reads. EMBnet.journal 17: 10–12.

69. Hackenberg M, Sturm M, Langenberger D, Falcón-Pérez JM, Aransay AM (2009) miRanalyzer: a microRNA detection and analysis tool for next-generation sequencing experiments. Nucleic Acids Res 37: W68–W76.

70. Hackenberg M, Rodríguez-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.

71. Axtell MJ (2013) ShortStack: Comprehensive annotation and quantification of small RNA genes. RNA 19: 740–751.

72. Meyers BC, Axtell MJ, Bartel B, Bartel DP, Baulcombe D, *et al*. (2008) Criteria for annotation of plant microRNAs. Plant Cell Online 20: 3186–3190.

73. Dai X, Zhao PX (2011) psRNATarget: a plant small RNA target analysis server. Nucleic Acids Res 39: W155–W159.

74. Addo-Quaye C, Miller W, Axtell MJ (2009) CleaveLand: a pipeline for using degradome data to find cleaved small RNA targets. Bioinformatics 25: 130–131.

75. Jeong D-H, Green PJ (2012) Methods for validation of miRNA sequence variants and the cleavage of their targets. Methods 58(2): 135–143.

76. Markham NR, Zuker M (2008) UNAFold: software for nucleic acid folding and hybridization. Methods Mol Biol Clifton NJ 453: 3–31.

77. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ (1990) Basic local alignment search tool. J Mol Biol 215: 403–410.

78. Jurka J, Kapitonov VV, Pavlicek A, Klonowski P, Kohany O, *et al*. (2005) Repbase Update, a database of eukaryotic repetitive elements. Cytogenet Genome Res 110: 462–467.

79. Kapitonov VV, Jurka J (2008) A universal classification of eukaryotic transposable elements implemented in Repbase. Nat Rev Genet 9: 411–412.

Chapter 4

High-throughput sequencing reveals small RNAs involved in ASGV infection

4.1 Abstract

Plant small RNAs (sRNAs) associated with virulent virus infections have been reported by previous studies, while the involvement of sRNAs in latent virus infection remains largely uncharacterised. Apple trees show a high degree of resistance and tolerance to viral infections. We analysed two sRNA deep sequencing datasets, prepared from different RNA size fractions, to identify sRNAs involved in *Apple stem grooving virus* (ASGV) infection. sRNA analysis revealed virus-derived siRNAs (vsiRNAs) originating from two ASGV genetic variants. A vsiRNA profile for one of the ASGV variants was also generated showing an increase in siRNA production towards the 3′ end of the virus genome. Virus-derived sRNAs longer than those previously analysed were also observed in the sequencing data. Additionally, tRNA-derived sRNAs were identified and characterised. These sRNAs covered a broad size-range and originated from both ends of the mature tRNAs as well as from their central regions. Several tRNA-derived sRNAs showed differential regulation due to ASGV infection. No changes in microRNA, natural antisense transcript siRNA, phased siRNA and repeat-associated siRNA levels were observed. This study is the first report on the apple sRNA-response to virus infection. The results revealed the vsiRNAs profile of an ASGV variant, as well as the alteration of the tRNA-derived sRNA profile in response to latent virus infection. It also highlights the importance of library preparation in the interpretation of high-throughput sequencing data.

4.2 Background

The domesticated apple, *Malus* x *domestica* (*M*. x *domestica*), has a wide range of infectious agents, which include fungi, bacteria, phytoplasma, viruses and viroids. One such virus, *Apple stem grooving virus* (ASGV), is the type member of the genus *Capillovirus* (family *Flexiviridae*) [1]. It is a positive-sense RNA virus with a genome of approximately 6.5 kb, which is organised into two overlapping open reading frames (ORFs) [2]. ASGV infection is mostly symptomless (latent) in apple cultivars, depending on the virus strain, however some cultivars are susceptible and may develop severe symptoms such as xylem pitting and grooving, phloem necrosis and the complete decay of the tree [3].

During infection the replication of RNA viruses generate long dsRNA intermediate molecules that triggers the synthesis of small interfering RNAs (siRNAs) [4]. Furthermore, the folded duplex regions of single stranded viral RNAs can also result in siRNA synthesis [5]. These virus-derived siRNAs (vsiRNAs) subsequently regulate viral RNA expression through a process known as RNA silencing. In addition to vsiRNA production, plants' endogenous small RNA (sRNA) pathways are also affected by viral infection [6–8].

With the introduction of next-generation sequencing the knowledge of sRNA species has been extended beyond the well-characterised miRNA, trans-acting siRNA (tasiRNA) and natural antisense transcript (NAT) siRNA (nat-siRNA) groups. Although sRNAs were shown to originate from tRNA before, Lee *et al*. (2009) was the first to illustrate that these molecules were not produced by non-systematic tRNA degradation [9]. Small RNAs associated with tRNAs have been divided into two categories based on the tRNA region they originate from. The first group, called tRNA halves (tsRNA/tiRNA), are derivatives of mature tRNAs cleaved in the anticodon loop, resulting in functional sRNAs of around 28 to 36 nucleotides in size. Enzymes involved in their biogenesis have been identified for humans [10], yeast [11] and bacteria [12], but are still unknown in plants.

Transfer RNA cleaved in the D or T loop give rise to a second group of sRNAs, called tRNA-derived RNA fragments (tRFs). This group can be further divided into sRNAs

stemming from (a) the 5' end of mature tRNAs, (b) the 3' end of mature tRNAs and (c) the 3' end of immature tRNAs, called 5'-tRFs, 3' CCA tRFs and 3' U tRFs respectively [13]. Several synonyms have been used for the different sub-groups [9,14].

In this study a next-generation sequencing approach was followed to identify sRNAs that are associated with a latent virus infection in apple plants. In addition to illustrating the vsiRNA profiles associated with an ASGV genetic variant the results from this study demonstrate the involvement of tRNA-derived sRNAs in plant-virus interaction. The lack of differential regulation of miRNAs, phasiRNAs, nat-siRNAs and rasiRNAs in leaf material is also shown.

4.3 Results and Discussion

4.3.1 sRNA sequencing libraries

Two library preparation approaches were followed. The first approach made use of total RNA to produce a broad range library (BRL) for each sample, with individual sequencing datasets comprising of between 7,543,861 and 11,648,479 reads. Reads of 27 nt and longer contributed to 73% of all BRL reads. Since sRNAs involved in gene silencing are mostly considered to fall within the 17 to 26 nt size-range, a second narrow range library (NRL) was prepared for each sample using size-selected sRNAs to increase the sequence depth of these sRNAs. These libraries generated 7,235,867 to 14,896,610 high quality reads per sample. The size-range 17 to 26 nt in length represented 97% of all the reads and were used for downstream analysis. Figure 4.1 illustrates the size distribution of the sRNA reads (1 to 50 nt in size) for the pooled BRL and pooled NRL datasets. The histogram not only highlights the increase in the percentage of reads 17 to 26 nt in length in the NRL datasets when compared to the BRL datasets, but also shows a change in ratios between the different size groups, in particular when comparing the ratio between the 21 and 24 nt groups. Since the same total RNA extract was used to prepare both libraries, this observation demonstrates the effect of library preparation on the final sequencing data and highlights the difficulty of comparing data generated by different protocols.

