Modulation of the *Vitis vinifera* cv. 'Chardonnay' microRNA and mRNA transcriptomes in response to aster yellows phytoplasma-infection

by

Marius Christian Snyman

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

Supervisor: Prof. Johan T. Burger
Co-supervisor: Dr. Dirk Stephan

April 2019

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.

Marius C. Snyman
April 2019

Abstract

Aster yellows (AY) phytoplasmas are part of a group of cell wall-less plant pathogenic bacteria responsible for a detrimental disease known as grapevine yellows (GY). The molecular mechanisms of AY phytoplasma pathogenicity on highly susceptible cultivars, such as Chardonnay, are still largely unknown. This has sparked considerable interest to gain knowledge about the basis of host susceptibility to GY in order to develop control strategies that may mitigate the scale of infection or even prevent spread. Leaf total RNA was extracted from both healthy and AY-infected plants to generate small RNA (sRNA) sequencing libraries, as well as mRNA sequencing libraries. These libraries were subjected to Illumina transcriptome sequencing (small RNA-seq and mRNA-seq, respectively), and comparative transcriptome profiling, to explore the involvement of microRNA (miRNA) and gene expression pathways in AY phytoplasma-infected Chardonnay. Multiple known miRNA sequence variants (isomiRs) were identified, and 13 known miRNAs were shown to be differentially expressed. A total of 175 novel miRNA precursor sequences, each derived from a previously uncharacterised genomic location, were identified, of which 23 were differentially expressed. Some of these novel miRNAs shared high sequence similarity with conserved miRNAs from other plant species, as well as known grapevine miRNAs. The relative expression of some of these known and novel miRNAs was confirmed with stem-loop RT-qPCR analysis, thereby validating the trend of miRNA expression in the normalised sRNA-seq read count data. miRNA target prediction, using a complementary-based *in silico* approach, followed by functional annotation, allowed the identification of potential target genes involved in plant morphology, hormone signalling, nutrient homeostasis, as well as plant stress. mRNA-seq results showed that 175 genes were differentially expressed in the AY phytoplasma-infected leaf material. Functional annotation of differentially expressed genes (DEGs) enabled the identification of mRNAs involved in plastid and cell wall metabolism/architecture, signalling, innate immunity, pathogen defence, secondary metabolism and photosynthesis. RT-qPCR analysis was used to validate the trend of expression of significant DEGs. Taken together, this study presents the first report on the modulation of miRNAs and genes associated with AY phytoplasma-infection in Chardonnay. The knowledge generated during this study may be crucial in understanding disease symptom development in AY phytoplasma-infected grapevines. Importantly, the findings of this study may also aid in developing GY disease control strategies and could provide added insight for future plant pathogenesis-related studies.

Opsomming

Fitoplasmas is selwandlose, patogeniese bakterieë waarvan die astervergelingfitoplasmas deel is. Astervergelingsfitoplasmas kan ‘n nadelige siekte in wingerd veroorsaak wat bekend staan as wingervergeling-siekte. Die onderliggende molekulêre meganismes betrokke by astervergelings-patogenisiteit in hoogsvatbare kultivars soos Chardonnay is meestal onbekend. Dit het gelei tot ‘n toenemende belangstelling om die basis van vatbaarheid deur die gasheerplant te begryp en sodoende voorkomingstrategieë teen die siekte te onwikkel. Blaar totale ‘RNA’ is geïsoleer uit gesonde, sowel as astervergelingsfitoplasma-geïnfekteerde plante, en gebruik om beide klein RNA (‘sRNA’) biblioteke asook boodskapper RNA (‘mRNA’) biblioteke te genereer vir volgende-generasie volgordebepaling. Illumina transkriptoom volgordebepaling (onderskeidelik ‘small RNA-seq’ en ‘mRNA-seq’) en relatiewe transkriptoom profielsamestellings is gebruik om die moontlike betrokkenheid van mikro-‘RNAs’ (‘miRNAs’) en gene in Chardonnay met wingervergeling-siekte te ondersoek. ‘n Groot aantal variante van bekende miRNAs (‘isomiRs’) is geïdentifiseer en daar is ook vasgestel dat 13 reeds bekende miRNAs differensieël uitgedruk is. 175 ‘nuwe’ miRNA volgordes is ook geïdentifiseer binne onbekende genoomareas waarvan 23 differensieël uitgedruk is. Nukleotiedvolgorde-analises is uitgevoer om eendersheid tussen hierdie ‘nuwe’ miRNAs en reeds bekende wingerd miRNAs asook ander plant miRNAs te bewys. ‘n Stam-lus tru-transkripsie kwantitatiewe polimerase kettingreaksie (‘stem-loop RT-qPCR’) metode is ingespan om die uitdrukkingspatrone van bekende en ‘nuwe’ miRNAs te beklemtoon. miRNA teikens is bepaal deur middel van ‘n komplementêr-gebaseerderde *in silico* metode. Daaropvolgende funksiebepalings het miRNA teikens geïdentifiseer wat betrokke is in plantmorfologie, hormoonregulering, voedingstofregulering, en plantstres. Die gebruik van ‘mRNA-seq’ het miljoene goeie-kwaliteit volgordes verskaf wat gebruik is om differensieële geenuitdrukking te bepaal. 175 differensieël uigedrukte gene (‘DEGs’) is deur middel van hierdie metode in the geïnfekteerde blaarmateriaal geïdentifiseer. Funksiebepalings het gewys dat die ‘DEGs’ verskillende rolle het in byvoorbeeld plastied- en selwandontwikkeling, boodskapverspreiding, immuniteit, patogeenbeskerming, sekondêre metabolisme en fotosintese. ‘RT-qPCR’ analise is gebruik om die uitdrukkingspatrone van die beduidende ‘DEGs’ te bevestig. Die bestaande studie is die eerste navorsingsprojek wat die identifisering en uitdrukkingsanalise van miRNAs en gene in astervergelingsfitoplasma-geïnfekteerde Chardonnay ondersoek. Die bogenoemde bevindinge kan gebruik word om kennis met betrekking tot simptoomontwikkeling weens wingerdvergeling-siekte te verbreed. Hierdie studie kan ook nuwe geleenthede en motiverings skep vir verdere diepgaande studies in hierdie veld en kan moontlik toegepas word in pogings om wingervergeling-siekteweerstand te bewerkstellig.

List of abbreviations

2-ODD	2-oxoglutarate/Fe(II)-dependent dioxygenases
ABA	abscisic acid
ABC	ATP-binding cassette
AFB	auxin signalling F-box proteins
AGO	Argonaute
AMFE	adjusted minimum folding free energy
amiRNA	artificial miRNA
AP2	APETALA2
ARF	auxin-responsive factor
AUSGY	Australian grapevine yellows
avr	avirulence
AY	aster yellows
AY-WB	aster yellows phytoplasma strain witches' broom
BAK1	brassinosteroid insensitive receptor kinase 1
bHLH	basic helix-loop-helix
BLAST	basic local alignment search tool
BN	‘Bois noir’
Ca.	*Candidatus*
CBC	cap-binding complex
TCP	Teosinte branched1/Cincinnata/Proliferating cell factor
CRISPR	clustered regularly interspaced short palindromic repeats
CSP	cold shock protein
CTAB	cetyl trimethylammonium bromide
DAPI	4',6-diamidino-2-phenylindole
DCL1	DICER-LIKE1
DDL	DAWDLE
DEG	differentially expressed gene
DFD	direct fluorescent detection
DLO2	DMR6-like oxygenase 2

DMR6	downy mildew resistance 6
dsRNA	double stranded RNA
EF1-α	elongation factor 1-alpha
EF-Tu	elongation factor thermo unstable
ELISA	enzyme-linked immunosorbent assay
ERF	ethylene-responsive factor
ETI	effector-triggered immunity
EY	elm yellows
F3H	flavanone 3-hydroxylase
FD	Flavescence dorée
FDR	false discovery rate
FPKM	fragments per kilobase of exon per million mapped reads
GAPDH	glyceraldehyde 3-phosphate dehydrogenase
GLRaV-3	Grapevine leafroll-associated virus 3
GO	gene ontology
GRSPaV	Grapevine rupestris stempitting-associated virus
GST	glutathione S-transferase
GTF	gene transfer format
GUI	graphical user interface
GVA	Grapevine virus A
GVE	Grapevine virus E
GY	grapevine yellows
H_2O_2	hydrogen peroxide
HD-ZIP	homeodomain leucine zipper
HEN1	HuaEnhancer 1
HPLC-DAD	high-performance liquid chromatography with a diode array detector
hpRNA	hairpin RNA
HR	hypersensitive response
HST	HASTY
HYL1	HYPONASTIC LEAVES1
IEM	immunosorbent electron microscopy
IMP	immunodominant membrane protein

ITS	internal transcribed spacer
JA	jasmonic acid
LNA	locked nucleic acids
MAM	mollicutes adhesin motif
MAMP	microbial-associated molecular pattern
MAPK	mitogen-activated protein kinase
MAPK	mitogen-activated protein kinase
MBSP	maize bushy stunt phytoplasma
MeSA	methyl salicylate
MFEI	minimal folding free energy index
miRNA	microRNA
MLO	mycoplasma-like organism
mRNA	messenger RNA
mRNA-seq	mRNA transcriptome sequencing
NAGY	North American grapevine yellows
nat-siRNA	natural antisense transcript siRNA
NBS-LRR	nucleotide-binding site leucine-rich repeat
NGS	next generation sequencing
NPR1	nonexpressor of pathogenesis-related genes 1
OY	onion yellows
PAMP	pathogen-associated molecular pattern
PARE	parallel analysis of RNA ends
PFGE	pulsed-field gel electrophoresis
PHAS	phasiRNA-producing loci
phasiRNA	phased, secondary siRNA
Pi	inorganic phosphate
PMU	potential mobile unit
Pol II	RNA polymerase II
POX	peroxidases
PPO	polyphenol oxidase
PR	pathogenesis-related
PR-1	pathogenesis-related protein 1

pre-miRNA	precursor miRNA
pri-miRNA	primary miRNA
PRR	pathogen recognition receptor
psaD	PSI reaction center subunit II protein
PSII	photosystem II
Pst	*Pseudomonas syringae* pv. tomato
PTGS	posttranscriptional gene silencing
PTI	PAMP-triggered immunity
PTS	phosphotransferase system
R gene	resistance gene
rasiRNA	repeat-associated siRNA
rbcL	RuBisCO large subunit
rDNA	ribosomal DNA
RDR6	RNA-dependent RNA polymerase 6
RFLP	restriction fragment length polymorphism
RIN	RNA integrity number
RISC	RNA-induced silencing complex
RLK	receptor-like kinase
RNAi	RNA interference
RNase	Ribonuclease
ROMT	trans-resveratrol di-O-methyltransferase
ROS	reactive oxygen species
RPKM	reads per kilobase of exon per million mapped reads
RPM	reads per million mapped reads
RT-qPCR	reverse transcription quantitative PCR
RuBisCO	ribulose 1,5-bisphosphate carboxylase/oxygenase
SA	salicylic acid
SAP	secreted AY-WB protein
SE	SERRATE
siRNA	small interfering RNA
SLY	strawberry lethal yellows
SOD	superoxide dismutase

SPL	squamosa-promoter binding-like
sRNA	small non-coding endogenous RNA
sRNA-seq	sRNA transcriptome sequencing
STOL	stolbur group
STS	stilbene synthase
SVM	sequence-variable mosaics
SVR	support vector regression
T_a	annealing temperature
TAS	trans-acting-siRNA-generating loci
tasiRNA	trans-acting siRNA
TCA	tricarboxylic acid
TEM	transmission electron microscopy
TENGU	tengu-su inducer
TF	transcription factor
TGS	transcriptional gene silencing
TIR1	transport inhibitor response 1
T_m	melting temperature
TMM	trimmed mean of M-values
TPM	transcripts per million
UPL	universal probe library
VK	Vergilbungskrankheit

Acknowledgements

I would like to express my gratitude to various individuals who each played an important role in the initiation and formation of this study.

First and foremost, thank you to Prof. Johan Burger for inspiring me through his passion for molecular biology and giving me the opportunity to be part of an amazing research group. Also, most sincere thanks to my co-supervisors, Dr. Dirk Stephan and Dr. Marie-Chrystine Solofoharivelo, for their outstanding mentorship, and valuable advice and teachings.

A special word of thanks to the following people for their generous contributions and/or technical assistance with either computational or experimental work: Marike Visser (Department of Genetics, SU), Beatrix Coetzee (Department of Genetics, SU), Ruhan Slabbert (Department of Genetics, SU), Shane Murray (CPGR, Cape Town), Jacqueline Meyer (CPGR, Cape Town), and Charl Möller (IT, SU).

Thank you to everyone at the Department of Genetics (SU) who have either given me advice or showed me friendship. I will miss all of you when I look back on my time in Stellenbosch.

The financial assistance of the National Research Foundation (NRF) towards this research is hereby acknowledged. Opinions expressed and conclusions arrived at, are those of the author and are not necessarily to be attributed to the NRF.

I am also very appreciative towards the Technology Innovation Agency (TIA) for financial support for the duration of this study.

Last but by no means least I would like to thank my parents and fiancé, Laurian Botha, for showing me unlimited love and constant support. I am eternally grateful.

Table of contents

List of figures

List of tables

Chapter 1

GENERAL INTRODUCTION

1.1 Introduction

Grapevine is recognised as one of the world's most economically-valuable woody perennial fruit crop species and is also known for its cultural impact on tradition, habit, art and even religion (Vivier and Pretorius, 2000; Martinelli and Mandoli, 2001). Grapevines are widely used for the production of wine, spirits, raisins, juice and jam, and many studies have focussed on their health-promoting polyphenols derived from their berries (Georgiev *et al.*, 2014). They thrive in temperate and tropical regions, and occupy a surface area of ~7.5 million hectares (ha) worldwide (OIV, 2016).

Grapevines belong to the genus *Vitis* that forms part of the *Vitaceae* family. *Vitis* spp. consist of two sub-genera, *Euvitis* and *Muscadinia*, each having diploid (2n) chromosome sets of 38 for *Euvitis* spp. and 40 for *Muscadinia* spp. (Antcliff, 1992; Mullins *et al.*, 1992; Jackson, 1994). The most widely cultivated grapevine species are classified as *Vitis vinifera* L. subsp. *vinifera* (or *sativa*) (Antcliff, 1992), and will henceforth be referred to as *Vitis vinifera*.

In South Africa, grapevines are cultivated over an area of 103,478 ha, of which wine grape areas constitute 95,775 ha (VinPro, 2017; SAWIS, 2017). South Africa's main wine producing regions include Paarl, Stellenbosch, Robertson, Swartland, Breedekloof, Olifants River, Worcester, Northern Cape, the Klein Karoo, and the Cape south coast (SAWIS, 2017). The country is the seventh largest wine producer, contributing ~4% of the world's wine, with an annual impact of more than R36.1 billion on growth domestic product (GDP) (VinPro, 2017).

Like all crop plants, grapevines are exposed to multiple stress factors and have developed specific mechanisms in order to cope with their ever-changing environment, often at the detriment of growth and yield (Herms and Mattson, 1992). Some researchers believe that selective breeding and reduction of the varietal range of commercial cultivars may lead to loss of genetic diversity which can have severe consequences for viticulture in the future, relating to climate change and plant-pathogen interactions (Bouquet, 2011). Advances in genetic and molecular techniques coupled with significant progress in understanding complex molecular and physiological mechanisms in plants may assist in facing challenges associated with biotic and abiotic stresses.

In addition to the different viruses that infect grapevines, phloem-limited bacterial pathogens known as phytoplasmas cause severe epidemics in vineyards of major grape producing countries (Lee *et*

al., 2000; Bertacinni, 2007; Martelli *et al.*, 2014). They are responsible for a detrimental disease known as grapevine yellows (GY), which pose a serious threat to the wine and table grape industries. In South Africa, GY is caused by aster yellows (AY) phytoplasma which is the world's most diverse and widespread phytoplasma group. Chardonnay is especially susceptible to the disease and displays an array of deleterious symptoms (Constable, 2010).

GY is known to disrupt developmental processes by causing hormonal imbalance and affecting the carbohydrate concentrations in the host. However, the molecular mechanisms of AY phytoplasma pathogenicity on highly susceptible cultivars, such as Chardonnay, are still largely unknown. This has sparked considerable interest within the viticultural community to gain knowledge about the basis of host susceptibility to GY in order to develop control strategies that may mitigate the scale of infection or even prevent spread.

Remarkable advances in next generation sequencing (NGS) technologies, and the resulting availability of high-quality grapevine genome sequences (Jaillon *et al.*, 2007; Velasco *et al.*, 2007), have been an important driving force for landmark studies on pathogenesis in grapevine. NGS, followed by comparative transcriptome profiling, has facilitated discovery of multiple gene regulatory elements, including the once-enigmatic small RNAs (sRNAs). In addition, the use of NGS approaches have been extremely valuable for elucidating complex plant defence responses which may potentially be correlated with symptom development.

Numerous NGS studies have supplied data that verify the importance of plant sRNAs during biotic stress responses caused by various pathogens. However, those which aim to characterise grapevine sRNA-mediated responses to AY phytoplasma-infection are still limited, and unknown for Chardonnay. Therefore, the involvement of a class of sRNAs, known as microRNAs (miRNA), in the regulation of developmental and resistance pathways is a pivotal topic of this work. Furthermore, high-throughput methods, such as mRNA transcriptome sequencing (mRNA-seq), can be incorporated to investigate plant-pathogen responses at the gene expression level. This has also not been attempted for AY phytoplasma-infected Chardonnay and could be essential towards better understanding phytoplasma-response pathways and their effect on the host. Our results may aid in developing GY disease control strategies and could provide added insight for future plant pathogenesis-related studies.

1.2 Aim and objectives

The aim of the study was to use NGS approaches to identify differentially expressed miRNAs, as well as differentially expressed genes, in order to explore pathogen response pathways in *Vitis vinifera* 'Chardonnay' upon AY phytoplasma-infection.

This aim was approached through the following objectives:

i. To identify plants that would serve as sources of healthy material and AY phytoplasma-infected material, using diagnostic PCR assays.
ii. To extract high-quality total RNA that would be used to generated sRNA and mRNA sequencing libraries for an Illumina NGS platform.
iii. To utilise bioinformatic software to dissect sRNA sequencing data in order to identify and characterise differentially expressed miRNAs.
iv. To validate differentially expressed miRNAs using a stem-loop RT-qPCR method.
v. To identify and functionally annotate possible targets for differentially expressed miRNAs.
vi. To utilise bioinformatic software to identify differentially expressed genes in the AY-phytoplasma-infected samples.
vii. To validate differential gene expression using RT-qPCR assays.
viii. To assign functional annotations to the differentially expressed genes.

1.3 Dissertation structure

This dissertation is comprised of five chapters:

1.3.1 Chapter 1: General introduction

A general introduction along with aims and objectives of the study, as well as the chapter layout of this dissertation are given. Titles of scientific outputs generated during the course of the study are provided and the involvement of M.C. Snyman in each output is specified.

1.3.2 Chapter 2: Literature review

The literature review provides an overview of phytoplasmas, phytoplasma pathogenesis, miRNAs, and computational methods used for the analysis of miRNA and mRNA high-throughput sequencing data. This chapter is also accompanied by a section on GY in South Africa.

1.3.3 Chapter 3: The use of high-throughput small RNA sequencing reveals differentially expressed microRNAs in response to aster yellows phytoplasma-infection in Vitis vinifera cv. 'Chardonnay'

This research chapter describes the use of sRNA sequencing for the genome-wide identification of differentially expressed miRNAs in AY phytoplasma-infected Chardonnay leaf material. In order to extend the miRNA knowledgebase, novel miRNAs were also identified using two different prediction algorithms. The functions of putative miRNA targets and their involvement in possible pathogen response pathways are also discussed.

1.3.4 Chapter 4: High-throughput mRNA transcriptome sequencing of aster yellows phytoplasma-infected Vitis vinifera cv. 'Chardonnay'

This research chapter describes the use of mRNA-seq to generate a gene expression profile for AY phytoplasma-infected Chardonnay leaf material. A detailed discussion regarding functional annotations of differentially expressed genes and potential involvement in pathogen response pathways are provided.

1.3.5 Chapter 5: Conclusion

This chapter provides final concluding remarks and future prospects of this work.

1.4 Research outputs

1.4.1 Publication:

- **Snyman M.C.**, Solofoharivelo M.-C., Souza-Richards R., Stephan D., Murray S., and Burger J.T. 2017. The use of high-throughput small RNA sequencing reveals differentially expressed microRNAs in response to aster yellows phytoplasma-infection in *Vitis vinifera* cv. 'Chardonnay'. *PLoS ONE* 12(8): e0182629. DOI: 10.1371/journal.pone.0182629 (Appendix 1.1; File attached separately).

  This paper matches Chapter 3 as a whole and is almost entirely the work of the first author who was responsible for the experimental design, sample preparation, data collection, formal analysis, original writing of the draft, as well as manuscript editing. Three non-authors mentioned in the Acknowledgment section assisted with the computational analysis.

1.4.2 Conference proceedings (Presenter underlined):

- **Snyman M.C.**, Solofoharivelo M.-C., Van der Walt A., Souza-Richards R., Stephan D., Murray S., and Burger J.T. Deep sequencing analysis reveals modulated gene expression in response to Aster yellows phytoplasma-infection in *Vitis vinifera* cv. Chardonnay. Proceedings of the 17th Conference of the International Council for the Study of Virus and Virus-like Diseases of the Grapevine (ICVG), October 2012. University of California, Davis. pp. 240-241. (Paper)

  This paper contains work described in Chapter 4 and is almost entirely the work of the first author.

- **Snyman M.C.,** Van der Walt A., Solofoharivelo M.-C., Stephan D., Murray S., and Burger J.T. Next generation sequencing reveals distinct microRNA and mRNA expression profiles in Aster yellows phytoplasma-infected *Vitis vinifera* cv. Chardonnay. South African Genetics, Bioinformatics and Computational Biology Society Conference (SASBi-SAGS), January 2012. Stellenbosch. PP73. (Poster)

  This presentation contains work described in Chapter 3 and Chapter 4 and is almost entirely the work of the first author.