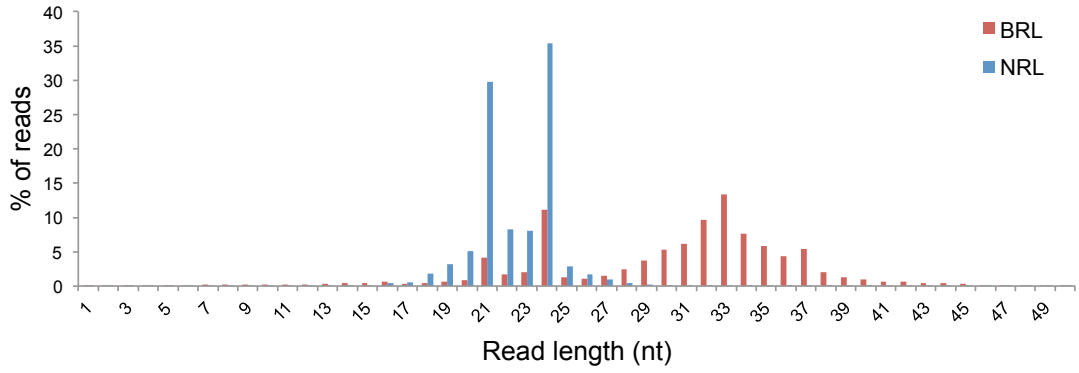


Figure 4.1. Size-distribution of the two sRNA sequencing library types. Histogram illustrating the number of reads, 1 to 50 nt in length, as a percentage of the reads in this size-range for the BRL and NRL data respectively.

4.3.2 vsiRNAs resulting from ASGV infection

The NRL data was first used to analyse the production of vsiRNAs, since these datasets were enriched for sRNAs in the size range known to be associated with vsiRNAs. Reads, which did not align to the apple nuclear, chloroplast or mitochondrial genomes, were mapped (allowing a single mismatch) onto the complete genomes of six ASGV isolates. sRNA read-mapping results for the pooled NRL sequencing data from the virus-infected samples are shown in Table 4.1. In total, 0.59% of all non apple-derived reads (17 to 26 nt in length) from the infected samples mapped onto at least one of the ASGV genomes. The large number of reads mapped onto the genome of the German isolate ASGV-AC, with 98% coverage. The majority of the virus-derived reads from the NRL data were 21 nt long followed by reads 22 nt in length (Figure 4.2), which is often seen for positive-sense RNA viruses [15,16].

Table 4.1. Results for the virus-infected NRL sRNA read-mapping against ASGV genomes.

Isolate	GenBank Accession number	Country	Host	Genome size (nt)	Total number of reads mapped	Non-redundant number of reads mapped	Genome coverage (%)
ASGV-AC	JX080201.1	Germany	*M. domestica*	6496	27069	5897	98.04
ASGVp12	HE978837.1	India	*M. domestica*	6478	25256	5297	96.34
ASGV P-209	NC_001749.2	Japan	*M. domestica*	6495	14341	3810	88.96
ASGV	D14995.2	Japan	*M. domestica*	6495	14341	3810	88.96
ASGV-HH	JN701424.1	China	*Pyrus pyrifolia*	6496	8591	1872	54.63
ASGV-CHN	JQ308181.1	China	*M. domestica*	6495	6555	1659	52.73

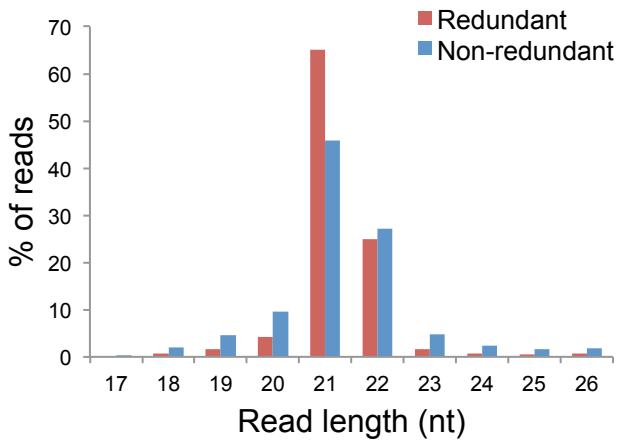


Figure 4.2. Size-distribution of NRL vsiRNA reads. Histogram illustrating the number of NRL vsiRNA reads, 17 nt to 26 nt in length, from the virus-infected samples, as a percentage of the reads in this size-range.

The occurrence of mixed ASGV infection was analysed using the genomes of three isolates (ASGV-AC, ASGV P-209 and ASGV-HH). These isolates each had an equal or higher sRNA read count (Table 4.1) than their closest relative (Figure 4.3). To determine the sRNA reads associated only with a specific variant genome, reads with a uniquely mapped position and genome were reported (Table 4.2). The large majority of variant-specific reads were associated with ASGV-AC, followed by the Japanese isolate (P-209). These variant-specific reads were distributed along the length of each of the three genomes (Figure 4.4), indicating that more than one ASGV variant was present with distinct genome sequences, rather than a single recombinant virus. Given their reasonably large number of total, as well as variant-specific reads, we suggest that at least two ASGV genetic variants, closely related to ASGV-AC and P-209 respectively, were present in the samples. Closer assessment of the read-mapping profiles for the individual samples suggested that two samples contained a mixed infection of the two variants, while the third was singly infected with a genetic variant of ASGV-AC.

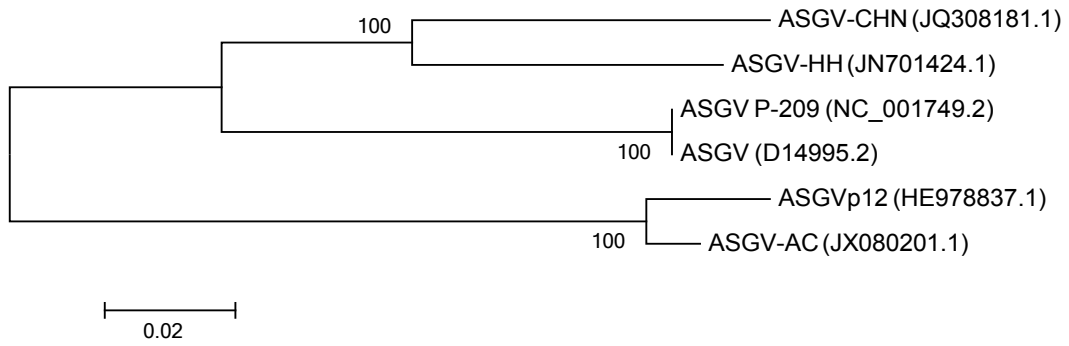


Figure 4.3. Phylogenetic tree based on the complete genome sequence of ASGV isolates. A neighbour joining method was applied and 1000 bootstrap replicates were used for the calculation of branch support. The branch length represents the number of substitutions per nucleotide position is indicated by the scale bar.

Table 4.2. Results for the vsiRNA variant-specific read-mapping. The number of reads which mapped only onto a specific ASGV genome for each virus-infected sample as well as for the pooled data are shown. The non-redundant read counts are given within brackets.

Isolate	GenBank Accession number	ASGV-infected sample 1	ASGV-infected sample 2	ASGV-infected sample 3	All ASGV-infected samples
ASGV-AC	JX080201.1	4074 (1326)	4905 (1544)	6507 (1832)	15486 (3111)
ASGV P-209	NC_001749.2	2469 (768)	2448 (703)	83 (20)	5000 (1120)
ASGV-HH	JN701424.1	269 (95)	267 (118)	141 (34)	677 (191)

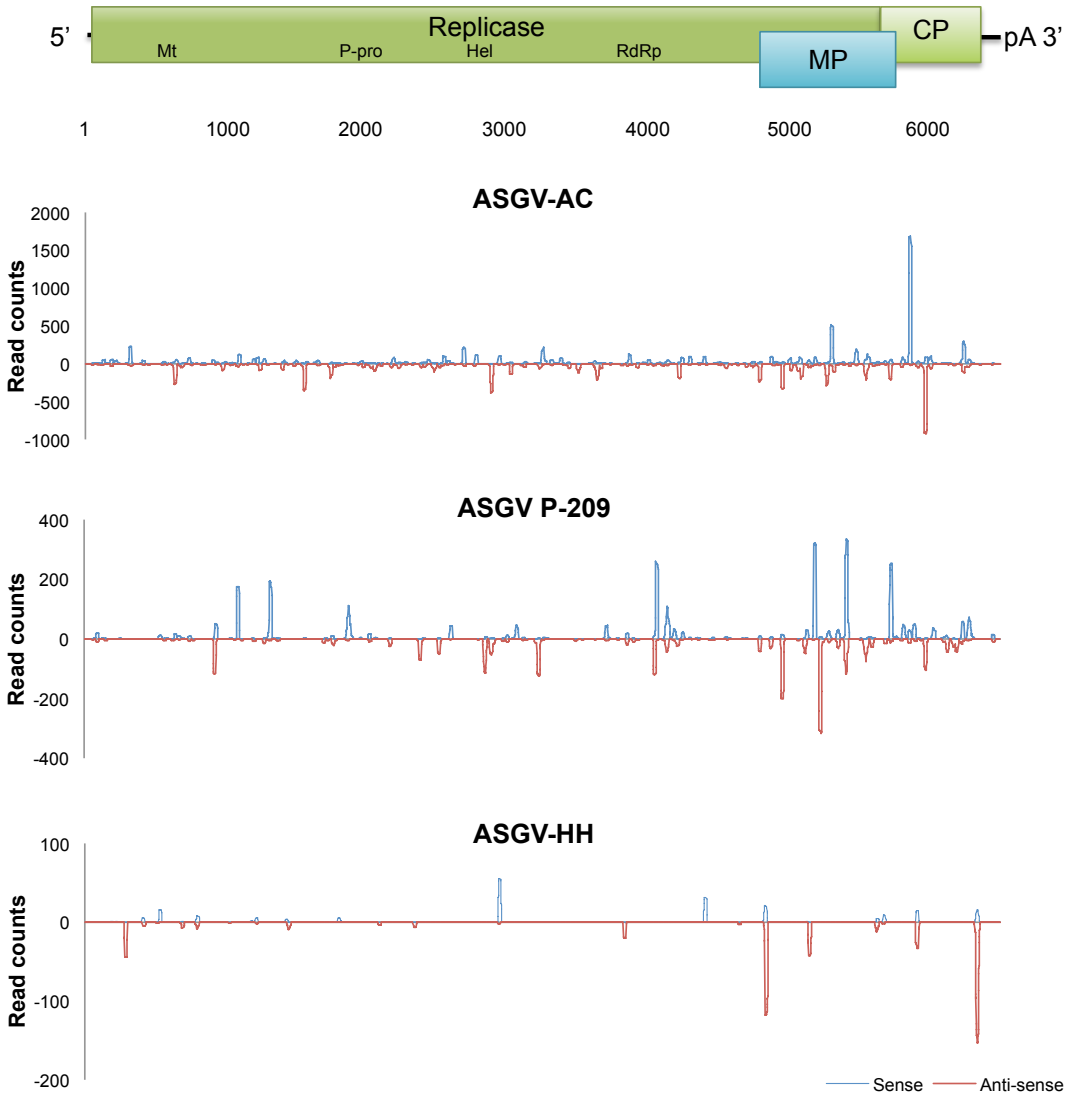


Figure 4.4. Distribution of variant-specific sRNA reads along ASGV genomes. The genomic positions of reads that could only map onto the genome of a single ASGV isolate are illustrated. Reads that mapped onto the positive or negative strand of the virus are represented in blue and red respectively. A schematic representation of the genome above the graphs illustrates the genomic positions of the vsiRNA reads.

Since only one sample was confirmed to be infected with a single ASGV variant, only reads from this sample could be used to generate a complete vsiRNA profile of this variant. Figure 4.5 shows the mapping distribution of the vsiRNA reads along the virus genome. In general the 3′ end of the genomes showed regions of higher genome coverage by vsiRNAs. The increase in vsiRNAs production toward the 3′ end of the genome has previously been ascribed to the presence of viral subgenomic RNAs (sgRNAs) [17,18]. Both the ASGV movement and coat proteins are expressed from 3′ sgRNAs [19–21] and may explain the increase in vsiRNA originating from the 3′ end. The non-redundant reads were also plotted onto the ASGV genomes (Figure 4.5). The majority of the genome sequences were associated with the production of more than one unique vsiRNA, illustrating that multiple Dicer-like (DCL) cleavage sites are in close proximity to each other on a virus genome.

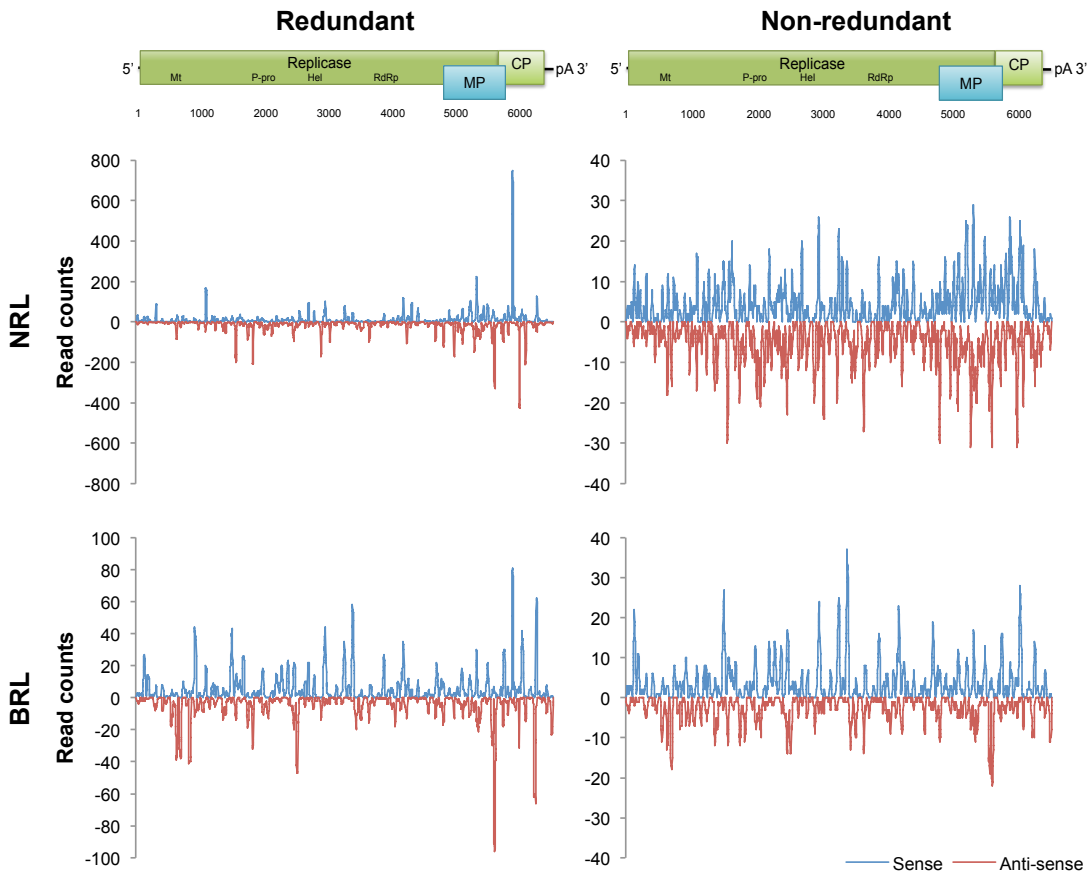


Figure 4.5. NRL and BRL vsiRNA profiles of an ASGV-AC genetic variant. vsiRNA profiles generated from redundant and non-redundant reads of both NRL and BRL datasets are depicted. Reads that mapped onto the positive or negative strand of the virus are represented in blue and red respectively. A schematic representation of the genome above the graphs illustrates the genomic position of the vsiRNA reads.