- Spinas N.L., **Snyman M.C.**, Visser M., Stephan D., Burger J.T. Can antimicrobial peptides be used to engineer resistance against the grapevine pathogen aster yellows phytoplasma? Proceedings of the 17th Conference of the International Council for the Study of Virus and Virus-like Diseases of the Grapevine (ICVG), October 2012. University of California, Davis. pp. 254-255. (Paper)

  This paper contains work on efficacy testing of certain antimicrobial peptides against AY phytoplasma in grapevine and is almost entirely the work of N.L. Spinas. M.C. Snyman and M. Visser were responsible for experimental design and optimisation of a semi-quantitative qPCR assay used to assess the AY phytoplasma-infection status of plant material.

- Visser M., **Snyman M.C.**, Stephan D., Burger J.T. Development of a real-time PCR for semi-quantitative detection of Aster yellows phytoplasma (16SrI). Second International Phytoplasmologist Working Group (IPWG), September 2011. Neustadt, Germany. (Poster)

This presentation contains work on the development of a SYBR® Green-based qPCR assay for the detection and quantification of AY phytoplasma (16SrI), originally designed by M. Visser. M.C. Snyman was responsible for sample preparation, assay optimisation and data analysis.

1.5 References

Antcliff A.I. 1992. Taxonomy - The grapevine as a member of the plant kingdom. *In: Viticulture, vol.1 - Winetitles.* Coombre, B. G. & Dry, P. R. (eds.). Underdale. pp.107-118.

Bertaccini A. 2007. Phytoplasmas: diversity, taxonomy, and epidemiology. *Frontiers in Bioscience* 12: 673-689.

Bouquet A. 2011. Grapevines and viticulture. *In: Genetics, Genomics and Breeding of Grapes.* Adam-Blondon, A.-F., Martinez-Zapater, J. M. & Kole, C. (eds.). Enfield, USA: Science Publishers. pp.1-29.

Constable F. 2010. Phytoplasma epidemiology: grapevines as a model. *In: PHYTOPLASMAS: Genomes, Plant Hosts and Vectors.* Weintraub, P. G. (ed.). pp.188-212.

Georgiev V., Ananga A. and Tsolova V. 2014. Recent advances and uses of grape flavonoids as nutraceuticals. *Nutrients* 6(1): 391-415.

Herms D.A. and Mattson W.J. 1992. The dilemma of plants: to grow or defend. *The Quarterly Review of Biology* 67(3): 283-335.

Jackson R.S. 1994. Grapevine species and varieties. *In: Wine Science: Principles and application.* San Diego: Academic Press. pp.11-31.

Jaillon O., Aury J.-M., Noel B., Policriti A., Clepet C., *et al.* 2007. The grapevine genome sequence suggests ancestral hexaploidization in major angiosperm phyla. *Nature* 449(7161): 463-467.

Lee I.-M., Davis R.E. and Gundersen-Rindal D.E. 2000. Phytoplasma: Phytopathogenic Mollicutes 1. *Annual Review of Microbiology* 54(1): 221-255.

Martelli G.P. 2014. Directory of virus and virus-like diseases of the grapevine and their agents. *Journal of Plant Pathology* 96(1): 1-136.

Martinelli L. and Mandolino G. 2001. Transgenic Grapes (*Vitis* Species). *In: Biotechnology in Agriculture and Forestry.* Bajaj, Y. P. S. (ed.). Berlin-Heidelberg: Springer-Verlag. pp.325-338.

Mullins M.G., Bouquet A. and Williams L.E. 1992. *In: Biology of the grapevine.* Cambridge: Cambridge University Press.

OIV 2016. International Organisation of Vine and Wine. Statistical Report on World Vitiviniculture. www.oiv.int

SAWIS 2017. South African Wine Industry Statistics nr 42. www.sawis.co.za

Velasco R., Zharkikh A., Troggio M., Cartwright D.A., Cestaro A., *et al.* 2007. A high quality draft consensus sequence of the genome of a heterozygous grapevine variety. *PLoS ONE* 2(12): e1326.

VinPro cost guide 2017. www.vinpro.co.za

Vivier M.A. and Pretorius I.S. 2000. Genetic improvement of grapevine: tailoring grape varieties for the third millennium - a review. *South African Journal for Enology and Viticulture* 21: 5-26.

Chapter 2

LITERATURE REVIEW

A. PHYTOPLASMAS

2.1 Introduction

Phytoplasmas are bacterial plant pathogens known to infect over 700 plant species worldwide and are responsible for devastating yield losses of many economically important crops, fruit trees, and ornamental plants (Lee *et al.*, 2000; Bertacinni, 2007). Formally known as mycoplasma-like organisms (MLOs), phytoplasmas were initially identified as etiological agents in plants that caused yellows diseases, and mistakenly thought to be viruses (Doi *et al.*, 1967; McCoy *et al*., 1989). They are obligate cell wall-less parasites and rely on plants and insects for biological dispersal. In plants, they are mainly restricted to the phloem tissue where they can move and multiply through the sieve tube elements (Hogenhout *et al.*, 2008). Plants infected with phytoplasmas exhibit a wide range of symptoms including stunting, yellowing, sterility of flowers, abnormal internode elongation or shortening, witches' broom (proliferation of axillary buds with small leaves), phyllody (formation of leaf-like structures instead of flowers), virescence (greening of floral organs), proliferation (growth of shoots from floral organs), purple top (reddening of leaves and stems), and phloem necrosis (Bertaccini, 2007; Maejima *et al.*, 2014).

Phytoplasmas are transmitted by phloem-feeding insect vectors known as Auchenorrhyncha (leafhoppers, planthoppers and psyllids), a suborder of the Hemiptera (Mitchell, 2004; Weintraub and Beanland, 2006). Phytoplasmas can invade insects by accumulating inside and outside cells of the guts, salivary glands and many other tissues (Hogenhout *et al.*, 2008). After infecting an insect, the bacteria traverse the intestinal tract wall, multiply in the hemolymph, and travel through the salivary glands where they accumulate further. During insect feeding on a new host plant, the phytoplasmas are introduced into the phloem tissue along with saliva (Hogenhout *et al.*, 2008) (Figure 2.1).

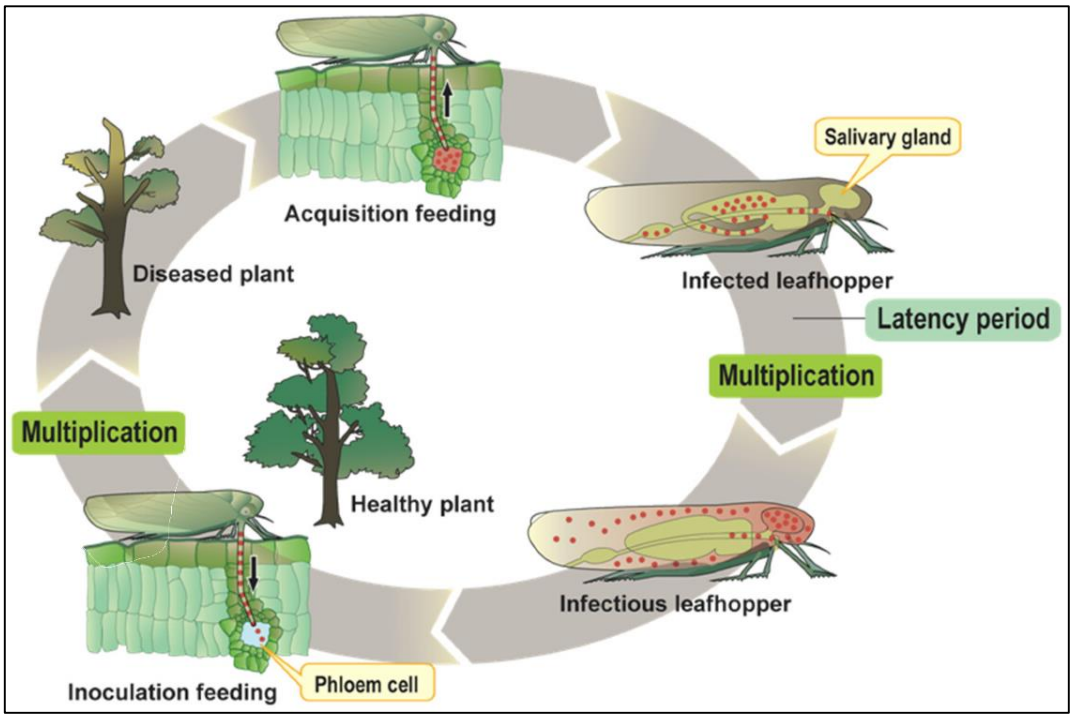

Figure 2.1: Biological dispersal of phytoplasmas involves replication in plants and insects. Following acquisition feeding, phytoplasmas multiply in secretory salivary gland cells from where they are transported to the salivary duct along with the saliva. They are then introduced back into the phloem tissue of host plants during inoculation feeding of leafhoppers (Image acquired from Oshima *et al.*, 2011).

Together with mycoplasmas, ureaplasmas, spiroplasmas and acholeplasmas, phytoplasmas belong to the class Mollicutes, which encompasses small pleiomorphic or spherical bacteria of 80-800 nm with single membranes. It is believed that they have diverged from Gram-positive bacteria in the *Clostridium*/*Lactobacillus* group, through genome reductions and the loss of the outer cell wall (Weisburg *et al.*, 1989; Woese, 1987; Razin *et al.*, 1998). These genomic reductions have caused phytoplasmas to rely heavily on acquiring essential metabolites from their hosts because they lack several pathways for the production of compounds necessary for survival (Bai *et al.*, 2006). In the last two decades much progress in phytoplasma research was made, following the optimisation of procedures to isolate and enrich phytoplasma DNA from infected hosts and vectors (Marcone, 2014). These methods have enabled closer inspection of the genetic diversity of phytoplasmas, by establishing a system for their taxonomic classification based on their phylogeny.

Only recently, axenic cultivation of phytoplasmas in complex media with broad applicability and a good repeatability were shown to support phytoplasma colony formation (Contaldo *et al.*, 2016), and will allow direct *in planta* investigation of molecular interactions postulated to exist between phytoplasmas and their plant hosts and insect vectors. In addition, a wealth of genetic data were generated after the genomes of several phytoplasma strains have been sequenced to completion

(Oshima *et al.*, 2004; Bai *et al.*, 2006; Kube *et al.*, 2008; Tran-Nguyen *et al.*, 2008; Andersen *et al.*, 2013; Kakizawa *et al.*, 2014; Lee *et al.*, 2015). Genomic data has enabled a better understanding of molecular mechanisms underlying virulence and host interactions (Oshima *et al.*, 2013).

Furthermore, high-throughput transcriptome analysis of next-generation sequencing (NGS) and microarray data, as well as proteomics, have served as valuable approaches for gaining new insights into physiological, biochemical and molecular mechanisms underlying disease symptom development caused by phytoplasma-infection in different plant species (Hren *et al.*, 2009; Albertazzi *et al.*, 2009; Margaria *et al.*, 2010; Mou *et al.*, 2013; Margaria *et al.*, 2013; Monavarfeshani *et al.*, 2013; Liu *et al.*, 2014).

2.2 Molecular detection of phytoplasmas

Before the application of molecular techniques, following their discovery, the detection of phytoplasma diseases were difficult due to their low concentrations, especially in woody plants, and their erratic titre distribution in infected plants (Berges *et al.*, 2000). Until the early 1980's, phytoplasma diseases were diagnosed by preparing ultrathin sections of phloem tissue and observing these sections using transmission electron microscopy (TEM) (Firrao *et al.*, 2007; Maejima *et al.*, 2014) (Figure 2.2).

Other traditional techniques used to diagnose phytoplasmoses were based on observation of plant symptoms and insect or dodder/graft transmission to healthy indicator plants (Jarauch *et al.*, 2000; Pastore *et al.*, 2001). These methods were soon replaced by more simple diagnostic techniques such as direct fluorescent detection (DFD) (Namba *et al.*, 1981) and 4',6-diamidino-2-phenylindole (DAPI) staining of A-T rich regions in DNA (Hiruki and Deng, 1992), both utilising fluorescent microscopy. The abovementioned techniques are known to be expensive, laborious and time-consuming, and often yielded inconclusive results. Dodder transmission, however, is still an effective method to preserve different types of phytoplasmas for future research (Přibylová and Spak, 2013).

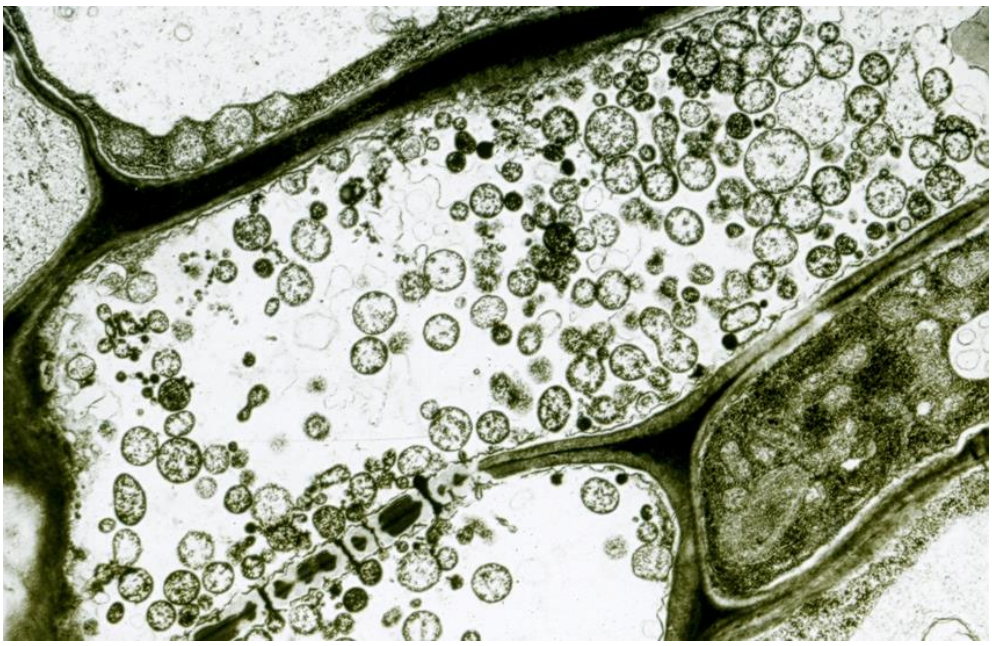

Figure 2.2: TEM micrograph showing maize bushy stunt phytoplasmas in an infected sieve element of a maize plant (Photo by courtesy of Prof. Lowell R. Nault).

Serological tests, such as immunofluorescence, immunosorbent electron microscopy (IEM) and enzyme-linked immunosorbent assay (ELISA), later became more prevalent methods for the specific detection and identification of phytoplasma diseases (Sinha and Benhamou, 1983; Chen *et al.*, 1992; Nejat and Vadamalai, 2013). These tests rely on the production of polyclonal and monoclonal antibodies and were successfully used in a few rare cases. Highly specific monoclonal antibodies have been developed against a limited number of phytoplasma strains, such as AY (Lin and Chen, 1985), apple proliferation (Loi *et al.*, 2002), and flavescence dorée (FD) (Seddas *et al.*, 1996). Despite the success of serological-based diagnostics, these methods were very time-consuming and labour-intensive due to difficulties in phytoplasma purification. Furthermore, they lack sensitivity, especially when the pathogen titre is low and when contaminant host proteins are present.

Nucleic acid-based techniques are now routinely used for phytoplasma detection. Since phytoplasma DNA was first cloned (Kirkpatrick *et al.*, 1987), nucleic acid-based probes (randomly cloned DNA or its cDNA) were widely used in dot and Southern hybridisation assays to detect several distinct phytoplasma groups (genomic strain clusters) and subgroups (subclusters) in plants and insects (Firrao *et al.*, 2007; Lee *et al.*, 2000). Hybridisation results combined with other analytical techniques such as restriction fragment length polymorphism (RFLP) analysis were the basis for establishing the first genotype-based classification of phytoplasma groups in several genomic strain clusters (Davis *et al.*, 1988; Lee *et al.*, 1992).

PCR-based assays utilising universal or phytoplasma group-specific oligonucleotide primers, based on conserved sequences (e.g. 16S rRNA; ribosomal protein, *tuf*, 16S-23S ITS region), are now regarded as the most suitable diagnostic techniques (Lee *et al.*, 1993; Lorenz *et al.*, 1995; Jarauch *et al.*, 1998; Bertin *et al.*, 2004). They have been shown to be much more sensitive than serological tests or DNA-DNA hybridisation assays and have enabled the amplification of conserved sequences from a broad spectrum of phytoplasma strains and from specific strains belonging to a given phytoplasma group. Nested-PCR protocols have been designed to increase both sensitivity and specificity for PCR amplification of phytoplasma DNA, especially for samples where low titres or inhibitors may interfere with PCR efficacy (Lee *et al.*, 1994).

In some laboratories, real-time PCR has recently replaced the traditional PCR assays in efforts to increase the speed and sensitivity of detection when dealing with a large number of samples (Galetto and Marzachi, 2010). Real-time PCR allows direct monitoring of phytoplasma-specific amplicon accumulation at each cycle by fluorescent detection, either utilising TaqMan® probe(s) or SYBR Green I® dye binding (Hren *et al.*, 2007; Angelini *et al.*, 2007; Hollingsworth *et al.*, 2008; Berger *et al.*, 2009). Therefore, there is no need for a post-PCR step necessary at the end of regular PCR (e.g. electrophoresis). The amount of fluorescence is proportional to the logarithm of the PCR target concentration, and can therefore also be used for quantification of specific phytoplasma DNA, thus allowing titre measurements in terms of copy number per reaction.

2.3 Classification of phytoplasmas

Since their initial discovery, four decades passed before the successful axenic cultivation of phytoplasmas by Contaldo *et al.* (2012). MLOs remain the most poorly understood phytopathogens in terms of their biology and taxonomy because their taxonomic status could not be determined using traditional methods applied to cultured prokaryotes. Since MLO classification was merely impossible, diagnosis of possible phytoplasma origin relied on close observation of characteristic symptoms and ultrathin sections of diseased plants (Lee *et al.*, 2000).

In 1989, the 16S ribosomal RNA gene sequence from an MLO (*Oenothera virescence* phytoplasma) was compared with the 16S rRNA gene sequences of other MLOs, as well as those of *Acholeplasma laidlawii*, *Spiroplasma citri* and several mycoplasmas (Lim and Sears, 1989). These analyses, including sequence analysis of other conserved genes, suggested that phytoplasmas comprise a large and distinct monophyletic clade within the class Mollicutes, but more closely

related to the *Acholeplasma* spp. than to the *Spiroplasma* spp.or to animal mycoplasmas (Kuske and Kirkpatrick, 1992; Namba *et al.*, 1993; Seemüller *et al.*, 1994; Toth *et al.*, 1994). On the basis of phylogenetic analyses of the highly-conserved 16S rRNA gene sequence, many new phytoplasma strains, belonging to approximately 20 phylogenetic groups or subclades, have been determined. This number is generally associated with "16Sr" groups established by RFLP analysis of PCR-amplified rDNA (Lee *et al.*, 1998, 2000; Seemüller *et al.*, 1998; Marcone, 2014). For finer differentiation of phytoplasmas, additional, less-conserved gene sequences e.g. ribosomal protein (*rp*), *secA*, *secY*, *tuf*, and the 16S-23S rRNA ITS region have been employed as supplementary markers (Bertaccini and Duduk, 2010; Marcone, 2014).

It was proposed that phytoplasmas be placed within the novel genus '*Candidatus* (*Ca.*) Phytoplasma' where each subclade (or corresponding 16Sr group) represents at least one distinct species under the provisional taxonomic status '*Candidatus*' (IRPCM, 2004). Basically, a novel '*Ca*. Phytoplasma' species rank can be assigned if a 16S rRNA gene sequence possesses less than 97.5% similarity to that of any previously described '*Ca*. Phytoplasma' species. To date, 37 '*Ca*. Phytoplasma' species have been formally described in accordance to the International Research Programme for Comparative Mycoplasmology (IRPCM, 2004) guidelines (Marcone, 2014).

2.4 The phytoplasma genome

Difficulties to culture phytoplasmas *in vitro* in the past hindered their molecular characterisation. However, the use of DNA techniques, such as genome sequencing, have enabled a better understanding of the molecular mechanisms underlying virulence and host interaction of phytoplasmas (Oshima *et al.*, 2013). Previous attempts to isolate pure phytoplasma preparations to study their genomes proved to be difficult until the development of pulsed-field gel electrophoresis (PFGE), which provided a wealth of genome size data for culturable mollicutes (Razin *et al.*, 1998). PFGE data demonstrated a wide range of genome sizes, ranging from 530 kb to 1350 kb, for over 100 phytoplasmas (Neimark and Kirkpatrick, 1993; Firrao *et al.*, 1996; Marcone *et al.*, 1999; Marcone *et al.*, 2001; Liefting and Kirkpatrick, 2003). To date, the STOLF tomato-infecting strain belonging to the stolbur phytoplasma group, contains the largest chromosome (~1350 kb) found in a phytoplasma. The smallest known mollicute chromosome (~530 kb) was found in isolates of the Bermuda grass white leaf agent '*Ca.* Phytoplasma cynodontis' (Marcone *et al.*, 1999).

It would be expected that a large genome and consequently a large number of genes would bestow a phytoplasma with less host-dependency and better growth in less complex media. However, previous studies on AY and stolbur phytoplasmas with large chromosomes, proved that no relation existed between symptom development or virulence, and genome sizes (Marcone *et al.*, 1999). In contrast, it was shown that a reduced onion yellows (OY) phytoplasma strain has a smaller chromosome (870 kb versus 1000 kb), and causes milder symptoms than the wild type (Oshima *et al.*, 2001).