Altogether, from the three BRL datasets of the infected samples, 0.24% of the non apple-derived reads (>16 nt in length) mapped onto the six ASGV genomes. Although less than the NRL viral reads, these reads still covered 97% and 82% of the ASGV-AC and P-209 genomes, respectively (Table 4.3).

Table 4.3. Results for the virus-infected BRL sRNA read-mapping against ASGV genomes.

Isolate	GenBank Accession number	Genome size (nt)	Total number of reads mapped	Non-redundant number of reads mapped	Genome coverage (%)
ASGV-AC	JX080201.1	6496	7795	3751	97.44
ASGVp12	HE978837.1	6478	6628	3201	93.96
ASGV P-209	NC_001749.2	6495	4456	2059	82.03
ASGV	D14995.2	6495	4456	2059	82.03
ASGV-HH	JN701424.1	6496	1945	886	42.13
ASGV-CHN	JQ308181.1	6495	1341	721	39.98

Similar to the NRL datasets, the 21 nt long reads also dominated the viral reads in the BRL datasets (Figure 4.6). The second most abundant size group was the 22-nt group, closely followed by reads 33 nt in length. To our knowledge this is the first report of plant virus-derived sRNA reads larger than 30 nt in length. These larger sRNAs contributed significantly to the number of virus-associated reads and may point towards their biological importance. Alternatively, these reads possibly represent remnants of siRNA-directed ASGV genome degradation. The distribution of BRL reads along the ASGV genome was also examined (Figure 4.5). The presence of a substantial number of larger sRNAs in the BRL data resulted in a change in the vsiRNA profiles. The dominant areas of higher coverage by the conventional vsiRNAs (as can been seen from the NRL data) are surpassed (in the BRL data) by the additional areas of higher coverage, which are generated by the longer vsiRNAs. Furthermore, the 3′ vsiRNA bias was also less evident in the BRL data, compared to the NRL data. This observation once again demonstrates the effect of library preparation on sequencing results and the interpretation thereof.

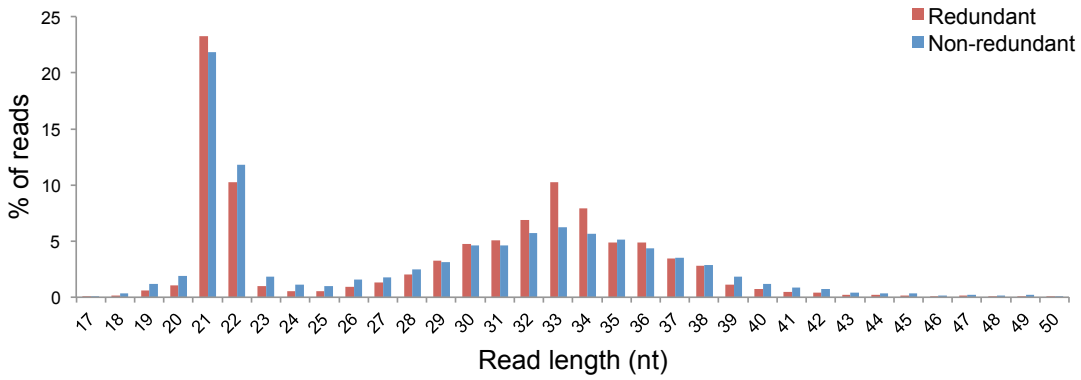


Figure 4.6. Size-distribution of BRL vsiRNA reads. Histogram illustrating the number of BRL vsiRNA reads, 17 nt and longer, from the virus-infected samples as a percentage of the reads in this size-range.

4.3.3 tRNA-derived sRNAs show differential regulation due to ASGV infection

Previous studies have established that sRNAs are generated from tRNAs in a non-random manner and that they play a regulatory role similar to other sRNA species [9,22,23]. In the BRL data the tRNA-derived sRNAs represented 23% and 19% of the reads larger than 16 nt for the infected and healthy samples respectively. These sRNAs varied in size from 17 to 59 nt in length, representing both tRFs and tRNA-halves. The broad size-range of the tRNA-derived sRNA reads in this study demonstrates that it is not always possible to clearly distinguish between these two classes only based on sequence length and origin. The larger species also stretched beyond the recognized tRNA-half size-range, spanning the anticodon loop, similar to previous reports [24,25]. The majority of tRNA-derived sRNAs in apple were 33 nt in length followed by reads of 32 and 37 nt (Figure 4.7). The dominant single tRNA-derived sRNA was a 5′ tRNA-half (33 nt long) originating from tRNA-Asp^{GTC}, and was represented by a total of 1,814,310 reads in the BRL datasets. In contrast, for the NRL dataset only 1.6% of all reads (17 tot 26 nt), originated from tRNAs.

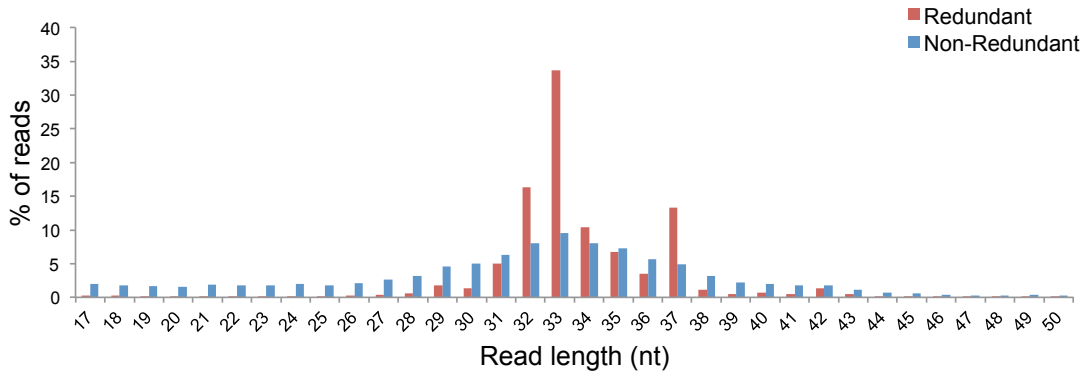


Figure 4.7. Size-distribution of BRL tRNA-derived sRNA reads. Histogram illustrating the number of BRL tRNA-derived sRNA reads, 17-50 nt in length, as a percentage of the reads in this size-range.

sRNAs, originating from both 5′ and 3′ ends of mature tRNAs, were identified in datasets from both library types (Additional file 1: Tables S1 and S2). Additionally, and in agreement with previous studies [23,26], sRNAs were also identified originating from the central part of tRNAs. These internal species were especially prominent in the cluster of sRNAs (in the BRL data) spawning from tRNA-Gln^{CTG}.

When the potential involvement of tRNA-derived sRNAs in ASGV infection was investigated, several tRNA-derived sRNAs showed significant variation in expression levels between infected and healthy samples (Additional file 2: Tables S3 to S6). Not only did individual sRNAs show differential expression, but the total number of sRNAs spawned by some of these tRNAs was found to significantly vary between the two groups. One tRNA, tRNA-Tyr^{GTA}, in particular displayed an interesting altered sRNA arrangement in the ASGV-infected samples (Figure 4.8). The BRL data revealed an increase in sRNAs derived from its 3′ end, extending into the variable region, while both BRL and NRL datasets showed a decrease in sRNAs that were generated from the central part of the tRNA. The 5′ ends of these internal sRNAs consistently coincided with the 5′ ends of the anticodon stems and extended into the variable regions. Due to the uniformity of their 5′ ends, these internal sRNAs cannot originate as a result of random tRNA degradation. From the inverse regulation of the two fragment types, it can be argued that a single tRNA molecule did not give rise to both species, but rather that they were generated through separate pathways. However, the possibility exists that these two processes are linked and that the production of the one species affects that of the other.

A

Chloroplast tRNA-Tyr^{GTA}

	Acc Stem	P8,P9	D Stem	D Loop	D Stem	p26	Ac Stem	Ac Loop	Ac Stem	V Region	T Stem	T Loop	T Stem	Acc Stem	p73	CCA	(counts)
	GGGTCGA	TG	CCCG	AGCGGTTAA	TGGG	G	ACGGA	CTGTAAA	TTCGT	TGGCAATATGTCTAC	GCTGG	TTCAAAT	CCAGC	TCGGCCC	A	CCA	
BRL										C	GCTGG	TTCAAAT	CCAGC	TCGGCCC	A	CCA	(217)
										AC	GCTGG	TTCAAAT	CCAGC	TCGGCCC	A	CCA	(158)
										TAC	GCTGG	TTCAAAT	CCAGC	TCGGCCC	A	CCA	(722)
										CTAC	GCTGG	TTCAAAT	CCAGC	TCGGCCC	A	CCA	(588)
										TCTAC	GCTGG	TTCAAAT	CCAGC	TCGGCCC	A	CCA	(838)
										TGTCTAC	GCTGG	TTCAAAT	CCAGC	TCGGCCC	A	CCA	(37313)
										TGTCTAC	GCTGG	TTCAAAT	CCAGC	TCGGCCC	A	CC	(747)
										ATGTCTAC	GCTGG	TTCAAAT	CCAGC	TCGGCCC	A	CCA	(686)
										TATGTCTAC	GCTGG	TTCAAAT	CCAGC	TCGGCCC	A	CCA	(927)
										ATATGTCTAC	GCTGG	TTCAAAT	CCAGC	TCGGCCC	A	CCA	(216)
							ACGGA	CTGTAAA	TTCGT	TGGCA (372)							
							ACGGA	CTGTAAA	TTCGT	TGGCAATATG (523)							
							ACGGA	CTGTAAA	TTCGT	TGGCAATATGT (287)							
							ACGGA	CTGTAAA	TTCGT	TGGCAATATGTCTAC	GC (275)						
	GGGTCGA	TG	CCCG	AGCGGTTAA	TGGG	G	ACGGA	CTGTAAA	TTCGT	TGGCAATA (786)							
NRL												AAT	CCAGC	TCGGCCC	A	CCA	(236)
												AAAT	CCAGC	TCGGCCC	A	CCA	(174)
												CAAAT	CCAGC	TCGGCCC	A	CCA	(867)
												TCAAAT	CCAGC	TCGGCCC	A	CCA	(473)
												TTCAAAT	CCAGC	TCGGCCC	A	CCA	(323)
											G	TTCAAAT	CCAGC	TCGGCCC	A	CCA	(583)
											AC	GCTGG	TTCAAAT	CCAGC	TCGGC (125)		
												G	TTCAAAT	CCAGC	TCGGC (486)		
							ACGGA	CTGTAAA	TTCGT (327)								
							ACGGA	CTGTAAA	TTCGT	TGGC (3260)							
							ACGGA	CTGTAAA	TTCGT	TGGCA (3227)							
							ACGGA	CTGTAAA	TTCGT	TGGCAA (215)							
							ACGGA	CTGTAAA	TTCGT	TGGCAAT (154)							
							ACGGA	CTGTAAA	TTCGT	TGGCAATA (487)							
	GGGTCGA	TG	CCCG	AGCGGTTAA	TG (589)												

B

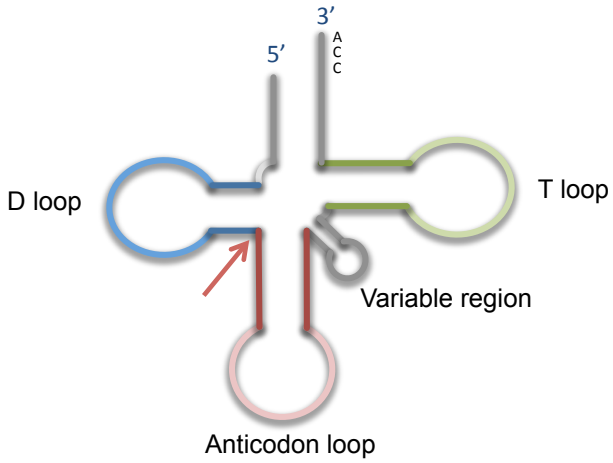


Figure 4.8. Variation in tRNA-derived sRNA profiles. A) Diagram showing the sRNA reads with the highest read count for each of the two types of data, which are associated with the chloroplast tRNA-Tyr^{GTA}. The sRNAs, which were up- or down-regulation due to ASGV infection, are indicated in red and blue respectively. The total read count of each sRNA is indicated within brackets. B) The red arrow illustrates the 5′ start position of a cluster of central sRNAs, originating from tRNA-Tyr^{GTA}, which are down-regulated during ASGV infection.

The biogenesis of tRNA-derived sRNAs, as well as the way in which they affect other molecular pathways remains to be elucidated. Earlier reports speculated that tRFs bind, to ribosomes resulting in a down-regulation of gene expression [27]. Through their association with argonaute proteins a possible role in post-transcriptional gene silencing was also suggested [23]. The biological function of the differentially regulated tRNA-derived sRNAs in the current study remains to be determined.

4.3.4 The involvement of other endogenous sRNAs in ASGV infection

Besides the vsiRNAs and tRNA-derived sRNAs involved in ASGV infection, differential expression analysis showed no variation in phasiRNA and miRNA levels as a result of ASGV infection; neither did the nat-siRNAs or rasiRNAs show any change in expression levels (Additional file 2: Tables S7 to S17). In addition to their regulatory role during plant development, miRNAs are often linked to stress response. The latent nature of ASGV may therefore explain what seems to be a lack of miRNA involvement during ASGV infection.

4.4 Conclusions

In this study next-generation sequencing of sRNAs was used to investigate plant responses to latent virus infection. Two different sRNA libraries were generated per sample. Both datasets illustrated the synthesis of virus-derived sRNAs in response to ASGV infection. Along with earlier reported tRNA-derived sRNAs of more than 30 nt in length, BRL data from this study additionally suggested virus-derived RNAs larger than the well-characterised vsiRNAs of around 21 nt. The vsiRNA profiles varied depending on the method of library preparation used, illustrating the importance of consistency when comparing different samples. Additionally, the results showed that ASGV-infection resulted in a change in the expression of tRNA-derived sRNAs, although the biological function of these sRNAs remains to be elucidated. This study is the first to report on sRNAs involved in ASGV-infection in the domesticated apple.