2.4.1 Genome properties of phytoplasmas: an introduction

To characterise the genome features of phytoplasmas and to better understand the molecular mechanism underlying virulence and host interaction, several phytoplasma genome projects were initiated in the late 1990's. To date, the complete assembled genome sequences of six phytoplasmas have been reported, among which genomic features of '*Ca.* Phytoplasma asteris' OY-M, '*Ca.* Phytoplasma asteris' AY-WB, '*Ca.* Phytoplasma australiense' SLY, '*Ca.* Phytoplasma australiense' Rp-A and '*Ca.* Phytoplasma mali' AT were discussed in detail (Oshima *et al.*, 2004; Bai *et al.*, 2006; Kube *et al.*, 2008; Tran-Nguyen *et al.*, 2008; Andersen *et al.*, 2013) (Table 2.2). The ~576 kbp whole genome of Maize bushy stunt phytoplasma (MBSP), isolate M3, however, was characterised in terms of polymorphisms associated with symptom severity on various maize genotypes (Orlovskis *et al.*, 2017). A further twelve draft genome sequences are also available (Saccardo *et al.*, 2012; Chung *et al.*, 2013; Chen *et al.*, 2014; Kakizawa *et al.*, 2014; Mitrovic *et al.*, 2014; Chang *et al.*, 2015; Quaglino *et al.*, 2015; Lee *et al.*, 2015; Sparks *et al.*, 2018). Phylogenetically, these belong to 16Sr-I, -II, -III, -X and -XII groups and have enabled the characterisation of genome content and organization of phytoplasmas. A phytoplasma genome generally consists of a single chromosome and small plasmids with a unique replication gene (*Rep*), involved in rolling-circle replication, as well as several other unknown proteins (Nishigawa *et al.*, 2001; Oshima *et al.*, 2001; Bai *et al.*, 2006). '*Ca.* Phytoplasma mali', however, harbors no plasmids (Kube *et al.* 2012).

Table 2.1: General features of phytoplasma genomes that were sequenced and assembled to completion[†]

'*Ca.* Phytoplasma' species	asteris	asteris	australiense	australiense	mali
Strain	OY-M	AY-WB	PAa	SLY	AT
16S rDNA group	I-B	I-A	XII-B	XII-B	X
Chromosome size (kb)	860,631	706,569	879,324	959,779	601,943
Chromosome organisation	Circular	Circular	Circular	Circular	Linear
G+C content (%)	28	27	27	27	21.4
Protein-coding regions (%)	73	72	74	78	78.9
Protein-coding genes with assigned function	446	450	502	528	338
Conserved hypothetical genes	51	149	214	249	72
Hypothetical genes	257	72	123	349	87
Total no. of genes	754	671	839	1126	497
rRNA operons	2	2	2	2	2
tRNA genes	32	31	35	35	32
Extrachromosomal DNAs	2	4	1	1	0
GenBank accession no.	AP006628	CP000061	AM422018	CP002548	CU469464

Data were obtained from Oshima *et al.* (2004), Bai *et al.* (2006), Tran-Nguyen *et al.* (2008), Kube *et al.* (2008) and Andersen *et al.* (2013).[†]

According to genome sequencing results, phytoplasma genome sizes range from 598 to 960 kb, with a low G+C content (less than 22% in '*Ca.* Phytoplasma mali'), similar to mycoplasmas (Glass *et al.*, 2000) (Figure 2.3). Other genomic features that are prominent among phytoplasmas include the presence of two rRNA operons and a spacer region of ~300 bp between the 16S and 23S ribosomal regions. Their genomes also contain large numbers of indels, and a family of repetitive palindromes, unique to these organisms (Bertaccini and Duduk, 2010; Marcone, 2014). Furthermore, large clusters of repeated sequences, which are mostly multicopy genes, are present and tend to congregate as tandem or multiple repeats in certain regions of phytoplasma chromosomes. These clusters (each ~20 kb) are known as potential mobile units (PMUs) because they are characteristicly similar to replicative composite transposons (Bai *et al.*, 2006; Hogenhout *et al.*, 2008). The formation of extrachromosomal elements by a PMU integrated in the chromosome of the '*Ca.* Phytoplasma asteris' AY-WB strain suggested it may replicate independently (Toruño *et al.*, 2010), suggesting its ability to transpose within the genome. Other genetic elements, apparently degenerated PMU-like sequences, known as sequence-variable mosaics (SVMs), have also been described, possibly originating from attacks from phages of the order *Caudovirales* (Jomantiene and Davis, 2006; Wei *et al.*, 2008). The presence of elements such as plasmids, phage-related sequences and PMUs may account for variation in a phytoplasma's genome size and arrangement, and thus may contribute to its adaptability in diverse host and vector environments.

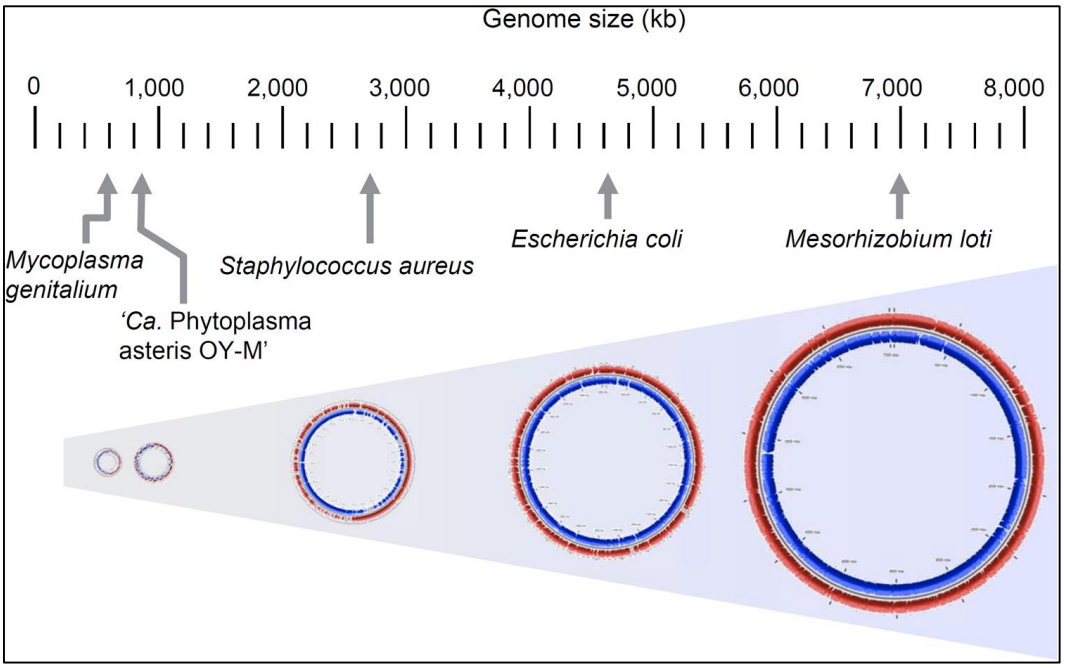

Figure 2.3: The average genome size of a phytoplasma compared to genomes of other bacteria (Image acquired from Oshima *et al.*, 2013).

2.4.2 Reductive genome evolution

As said before, phytoplasmas are phloem-inhabiting plant pathogens transmitted by insect vectors, which in turn are also colonised. During the course of reductive evolution phytoplasmas lost most genes essential for basic metabolism but retained the set of functions necessary for survival in these habitats. It is for this reason that phytoplasmas have reduced biosynthetic capabilities, lacking genes involved in oxidative phosphorylation, tricarboxylic acid (TCA) cycle, pentose phosphate pathway, fatty acid biosynthesis, amino acid biosynthesis, and nucleotide metabolism. Phytoplasma genomes are also devoid of genes encoding the phosphotransferase system (PTS), which most bacteria use as an energy-efficient way to simultaneously import and phosphorylate sugars to be fed to the glycolytic pathway (Razin *et al.*, 1998; Oshima *et al.*, 2004; Bai *et al.*, 2006; Kube *et al.*, 2008; Kube *et al.*, 2012). These genomic features have made phytoplasmas entirely dependent on their hosts for survival.

Another pronounced feature of phytoplasma genomes is the lack of genes encoding any of the subunits for F_0F_1-type ATP synthase, responsible for generating a transmembrane potential and producing ATP (Oshima *et al.*, 2013). The lack of F_0F_1-type ATP synthases and a previous report showing an increase in glycolytic turnover in *Bacillus subtilis atp*-operon mutant strains (Santana *et al.*, 1994), suggests that ATP synthesis in phytoplasmas may strongly depend on glycolysis. The

enzymatic steps of glycolysis, involving the conversion of glucose to pyruvate, is believed to be the major energy-yielding pathway of phytoplasmas (Oshima *et al.*, 2004; Bai *et al.*, 2006; Tran-Nguyen *et al.*, 2008).

2.4.3 Genomic properties for dual host interactions

Collectively there are coding sequences for at least six types of ATP-binding cassette (ABC) transporter-like subunits conserved in the evaluated phytoplasma genomes. ABC transporters are known to shuttle metabolites across bacterial membranes, and are predicted to allow nutrient and metabolite uptake from the host (Kube *et al.*, 2012; Kube *et al.*, 2014). Phytoplasma genomes also contain the *sodA* gene encoding superoxide dismutase (SOD), possibly used to inactivate reactive oxygen species deployed through oxidative burst by hosts upon pathogen attack (Miuru *et al.*, 2012). The high number of *HflB* genes present in phytoplasma genomes that encode zinc-dependent HflB metalloproteases suggest these enzymes play a prominent role in phytoplasmas. Half of these proteins show a predicted extracellular orientation in '*Ca.* Phytoplasma mali' and may be involved in pathogen-host interactions resulting in compromised phloem function (Siewert *et al.*, 2013). Recent studies have also suggested that HflB proteases are associated with strain virulence (Wang *et al.*, 2014; Siewert *et al.*, 2013). Recently a conserved 'Mollicutes adhesin motif' (MAM) was identified in a putative membrane protein (P38) of the '*Ca.* Phytoplasma asteris' OY-M genome. Binding assays showed that P38 interacts with crude insect extracts and weakly with plants extracts. The host factors targeted by P38, however, have not yet been identified (Neriya *et al.*, 2014).

The phytoplasma genome lacks homologs of the type III secretion system, which play an essential part in bacterial viability (Maejima *et al.*, 2014). Homologs for type II secretion system genes encoding SecA, SecY, and SecE translocases, required for protein translocation in *Escherichia coli* (Economou, 1999), were identified in different phytoplasma genomes (Bai *et al.*, 2006; Kube *et al.*, 2008; Tran-Nguyen *et al.*, 2008). This secretion system allows the delivery of functionally distinct proteins with a characteristic *n*-terminal signal peptide to the bacterial membrane.

Secreted phytoplasma proteins can act as effectors which may alter host functions (Bai *et al.*, 2009). To date, only four of these effector proteins *viz.* TENGU and PHYL1 from '*Ca.* Phytoplasma asteris' OY wild type strain (OY-W), as well as SAP11 and SAP54 from '*Ca.* Phytoplasma asteris' AY-WB, have been functionally characterised (Minato *et al.*, 2014). Phytoplasmas inhabit phloem sieve cells where they reside intracellularly and secrete effector proteins via their type-II (Sec)

protein translocation system (Bai *et al*., 2006; Kube *et al*., 2008; Tran-Nguyen *et al*., 2008). These effectors have been shown to disrupt host developmental processes through interaction with transcription factors and modulation of phytohormones (auxin and JA) that are crucial for both plant development and defence signalling (Sugio *et al.*, 2011; Minato *et al*. 2014; Orlovskis and Hogenhout, 2016). TENGU has been reported to induce symptoms such as witches' broom, dwarfism and sterility after ectopic expression in *Arabidopsis thaliana* and *Nicotiana benthamiana* (Hoshi *et al.*, 2009; Sugawara *et al.*, 2013; Minato *et al.*, 2014). Microarray expression analysis revealed down-regulation of auxin-responsive factor (*ARF*) and auxin efflux carrier genes in *TENGU*-expressing *Arabidopsis* lines, thereby affecting normal plant development (Hoshi *et al.*, 2009). The '*Ca.* Phytoplasma asteris' AY-WB genome can encode over 50 secreted proteins and a number of these have been functionally characterised (Bai *et al.*, 2009).

SAP11 contains a eukaryotic nuclear signalling peptide and localises in plant cell nuclei (Bai *et al.*, 2009). *Arabidopsis* expression of SAP11 exhibits a crinkled-leaf phenotype and induces stem production, thereby resembling witches' broom symptoms (Sugio *et al.*, 2011). SAP11 destabilises Teosinte branched1/Cincinnata/Proliferating cell factor (TCP) transcription factors 1 and 2, resulting in the suppression of jasmonate (JA) production that create favourable conditions for insect vector proliferation (Hoshi *et al.*, 2009; Sugio *et al.*, 2011), thus assisting the spread of AY-WB by its insect vector. SAP54/PHYL interacts with floral transcription factors and promotes degradation of the MADS-box proteins. MADS-box proteins are critical for floral meristem development and plants expressing SAP54/PHYL flower abnormally (MacLean *et al.*, 2014; Maejima *et al.*, 2014).

A second group, known as the immunodominant membrane proteins (IMPs), is also delivered by the Sec-secretion system. IMPs are unique for phytoplasmas and remain anchored on the outer membrane. Three subgroups of IMPs exist, namely Amp, IdpA and Imp, and have been classified based on the N- or C-terminus of the protein that is exposed extracellularly (Kakizawa *et al.*, 2006).

2.5 Phytoplasma in grapevine

Phytoplasma-infection in grapevine (*Vitis vinifera* L.) is associated with a severe disease known as grapevine yellows (GY) (Alma *et al.*, 1996). The disease was first described as Flavescence dorée (FD), which appeared in south-western France in the 1950's, from where it spread to other viticultural districts of France, northern Italy and neighbouring European countries. FD phytoplasmas are members of the elm yellows (EY) group (Seemüller *et al.*, 1998; Angelini *et al.*,

2001; Belli *et al.*, 2010). Other early reports of GY diseases, characterised by similar symptoms to those of FD, known as Bois noir (BN) and Vergilbungskrankheit (VK), were first reported in north-eastern France and Germany, respectively, in the 1960's before it spread to other viticultural areas of Europe (Caudwell, 1961; Gärtel, 1965). BN and VK are caused by phytoplasmas that belong to the stolbur group (STOL) of phytoplasmas (Belli *et al.*, 2010).

Today, GY has been reported in all major grape-growing countries of the world and pose a serious threat to the wine and table grape industries. With the use of molecular techniques, several distinct GY-causing phytoplasmas have been described worldwide and represent various phylogenetic groups, *viz.* elm yellows (EY/16SrV), stolbur (STOL/16SrXII-A), X-disease (16SrIII), aster yellows (AY/16SrI), Western X (WX/16SrIII-I), Australian grapevine yellows (AUSGY/16SrXII or Tomato big bud/16SrII) and faba bean phyllody (FBP/16SrII) phytoplasmas, respectively (Boudon-Padieu, 2003; Constable, 2010).

2.5.1 Grapevine yellows epidemiology

Kranz (1990) described epidemiology within plant pathology as "*the science of populations of pathogens in populations of host plants, and the diseases resulting therefrom under the influence of the environment and human interferences*". Epidemiology aims to clearly describe the disease triangle in terms of interactions between the host, the pathogen and the environment and applies this information to develop control strategies. Control of phytoplasma-associated diseases relies on prevention rather than cure. Consequently, the epidemiology of many phytoplasma-associated diseases has been well studied, particularly in economically important crops such as pome, stone fruit and grapevines. Currently, the only available control strategies include early eradication of infected crops, early eradication of infected source plants (weed control), and chemical control of vectors through regular insecticide treatments (Maixner *et al.*, 2006).

According to Constable (2010) the epidemiology of phytoplasma-associated diseases is intrinsically linked to the biology of their insect vectors. Knowledge of vectors, their biology and behaviour and choice of target plants are prerequisites for an understanding of GY epidemiology and the development of well-adjusted control strategies (Figure 2.4; Table 2.2). High sensitivity of molecular diagnostic tools has allowed reliable detection of phytoplasmas in various vector species after feeding. Subsequent transmission trials are an essential requirement to prove the vectoring ability of a particular species. Principal transmission experiments can be tested through vector

feeding on artificial media (Tanne *et al.*, 2001) and experimental plants, such as *Clematis vitalba* and *Catharanthus roseus* (Filippin *et al.*, 2009; Spinas, 2013). Vector-transmission of phytoplasmas to grapevine is also important to understand the specific vector-grapevine relationship and to demonstrate the ability of a particular vector species to transmit GY to field-grown grapevines.

Extensive epidemiological studies have been carried out for four of the grapevine yellows (GY) diseases, including FD, BN, AUSGY and North American grapevine yellows (NAGY) (Maixner *et al.*, 2006; Constable, 2010). Although the symptomatology of the GY diseases is nearly identical regardless of the location, the epidemiology associated with each phytoplasma species, and even amongst strains of the same phytoplasma, can vary. These diseases provide a unique opportunity to compare and highlight biological attributes that are important to the epidemiology of phytoplasma-associated disease (Constable, 2010). Table 2.2 provides a summary of epidemiological information available for each of the GY diseases reported worldwide.

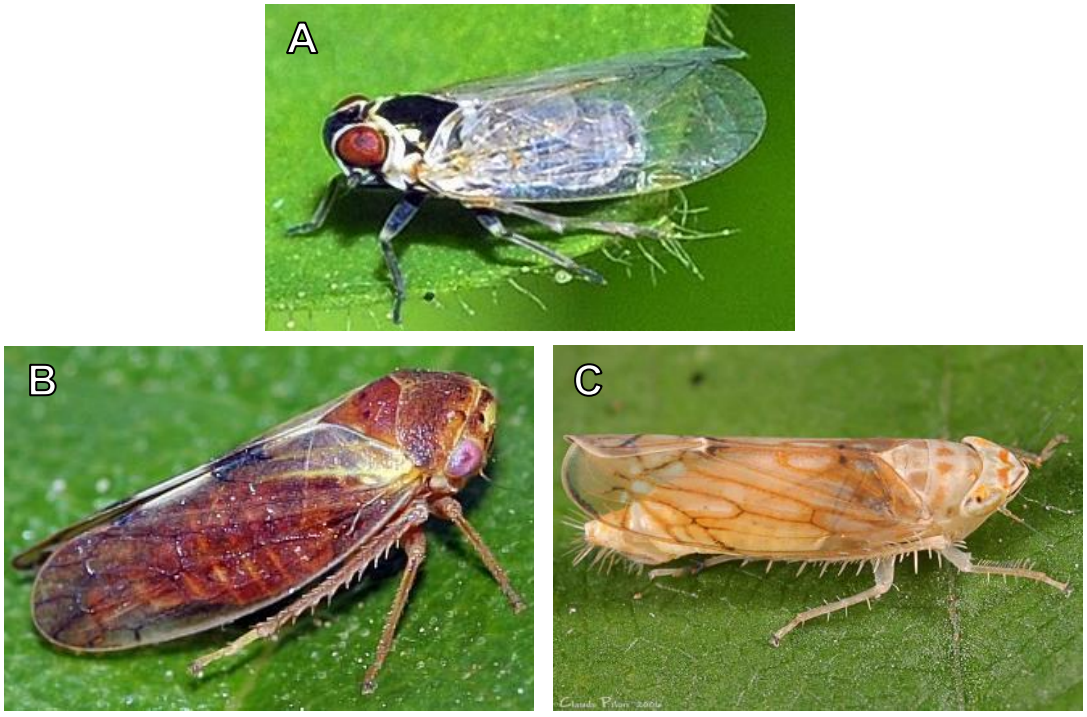

Figure 2.4: Known insect vectors responsible for the transmission and spread of quarantine GY diseases. A.) *Hyalesthes obsoletus* Signoret (Photo: José Luis García; http://faluke.blogspot.com/2015/02/hyalesthes-obsoletus.html); B.) *Oncopsis alni* Shrank (Photo: Gabriel Seljak; http://www1.pms-lj.si/animalia/galerija.php?load=2548); C.) *Scaphoideus titanus* Ball (Photo: Claude Pilon; https://cityportal.hr/strucnjaci-upozoravaju-zaustavite-ovog-cvrcka-prije-nego-vam-potpuno-unisti-vinograd/).

The expression and severity of GY symptoms differ among grapevine varieties. Typical symptoms associated with GY include yellow or bright red discoloration and necrosis of leaf veins and leaf blades, abnormal leaf shape and size, downward curling of leaves, incomplete lignification of

shoots, stunting and necrosis of shoots, abortion of flowers and berry withering (Figure 2.5) (Maixner *et al.*, 2006; Belli *et al.*, 2010). In comparison with other grapevine varieties, Chardonnay and Riesling are more severely affected by GY, whereas some rootstocks may be infected by phytoplasmas but do not show symptoms (Constable, 2010).