4.5 Methods

4.5.1 Sequencing library construction and data preparation

Sample material was collected from three healthy and three asymptomatic ASGV-infected (as confirmed by RT-PCR), greenhouse-grown, *M*. x *domestica* cv. 'Golden Delicious' (NIVV) seedlings, grafted onto MM.109 rootstocks. The viral status was confirmed by two multiplex RT-PCR reactions described in Menzel *et al*. [28]. The primers for *Apple mosaic virus* detection were replaced with those described in Hassan *et al*. [29]. See Table S18 in Additional file 3 for primer information. Total RNA

was extracted from mature leaf material using the Plant RNA Reagent Kit (Invitrogen) and used for library (BRL) preparation by means of the TruSeq Small RNA library preparation kit from Illumina. For each sample a second library (NRL) was prepared using the small RNA fraction (17-29 nt) purified from total RNA using a 15% TBE-urea polyacrylamide gel. The final BRL and NRL libraries were size-selected by means of a 3% Pippin Prep cassette (Sage) and a 6% polyacrylamide gel (Invitrogen), respectively, and sequenced on an Illumina HiScan SQ instrument. The software cutadapt (V 1.0) [30] was used to remove adapter sequences and the reads were filtered for quality (phred score ≥ 20) using FASTX-toolkit (V 0.0.13, *http://hannonlab.cshl.edu/fastx_toolkit/index.html*). For the NRL, reads less than 17 or longer than 26 nt in length were discarded, while all filtered reads 17 nt and longer were used for the analysis of the BRL data.

4.5.2 vsiRNA analysis

Reads from the three NRL virus-infected datasets were combined for vsiRNA analyses. Reads that could map with less than two mismatches onto the apple nuclear, chloroplast or mitochondrial genomes, obtained from the Genome Database for Rosaceae [31,32], were removed. Bowtie (V 0.12.7) [33] was used to perform all read-mapping analyses. The filtered reads were then mapped onto six ASGV genomes, allowing only a single mismatch. Similar analyses were performed for the pooled BRL virus-infected samples. Variant-specific reads were identified as those reads that uniquely mapped (using Bowtie) onto one of the six ASGV genomes, when only allowing perfect matches between the sRNA read and the genome.

4.5.3 tRF and tRNA-half identification

Mature tRNA sequences of five angiosperms (*Arabidopsis thaliana*, *Brachypodium distachyon*, *Medicago truncatula*, *Oryza sativa* and *Populus trichocarpa*) were retrieved from the PlantRNA database [34]. To identify apple tRFs present, the six NRL datasets were combined and mapped, with Bowtie, onto the retrieved mature tRNA sequences, allowing two mismatches. tRNA-halves were correspondingly identified using the pooled BRL datasets.

4.5.4 Differential expression analysis of apple sRNA species

The standalone differential expression tool of miRanalyzer [35,36], which implements the R package, DESeq2 [37], was used to determine variation in sRNA expression levels between the healthy and the ASGV-infected samples. Five distinct sRNA species were investigated using the NRL data, namely miRNAs, phasiRNAs, nat-siRNAs, rasiRNAs and tRFs. The BRL data was used for tRNA-halves differential expression analysis. miRNA analysis was based on miRBase (version 20) [38–41] apple entries, as well as recently predicted novel miRNAs [42]. The phasiRNAs, nat-siRNAs and rasiRNAs analysed were also previously identified [42], while the tRFs and tRNA-haves were identified during the current study. The phasiRNAs included a group of apple tasiRNAs available on the tasiRNAdb [43–45].

4.6 Availability of supporting data

The datasets supporting the results of this article are available in the BioProject repository of the National Centre for Biotechnology Information, BioProject: PRJNA235941 in http://www.ncbi.nlm.nih.gov/bioproject/.

4.7 Additional files

Additional file 1 – Supplemental Tables S1 and S2

Table S1. tRNA-derived sRNAs identified in the BRL data. The 15 sRNA reads with the highest read counts associated with a tRNA are shown and those differentially regulated (|log2fold change|>=1 and padj<=0.05) as a result of ASGV infection are indicated. **Table S2.** tRNA-derived sRNAs identified in the NRL data. The 15 sRNA reads with the highest read counts associated with a tRNA are shown and those differentially regulated (|log2fold change|>=1 and padj<=0.05) as a result of ASGV infection are indicated.
This document can be found online at:
www.biomedcentral.com.ez.sun.ac.za/content/supplementary/1471-2164-15-568-s1.xlsx

Additional file 2 – Supplemental Tables S3 to S17

Table S3. Results for the differential expression analysis of clusters of sRNAs originating from tRNAs, based on BRL data. **Table S4.** Results for the differential expression analysis of the individual tRNA-derived sRNAs, based on BRL data. **Table S5.** Results for the differential expression analysis of clusters of sRNAs originating from tRNAs, based on NRL data. **Table S6.** Results for the differential expression analysis of the individual tRNA-derived sRNAs, based on NRL data. **Table S7.** Results for the differential expression analysis of apple miRNAs present in miRBase. **Table S8.** Results for the differential expression analysis of recently predicted apple miRNAs. **Table S9.** Results for the differential expression analysis of the cluster of nat-siRNAs originating from both strands of the overlapping region of NAT pairs. **Table S10.** Results for the differential expression analysis of the cluster of nat-siRNAs, from the first transcript, originating from the overlapping region of NAT pairs. **Table S11.** Results for the differential expression analysis of the cluster of nat-siRNAs, from the second transcript, originating from the overlapping region of NAT pairs. **Table S12.** Results for the differential expression analysis of all the sRNAs originating from a phased cluster. **Table S13.** Results for the differential expression analysis of phasiRNAs. **Table S14.** Results for the differential expression analysis of tasiRNAs. **Table S15.** Results for the differential expression analysis of the cluster of rasiRNAs originating from both strands of a repetitive sequence. **Table S16.** Results for the differential expression analysis of the cluster of rasiRNAs originating from the forward strand of a repetitive sequence. **Table S17.** Results for the differential expression analysis of the cluster of rasiRNAs originating from the reverse strand of a repetitive sequence.

This document can be found online at:

www.biomedcentral.com.ez.sun.ac.za/content/supplementary/1471-2164-15-568-s2.xlsx

Additional file 3 – Supplemental Table S18

Table S18. Diagnostic RT-PCR primers. Multiplex-primers used to determine the viral status of apple plants.

This document can be found online at:

www.biomedcentral.com.ez.sun.ac.za/content/supplementary/1471-2164-15-568-s3.xlsx

4.8 Acknowledgements

The authors would like to acknowledge Bernard Visser and Michael Hackenberg for bioinformatics support, as well as the National Research Foundation (NRF) for their financial assistance towards this research. Opinions expressed and conclusions arrived at, are those of the authors and are not necessarily to be attributed to the NRF.

References

1. Adams MJ, Antoniw JF, Bar-Joseph M, Brunt AA, Candresse T, *et al*. (2004) The new plant virus family *Flexiviridae* and assessment of molecular criteria for species demarcation. Arch Virol 149: 1045–1060.
2. Yoshikawa N, Sasaki E, Kato M, Takahashi T (1992) The nucleotide sequence of apple stem grooving capillovirus genome. Virology 191: 98–105.
3. Nickel O, Fajardo TV, Jelkmann W, Kuhn GB (2001) Sequence analysis of the capsid protein gene of an isolate of *Apple stem grooving virus*, and its survey in Southern Brazil. Fitopatol Bras 26: 655–659.
4. Hamilton AJ, Baulcombe DC (1999) A species of small antisense RNA in posttranscriptional gene silencing in plants. Science 286: 950–952.
5. Molnar A, Csorba T, Lakatos L, Varallyay E, Lacomme C, *et al*. (2005) Plant virus-derived small interfering RNAs originate predominantly from highly structured single-stranded viral RNAs. J Virol 79: 7812–7818.
6. Romanel E, Silva TF, Corrêa RL, Farinelli L, Hawkins JS, *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.
7. Shivaprasad PV, Chen H-M, Patel K, Bond DM, Santos BACM, *et al*. (2012) A microRNA superfamily regulates nucleotide binding site-leucine-rich repeats and other mRNAs. Plant Cell 24: 859–874.
8. Xiao B, Yang X, Ye C-Y, Liu Y, Yan C, *et al*. (2014) A diverse set of miRNAs responsive to begomovirus-associated betasatellite in *Nicotiana benthamiana*. BMC Plant Biol 14: 60.
9. 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.

10. Fu H, Feng J, Liu Q, Sun F, Tie Y, *et al*. (2009) Stress induces tRNA cleavage by angiogenin in mammalian cells. FEBS Lett 583: 437–442.

11. Thompson DM, Parker R (2009) Stressing out over tRNA cleavage. Cell 138: 215–219.

12. Ogawa T, Tomita K, Ueda T, Watanabe K, Uozumi T, *et al*. (1999) A cytotoxic ribonuclease targeting specific transfer RNA anticodons. Science 283: 2097–2100.

13. Sobala A, Hutvagner G (2011) Transfer RNA-derived fragments: origins, processing, and functions. Wiley Interdiscip Rev RNA 2: 853–862.

14. Haussecker D, Huang Y, Lau A, Parameswaran P, Fire AZ, *et al*. (2010) Human tRNA-derived small RNAs in the global regulation of RNA silencing. RNA N Y N 16: 673–695.

15. Donaire L, Wang Y, Gonzalez-Ibeas D, Mayer KF, Aranda MA, *et al*. (2009) Deep-sequencing of plant viral small RNAs reveals effective and widespread targeting of viral genomes. Virology 392: 203–214.

16. 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.

17. Ruiz-Ruiz S, Navarro B, Gisel A, Peña L, Navarro L, *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.

18. Silva TF, Romanel EA, Andrade RR, Farinelli L, Østerås M, *et al*. (2011) Profile of small interfering RNAs from cotton plants infected with the polerovirus *Cotton leafroll dwarf virus*. BMC Mol Biol 12: 40.

19. Hirata H, Lu X, Yamaji Y, Kagiwada S, Ugaki M, *et al*. (2003) A single silent substitution in the genome of *Apple stem grooving virus* causes symptom attenuation. J Gen Virol 84: 2579–2583.

20. Tatineni S, Afunian MR, Gowda S, Hilf ME, Bar-Joseph M, *et al*. (2009) Characterization of the 5′- and 3′-terminal subgenomic RNAs produced by a capillovirus: Evidence for a CP subgenomic RNA. Virology 385: 521–528.

21. Komatsu K, Hirata H, Fukagawa T, Yamaji Y, Okano Y, *et al*. (2012) Infection of capilloviruses requires subgenomic RNAs whose transcription is controlled by promoter-like sequences conserved among flexiviruses. Virus Res 167: 8–15.

22. Li Z, Ender C, Meister G, Moore PS, Chang Y, *et al*. (2012) Extensive terminal and asymmetric processing of small RNAs from rRNAs, snoRNAs, snRNAs, and tRNAs. Nucleic Acids Res 40: 6787–6799.

23. 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: 6.

24. 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.

25. Nowacka M, Strozycki PM, Jackowiak P, Hojka-Osinska A, Szymanski M, *et al*. (2013) Identification of stable, high copy number, medium-sized RNA degradation intermediates that accumulate in plants under non-stress conditions. Plant Mol Biol 83: 191–204.

26. Wang Q, Lee I, Ren J, Ajay SS, Lee YS, *et al*. (2013) Identification and functional characterization of tRNA-derived RNA fragments (tRFs) in respiratory syncytial virus infection. Mol Ther 21: 368–379.

27. 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. Article ID 260909.

28. Menzel W, Jelkmann W, Maiss E (2002) Detection of four apple viruses by multiplex RT-PCR assays with coamplification of plant mRNA as internal control. J Virol Methods 99: 81–92.

29. Hassan M, Myrta A, Polak J (2006) Simultaneous detection and identification of four pome fruit viruses by one-tube pentaplex RT-PCR. J Virol Methods 133: 124–129.

30. Martin M (2011) Cutadapt removes adapter sequences from high-throughput sequencing reads. EMBnet.journal 17: 10–12.

31. Jung S, Jesudurai C, Staton M, Du Z, Ficklin S, *et al*. (2004) GDR (Genome Database for Rosaceae): integrated web resources for Rosaceae genomics and genetics research. BMC Bioinformatics 5: 130.

32. Jung S, Staton M, Lee T, Blenda A, Svancara R, *et al*. (2008) GDR (Genome Database for Rosaceae): integrated web-database for Rosaceae genomics and genetics data. Nucleic Acids Res 36: D1034–1040.

33. 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.

34. Cognat V, Pawlak G, Duchene A-M, Daujat M, Gigant A, *et al*. (2012) PlantRNA, a database for tRNAs of photosynthetic eukaryotes. Nucleic Acids Res 41: D273–D279.

35. Hackenberg M, Sturm M, Langenberger D, Falcón-Pérez JM, Aransay AM (2009) miRanalyzer: a microRNA detection and analysis tool for next-generation sequencing experiments. Nucleic Acids Res 37: W68–W76.

36. Hackenberg M, Rodríguez-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.

37. Love MI, Huber W, Anders S (2014) Moderated estimation of fold change and dispersion for RNA-Seq data with DESeq2. bioRxiv: http://dx.doi.org/10.1101/002832.

38. Griffiths-Jones S, Grocock RJ, Dongen S van, Bateman A, Enright AJ (2006) miRBase: microRNA sequences, targets and gene nomenclature. Nucleic Acids Res 34: D140–D144.

39. Griffiths-Jones S, Saini HK, Dongen S van, Enright AJ (2008) miRBase: tools for microRNA genomics. Nucleic Acids Res 36: D154–D158.

40. Kozomara A, Griffiths-Jones S (2011) miRBase: integrating microRNA annotation and deep-sequencing data. Nucleic Acids Res 39: D152–D157.

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

42. Visser M, van der Walt AP, Maree HJ, Rees DJG, Burger JT (2014) Extending the sRNAome of apple by next-generation sequencing. PLoS ONE 9: e95782.

43. Xia R, Zhu H, An Y, Beers EP, Liu Z (2012) Apple miRNAs and tasiRNAs with novel regulatory networks. Genome Biol 13: R47.

44. Xia R, Meyers BC, Liu Z, Beers EP, Ye S, *et al*. (2013) MicroRNA superfamilies descended from miR390 and their roles in secondary small interfering RNA biogenesis in eudicots. Plant Cell 25: 1555–1572.

45. Zhang C, Li G, Zhu S, Zhang S, Fang J (2014) tasiRNAdb: a database of ta-siRNA regulatory pathways. Bioinforma Oxf Engl 30: 1045–1046.