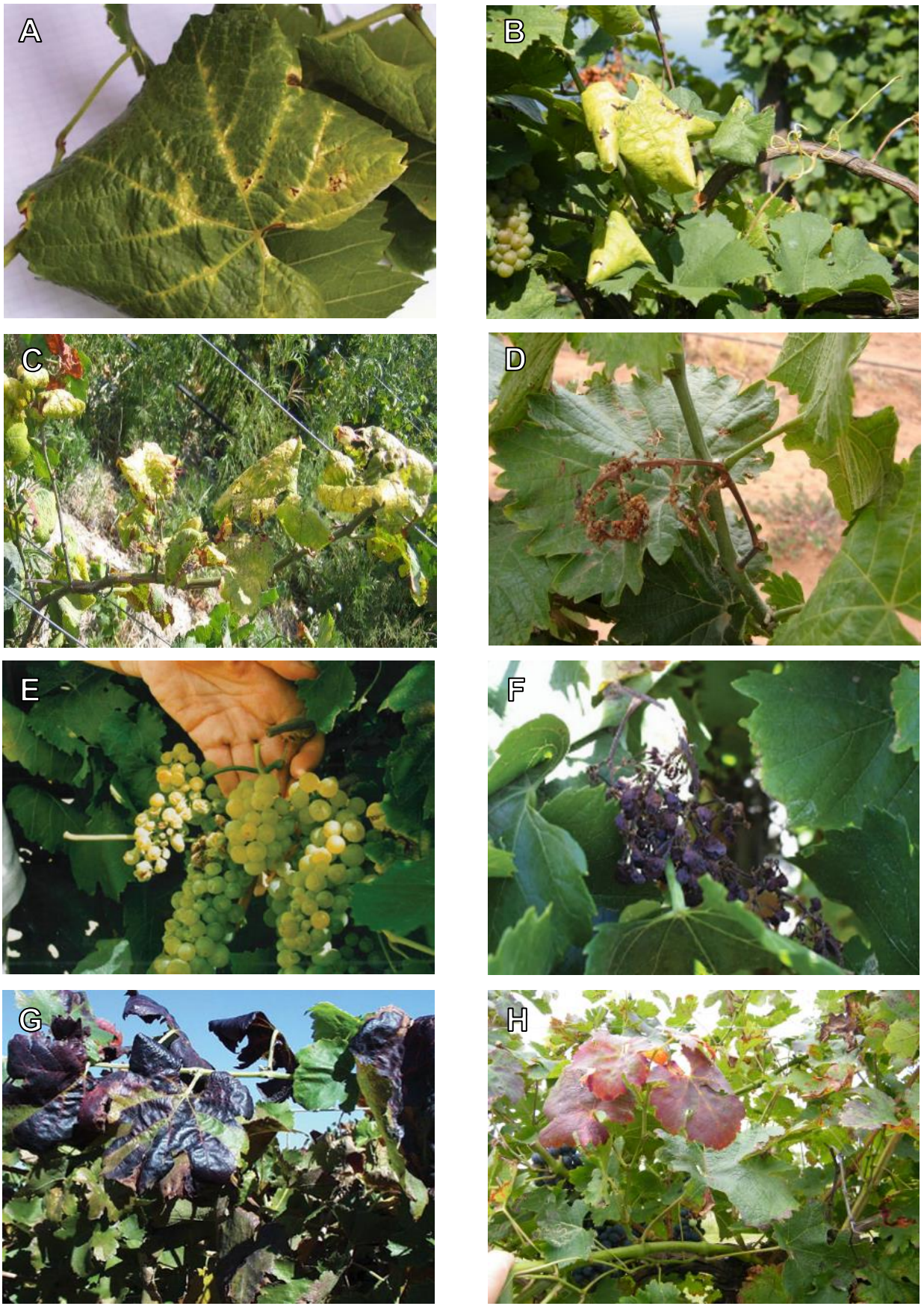

Figure 2.5: Images depicting typical GY disease symptoms. They include leaves with a wrinkled appearance, downward curling, yellowing and necrosis of leaves (A, B), incomplete lignification of shoots (C), flower abortion (D), shrivel and withering of berries (E, F), as well as leaf reddening in red cultivars (G, H) (Photos A, E-G: acquired from Constable and Radoni, 2011; Photos B, C, H: courtesy of Jeff Joubert and Roelene Carstens).

Table 2.2: Current status of molecular characterization, biology and vectors of phytoplasmas associated with GY diseases (Constable, 2010).

Grapevine yellows disease	Phytoplasma name	Ribosomal group (subgroup)	Known insect vector to grapevine	Preferred host plants of vector	Alternative hosts of the phytoplasma	Occurrence
Flavescence dorée	Flavesence dorée (FD; '*Ca*. Phytoplasma *Vitis*'*)	16SrV (-C, -D) or EY	*Scaphoideus titanus* Ball	*Vitis* spp.	*Clematis alba*	France, Italy, Spain, Serbia, Slovenia, Switzerland
Palatinate grapevine yellows	Palatinate grapevine yellows (PGY)	16SrV or EY	*Oncopsis alni* (Schrank)	*Alnus glutinosa*		Germany
Bois noir, Legno nero, Vergilbungskrankheit, Schwarzholzkrankheit	Stolbur (STOL, '*Ca*. Phytoplasma solani'*)	16SrXII-A or stolbur	*Hyalesthes obsoletus* (Signoret)	*Convolvulus arvensis,Urtica dioica,Ranunculus* spp., *Solanum* spp., *Lavandula* spp.	*C. arvensi, U. dioic, Ranunculus* spp., *Solanum* spp., *Lavandula* spp.	Europe, Israel, Lebanon
Australian grapevine yellows (AGY)	'*Ca*. Phytoplasma australiense'	16SrXII-B	ND†	ND	*Maireana brevifolia*	Australia
	Tomato big bud (TBB)	16SrII -D	ND	ND		Australia
Buckland Valley grapevine yellows	Buckland Valley grapevine yellows (BVGY)	16SrI-related or AY	ND	ND		Australia
Grapevine yellows	Aster yellows	16SrI (-B, -C) or AY	ND	ND		Italy, Chile, Tunisia
North American grapevine yellows (NAGY)	Virginia grapevine yellows I (NAGY I)	16SrI-A or AY	ND	ND	*Vitis* spp., Various herbaceous hosts	Virginia (USA)
	Western X Virginia grapevine yellows III (NAGYIII)	16SrIII-I or WX	ND	ND	*Vitis* spp., *Prunus* spp.	New York (USA), Virginia (USA)
Grapevine yellows	'*Ca*. Phytoplasma fraxini'	16SrVII	ND	ND	ND	Chile
Grapevine yellows	X-disease	16SrIII	ND	ND	ND	Italy, Israel

*Suggested *Candidatus* phytoplasma names; however, the species are still to be described. †ND = not determined.

2.5.2 Grapevine yellows in South Africa

Botti *et al.* (2006) first reported typical GY symptoms in South African grapevines and identified a mixed infection of phytoplasmas belonging to groups 16SrXII-A and 16SrII-B in symptomatic *V. vinifera* cv. 'Cabernet'. Later GY symptoms were also observed in vineyards of the Olifants River Valley (Western Cape) in 2006, and were shown to be caused by AY phytoplasma ('*Ca.* Phytoplasma asteris'), based on diagnostic nested-PCR results (Figure 2.6) (Engelbrecht *et al.*, 2010; GenBank: GQ365729.1). According to *in silico* classification with *i*PhyClassifier (Zhao *et al.*, 2009), this AY phytoplasma belong to the 16Sr group I, subgroup B (16SrI-B). Initially the occurrence of AY phytoplasma-infection was reported in Vredendal (Olifants River) and the Wabooms River area (Breedekloof) (Burger, 2008), and since GY symptoms were reported for grapevines in the Robertson, Trawal (Olifants River) and Montagu (Klein Karoo) wine production areas of the Western Cape Province (Carstens, 2014).

Figure 2.6: Area within a Chardonnay vineyard in Vredendal where grapevines suffering from AY phytoplasma-infection display stunted growth. Stunted growth is usually caused by delayed bud burst and dwarfed shoot growth.

The AY phytoplasma group is the world's most diverse and widespread phytoplasma subclade (Lee *et al.*, 2004). AY phytoplasmas are vectored by at least 30, often polyphagous, insect species, and consequently able to infect more than 80 plant species (Firrao *et al.*, 2007), including many weeds that surround important crops (Marcone *et al.*, 2000). Microscopic analysis was previously used to

determine that *Mgenia fuscovaria* (Hemiptera: Cicadellidae) is the predominant leafhopper in vineyards of the Vredendal and Waboomsrivier in 2009 (Figure 2.7) (de Klerk and Carstens, 2016). Transmission experiments conducted on vineyards in the vicinity of Vredendal (Western Cape) confirmed that *M. fuscovaria* is a vector of AY phytoplasma in South Africa (Krüger *et al.*, 2011). The experiments were conducted by placing bait plants, i.e. *N. benthamiana*, periwinkle (*C. roseus*) and certain grapevine plants, in an infected vineyard. Insects were collected using sticky traps, sweep-netting and vacuum samplers (pooters), and then stored in 95% ethanol until being examined microscopically (Krüger, 2012). Following controlled transmission analyses of AY by leaf-/planthoppers, the presence of AY phytoplasma in plants was tested using a CTAB extraction method for DNA isolation, followed by a sensitive TaqMan-based real-time PCR detection assay (Angelini *et al.*, 2007; Krüger, 2012).

Figure 2.7: Micrograph of a *Mgenia fuscovaria* specimen (Photo by courtesy of Michael Stiller).

As mentioned earlier, the epidemiology of four grapevine yellows (GY) diseases (FD, BN, AUSGY, NAGY) have been studied extensively (Maixner *et al.*, 2006; Constable, 2010). To gain more knowledge on the epidemiology of GY in South Africa, a survey that spanned four years was conducted on selected vineyards within the Vredendal wine producing district. The survey was used to determine incidence and spatial distribution of the disease which would assist in developing control measures to combat the disease (Carstens *et al.*, 2011; Carstens, 2014). Different cultivars (Chenin Blanc, Shiraz, Chardonnay, Cabernet Franc, Sauvignon Blanc, Pinotage and Colombar), were monitored and revealed that Chardonnay is especially susceptible to AY phytoplasma infection. Cultivars showing the highest mean cumulative disease incidences over four years were Pinotage (10.87%), Chenin Blanc (32.31%) and Chardonnay (37.77%) (Carstens, 2014). Such an infection rate may infer Chardonnay vineyards to be 100% infected with AY phytoplasma after 10 years which could have ruinous consequences for wine production in the future. Furthermore, the

survey also proved that symptomless plants can be infected with AY, and also corroborated data demonstrating the uneven distribution of the pathogen in grapevines (Spinas, 2013; Carstens, 2014).

For GY disease control in South Africa, it has been suggested that infected cordons as well as shoots with yellow leaves must be removed regularly throughout the growing season. The removal of the entire vine with its roots following harvest will also considerably reduce transmission of AY phytoplasma. Recent field trials in three viticultural areas, currently hampered by GY, showed that the contact insecticides Steward (active ingredient: indoxacarb) and Dursban (active ingredient: chlorpyrifos), as well as the systemic insecticide Kohinor (active ingredient: imidacloprid) provided excellent control of leafhoppers. Dursban, however, should not be applied within four weeks after bud burst, due to its phytotoxic effect on young leaves (de Klerk and Carstens, 2016).

B. MicroRNAs

2.6 Introduction

During the course of evolution plants have developed complex regulatory processes for development and to cope with changes in their environment. One such process involves the production of small non-coding endogenous RNAs (sRNAs) that can potentially affect different levels of gene expression. The use of high-throughput sequencing approaches allowed the discovery of a multitude of 20-26 nt small RNA species that accumulate in plant tissues (Lu *et al.*, 2005; Brodersen and Voinnet, 2006; Liu *et al.*, 2014). Although types of small RNAs may differ in terms of size, sequence, genomic distribution, biogenesis, and action, most of these molecules play an important role in mediating gene regulation through ‘RNA silencing’ or ‘RNA interference’ (RNAi) (Brodersen and Voinnet, 2006).

The biogenesis and function of small RNAs have been studied in great detail and revealed their role in transcriptional gene silencing (TGS), post-transcriptional gene silencing (PTGS), as well as translational repression (Hamilton and Baulcombe, 1999; Mette *et al.*, 2000; Brodersen *et al.*, 2008). These processes share three common biochemical features: (i) formation of double stranded RNA (dsRNA); (ii) processing of dsRNA to small 20-26 nt small interfering RNAs (siRNAs) with overlapping ends; and (iii) inhibitory action of a selected siRNA strand within effector complexes acting on partially or fully complementary RNA (Brodersen and Voinnet, 2006). Axtell (2013), recently classified RNAs derived from single-stranded hairpins as hairpin RNAs (hpRNAs), and proposed a primary distinction between siRNAs and hpRNAs when classifying small RNAs. According to this classification scheme, hpRNAs can be divided into microRNAs (miRNAs) and all

other hpRNAs, while siRNAs are divided into heterochromatic siRNAs, secondary siRNAs, and natural antisense transcript siRNAs (nat-siRNAs). Henceforth miRNAs will be discussed in greater detail.

miRNAs were first discovered as components of the heterochronic pathway in *Caenorhabditis elegans* (Lee *et al.*, 1993), and have since been identified in diverse species of the four eukaryote kingdoms (protists, fungi, plants, and animals). Mature miRNAs are typically 20 to 24 nt in length and play a major role in PTGS and translational repression during plant development, as well as responses to biotic and abiotic stresses (Ruiz-Ferrer and Voinnet, 2009; Sunkar *et al.*, 2012). MiRBase (Kozomara and Griffiths-Jones, 2011) is currently the most useful miRNA registry with a total of 28645 high confidence miRNA entries (release 21), which include entries for 73 different plant species. To date, 186 mature grapevine miRNA sequences with 163 known precursor sequences have been deposited in miRBase v21 (release 21), and represent 47 different miRNA families.

2.7 miRNA Biogenesis

The miRNA pathway is an essential part of the gene silencing machinery of plants since it modulates homeostasis through the inactivation of specific mRNAs, especially those encoding transcription factors. Unlike animals, plant miRNA processing is completed exclusively within plant nuclei (Park *et al.*, 2005), where primary miRNAs (pri-miRNAs) originate from *MIR* genes located within intergenic spacer regions of the genome. Pri-miRNAs are predominantly transcribed by RNA Polymerase II (Pol II) and harbour imperfect, self-complementary stem-loop regions to form characteristic hairpin structures (Bartel, 2004; Lee *et al.*, 2004).

Before further processing, pri-miRNAs are capped at their 5'-end and polyadenylated at their 3'-end, similar to most Pol II-derived transcriptional events (Chen, 2009). The stem-loop region is then recognised by the RNase III-type enzyme DICER-LIKE1 (DCL1), which cleaves the pri-miRNA near the base of the stem region to produce a shorter hairpin structure known as a precursor miRNA (pre-miRNA) (Park *et al.*, 2002; Kim, 2005). The RNA-binding protein DAWDLE (DDL) presumably stabilizes pri-miRNAs for their conversion in nuclear processing centres called D-bodies to stem-loop pre-miRNAs (Yu *et al.*, 2008). Excision of pre-miRNAs from pri-miRNA involves the joint action and physical interaction of the C2H2-zinc finger protein SERRATE (SE), the double-stranded RNA-binding protein HYPONASTIC LEAVES1 (HYL1), DCL1, and nuclear cap-binding complex (CBC) (Voinnet, 2009).

Each pre-miRNA stem region contains one or more short, complementary, 20-24 nt miRNA-miRNA* duplexes. A consecutive DCL1-mediated cleavage step then follows where the miRNA-miRNA* duplexes, which contain typical 3'-nucleotide overhangs, are liberated from the stem region (Kurihara and Watanabe, 2004; Xie *et al.*, 2005). The S-adenosyl methionine-dependent methyltransferase HuaEnhancer 1 (HEN1) then stabilises the mature miRNA duplexes through 3'-methylation in order to block their uridylation and subsequent degradation (Li *et al.*, 2005; Yang *et al.*, 2006). The miRNA duplexes are then exported to the cytoplasm by a plant homolog of Exportin-5 called HASTY (HST) (Park *et al.*, 2005) where one strand of the duplex is incorporated into an Argonaute (AGO) protein, the catalytic component of the RNA-induced silencing complex (RISC) (Baumberger and Baulcombe, 2005). Association of plant miRNAs with AGO1 promotes guided binding to its target mRNA through perfect or near-perfect complementarity within RISC (Rhoades *et al.*, 2002; Schwab *et al.*, 2005). This assembly allows regulation of gene expression through either translational repression or endonucleolytic cleavage of the target mRNA by AGO1 leading to mRNA degradation (Bartel 2004; Jones-Rhoades *et al.*, 2006).

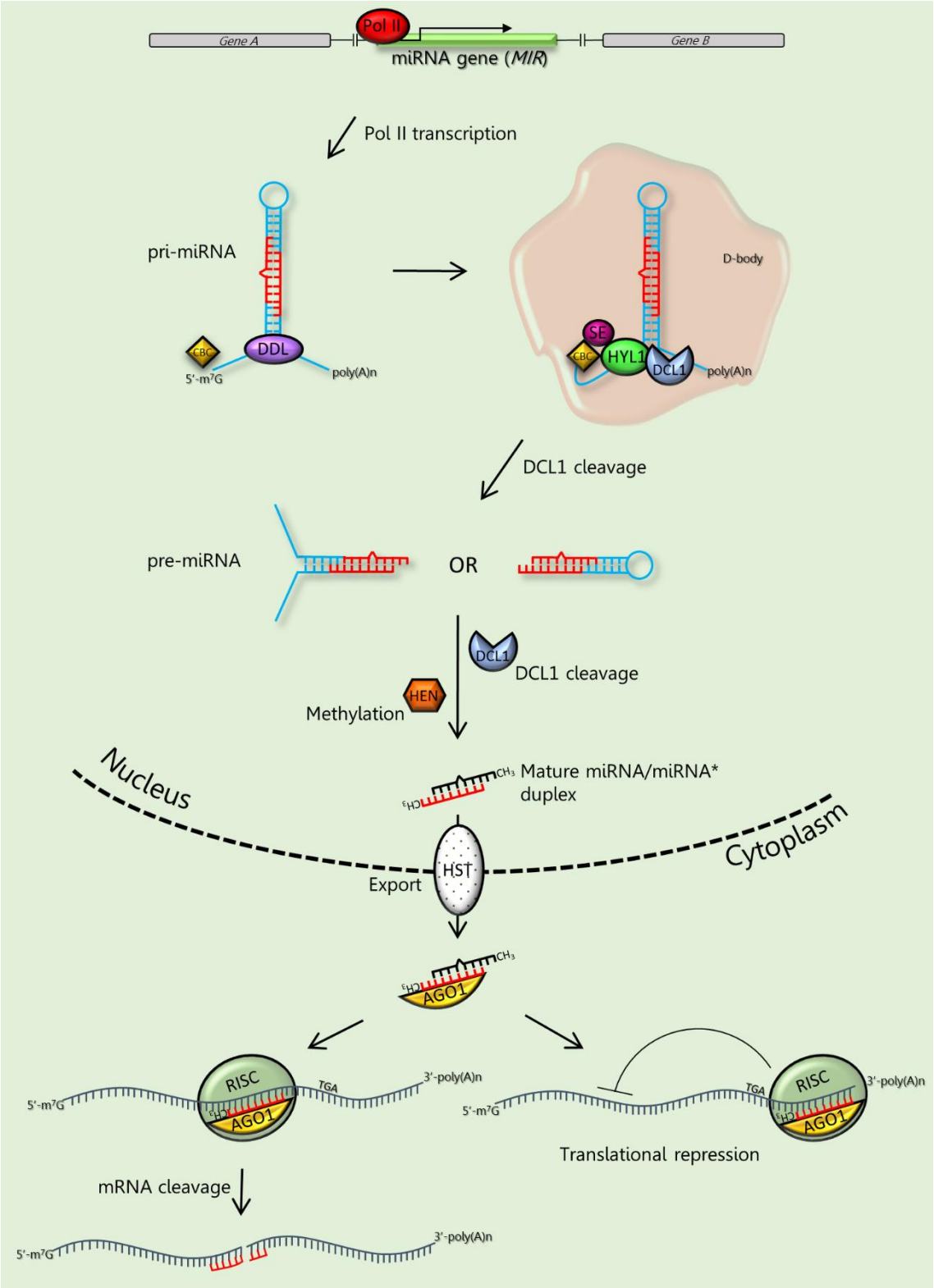

Figure 2.8: Schematic diagram of miRNA biogenesis and miRNA-mediated target regulation. The transcript of the *MIR* gene, pri-miRNA, folds back to form a hairpin structure and is stabilised by the DAWDLE (DDL) protein. Splicing and processing (within nuclear dicing bodies) of the pri-miRNA and pre-miRNA is facilitated by interaction with the proteins SERRATE (SE) HYPONASTIC LEAVES1 (HYL1), DICER-LIKE1 (DCL1) and cap-binding proteins (CBP). This results in one or more miRNA/miRNA* duplexes. These duplexes are then methylated on the 3'-end by HEN1 before being transported to the cytoplasm by HASTY (HST). The Argonaute 1 (AGO1) protein facilitates selection, incorporation and stabilisation of one strand of the duplex. Association with AGO1 within RISC promotes mRNA-binding with the guide miRNA strand through complementarity, leading to regulation of gene expression through either translational repression or endonucleolytic cleavage of the target mRNA.

2.8 Role of miRNAs in plant stress responses

Most plants have evolved sophisticated mechanisms essential to their survival. These mechanisms include those leading to normal development, as well as those involving regulatory pathways that enable stress tolerance. Stress-related pathways are extremely important in cases where plants are infected with a biotic agent (bacterium, virus, fungus or nematode), or during environmental (abiotic) changes. There is a vast number of studies that encompass the identification and characterisation of miRNAs in dozens of plants species exposed to stressful conditions. Comprehensive reviews are available that cover the role of miRNAs and siRNAs in plant development, as well as biotic and abiotic stress responses (Jones-Rhoades *et al.*, 2006; Jung *et al.*, 2009; Chen, 2010; Khraiwesh *et al.*, 2012; Sunkar *et al.*, 2012; Jin *et al.*, 2013; Gupta *et al.*, 2014; Kumar, 2014; Weiberg and Jin, 2015; Chaloner *et al.*, 2016; Couzigou and Combier, 2016; Li and Zhang, 2016).

Plant abiotic stress can be referred to as any adverse impact of non-living factors from the environment on living plants. The use of bioinformatics data analyses have revealed differential miRNA expression induced by abiotic stresses such as: 1) water stress (drought, acute water stress, water salinity) (Liu *et al.*, 2008; Zhou *et al.*, 2010; Trindade *et al.*, 2010; Carnavale-Bottino *et al.*, 2013); 2) UV-B stress (Zhou *et al.*, 2007; Jia *et al.*, 2009); 3) temperature stress (heat or chilling) (Liu *et al.*, 2008; Zhang *et al.*, 2009; Xin *et al.*, 2010); 4) oxidative stress caused by reactive oxygen species (ROS) (Sunkar *et al.*, 2006; Li *et al.*, 2011); 5) nutrient stress during copper, sulphate, nitrogen, aluminium and phosphate starvation (Pant *et al.*, 2008; Yamasaki *et al.*, 2009; Buhtz *et al.*, 2010; Liang *et al.*, 2012; Zeng *et al.*, 2012); 6) phytohormone stress (Sunkar and Zhu 2004; Mallory *et al.*, 2005); and 7) mechanical stress (Lu *et al.*, 2005). The functional role of miRNA-mediated regulation during biotic stress, however, was important for this study.

Many studies have revealed the complexity and overlapping nature of plant responses to different stresses. Understanding the complexity of small RNA-guided stress regulatory networks may provide us with new insights that could be invaluable to the genetic improvement of stress tolerance in agricultural plants (Liu and Chen, 2010; Zhang and Wang, 2014).

2.8.1 miRNA-mediated responses to biotic stress

Plants have evolved an immune system with multiple layers of protection in response to various pathogen attacks. These include: 1) non-host resistance via physical barriers; 2) PAMP-triggered immunity (PTI); and 3) effector-triggered immunity (ETI) (Jones and Dangl, 2006; Chisholm *et al.*, 2006; Jin, 2008; Ruiz-Ferrer and Voinnet, 2009; Budak *et al.*, 2015). Following infection,

pathogens encounter the first layer of defence which involves cell wall reinforcement and deprivation of nutrients and other factors required for pathogen growth and multiplication. Pathogens that successfully overcome this barrier and gain access to the cells, encounter the PTI defence layer. This involves pathogen recognition receptors (PRRs), which senses microbial or pathogen-associated molecular patterns (MAMPs or PAMPs), such as bacterial flagella. PTI is usually a basal defence cascade, which involves mitogen-activated protein kinases (MAPK), production of ROS and nitric oxide (oxidative burst), cell wall reinforcement, and salicylic acid (SA) synthesis and signalling.

However, many pathogens can counter PTI by delivering certain virulence factors called effectors. In turn, plants have evolved another counter-defensive response by producing resistance (R) proteins which triggers ETI signalling. This occurs when the R proteins recognise specific pathogen effectors, such as bacterial avirulence (avr) proteins, and rapidly re-program expression of crucial genes to inhibit pathogen growth. ETI often culminates to hypersensitive response (HR) in the form of programmed cell death, which is accompanied by a potent SA-mediated systemic defence response (Jones and Dangl, 2006; Chisholm *et al.*, 2006; Jin, 2008; Ruiz-Ferrer and Voinnet, 2009; Budak *et al.*, 2015).

Numerous studies have linked miRNAs and siRNAs to biotic stress responses in plants (PTI and ETI), and their role in plants infected by pathogenic bacteria, viruses, nematodes, and fungi has been reported (Khraiwesh *et al.*, 2012; Weiberg and Jin, 2015). Normally, these sRNAs are differentially expressed upon pathogen attack in order to allow or inhibit target expression (Ruiz-Ferrer and Voinnet, 2009). This may lead to activation and suppression of a large array of genes. For example, over 2000 genes are differentially expressed in *Arabidopsis* in response to *Pseudomonas syringae* infection upon ETI signalling by two nucleotide-binding site leucine-rich repeats (NBS-LRR)-type *R* genes, *RPS2* and *RPM1* (Tao *et al.*, 2003).

So far, the best-understood plant-pathogen model of miRNAs and their role in antibacterial PTI was formed from studies where *Arabidopsis* was infected with *P. syringae* pv. *tomato* (*Pst*) (Navarro *et al.*, 2006; Fahlgren *et al.*, 2007; Zhang *et al.*, 2011). It was demonstrated that the PAMP peptide, flg22, caused induction of miR393 expression. MiR393 can target mRNAs encoding the F-box auxin receptor transport inhibitor response 1 (TIR1) and Auxin signalling F-box proteins (AFB) 1, 2 and 3, thereby repressing auxin signalling (Jones-Rhoades and Bartel, 2004). Conversely, down-regulation of auxin signalling resulted in increased bacterial resistance to virulent *Pst* DC3000 (Navarro *et al.*, 2006). Similarly, based on a sRNA-profiling study, Fahlgren *et al. (*2007) showed that miR160, miR167 and miR393 were significantly up-regulated in *Arabidopsis* challenged with

Pst DC3000 *hrcC*. MiR160 and miR167 each target mRNAs encoding members of the auxin response factor (ARF) family of transcription factors (Rhoades *et al.*, 2002).

The role of miRNAs in plant basal defence was further supported by the finding that *Pst* DC3000 *hrcC* growth was consistently higher in the miRNA-deficient mutants *dcl1* and *hen1*. Furthermore, *dcl1* also sustained growth from other bacteria that are non-pathogenic to *Arabidopsis*, indicating miRNA contribution to non-host resistance (Navarro *et al.*, 2008). In another study where *Arabidopsis* leaves were infected with *Pst* DC3000 *hrcC*, sRNA-profiling data showed differential expression of miR160, miR167, miR390 and miR393, also suggesting regulation of genes involved in the auxin signalling pathway (Zhang *et al.*, 2011).

miR319 and miR159 were also induced in *Pst* DC3000-infected *Arabidopsis*, possibly promoting SA-mediated defence responses due to suppression of genes encoding TCP and MYB transcription factors, respectively (Reyes and Chua, 2007; Zhang *et al.*, 2011). The abovementioned findings imply that negative regulators of plant immune responses can be targeted by miRNAs leading to up-regulation of resistance pathways.

There are also many studies on the role of miRNA interactions in disease signalling during fungal invasion of plants (Gupta *et al.*, 2014). The use of high-throughput technologies and bioinformatics has allowed the identification and expression profiling of a vast number of miRNAs involved in fungal infection. For example, Zhao *et al. (*2012) used a microarray/qPCR-based approach to identify fungus-responsive miRNAs in *Populus trichocarpa* inoculated with the polyphagous fungus *Botryosphaeria dothidea*, which causes stem bark disease. Forty-one microarray probes, representative of 12 miRNA families showed significant altered expression. miR159, miR168, miR172, miR319, miR1450, and 13 members of miR166 showed continuous increase at 3, 5, and 7 days after infection (DAI) while miR156, miR160, miR164, miR1448, miR398, miR408, and three members of miR166 showed increased expression only at 5 DAI than that at 3 or 7 DAI. miRNA target prediction indicated gene products involved in regulating a broad spectrum of cellular processes e.g. defence response, regulatory cascades and metabolic pathways.

miRNAs targeting *R* gene transcripts are also observed in different plants infected with bacteria, fungi and viruses. For instance, an *NBS-LRR* gene was suppressed by miR482 four hours after inoculation with *P. syringae* DC3000 in tomato (Shivaprasad *et al.*, 2012). Differential miRNA expression was also observed in pines in response to infection by the rust fungus *Cronartium quercuum* (Lu *et al.*, 2007). Ten out of 11 miRNA families, including seven pine-specific miRNAs, were down-regulated. Remarkably, most of the pine-specific miRNA families target defence-related genes that encode R proteins and receptor-like kinases (RLKs), as well as genes (targeted by pbe-

miR156), responsible for organ development. This suggested that miRNA suppression induces growth- and *R* gene expression to restrict fungal growth.

In another study, two miRNAs (nta-miR6019 and nta-miR6020) that guide sequence-specific cleavage of transcripts of a NB-LRR immune receptor, which confers resistance to tobacco mosaic virus (TMV), were identified (Li *et al.*, 2012a). This is one of many examples where miRNAs triggered the biogenesis of 21 nt secondary (potentially trans-acting) siRNAs (tasiRNAs) which are "in phase" with the miRNA binding site (phasiRNAs) (Zhai *et al.*, 2011). The production of tasi-RNAs requires either the so-called "one-hit" model (single-target site) by 22-nt miRNAs, or "two-hit" model of dual miRNA target sites (Figure 2.9). phasiRNAs are produced from an mRNA that are converted by RNA-dependent RNA polymerase 6 (RDR6) to dsRNA which is processed by DCL4. They have the ability to direct AGO1-dependent slicing of *NB-LRR* transcripts, either in *cis* or *trans* at other *NB-LRR* loci, representing a self-augmenting regulatory network (Figure 2.9) (Zhai *et al.*, 2011; Fei *et al.*, 2013). Recently, Källman *et al. (*2013) reported the regulation of *NB-LRR* transcripts by secondary siRNAs in a wide variety of plant species, indicating the high conservation and ancient origin of phasiRNAs and their role in plants. NBS-LRRs were up-regulated to induce pathogen resistance in *Solanum lycopersicum* (tomato) and *Gossypium raimondii* (cotton) following suppression of the miR482-mediated silencing cascade (Shivaprasad *et al.*, 2012; Zhu *et al.*, 2014).

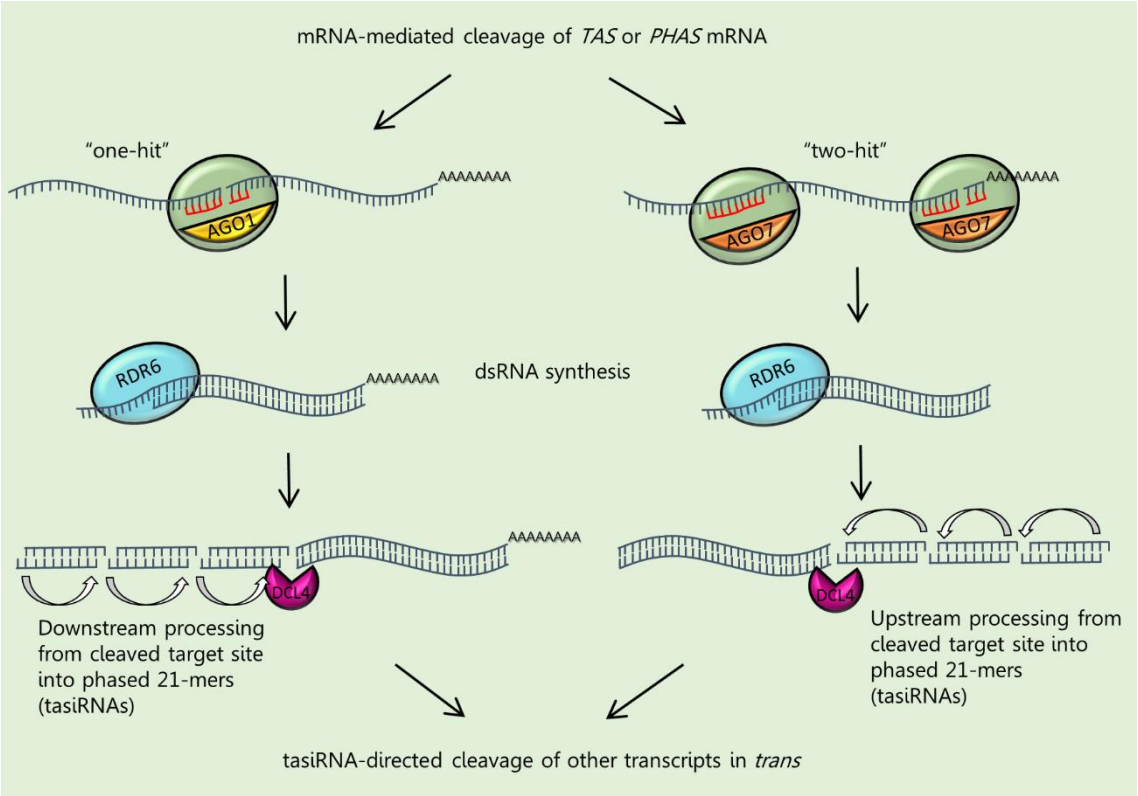

Figure 2.9: Schematic diagram illustrating the "one-hit" and "two-hit" models of phasiRNA biogenesis. Abbreviations: Trans-acting-siRNA-generating loci (*TAS* genes); phasiRNA-producing loci (*PHAS*); Argonaute (AGO); RNA-dependent RNA polymerase 6 (RDR6); DICER-LIKE4 (DCL4) (modified from Fei *et al.*, 2013).

When considering previous studies on miRNA expression during plant disease, we notice that pathogen responsive miRNAs can target several genes at once and each target gene can regulate numerous biochemical and physiological processes. Therefore, regulation and crosstalk of gene expression during disease development is an actively growing area to develop a better understanding of disease pathogenesis. It has been noticed that different miRNAs target the same gene but their expression pattern varies with the type of plants and pathogen under study.

2.9 Methods for the discovery, characterisation and quantification of plant miRNAs

miRNAs and the targets they regulate have a profound impact on plant physiology and there has been a growing desire to understand their roles in cellular processes. Since miRNAs are either conserved or non-conserved, found in plants, animals and invertebrates, and show tissue-specific differential expression under different conditions, tools are needed to discover, characterise and quantify them. Over the last two decades different approaches have been used, some of which have been replaced or obsoleted by alternative high-throughput technologies to overcome their limitations.

2.9.1 Methods for the discovery of miRNAs

The identification of known and novel miRNAs in a biological system lays the foundation for miRNA research. Methods for miRNA identification were developed based of the availability or unavailability of prior sequence information. A conventional method known as forward genetic screening, whereby mutant genes are isolated from organisms showing abnormal phenotypes, was used to identify the first known miRNA, lin-4, in *C. elegans*, as well as ath-miR164c responsible for the extra petal phenotype of mutant *Arabidopsis* plants (Lee *et al.*, 1993; Baker *et al.*, 2005). Despite its use in functional characterisation, this method is costly, time-consuming and only effective on a small scale. In the past, cloning and direct Sanger sequencing of small RNA molecules lead to a breakthrough since hundreds of miRNAs could be identified (Llave *et al.*, 2002; Ambros and Lee, 2004). This approach can be applied to any organism, with little or no genomic information, and miRNAs can be identified independent of their function, thus allowing the identification of redundant miRNAs. The only limitation of this approach is that miRNAs that are expressed at a low level in a specific condition or cell type are difficult to detect.

Currently, high-throughput sequencing technologies, also known as NGS or deep sequencing has become a widely used strategy for plant miRNA research on a genomic scale. Reasons why NGS is such a promising tool include: (1) the production of millions of sequence reads at lower cost in a shorter time; (2) it delivers greater sensitivity and accuracy than previous technologies such as microarray hybridisation techniques; (3) it does not rely on target-probe hybridization, permitting sequencing of the exact transcript on a single nucleotide resolution; (4) it requires no previous sequence information, utilising relevant databases for comparison of sequencing information; (5) it provides high depth of coverage for any library type since it can be modified to study specific nucleic acid components, e.g. small RNA-seq (sRNA-seq); (6) data analysis can be done independent of reference genome sequences; and (7) post-transcriptional modifications can be detected (Veneziano *et al.*, 2015).

Popular NGS platforms that can be used for sRNA deep sequencing include the iSeq100, MiniSeq, MiSeq series, and HiSeq series (Illumina/Solexa) (https://www.illumina.com/systems/sequencing-platforms.html), as well as the Sequencing by Oligonucleotide Ligation and Detection (SOLiD) system (ABI/Life Technologies) (Norden-Krichmar *et al.*, 2011) (Table 2.3). A single NGS run can provide a plethora of sequence data that can be extensively interrogated to detect millions of sRNA sequences in a wide variety of organisms with a high degree of reliability. For instance, an Illumina HiSeq 2500 instrument can deliver around one billion sRNA sequence reads in less than 2 days (Kang and Friedländer, 2015).

Table 2.3: Popular NGS platforms used for sRNA-seq (Kulski, 2016; www.illumina.com)

NGS Platform	Maximum read length	Maximum reads per run	Maximum output	Run time	Developer
iSeq100 system	2 × 150 bp	4 million	1.2 Gb	9-17.5 hours	Illumina Inc.
MiniSeq system	2 × 150 bp	25 million	7.5 Gb	4-24 hours	Illumina Inc.
Miseq series	2 × 300 bp	25 million (v3 kits)	15 Gb	4-55 hours	Illumina Inc.
HiSeq series	2 × 150 bp	5 billion	1500 Gb	< 1-3.5 days (HiSeq 3000/HiSeq 4000); 7 hours-6 days (HiSeq 2500)	Illumina Inc.
SOLiD system	~50 bp	1 billion	1-3 Gb	14 days	ABI/Life Technologies

All currently available NGS platforms require pre-processing of total RNA into a sRNA library suitable for sequencing (van Dijk *et al.*, 2014). For example, the TruSeq® Small RNA Library Prep Kit protocol used for Illumina sRNA-seq is comprised of 3' and 5' adaptor ligation to total RNA, enrichment of adaptor-ligated RNA by reverse transcription and library amplification, gel purification of the amplified cDNA constructs, selective gel excision of adaptor-ligated-size constructs derived from sRNA fragments, and gel purification of enriched cDNA that will serve as template for subsequent cluster generation by bridge amplification (https://support.illumina.com). During bridge amplification, massively parallel sequencing occurs which uses a proprietary reversible terminator-based method where single bases are detected during incorporation into growing DNA strands (Moorthie *et al.*, 2011), and has also been termed sequencing by synthesis (SBS) (Figure 2.10; Figure 2.11)

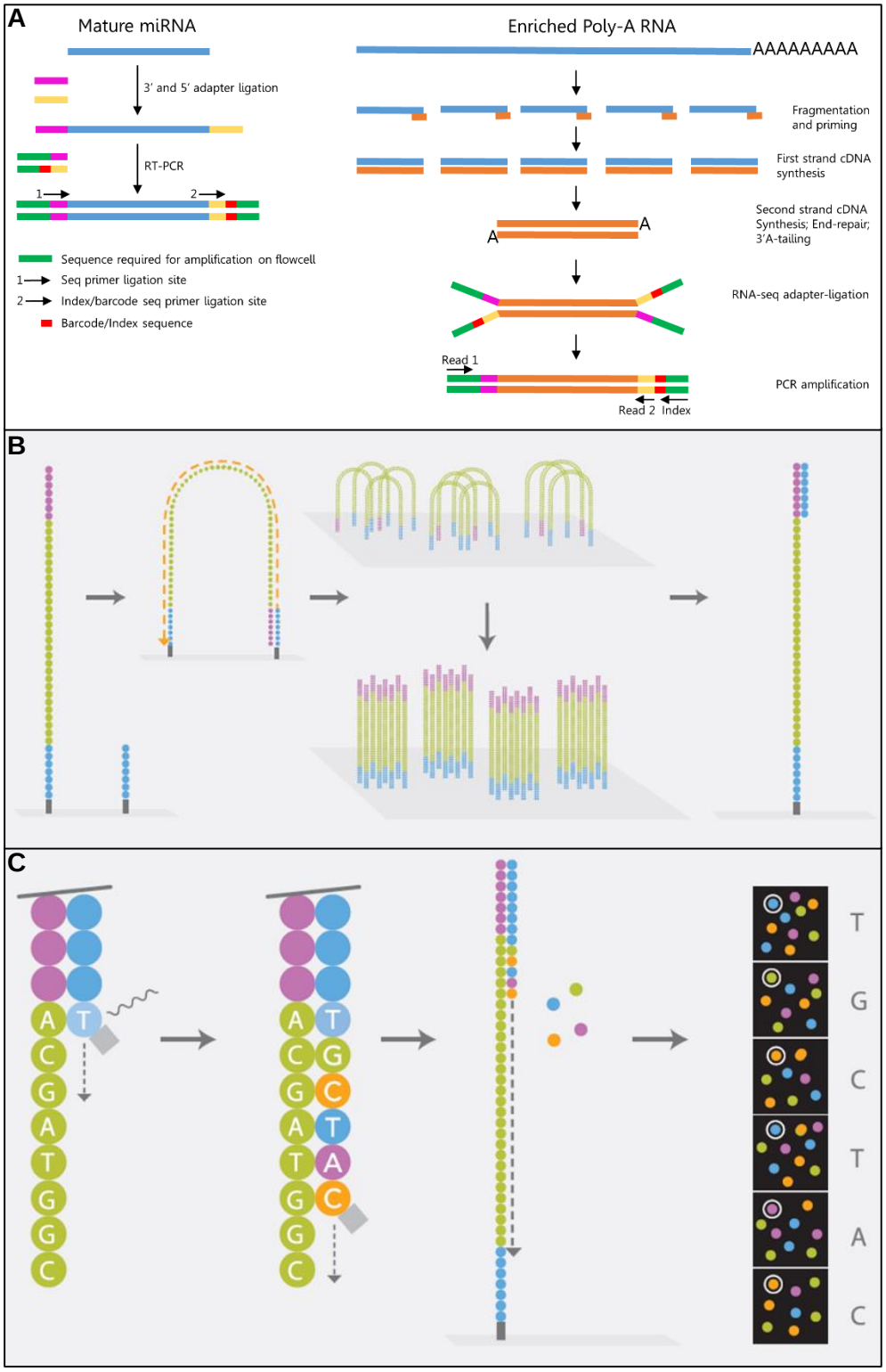

Figure 2.10: Simplified representation of steps for cDNA library construction and Illumina sequencing by synthesis. cDNA sequencing library preparation yields enriched sequencing constructs, each comprised of a dual-index adapter-ligated cDNA insert (A). The cDNA libraries are attached to a flowcell and undergoes cluster generation (B), followed by sequencing by synthesis with reversible terminators (C) (www.illumina.com).

2.9.2 Computational tools and resources for miRNA identification

The advent of sRNA-seq has not only reduced the cost for miRNA discovery but the magnitude of its data outputs have also allowed the detection of millions of transcript reads, including reads for lowly expressed miRNAs. Computational prediction of novel miRNAs is based on miRNA gene identification that pertains to their origin in the genome. This is not a trivial task since hairpin structures in eukaryotic genomes are not necessarily miRNA-related (Berezikov *et al.*, 2006). Accurate miRNA prediction from NGS data, depending on the organism being studied, requires computational tools involving sophisticated algorithms and proper computational infrastructure (Figure 2.11). Some of them also integrate expression analysis, functional analysis, as well as target prediction of miRNAs. A big advantage of such tools is that the prediction of potential candidate miRNAs and their putative targets can subsequently be validated directly or indirectly by experimental approaches.

miRNA prediction tools rely on incorporating several miRNA characteristics necessary for sRNA post-filtering steps. These include the high conservation of miRNA genes in the genomes of related species, the genomic distribution, location and length of miRNA genes, differences between plant and animal miRNAs, significant structural pre-miRNA features such as the stem-loop hairpin conformation, and the extent of base pairing and nucleotide composition within the duplex region(s) (Bartel *et al.*, 2004; Gomes *et al.*, 2013). Primary criteria described by Meyers *et al. (*2008) for annotating duplex-forming plant pre-miRNAs are also often applied by these prediction tools. Statistical methods incorporating pre-miRNA thermodynamics related to high negative minimum folding energy (MFE; ΔG) of potential pre-miRNAs (Zhang *et al.*, 2006), may also be integrated.

miRNA prediction pipelines are generally comprised of three steps (Figure 2.11). First, mapping of reads onto a reference genome to identify sRNA read clusters. Second, loci associated with these read clusters are extended to include defined flanking regions, followed by sequence extraction of these extended genomic regions. Lastly, evaluating transcript sequences in terms of the structural pre-miRNA features mentioned above. Besides the core prediction methods, the choice of the prediction tool may rely on other factors. These include the mapping tool, whether read pre-processing is required, whether the tool has a graphical user interface (GUI) or is command-line driven, and whether additional analyses, e.g. expression analysis and target prediction, are supported (Kang and Friedländer, 2015).