Chapter 5

Conclusion

The global importance of apples warrants the study of all factors that could influence production, including pathogens. Although *Apple stem grooving virus* (ASGV) infection is mostly latent in commercial apples, understanding the nature of its latency can lead to a better understanding of pathogen resistance. sRNAs are small functional RNA molecules which play an important role in gene regulation. The ultimate aim of this study was to identify sRNA species that are potentially associated with apple-virus infection. To evaluate the differential expression of sRNAs, a comprehensive database of known apple sRNAs had to be established. NGS and subsequent bioinformatic analysis were used to identify several members of different sRNA species in apple leaves. These included miRNAs, nat-siRNAs, phasiRNAs, rasiRNAs, tRNA-derived sRNAs and vsiRNAs.

MicroRNAs (miRNAs) were analysed by predicting *MIR* genes as well as their associated mature miRNAs. The sequences of some of the mature miRNAs predicted for known *MIR* genes differed from that of the miRBase entry, suggesting tissue- or environmental condition-specific processing of mature miRNAs from their precursors. Furthermore, for a few *MIR* genes more than one mature miRNA were predicted. Taken together, the miRNA prediction results have added a number of novel apple *MIR* genes and mature miRNAs to the current database (miRBase 21) and have highlighted aspects of their differential processing. Understanding the effect of different biological conditions on the miRNA biogenesis pathway may support the development of robust artificial miRNAs from which the desired mature miRNAs will be synthesised. Targets were also predicted for the novel miRNAs identified in this study, some of which could be validated using a publicly available apple-specific degradome dataset. This dataset, however, was generated from different source material (a combination of diverse apple tissues), which appeared to be suboptimal for the validation of targets predicted for leaf material, in the current study. While

some miRNAs may not guide the cleaving of their targets, other less-conserved miRNAs may not even have a target, which could also explain the lack of target validation. The characterisation of non-canonical miRNA-target interactions may in future shed additional light unto the investigation of these miRNA targets.

The evaluation of natural antisense transcripts (NATs) revealed a complex network of apple genes that can potentially interact with each other on a transcript level. Future investigations can determine which of the predicted NAT pairs are concurrently expressed in the same tissue(s) and therefore can affect the expression of others under specific biological conditions. While the role of sRNAs, produced from NATs, in the regulation of the gene expression is still slightly controversial, the current study characterised the expression patterns of nat-siRNAs originating from the overlapping regions of cis- and trans-NATs. From the results of this study, and in accordance with other studies, it seemed that cis-NATs produce a low number of siRNAs. The overlaps of these cis-NATs were also not enriched in nat-siRNAs when compared to the non-overlapping regions of the transcripts. In contrast, the overlaps of trans-NAT pairs were highly enriched for siRNAs. A thorough analysis of transcript expression, such as identifying concurrently expressed transcripts using NGS, may determine the cause of this observed difference in siRNA synthesis and overlap-siRNA enrichment between the two kinds of NATs. The majority of NAT pairs, found in this study, showed a siRNA-transcript-bias, which might strengthen the argument for siRNA involvement in the regulation of gene expression.

phasiRNA-producing regions were identified on genome as well as transcriptome level. miRNAs that could induce phasing were identified for a subset of these regions, some of which could be validated with the degradome dataset. Phasing in the absence of an in-phase miRNA target site may have been initiated through siRNA-directed cleaving or through the influence of secondary siRNA-directed cleaving on the dominant phasing register. In agreement with previous studies, many of the phased regions were associated with NBS-LRR disease resistance proteins, and are indicative of the potential importance of phasiRNAs in the regulation of plant disease responses. Once the criteria for phasing, along with the process of gene regulation resulting from phasing, are completely understood, these systems could also be manipulated to direct the down-regulation of specific genes.

sRNAs associated with nearly all apple repeat sequences (in Repbase) were identified. These rasiRNAs originated from satellite DNA, integrated virus sequences and TEs, and showed no strand-bias. Likewise, almost all mature tRNAs were found to give rise to sRNAs. Two sRNA datasets were used to analyse the two known types of tRNA-derived sRNAs, namely tRFs and tRNA-halves. The one dataset was generated from sRNAs of a narrow size range while the second applied a broader size range for library preparation. The tRNA-derived sRNAs identified originated predominately from the terminal regions of the tRNAs, while fragments related to the central regions were also present. Fragments longer than the conventional tRNA-halves were also identified, which is in agreement with previous studies. Together, the results from the tRNA-derived sRNAs analysis demonstrate the diversity of tRNA regions that can spawn sRNAs.

The analysis of vsiRNAs in ASGV-infected material suggested the presence of two ASGV genetic variants in two of the infected replicate samples, while the third seemed to be infected with only one variant. These analyses demonstrate the ability of sRNA sequencing to not only identify different virus species, but also to discriminate genetic variants within a species. Analysis of single-infections provides insight into viral RNA silencing strategies used in defence against different variants. With the aid of genomic information, sRNA sequencing also provides a potential diagnostic tool even in the case of infection by multiple variants. There are currently no complete genome sequences available for any South African ASGV isolates and therefore the vsiRNA analyses were based on whole genome sequences of isolates from other geographic regions. This study provides the first sequence evidence of mixed infections of ASGV genetic variants in South African apple orchards. Future studies remain to determine the extent of variation amongst local ASGV isolates. The two different data sets (broad and narrow range) were also applied to investigate distribution of vsiRNAs on an ASGV genome. The genome areas of higher read coverage varied between the two datasets. These results, along with variation in read size-distribution, emphasised the influence which sample preparation has on the interpretation of data, as well as the importance of applying consistent protocols when different samples are compared. While vsiRNAs are considered to be roughly 21 nt in length, virus-derived sRNAs that were considerably longer (~33 nt), were

also consistently detected in this study. Since sRNAs of similar size derived from tRNAs are considered to be functional sRNAs, these longer virus-derived sRNAs may potentially also be of functional significance. Although the pathways, which involve these longer sRNAs types, remain to be elucidated, their similar size would suggest that some aspects of their biogenesis and function might agree.

Following the identification of the different sRNA species in apple, differential expression analysis was performed to identify sRNAs, other than vsiRNAs, which are potentially associated with ASGV infection. In this study, tRNA-derived sRNAs were the only sRNA species that showed altered levels of expression as a result of virus infection. Even phasiRNAs generated from NBS-LRR disease resistance proteins did not significantly vary between the healthy and the infected samples. Despite the fact that the function of tRNA-derived sRNAs is not well characterised in plants, results from this study suggest that, along with vsiRNA synthesis, altered tRNA-derived sRNA levels are of importance during virus infection.

Next-generation sequencing allows for the analysis of sRNA species on a genomic scale. It generates a vast amount of sequence data, of which currently only a relatively small subset of the reads can be annotated. This poses questions about the origin of the majority of reads. Studies are still identifying novel sRNA species, which in future may explain many of the sRNA sequences in NGS data sets that cannot be grouped with any of the currently known species.

In future, functional studies will have to be performed in order to determine the exact role of tRNA-derived sRNAs during virus infection. Comparison of symptomatic and asymptomatic ASGV infected plants may shed more light on sRNAs that are specifically responsible for inhibiting symptom development. Since different plant species show varying degrees of symptom development, the sRNA expression during ASGV infection for different species and rootstock-scion combinations also remains open for investigation. This study presents the first report on different sRNAs associated with ASGV infection in apple. These results have contributed significantly to our knowledge on sRNAs produced in apples. With the importance of sRNAs in different biological processes continuously being demonstrated, the comprehensive

list of apple sRNAs generated, lays the foundation for a variety of future functional studies in apple.