Numerous bioinformatic tools have been developed for the identification of known and novel miRNAs. Lukasik *et al. (*2016) recently compiled a database i.e. Tool4miRs, which present a comprehensive collection of more than 170 methods for miRNA analysis (https://tools4mirs.org).

These tools usually exist as integrated pipelines in the form of standalone software or launched from web/cloud-based servers. Initially, computational identification of plant miRNAs relied on comparative approaches which were based on sequence and secondary structure conservation of known miRNAs, e.g. MIRcheck (Jones-Rhoades and Bartel, 2004) and MiRAlign (Wang *et al.*, 2005). More recent miRNA prediction tools that use complex algorithms to analyse sRNA-seq data does not rely on phylogenetic conservation of miRNAs and are therefore beneficial for the discovery of novel miRNAs. These tools rely on nucleotide sequence characteristics along with other structural and thermodynamic parameters of pre-miRNAs. Popular examples include miRDeep-P (Friedländer *et al.*, 2008; Yang and Li, 2011), miRCat (Stocks *et al.*, 2012), miRanalyzer (Hackenberg *et al.*, 2009), miRExpress (Wang *et al.*, 2009), sRNAbench (Barturen *et al.*, 2014; Rueda *et al.*, 2015) and Shortstack (Axtell, 2013).

miRDeep-P is the plant-specific version of miRDeep since the output (from the core algorithm, miRDeep) is filtered with pre-defined plant-specific criteria (Friedländer *et al.*, 2008; Meyers *et al.*, 2008; Yang and Li, 2011). The pipeline was developed in Perl and consists of mapping of sRNA reads to a reference sequence using Bowtie (Langmead *et al.*, 2009), and secondary structure prediction using the Vienna RNA package (Hofacker, 2003). No prior sequence information is required for miRDeep-P and outputs include information on novel miRNAs, stem-loop structures, genomic location, and quantification of the signature distribution of sRNA reads. ShortStack involves similar steps for miRNA predictions, but allows use of alternative read mappers and permits extensive flexibility in analysis since users have freedom to set different parameters (Axtell, 2013). Users may also provide an input file containing a set of genomic loci used to flag for overlaps with known small RNA loci. A key feature of ShortStack is the detailed analysis of the size distributions of sRNAs within sRNA genes.

miRanalyzer and sRNAbench are examples of miRNA prediction tools that employ a machine-learning algorithm (Hackenberg *et al.*, 2009; Barturen *et al.*, 2014). miRanalyzer employs a random forest prediction algorithm, while sRNAbench uses hierarchical clustering to predict miRNAs. sRNAbench maintains the main features implemented in its predecessor program, miRanalyzer, but include updated features such as multi-species support, genome and library mapping approaches, and improved prediction of novel miRNAs and isomiRs. Both pipelines can perform differential expression analysis of profiled miRNAs as well as miRNA target prediction.

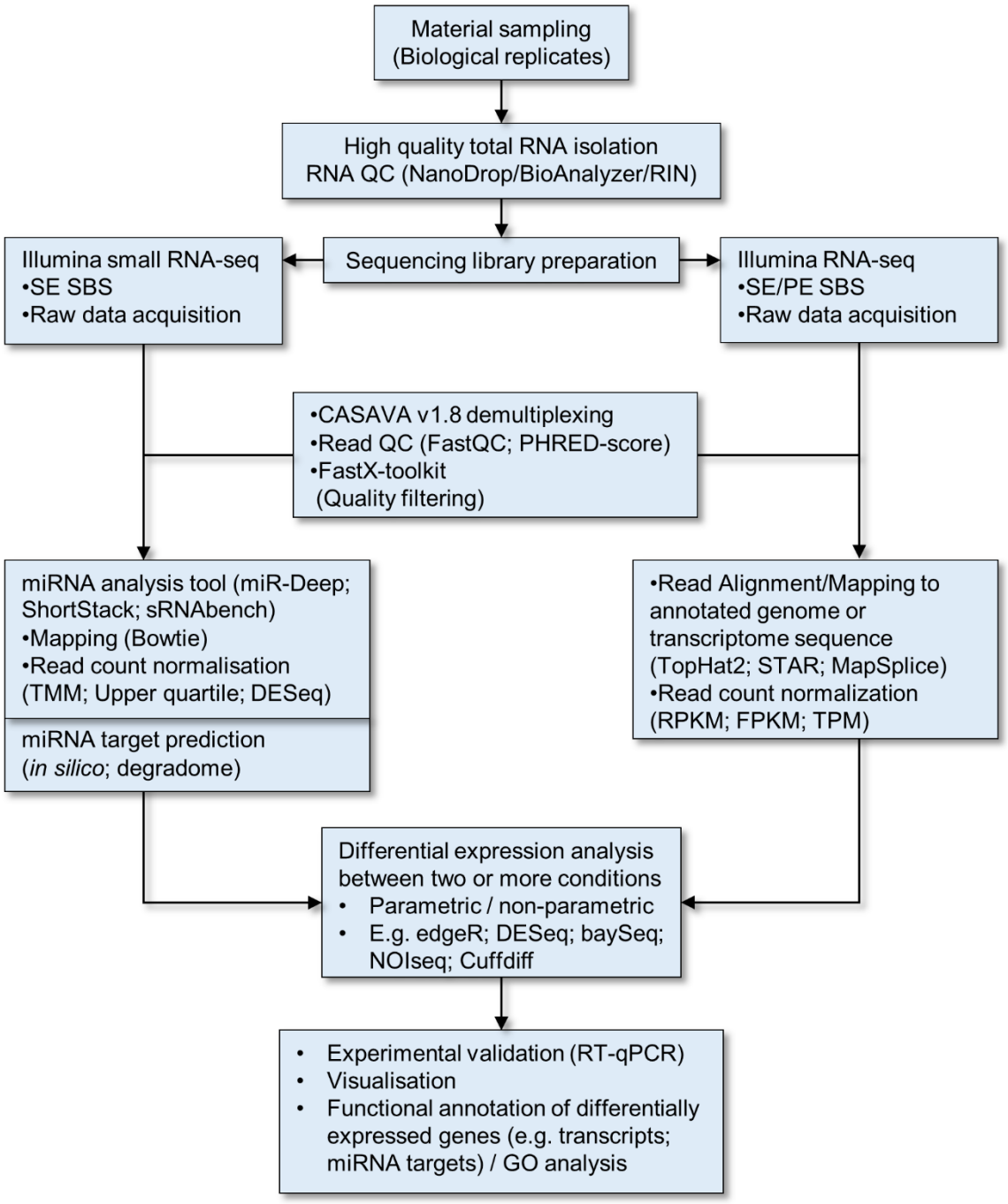

Figure 2.11: Schematic diagram of a simplified sRNA and RNA-seq workflow, outlining experimental design and data analysis procedures.

2.9.3 miRNA target prediction and validation

In order to study miRNA-target interactions, sequences of miRNAs and transcripts are required (Figure 2.11). MiRBase is currently the registry with the largest collection of miRNA sequence data for various species (Kozomara and Griffiths-Jones, 2011). Besides miRBase, other well-annotated databases, i.e. PMRD (Zhang *et al.*, 2010), PmiRKB (Meng *et al.*, 2011), MicroPC (Mhuantong *et*

al., 2009), miRNEST (Szcześniak *et al.*, 2011), Rfam (Nawrocki *et al.*, 2015), and miRVIT (Chitarra *et al.*, 2018) are useful in plant miRNA studies.

To study the function of novel miRNAs, they need to be experimentally-validated in terms of their expression and the target(s) they regulate. Well-established techniques for miRNA target validation includes reverse transcription quantitative PCR (RT-qPCR) for mRNA quantification, and parallel analysis of RNA ends (PARE) or degradome sequencing (Addo-Quaye *et al.*, 2008; German *et al.*, 2008; Ding *et al.*, 2012). The latter is a direct approach for combining miRNA target identification and confirmation, and is only useful for detecting mRNAs regulated by cleavage. Bioinformatic tools such as CleaveLand (Addo-Quaye *et al.*, 2009), SeqTar (Zheng *et al.*, 2012), and PAREsnip (Folkes *et al.*, 2012) are convenient pipelines for the analysis of sRNA-mediated cleavage products obtained from degradome data.

Computational methods, based on miRNA-target interaction, for miRNA target prediction are either statistical prediction or machine-learning approaches (Meng *et al.*, 2014). Statistical approaches are widely used for plant miRNA research. Their algorithms are based on different miRNA-mRNA duplex characteristics: (1) perfect or near-perfect complementarity and base-pairing pattern; (2) thermodynamic stability; (3) binding site evolutionary conservation; and (4) target site accessibility (Salim and Chandra, 2014). Representative tools that incorporate some or all of these properties include miRU (Zhang, 2005), psRNATarget (Dai and Zhao, 2011), UEA sRNA toolkit (Moxon *et al.*, 2008), TargetFinder (Fahlgren and Carrington, 2010), and TAPIR (Bonnet *et al.*, 2010). P-TAREF is a tool that implements a machine-learning algorithm *viz.* support vector regression (SVR), for the identification of plant miRNA targets (Jha and Shankar, 2011).

2.9.4 Expression analysis of sRNA-seq data

Quantification of miRNA expression levels are required to functionally characterise miRNAs that are differentially expressed under certain conditions, for example, during pathogen attack or environmental change(s) (Figure 2.11). When comparing two sample groups (e.g. diseased versus healthy control), having different miRNA expression profiles, the ‘fold-change’ of certain miRNAs can be determined. The fold-change can also be described as the ratio of the group averages. Quantification of sRNA-seq data is based on read counts that are absolutely assigned to transcripts, and represents a measure of relative abundance. Following alignment, miRNA read counts need to be normalised in order for variation in data to be removed. Such variation, which stems from experimental procedure, can affect measured abundance levels, even for replicate experiments. A

good normalisation method should ultimately minimise technical and experimental bias without introducing noise (Tam *et al.*, 2015).

Two frequently used Bioconductor packages (Gentlemen *et al.*, 2004) that allow empirical analysis of gene expression count data in R, are edgeR and DESeq (Anders and Huber, 2010; Robinson *et al.*, 2010). Both methods utilise negative binomial distribution to model discrete count data. In edgeR the read counts are normalised for both compositional bias in sequenced libraries, as well as differences between libraries in sequencing depth. The data is first scaled to library size, followed by normalisation with weighted trimmed mean of the log expression ratios, a method also known as trimmed mean of M values (TMM) (Robinson and Oshlack, 2010). DESeq follows as similar approach, but it extends the model to allow a better fit for the data. A pre-processing step incorporates a scaling factor to adjust the data to a common scale in order to normalise data according to library size. According to Tam *et al.* (2015), this size factor is defined as the median of the ratios of observed counts to the geometric mean of each corresponding target over all samples. Finally, differential miRNA expression and estimate significance (*p*)-values are calculated. *P*-values are usually adjusted for multiple testing using the Benjamini and Hochberg (1995) approach for false discovery rate (FDR).

2.9.5 Experimental detection and quantification of miRNAs

Beside NGS-based approaches, other detection methods that rely on prior sequence information have been developed for miRNA detection and/or quantification. These methods have both their advantages and limitations, depending on the application they are used for and require careful adjustments to deliver sound and reliable results. Like with sRNA-seq, special care must be paid to optimally preserve total RNA when applying any of these methods. When attempting to use good quality RNA, an RNA integrity number (RIN) of 7 or more would be ideal (Schroeder *et al.*, 2006; Ibberson *et al.*, 2009). Northern blot analysis is still a commonly-used robust technique for targeted miRNA research, and can provide information on the size and expression of predicted pre-miRNAs and mature miRNAs (Lee *et al.*, 1993; Pasquinelli *et al.*, 2000; Zhang *et al.*, 2011). The use of "locked nucleic acids" (LNA)-modified oligonucleotide probes were successfully used for the 10-fold increase in sensitivity in miRNA northern blot analysis (Válóczi *et al.*, 2004; Várallyay *et al.*, 2008). Although this technique is time-consuming, it is beneficial when sample quantities are limited, when miRNA expression is low, or when subtle discrimination between related miRNAs is necessary.

Another probe-based technique, known as microarray hybridisation analysis, is an effective approach to profile a large number of known miRNAs. The method depends on specific binding of fluorophore-labelled miRNAs to their corresponding complementary probes. Consequently, relative miRNA quantification can be performed by analysing fluorescence signal data. Due to the short nature of miRNAs and the need to design probe sets with homogeneous melting temperatures (T_m), probe lengths were adjusted to detect aberrant expression of miRNAs after different treatments (Li and Ruan, 2009). Despite many improvements to the method, microarrays still have certain limitations such as narrow sensitivity range together with limited discrimination between miRNAs with similar sequences.

Currently, RT-qPCR is the method of choice for accurate and sensitive detection of plant miRNAs and other sRNAs (Figure 2.11). It is also becoming the gold standard for validating miRNA expression profiles obtained from microarray and sRNA-seq studies (Benes and Castoldi, 2010). A popular reverse transcription step using a stem-loop miRNA-specific primer has been extensively used to produce cDNA enriched for a specific miRNA (Chen *et al.*, 2005). This step can then be followed with a SYBR Green assay with a universal reverse primer and miRNA-specific forward primer to detect the miRNA of interest. An even more sensitive version of the stem-loop approach is the use of a TaqMan (Applied Biosystems) or Universal ProbeLibrary (UPL) (Roche) probe-based assay which can be used to distinguish a single nucleotide change in a miRNA sequence (Varkonyi-Gasic *et al.*, 2007). Another sensitive, yet very expensive, miRNA RT-qPCR method involves the use of LNA-modified primers (Kauppinen *et al.*, 2006). An advantage of this method is that LNA-modified primers can be designed to acquire a desired T_m-value and can increase duplex stability (Raymond *et al.*, 2005; Kauppinen *et al.*, 2006).

2.10 Integrating mRNA-seq transcriptome analysis

As with sRNA-seq, 'typical' Illumina mRNA-seq is performed using a similar approach (See section 2.9.1), and can be used solely for the purpose of transcriptome profiling. RNA-seq can also be coupled with different biochemical assays to investigate other aspects of molecular biology, such as RNA-structure, RNA-protein interaction, and RNA-RNA (e.g. miRNA-mRNA) interaction. Multiple applications for Illumina RNA sequencing methods exist and examples have been captured in a recent overview of RNA-seq publications (RNA sequencing method collection: www.illumina.com). The majority of these applications, however, are beyond the scope of this study as we focussed more on gene expression profiling and its use for investigating host-pathogen interactions.

RNA-seq on an Illumina platform is currently the most popular approach to study such interactions in humans, plants and animals and has been reviewed (Greenwood *et al.*, 2016). Several studies have integrated both mRNA-seq and sRNA-seq to 1) elucidate molecular mechanisms underlying pathogenicity in plants, for example, during virus-infection (Yang *et al.*, 2016; Guo *et al.*, 2017 Li *et al*., 2017) and fungus-infection (Burkhardt and Day, 2016; Lin *et al.*, 2016; Sarkar *et al.*, 2017); 2) in order to understand plant developmental mechanisms (He *et al.*, 2013; An *et al.*, 2015); and 3) for detection and validation of the expression level of miRNA targets (Baksa *et al.*, 2015). Here, we aim to provide a brief summary of experimental design and data analysis when using standard Illumina mRNA-seq to study host-pathogen interactions.

2.10.1 Experimental design

In order to achieve a good experimental design, prerequisites must be identified to answer the biological questions of interest. In an "infection experiment" these questions should generally allow the determination of the expression levels of genes or splice variants in a plant sample upon infection. The discovery of novel mRNA transcripts and alternative splice variants, however, was not of interest for this study. In formulating an experimental design, one must choose the type of library, desired depth of sequencing, include biological replicates for each condition, and adequately plan and execute the sequencing experiment (Conesa *et al.*, 2016).

The RNA isolation method is an important aspect and relies on the removal of highly abundant rRNA, which constitute more that 90% of cellular total RNA, while one to two percent constitute the mRNA of interest. Therefore, for eukaryotic transcriptomes, RNA-seq procedures generally starts with enrichment of polyadenylated mRNAs using extraction with oligo-dT beads, or alternatively rRNAs can be selectively depleted by exonucleases, while mRNAs are protected by their 5' cap structures. The remaining poly(A) fraction is required to contain a high proportion of mRNA with little or no degradation which is measured by an RNA integrity number (RIN). RNA is then fragmented using either sonication or enzymatic digestion into shorter RNA molecules (usually less than 500 bp) to allow good read coverage over the length of transcripts. To produce a cDNA template library these RNA fragments are then reverse transcribed to double-stranded cDNA using random primers and RNase H, followed by end-repair, 3'A-tailing, 5' -and 3' dsDNA adapter-ligation, PCR amplification, and cDNA library purification (Van Dijk *et al.*, 2014) (Figure 2.10).

RNA-seq can involve single-end (SE) or paired-end (PE) reads, with the latter commonly used for *de novo* transcript discovery or expression analysis of isoforms (Katz *et al.*, 2010; Garber *et al.*, 2011). Single-end reads are usually shorter and sufficient for determining gene expression levels in

well-annotated organisms, while longer paired-end reads are used to characterize poorly annotated transcriptomes (Garber *et al.*, 2011; Łabaj *et al.*, 2011. Sequencing depth or library size is another important factor that is defined by the number of available sequence reads for a given sample. The higher the depth of sequencing the more transcripts will be detected and accurately quantified (Mortazavi *et al.*, 2008). Some researchers will argue that a sequencing run generating up to a 100 million reads will allow accurate quantification of eukaryotic transcripts with low levels of expression (Sims *et al.*, 2014).

Finally, the number of biological replicates used for a certain condition depends on the degree of technical variability introduced during experimental procedures as well as the amount of biological variability in a system. In addition, more biological replicates for each condition will also increase the extent of statistical meaningful differential expression analysis (Auer and Doerge, 2010).

2.10.2 mRNA-seq data analysis

During RNA-seq of a transcriptome library millions of raw sequence reads are generated. Analysis of read data consists of several steps, including quality control of raw reads, read alignment and quantification (Figure 2.11). Quality control analysis of raw read data involves checking sequence quality, the presence of adaptor sequences, overrepresented *k*-mers, GC content and duplicated sequences in order to detect possible sequencing errors. FastQC (www.bioinformatics.babraham.ac.uk/projects/fastqc/) can be used as a tool to visualise these quality control measurements of Illumina reads. The decrease of read quality towards the 3' end of reads occurs in general. The lower the quality, the lower the confidence level of base-calling becomes and therefore removal of 3' bases are required to improve read mappability. FASTX-Toolkit can be used for adaptor-trimming, removal of low quality reads and trimming of low quality bases (hannonlab.cshl.edu/fastx_toolkit/).

2.10.2.1 Read mapping

In order to analyse transcriptomic differences between experimental conditions, reads must first be mapped to a reference genome sequence or an annotated transcriptome sequence. Raw reads can be directly mapped to an organism's genome if the genome sequence is of high quality. Mapping software can allow the identification of novel genes or alternative splice variants with the use of a gapped or spliced mapper (also known as splice-aware aligners) in cases where reads may span splice junctions. Examples of such programmes include TopHat2, MapSplice, GSNAP, and STAR

(Wang *et al.*, 2010; Wu and Nacu, 2010; Dobin *et al.*, 2013; Kim *et al.*, 2013). Mapping parameters that should be considered are number of allowed mismatches, strandedness of the library and the length and type (SE or PE) of reads. Furthermore, leverage in the form of an available annotation file, also known as a gene transfer format (GTF) file, can allow accurate mapping within exon coordinates and assist in identifying splice variants (http://genome.ucsc.edu/FAQ/FAQformat.html#format4). In cases where a genome has an existing annotation of good quality, or the researcher is only interested in known gene or transcript sequences, expression levels can simply be determined for each annotated gene/transcript following mapping (Conesa *et al.*, 2016). However, where an annotation is unavailable or incomplete, transcriptome assembly can occur prior to quantification. This provides researchers the opportunity to identify novel features and allow more accurate expression analysis (Trapnell *et al.*, 2012).

2.10.2.2 Differential expression analysis

The use of read count data to perform differential expression analysis is the most common application of RNA-seq, which aims to determine the significant expression of genes that will differ across two or more conditions. To obtain confident results raw read counts are required to be normalised before these expression values can be compared among samples. Normalisation are used to account for inter-sample differences in sequencing depth, pertaining to number of reads, transcript length, and sequencing biases (Conesa *et al.*, 2016). Frequently-used normalisation methods include those that measure RPKM (reads per kilobase of exon per million mapped reads) (Mortazavi *et al.*, 2008), and its derivatives: FPKM (fragments per kilobase of exon per million mapped reads), and TPM (transcripts per million) (Conesa *et al.*, 2016). These methods rely on total or effective read counts and tend to provide inaccurate results when analysing samples with high transcript heterogeneity (Bullard *et al.*, 2010). Different normalisation strategies have been evaluated (Dillies *et al.*, 2013), and take this into consideration, including upper quartile (Bolstad *et al.*, 2003), trimmed mean of M-values (TMM) (Robinson and Oshlack, 2010), PoissonSeq (Li *et al.*, 2012b), and DESeq (Anders and Huber, 2010).

Another normalisation strategy involves the use of Cufflinks (Trapnell *et al.*, 2010), which was designed to account for problems surrounding transcript length. Transcript length may interfere with accurate ranking of transcript expression values due to the fact that transcript isoforms tend to share majority of reads and that reads tend to map to longer transcripts. Following reads mapping to a genome using TopHat, for example, Cufflinks make use of an expectation-maximum approach which takes into account positional biases associated with non-uniform distribution of mapped

reads along the length of a transcript. It utilises GTF information and PE reads to identify expressed transcripts, or determines transcripts *de novo* from mapping information.

Following normalisation of mapped data, differentially expressed genes (DEGs) can be determined (Figure 2.11). This area of bioinformatics is still developing and comparisons between the different software and pipelines have been reviewed (Soneson and Delorenzi, 2013; Costa-Silva *et al.*, 2017; Spies *et al.*, 2017). Tools such as edgeR, DESeq and baySeq (Anders and Huber, 2010; Hardcastle and Kelly, 2010; Robinson *et al.*, 2010), make use of the negative binomial model for analysis, while NOIseq and SAMseq adopt non-parametric methods (Li and Tibshirani, 2013; Tarazona *et al.*, 2015). Some differential expression methods, such as Cuffdiff2 (Trapnell *et al.*, 2013), and EBSeq (Leng *et al.*, 2013), were designed for unknown transcript and isoform detection and can also be used for the identification of DEGs.

2.10.2.3 Functional analysis of DEGs

This is usually the last step in a standard transcriptome analysis study and involves the functional characterisation of DEGs in terms of molecular function and pathways in which the DEGs are involved (Figure 2.11). Two informative approaches used for functional characterisation, initially designed to interrogate microarray data, are: gene set enrichment analysis (GSEA); and determining overrepresented functions by comparing DEGs against the rest of a genome.

Functional analysis relies on the availability of annotation data. Popular resources which provide annotation data for model species include Bioconductor (Huber *et al.*, 2015), Gene Ontology (Ashburner *et al.*, 2000), DAVID (Huang *et al.*, 2009) and Babelomics (Medina *et al.*, 2010). The use of an orthology-based search can allow the identification of similar protein-coding transcripts using databases such as SwissProt (Bairoch *et al.*, 2004), while databases such as pFam (Finn *et al.*, 2014), and InterPro (Finn *et al.*, 2017) can be used to identify protein-coding transcripts based on conserved domains. Blast2GO is a widely used GUI and allows functional annotation of massive transcriptome datasets, as well as pathway enrichment analysis (Conesa and Götz, 2008).

C. Conclusion

The AY phytoplasma group are responsible for a detrimental disease known as GY which cause dramatic yield losses in vineyards across the world. This poses a serious threat to the sustainability of the table grape and wine industry. Current methods to control the disease rely on preventative measures relating to eradication of infected plant material, as well as insecticide treatments.

Extensive research has been conducted on GY symptomology, epidemiology, and the detection and classification of phytoplasmas, with limited studies on the mechanism of phytoplasma pathogenicity in highly susceptible cultivars.

Plants have developed complex regulatory processes to enable their normal development and to cope with changes in their environment. These processes include the production of sRNAs known to affect different levels of gene expression. The miRNAs are a well characterised class of sRNAs that play major roles in development and responses to biotic and abiotic stress. Integration of NGS approaches such as mRNA-seq and sRNA-seq have facilitated comprehensive discovery and expression profiling of miRNAs and mRNAs. Such data can be used to unravel the basis of host susceptibility and may assist in the development of disease control strategies.

The aim of the study was to use NGS approaches to identify differentially expressed miRNAs, as well as differentially expressed genes in order to explore pathogen response pathways in *V. vinifera* 'Chardonnay' upon AY phytoplasma-infection.

D. References

Addo-Quaye C., Eshoo T.W., Bartel D.P. and Axtell M.J. 2008. Endogenous siRNA and miRNA targets identified by sequencing of the *Arabidopsis* degradome. *Current biology : CB* 18(10): 758-762.

Addo-Quaye C., Miller W. and Axtell M.J. 2009. CleaveLand: a pipeline for using degradome data to find cleaved small RNA targets. *Bioinformatics (Oxford, England)* 25(1): 130-131.

Albertazzi G., Milc J., Caffagni A., Francia E., Roncaglia E., *et al.* 2009. Gene expression in grapevine cultivars in response to Bois Noir phytoplasma infection. *Plant Science* 176(6): 792-804.

Alma A., Davis R.E., Vibio M., Danielli A., Bosco D., *et al.* 1996. Mixed infection of grapevines in northern Italy by phytoplasmas including 16S rRNA RFLP subgroup 16SrI-B strains previously unreported in this host. *Plant Disease* 80(4): 418-421.

Ambros V. and Lee R.C. 2004. Identification of microRNAs and other tiny noncoding RNAs by cDNA cloning. *In: RNA Interference, Editing, and Modification: Methods and Protocols.* Gott J.M. (ed.). Totowa, NJ: Humana Press. pp.131-158. https://doi.org/10.1385/1-59259-775-0:131

An W., Gong W., He S., Pan Z., Sun J., *et al.* 2015. MicroRNA and mRNA expression profiling analysis revealed the regulation of plant height in *Gossypium hirsutum*. *BMC Genomics* 16: 886-886.

Anders S. and Huber W. 2010. Differential expression analysis for sequence count data. *Genome Biology* 11(10): R106.

Andersen M.T., Liefting L.W., Havukkala I. and Beever R.E. 2013. Comparison of the complete genome sequence of two closely related isolates of '*Candidatus* Phytoplasma australiense' reveals genome plasticity. *BMC Genomics* 14(1): 529.

Angelini E., Luca Bianchi G., Filippin L., Morassutti C. and Borgo M. 2007. A new TaqMan method for the identification of phytoplasmas associated with grapevine yellows by real-time PCR assay. *Journal of Microbiological Methods* 68(3): 613-622.

Ashburner M., Ball C.A., Blake J.A., Botstein D., Butler H., *et al.* 2000. Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. *Nature Genetics* 25(1): 25-29.

Auer P.L. and Doerge R.W. 2010. Statistical design and analysis of RNA sequencing data. *Genetics* 185(2): 405-416.

Axtell M.J. 2013. ShortStack: Comprehensive annotation and quantification of small RNA genes. *RNA* 19(6): 740-751.

Bai X., Correa V.R., Toruño T.Y., Ammar E.-D., Kamoun S., *et al.* 2008. AY-WB phytoplasma secretes a protein that targets plant cell nuclei. *Molecular Plant-Microbe Interactions* 22(1): 18-30.

Bai X., Zhang J., Ewing A., Miller S.A., Jancso Radek A., *et al.* 2006. Living with genome instability: the adaptation of phytoplasmas to diverse environments of their insect and plant hosts. *Journal of Bacteriology* 188: 3682-3696.

Bairoch A., Boeckmann B., Ferro S. and Gasteiger E. 2004. Swiss-Prot: juggling between evolution and stability. *Briefings in Bioinformatics* 5(1): 39-55.

Baker C.C., Sieber P., Wellmer F. and Meyerowitz E.M. 2005. The early extra *petals1* mutant uncovers a role for microRNA miR164c in regulating petal number in *Arabidopsis*. *Current Biology* 15(4): 303-315.

Baksa I., Nagy T., Barta E., Havelda Z., Várallyay É., *et al.* 2015. Identification of *Nicotiana benthamiana* microRNAs and their targets using high throughput sequencing and degradome analysis. *BMC Genomics* 16: 1025-1025.

Baksa I., Nagy T., Barta E., Havelda Z., Várallyay É., *et al.* 2015. Identification of Nicotiana benthamiana microRNAs and their targets using high throughput sequencing and degradome analysis. *BMC Genomics* 16: 1025-1025.

Bartel D.P. 2004. MicroRNAs. *Cell* 116(2): 281-297.

Barturen G., Rueda A., Hamberg M., Alganza A., Lebron R., *et al.* 2014. sRNAbench: profiling of small RNAs and its sequence variants in single or multi-species high-throughput experiments. *Methods in Next Generation Sequencing* 1(1): 21-31.

Belli G., Bianco P.A. and Conti M. 2010. Grapevine yellows in Italy: past, present and future. *Journal of Plant Pathology* 92(2): 303-326.

Benes V. and Castoldi M. 2010. Expression profiling of microRNA using real-time quantitative PCR, how to use it and what is available. *Methods* 50(4): 244-249.

Benjamini Y.H., Y 1995. Controlling the false discovery rate: A practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society* 57(1): 289-300.

Berezikov E., Cuppen E. and Plasterk R.H.A. 2006. Approaches to microRNA discovery. *Nature Genetics* 38: S2.

Berger J., Dalla Via J. and Baric S. 2009. Development of a TaqMan allelic discrimination assay for the distinction of two major subtypes of the grapevine yellows phytoplasma Bois noir. *European Journal of Plant Pathology* 124(3): 521.

Berges R., Rott M. and Seemüller E. 2000. Range of phytoplasma concentrations in various plant hosts as determined by competitive polymerase chain reaction. *Phytopathology* 90: 1145-1152.

Bertaccini A. 2007. Phytoplasmas: diversity, taxonomy, and epidemiology. *Frontiers in Bioscience* 12: 673-689.

Bertaccini A. and Duduk B. 2010. Phytoplasma and phytoplasma diseases: a review of recent research. *Phytopathologia Mediterranea* 48(3): 355-378.

Bertin S., Palermo S., Marzachì C. and Bosco D. 2004. Note: A comparison of molecular diagnostic procedures for the detection of aster yellows phytoplasmas (16Sr-I) in leafhopper vectors. *Phytoparasitica* 32(2): 141-145.

Bolstad B.M., Irizarry R.A., Astrand M. and Speed T.P. 2003. A comparison of normalization methods for high density oligonucleotide array data based on variance and bias. *Bioinformatics* 19(2): 185-193.

Bonnet E., He Y., Billiau K. and Van De Peer Y. 2010. TAPIR, a web server for the prediction of plant microRNA targets, including target mimics. *Bioinformatics* 26(12): 1566-1568.

Botti S. and Bertaccini A. 2006. First report of phytoplasmas in grapevine in South Africa. *Plant Disease* 90(10): 1360.

Boudon-Padieu E. 2003. The situation of Grapevine yellows and current research directions: distribution, diversity, vectors, diffusion and control. 14th meeting of the international council for the study of virus and virus-like diseases of the grapevine (ICVG). Locorotondo (BARI), Italy.

Brodersen P. and Voinnet O. 2006. The diversity of RNA silencing pathways in plants. *Trends in Genetics* 22(5): 268-280.

Brodersen P., Sakvarelidze-Achard L., Bruun-Rasmussen M., Dunoyer P., Yamamoto Y.Y., *et al.* 2008. Widespread translational inhibition by plant miRNAs and siRNAs. *Science* 320(5880): 1185.

Budak H. and Akpinar B.A. 2015. Plant miRNAs: biogenesis, organization and origins. *Functional & Integrative Genomics* 15(5): 523-531.

Buhtz A., Pieritz J., Springer F. and Kehr J. 2010. Phloem small RNAs, nutrient stress responses, and systemic mobility. *BMC Plant Biology* 10: 64-64.

Bullard J.H., Purdom E., Hansen K.D. and Dudoit S. 2010. Evaluation of statistical methods for normalization and differential expression in mRNA-Seq experiments. *BMC Bioinformatics* 11: 94-94.

Burger J.T. Impact and extent of phytoplasma disease still a grey area. Winetech Technical. July 2008. https://www.wineland.co.za/mpact-and-extent-of-phytoplasma-disease-still-a-grey-area/

Burkhardt A. and Day B. 2016. Transcriptome and small RNAome dynamics during a resistant and susceptible interaction between cucumber and downy mildew. *The Plant Genome* 9(1).

Carnavale Bottino M., Rosario S., Grativol C., Thiebaut F., Rojas C.A., *et al.* 2013. High-throughput sequencing of small RNA transcriptome reveals salt stress regulated microRNAs in sugarcane. *PLoS ONE* 8(3): e59423-e59423.

Carstens R. 2014. *The incidence and distribution of grapevine yellows disease in South African vineyards.* Master's degree thesis, Department of Genetics, Stellenbosch University. http://scholar.sun.ac.za/handle/10019.1/86683

Carstens R., Petersen Y., Stephan D. and Burger J.T. 2011. Current status of Aster yellows disease in infected vineyards in the Vredendal grape producing area of South Africa. *Phytopathogenic Mollicutes* 1(2): 83-85.

Caudwell A. 1961. A study of black wood disease of vines: its relationship to Flavescence dorée. *Annales des Epiphyties* 12: 241-262.

Chaloner T., Van Kan J.a.L. and Grant-Downton R.T. 2016. RNA 'Information Warfare' in pathogenic and mutualistic interactions. *Trends in Plant Science* 21(9): 738-748.

Chang S.-H., Cho S.-T., Chen C.-L., Yang J.-Y. and Kuo C.-H. 2015. Draft genome sequence of a 16SrII-A subgroup phytoplasma associated with purple coneflower (*Echinacea purpurea*) witches' broom disease in Taiwan. *Genome Announcements* 3(6): e01398-15.

Chen K.H., Guo J.R., Wu X.Y., Loi N., Carraro L., *et al.* 1993. Comparison of monoclonal antibodies, DNA probes, and PCR for detection of the grapevine yellows disease agent. *Phytopathology* 83: 915-922.

Chen W., Li Y., Wang Q., Wang N. and Wu Y. 2014. Comparative genome analysis of wheat blue dwarf phytoplasma, an obligate pathogen that causes wheat blue dwarf disease in China. *PLoS ONE* 9(5): e96436.

Chen X. 2009. Small RNAs and their roles in plant development. *Annual Review of Cell and Developmental Biology* 25: 21-44.

Chen X. 2010. Small RNAs - secrets and surprises of the genome. *The Plant journal* 61(6): 941-958.

Chisholm S.T., Coaker G., Day B. and Staskawicz B.J. 2006. Host-microbe interactions: shaping the evolution of the plant immune response. *Cell* 124(4): 803-814.

Chitarra W., Pagliarani C., Abbà S., Boccacci P., Birello G., *et al.* 2018. miRVIT: a novel miRNA database and its application to uncover *Vitis* responses to Flavescence dorée infection. *Frontiers in plant science* 9: 1034-1034.

Chung W.-C., Chen L.-L., Lo W.-S., Lin C.-P. and Kuo C.-H. 2013. Comparative analysis of the peanut witches'-broom phytoplasma genome reveals horizontal transfer of potential mobile units and effectors. *PLoS ONE* 8(4): e62770.

Conesa A. and Götz S. 2008. Blast2GO: A comprehensive suite for functional analysis in plant genomics. *International Journal of Plant Genomics* 2008: 619832.

Conesa A., Madrigal P., Tarazona S., Gomez-Cabrero D., Cervera A., *et al.* 2016. A survey of best practices for RNA-seq data analysis. *Genome biology* 17: 13-13.

Constable F. 2010. Phytoplasma epidemiology: grapevines as a model. *In: PHYTOPLASMAS: Genomes, Plant Hosts and Vectors.* Weintraub P.G. (ed.). pp.188-212.

Constable F. and Rodoni B. Australian grapevine yellows. *Wine Australia for Australian Wine - Factsheet*. July 2011. pp.1-4.

Contaldo N., Bertaccini A., Paltrinieri S., Windsor H.M. and Windsor G.D. 2012. Axenic culture of plant pathogenic phytoplasmas. *Phytopathologia Mediterranea* 51(3): 607-617.

Costa-Silva J., Domingues D. and Lopes F.M. 2017. RNA-Seq differential expression analysis: An extended review and a software tool. *PLoS ONE* 12(12): e0190152-e0190152.

Couzigou J.-M. and Combier J.-P. 2016. Plant microRNAs: key regulators of root architecture and biotic interactions. *New Phytologist* 212(1): 22-35.

Dai X. and Zhao P.X. 2011. psRNATarget: a plant small RNA target analysis server. *Nucleic Acids Research* 39: W155-W159.

Davis R.E., Lee I.M., Dally E.L., Dewitt N. and Douglas S.M. 1988. Cloned nucleic acid hybridization probes in detection and classification of mycoplasmalike organisms (MLOs). *Acta Horticulturae* 234: 115-122.

de Klerk A. and Carstens R. Aster yellows and leafhoppers. Winetech Technical. June 2016. https://www.wineland.co.za/aster-yellows-and-leafhoppers/

Dillies M.-A., On Behalf of the French Statomique C., Rau A., Aubert J., Hennequet-Antier C., *et al.* 2013. A comprehensive evaluation of normalization methods for Illumina high-throughput RNA sequencing data analysis. *Briefings in Bioinformatics* 14(6): 671-683.

Ding J., Zhou S. and Guan J. 2012. Finding microRNA targets in plants: current status and perspectives. *Genomics, proteomics & bioinformatics* 10(5): 264-275.

Dobin A., Davis C.A., Schlesinger F., Drenkow J., Zaleski C., *et al.* 2013. STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* 29(1): 15-21.

Doi, M., Tetranaka, M., Yora, K. and Asuyama, H. 1967. Mycoplasma or PLT-grouplike organism found in the phloem elements of plants infected with mulberry dwarf, potato witches' broom, aster yellow, or paolownia witches' broom. *Annals of the Phytopathological Society of Japan* 33: 259-266.

Economou A. 1999. Following the leader, bacterial protein export through the Sec pathway. *Trends in Microbiology* 7: 315-320.

Engelbrecht M., Joubert J. and Burger J.T. 2010. First report of aster yellows phytoplasma in grapevines in South Africa. *Plant Disease* 94(3): 373-373.

Fahlgren N. and Carrington J.C. 2010. miRNA target prediction in plants. *In: Plant MicroRNAs: Methods and Protocols.* Meyers B.C. and Green P.J. (eds.). Totowa, NJ: Humana Press. pp.51-57. https://doi.org/10.1007/978-1-60327-005-2_4

Fahlgren N., Howell M.D., Kasschau K.D., Chapman E.J., Sullivan C.M., *et al.* 2007. High-throughput sequencing of *Arabidopsis* microRNAs: evidence for frequent birth and death of MIRNA genes. *PLoS ONE* 2(2): e219-e219.

Fei Q., Xia R. and Meyers B.C. 2013. Phased, secondary, small interfering RNAs in posttranscriptional regulatory networks. *The Plant Cell* 25(7): 2400-2415.

Filippin L., Jović J., Cvrković T., Forte V., Clair D., *et al.* 2009. Molecular characteristics of phytoplasmas associated with Flavescence dorée in clematis and grapevine and preliminary results on the role of *Dictyophara europaea* as a vector. *Plant Pathology* 58(5): 826-837.

Finn R.D., Attwood T.K., Babbitt P.C., Bateman A., Bork P., *et al.* 2017. InterPro in 2017 - beyond protein family and domain annotations. *Nucleic Acids Research* 45: D190-D199.

Firrao G., Garcia-Chapa M. and Marzachì C. 2007. Phytoplasmas: genetics, diagnosis and relationships with the plant and insect host. *Frontiers in Bioscience* 12: 1353-1375.

Firrao G., Smart C.D. and Kirkpatrick B.C. 1996. Physical map of the Western X disease phytoplasma chromosome. *Journal of Bacteriology* 178(13): 3985-3988.

Folkes L., Moxon S., Woolfenden H.C., Stocks M.B., Szittya G., *et al.* 2012. PAREsnip: a tool for rapid genome-wide discovery of small RNA/target interactions evidenced through degradome sequencing. *Nucleic Acids Research* 40(13): e103-e103.

Friedländer M.R., Chen W., Adamidi C., Maaskola J., Einspanier R., *et al.* 2008. Discovering microRNAs from deep sequencing data using miRDeep. *Nature Biotechnology* 26: 407.

Galetto L. and Marzachi C. 2010. Real-time PCR diagnosis and quantification of phytoplasmas. *In: PHYTOPLASMAS: Genomes, Plant Hosts and Vectors*. Weintraub P.G. (ed.). pp.1-18.

Garber M., Grabherr M.G., Guttman M. and Trapnell C. 2011. Computational methods for transcriptome annotation and quantification using RNA-seq. *Nature Methods* 8: 469.

Gartel W. 1965. Untersuchungen nber das Auftreten und das Verhalten des flavescence dorTe in den Weinbaugebieten am Mosel und Rhein. Weinbeg und Keller 12: 347-376.

Gentleman R.C., Carey V.J., Bates D.M., Bolstad B., Dettling M., *et al.* 2004. Bioconductor: open software development for computational biology and bioinformatics. *Genome Biology* 5(10): R80.

German M.A., 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. *Nature Biotechnology* 26: 941.

Glass J.I., Lefkowitz E.J., Glass J.S., Heiner C.R., Chen E.Y., *et al.* 2000. The complete sequence of the mucosal pathogen *Ureaplasma urealyticum*. *Nature* 407: 757.

Gomes C., Cho J.-H., Hood L., Franco O., Pereira R., *et al.* 2013. A review of computational tools in microRNA discovery. *Frontiers in Genetics* 4: 81.

Greenwood J.M., Ezquerra A.L., Behrens S., Branca A. and Mallet L. 2016. Current analysis of host-parasite interactions with a focus on next generation sequencing data. *Zoology* 119(4): 298-306.

Guo Y., Jia M.-A., Yang Y., Zhan L., Cheng X., *et al.* 2017. Integrated analysis of tobacco miRNA and mRNA expression profiles under PVY infection provids insight into tobacco-PVY interactions. *Scientific reports* 7(1): 4895-4895.

Gupta O.P., Meena N.L., Sharma I. and Sharma P. 2014. Differential regulation of microRNAs in response to osmotic, salt and cold stresses in wheat. *Molecular Biology Reports* 41(7): 4623-4629.

Hackenberg M., Sturm M., Langenberger D., Falcón-Pérez J.M. and Aransay A.M. 2009. miRanalyzer: a microRNA detection and analysis tool for next-generation sequencing experiments. *Nucleic Acids Research* 37: W68-W76.

Hamilton A.J. and Baulcombe D.C. 1999. A species of small antisense RNA in posttranscriptional gene silencing in plants. *Science* 286(5441): 950.

Hardcastle T.J. and Kelly K.A. 2010. baySeq: empirical Bayesian methods for identifying differential expression in sequence count data. *BMC Bioinformatics* 11: 422-422.

He C.-Y., Cui K., Zhang J.-G., Duan A.-G. and Zeng Y.-F. 2013. Next-generation sequencing-based mRNA and microRNA expression profiling analysis revealed pathways involved in the rapid growth of developing culms in Moso bamboo. *BMC Plant Biology* 13: 119-119.

Hiruki C. and Deng S. 1992. Distribution of plant pathogenic Mollicutes in Catharanthus roseus determined by in situ molecular hybridization and DNA staining with DAPI. *Proceedings of the Japan Academy* 68(10): 187-190.

Hofacker I.L. 2003. Vienna RNA secondary structure server. *Nucleic Acids Research* 31(13): 3429-3431.

Hogenhout S.A., Oshima K., Ammar E.-D., Kakizawa S., Kingdom H.N., *et al.* 2008. Phytoplasmas: bacteria that manipulate plants and insects. *Molecular Plant Pathology* 9(4): 403-423.

Hollingsworth C.R., Atkinson L.M., Samac D.A., Larsen J.E., Motteberg C.D., *et al.* 2008. Region and field level distributions of aster yellows phytoplasma in small grain crops. *Plant Disease* 92: 623-630.

Hoshi A., Oshima K., Kakizawa S., Ishii Y., Ozeki J., *et al.* 2009. A unique virulence factor for proliferation and dwarfism in plants identified from a phytopathogenic bacterium. *Proceedings of the National Academy of Sciences of the United States of America* 106(15): 6416-6421.

Hren M., Boben J., Rotter A., Kralj P., Gruden K., *et al.* 2007. Real-time PCR detection systems for Flavescence dorée and Bois noir phytoplasmas in grapevine: comparison with conventional PCR detection and application in diagnostics. *Plant Pathology* 56: 785-796.

Hren M., Nikolić P., Rotter A., Blejec A., Terrier N., *et al.* 2009. 'Bois noir' phytoplasma induces significant reprogramming of the leaf transcriptome in the field grown grapevine. *BMC Genomics* 10(1): 460.

Huang D.W., Sherman B.T. and Lempicki R.A. 2008. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nature Protocols* 4: 44.

Huber W., Carey V.J., Gentleman R., Anders S., Carlson M., *et al.* 2015. Orchestrating high-throughput genomic analysis with Bioconductor. *Nature Methods* 12(2): 115-121.

Ibberson D., Benes V., Muckenthaler M.U. and Castoldi M. 2009. RNA degradation compromises the reliability of microRNA expression profiling. *BMC Biotechnology* 9: 102-102.

IRPCM Phytoplasma/Spiroplasma Working Team - Phytoplasma Taxonomy Group 2004. '*Candidatus* Phytoplasma', a taxon for the wall-less, non-helical prokaryotes that colonise plant phloem and insects. *International Journal of Systematic and Evolutionary Microbiology* 54: 1243-1255.

Jarausch W., Lansac M., Saillard C., Broquaire J.M. and Dosba F. 1998. PCR Assay for specific detection of European stone fruit yellows phytoplasmas and its use for epidemiological studies in France. *European Journal of Plant Pathology* 104(1): 17-27.

Jarausch W., Saillard C., Helliot B., Garnier M. and Dosba1 F. 2000. Genetic variability of apple proliferation phytoplasmas as determined by PCR-RFLP and sequencing of a non-ribosomal fragment. *Molecular and Cellular Probes* 14: 17-24.

Jha A. and Shankar R. 2011. Employing machine learning for reliable miRNA target identification in plants. *BMC Genomics* 12: 636-636.

Jia X., Ren L., Chen Q.-J., Li R. and Tang G. 2009. UV-B-responsive microRNAs in *Populus tremula*. *Journal of Plant Physiology* 166(18): 2046-2057.

Jin D., Wang Y., Zhao Y. and Chen M. 2013. MicroRNAs and their cross-talks in plant development. *Journal of Genetics and Genomics* 40(4): 161-170.

Jin H. 2008. Endogenous small RNAs and antibacterial immunity in plants. *FEBS Letters* 582(18): 2679-2684.

Jomantiene R., Zhao Y. and Davis R.E. 2007. Sequence-variable mosaics: composites of recurrent transposition characterizing the genomes of phylogenetically diverse phytoplasmas. *DNA and Cell Biology* 26(8): 557-564.

Jones J.D.G. and Dangl J.L. 2006. The plant immune system. *Nature* 444: 323.

Jones-Rhoades M.W. and Bartel D.P. 2004. Computational identification of plant microRNAs and their targets, including a stress-induced miRNA. *Molecular Cell* 14(6): 787-799.

Jones-Rhoades M.W., Bartel D.P. and Bartel B. 2006. MicroRNAS and their regulatory roles in plants. *Annual Review of Plant Biology* 57(1): 19-53.

Jung J.-H., Seo P.J. and Park C.-M. 2009. MicroRNA biogenesis and function in higher plants. *Plant Biotechnology Reports* 3(2): 111-126.

Kakizaw S., Oshima K. and Namba S. 2006. Diversity and functional importance of phytoplasma membrane proteins. *Trends in Microbiology* 14(6): 254-256.

Kakizawa S., Makino A., Ishii Y., Tamaki H. and Kamagata Y. 2014. Draft genome sequence of "*Candidatus* Phytoplasma asteris" strain OY-V, an unculturable plant-pathogenic bacterium. *Genome Announcements* 2(5).

Källman T., Chen J., Gyllenstrand N. and Lagercrantz U. 2013. A significant fraction of 21-nucleotide small RNA originates from phased degradation of resistance genes in several perennial species. *Plant Physiology* 162(2): 741-754.

Kang W. and Friedländer M.R. 2015. Computational prediction of miRNA genes from small RNA sequencing data. *Frontiers in Bioengineering and Biotechnology* 3: 7-7.

Katz Y., Wang E.T., Airoldi E.M. and Burge C.B. 2010. Analysis and design of RNA sequencing experiments for identifying isoform regulation. *Nature Methods* 7(12): 1009-1015.

Kauppinen S., Vester B. and Wengel J. 2006. Locked nucleic acid: high-affinity targeting of complementary RNA for RNomics. *Handbook of Experimental Pharmacology* 173: 405-22.

Khraiwesh B., Zhu J.-K. and Zhu J. 2012. Role of miRNAs and siRNAs in biotic and abiotic stress responses of plants. *Biochimica et Biophysica Acta (BBA) - Gene Regulatory Mechanisms* 1819(2): 137-148.

Kim D., Pertea G., Trapnell C., Pimentel H., Kelley R., *et al.* 2013. TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. *Genome biology* 14(4): R36-R36.

Kim V.N. 2005. MicroRNA biogenesis: coordinated cropping and dicing. *Nature Reviews Molecular Cell Biology* 6: 376.

Kirkpatrick B.C., Stenger D.C., Morris T.J. and Purcell A.H. 1987. Cloning and detection of DNA from a nonculturable plant pathogenic mycoplasma-like organism. *Science* 238(4824): 197.

Kozomara A. and Griffiths-Jones S. 2011. miRBase: integrating microRNA annotation and deep-sequencing data. *Nucleic Acids Research* 39: D152-D157.

Krüger K., De Klerk A., Douglas-Smit N., Joubert J., Pietersen G., *et al.* 2011. Aster yellows phytoplasma in grapevines: identification of vectors in South Africa. *Bulletin of Insectology* 64: S137-S1387.

Krüger, K. 2017 Transmission of Aster yellows to grapevine by insect vectors. Winetech Final Report for 2012. http://www.sawislibrary.co.za/dbtextimages/Kr%C3%BCgerK1.pdf

Kube M., Mitrovic J., Duduk B., Rabus R. and Seemüller E. 2012. Current view on phytoplasma genomes and encoded metabolism. *The Scientific World Journal* 2012: 185942.

Kube M., Schneider B., Kuhl H., Dandekar T., Heitmann K., *et al.* 2008. The linear chromosome of the plant-pathogenic mycoplasma '*Candidatus* Phytoplasma mali'. *BMC Genomics* 9(1): 306.

Kulski J.K. 2016. Next-generation sequencing - an overview of the history, tools, and "Omic" applications. *In: Next Generation Sequencing - Advances, Applications and Challenges.* Kulski J.K. (ed.). IntechOpen. pp.3-60. https://www.intechopen.com/books/next-generation-sequencing-advances-applications-and-challenges/next-generation-sequencing-an-overview-of-the-history-tools-and-omic-applications

Kumar R. 2014. ole of microRNAs in biotic and abiotic stress responses in crop plants. *Applied Biochemistry and Biotechnology* 174(1): 93-115.

Kurihara Y. and Watanabe Y. 2004. *Arabidopsis* micro-RNA biogenesis through Dicer-like 1 protein functions. *Proceedings of the National Academy of Sciences of the United States of America* 101(34): 12753-12758.

Łabaj P.P., Leparc G.G., Linggi B.E., Markillie L.M., Wiley H.S., *et al.* 2011. Characterization and improvement of RNA-Seq precision in quantitative transcript expression profiling. *Bioinformatics* 27(13): i383-i391.

Langmead B., Trapnell C., Pop M. and Salzberg S.L. 2009. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biology* 10(3): R25.

Lee I.M. 2004. '*Candidatus* Phytoplasma asteris', a novel phytoplasma taxon associated with aster yellows and related diseases. *International Journal of Systematic and Evolutionary Microbiology* 54(4): 1037-1048.

Lee I.-M. Gundersen-Rindal D.E., Davis R.E. and Bartoszyk, I.M. 1998. Revised classification scheme of phytoplasmas based on RFLP analyses of 16S rRNA and ribosomal protein gene sequences. *International Journal of Systematic Bacteriology* 48: 1153-1169.

Lee I.-M., Davis R.E. and Gundersen-Rindal D.E. 2000. Phytoplasma: Phytopathogenic Mollicutes 1. *Annual Review of Microbiology* 54(1): 221-255.

Lee I.M., Gundersen D.E., Davis R.E. and Chiykowski L.N. 1992. Identification and analysis of a genomic strain cluster of mycoplasmalike organisms associated with Canadian peach (eastern) X disease, western X disease, and clover yellow edge. *Journal of Bacteriology* 174(20): 6694-6698.

Lee I.-M., Gundersen D.E., Hammond R.W. and Davis R.E. 1994. Use of mycoplasmalike organisms (MLO) group-specific oligonucleotide primers for nested PCR assays to detect mixed-MLO infections in a single host plant. *Phytopathology* 84: 559-566.

Lee I.M., Hammond R.W., Davis R.E. and Gundersen D.E. 1993. Universal amplification and analysis of pathogen 16S rDNA for classification and identification of mycoplasmalike organisms. *Phytopathology* 83(8): 834-842.

Lee I.M., Shao J., Bottner-Parker K.D., Gundersen-Rindal D.E., Zhao Y., *et al.* 2015. Draft genome sequence of "*Candidatus* Phytoplasma pruni" strain CX, a plant-pathogenic bacterium. *Genome Announcements* 3(5): e01117-15.

Lee R.C., Feinbaum R.L. and Ambros V. 1993. The C. elegans heterochronic gene *lin-4* encodes small RNAs with antisense complementarity to *lin-14*. *Cell* 75(5): 843-854.

Lee Y., Kim M., Han J., Yeom K.-H., Lee S., *et al.* 2004. MicroRNA genes are transcribed by RNA polymerase II. *The EMBO journal* 23(20): 4051-4060.

Leng N., Dawson J.A., Thomson J.A., Ruotti V., Rissman A.I., *et al.* 2013. EBSeq: an empirical Bayes hierarchical model for inference in RNA-seq experiments. *Bioinformatics* 29(8): 1035-1043.

Li C. and Zhang B. 2015. MicroRNAs in Control of Plant Development. *Journal of Cellular Physiology* 231(2): 303-313.

Li F., Pignatta D., Bendix C., Brunkard J.O., Cohn M.M., *et al.* 2012a. MicroRNA regulation of plant innate immune receptors. *Proceedings of the National Academy of Sciences of the United States of America* 109(5): 1790-1795.

Li J. and Tibshirani R. 2013. Finding consistent patterns: a nonparametric approach for identifying differential expression in RNA-Seq data. *Statistical methods in medical research* 22(5): 519-536.

Li J., Witten D.M., Johnstone I.M. and Tibshirani R. 2012b. Normalization, testing, and false discovery rate estimation for RNA-sequencing data. *Biostatistics* 13(3): 523-538.

Li J., Yang Z., Yu B., Liu J. and Chen X. 2005. Methylation protects miRNAs and siRNAs from a 3'-end uridylation activity in *Arabidopsis*. *Current biology : CB* 15(16): 1501-1507.

Li T., Li H., Zhang Y.-X. and Liu J.-Y. 2011. Identification and analysis of seven H_2O_2-responsive miRNAs and 32 new miRNAs in the seedlings of rice (*Oryza sativa* L. ssp. indica). *Nucleic Acids Research* 39(7): 2821-2833.

Li W. and Ruan K. 2009. MicroRNA detection by microarray. *Analytical and Bioanalytical Chemistry* 394(4): 1117-1124.

Li Y., Hu X., Chen J., Wang W., Xiong X., *et al.* 2017. Integrated mRNA and microRNA transcriptome analysis reveals miRNA regulation in response to PVA in potato. *Scientific Reports* 7(1): 16925.

Liang G., He H. and Yu D. 2012. Identification of nitrogen starvation-responsive microRNAs in *Arabidopsis thaliana*. *PLoS ONE* 7(11): e48951-e48951.

Liefting L.W. and Kirkpatrick B.C. 2006. Cosmid cloning and sample sequencing of the genome of the uncultivable mollicute, Western X-disease phytoplasma, using DNA purified by pulsed-field gel electrophoresis. *FEMS Microbiology Letters* 221(2): 203-211.

Lim P.O. and Sears B.B. 1992. Evolutionary relationships of a plant-pathogenic mycoplasmalike organism and Acholeplasma laidlawii deduced from two ribosomal protein gene sequences. *Journal of Bacteriology* 174(8): 2606.

Lin R., He L., He J., Qin P., Wang Y., *et al.* 2016. Comprehensive analysis of microRNA-Seq and target mRNAs of rice sheath blight pathogen provides new insights into pathogenic regulatory mechanisms. *DNA Research* 23(5): 415-425.

Lin, C.P. and Chen, T.A. 1985. Monoclonal antibodies against the aster yellows agent. *Science* 227, 1233-1235.

Liu H.-H., Tian X., Li Y.-J., Wu C.-A. and Zheng C.-C. 2008. Microarray-based analysis of stress-regulated microRNAs in *Arabidopsis thaliana*. *RNA (New York, N.Y.)* 14(5): 836-843.

Liu L.Y.D., Tseng H.I., Lin C.P., Lin Y.Y., Huang Y.H., *et al.* 2014. High-throughput transcriptome analysis of the leafy flower transition of *Catharanthus roseus* induced by peanut witches'-broom phytoplasma infection. *Plant and Cell Physiology* 55(5): 942-957.

Liu Q. and Chen Y.-Q. 2010. A new mechanism in plant engineering: The potential roles of microRNAs in molecular breeding for crop improvement. *Biotechnology Advances* 28(3): 301-307.

Liu Y.-X., Wang M. and Wang X.-J. 2014. Endogenous small RNA clusters in plants. *Genomics, Proteomics & Bioinformatics* 12(2): 64-71.

Llave C., Kasschau K.D., Rector M.A. and Carrington J.C. 2002. Endogenous and silencing-associated small RNAs in plants. *The Plant cell* 14(7): 1605-1619.

Loi N., Ermacora P., Carraro L., Osler R. and An Chen T. 2002. Production of monoclonal antibodies against apple proliferation phytoplasma and their use in serological detection. *European Journal of Plant Pathology* 108(1): 81-86.

Lorenz K.H., Schneider B., Ahrens U. and Seemuller E. 1995. Detection of the apple proliferation and pear decline phytoplasmas by PCR amplification of ribosomal and non-ribosomal DNA. *Phytopathology* 85: 771-776.

Lu C., Tej S.S., Luo S., Haudenschild C.D., Meyers B.C., *et al.* 2005. Elucidation of the small RNA component of the transcriptome. *Science* 309(5740): 1567.

Lu S., Sun Y.-H., Amerson H. and Chiang V.L. 2007. MicroRNAs in loblolly pine (*Pinus taeda* L.) and their association with fusiform rust gall development. *The Plant Journal* 51(6): 1077-1098.

Lu S., Sun Y.-H., Shi R., Clark C., Li L., *et al.* 2005. Novel and mechanical stress-responsive microRNAs in *Populus trichocarpa* that are absent from *Arabidopsis*. *The Plant cell* 17(8): 2186-2203.

Lukasik A., Wójcikowski M. and Zielenkiewicz P. 2016. Tools4miRs - one place to gather all the tools for miRNA analysis. *Bioinformatics* 32(17): 2722-2724.

Maclean A.M., Orlovskis Z., Kowitwanich K., Zdziarska A.M., Angenent G.C., *et al.* 2014. Phytoplasma effector SAP54 hijacks plant reproduction by degrading MADS-box proteins and promotes insect colonization in a RAD23-dependent manner. *PLoS Biology* 12(4): e1001835.

Maejima K., Oshima K. and Namba S. 2014. Exploring the phytoplasmas, plant pathogenic bacteria. *Journal of General Plant Pathology* 80(3): 210-221.

Maixner M. 2006. Grapevine yellows - current developments and unsolved questions. 15th meeting of the international council for the study of virus and virus-like diseases of the grapevine (ICVG), Stellenbosch, South Africa, 86-88.

Mallory A.C., Bartel D.P. and Bartel B. 2005. MicroRNA-directed regulation of *Arabidopsis* AUXIN RESPONSE FACTOR17 is essential for proper development and modulates expression of early auxin response genes. *The Plant cell* 17(5): 1360-1375.

Marcone C. 2014. Molecular biology and pathogenicity of phytoplasmas. *Annals of Applied Biology* 165(2): 199-221.

Marcone C., Neimark H., Ragozzino A., Lauer U. and Seemüller E. 1999. Chromosome sizes of phytoplasmas composing major phylogenetic groups and subgroups. *Phytopathology* 89(9): 805-810.

Marcone C., Ragozzino A., Camele I., Rana G.L., Seem, *et al.* 2001. Updating and extending genetic characterization and classification of phytoplasmas from wild and cultivated plants in southern Italy. *Journal of Plant Pathology* 83(2): 133-138.

Margaria P. and Palmano S. 2010. Response of the *Vitis vinifera* L. cv. 'Nebbiolo' proteome to Flavescence dorée phytoplasma infection. *Proteomics* 11(2): 212-224.

Margaria P., Abbà S. and Palmano S. 2013. Novel aspects of grapevine response to phytoplasma infection investigated by a proteomic and phospho-proteomic approach with data integration into functional networks. *BMC Genomics* 14(1): 38.

McCoy, R.E., Caudwell, A., Chang, C.J., Chen, T.A., Chiykowski, L.N., *et al.* 1989. Plant diseases associated with mycoplasma-like organisms. *In: The Mycoplasmas vol. V.* Whitcomb R.F. and Tully J.G. (eds.). Academic Press, San Diego, California, pp.545-640.

Medina I., Carbonell J., Pulido L., Madeira S.C., Goetz S., *et al.* 2010. Babelomics: an integrative platform for the analysis of transcriptomics, proteomics and genomic data with advanced functional profiling. *Nucleic Acids Research* 38: W210-W213.

Meng Y., Gou L., Chen D., Mao C., Jin Y., *et al.* 2011. PmiRKB: a plant microRNA knowledge base. *Nucleic Acids Research* 39: D181-D187.

Mette M.F., Aufsatz W., Van Der Winden J., Matzke M.A. and Matzke A.J. 2000. Transcriptional silencing and promoter methylation triggered by double-stranded RNA. *The EMBO journal* 19(19): 5194-5201.

Mhuantong W. and Wichadakul D. 2009. MicroPC (μPC): A comprehensive resource for predicting and comparing plant microRNAs. *BMC Genomics* 10(1): 366.

Minato N., Himeno M., Hoshi A., Maejima K., Komatsu K., *et al.* 2014. The phytoplasmal virulence factor TENGU causes plant sterility by downregulating of the jasmonic acid and auxin pathways. *Scientific Reports* 4: 7399.

Mitchell P.L. 2004. Heteroptera as vectors of plant pathogens. *Neotropical Entomology* 33(5): 519-545.

Mitrović J., Siewert C., Duduk B., Hecht J., Mölling K., *et al.* 2014. Generation and analysis of draft sequences of 'stolbur' phytoplasma from multiple displacement amplification templates. *Journal of Molecular Microbiology and Biotechnology* 24(1): 1-11.

Miura C., Sugawara K., Neriya Y., Minato N., Keima T., *et al.* 2012. Functional characterization and gene expression profiling of superoxide dismutase from plant pathogenic phytoplasma. *Gene* 510(2): 107-112.

Monavarfeshani A., Mirzaei M., Sarhadi E., Amirkhani A., Khayam Nekouei M., *et al.* 2013. Shotgun proteomic analysis of the Mexican lime tree infected with '*Candidatus* Phytoplasma aurantifolia'. *Journal of Proteome Research* 12(2): 785-795.

Moorthie S., Mattocks C.J. and Wright C.F. 2011. Review of massively parallel DNA sequencing technologies. *The HUGO Journal* 5(1-4): 1-12.

Mortazavi A., Williams B.A., Mccue K., Schaeffer L. and Wold B. 2008. Mapping and quantifying mammalian transcriptomes by RNA-Seq. *Nature Methods* 5: 621.

Mou H.-Q., Lu J., Zhu S.-F., Lin C.-L., Tian G.-Z., *et al.* 2013. Transcriptomic analysis of *Paulownia* infected by Paulownia witches'-broom phytoplasma. *PLoS ONE* 8(10): e77217.

Moxon S., Schwach F., Dalmay T., Maclean D., Studholme D.J., *et al.* 2008. A toolkit for analysing large-scale plant small RNA datasets. *Bioinformatics* 24(19): 2252-2253.

Namba S., Oyaizu H., Kato S., Iwanami S., and Tsuchizaki T. 1993. Phylogenetic diversity of phytopathogenic mycoplasmalike organisrns. *International Journal of Systematic Bacteriology* 43: 461-467.

Namba S., Yamashita S., Doi Y. and Yora K. 1981. Direct fluorescence detection method (DFD method) for diagnosing yellow-type virus diseases and mycoplasma diseases of plants (in Japanese with English summary). *Annals of the Phytopathological Society of Japan* 47: 258-263.

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(5772): 436-439.

Nawrocki E.P., Burge S.W., Bateman A., Daub J., Eberhardt R.Y., *et al.* 2015. Rfam 12.0: updates to the RNA families database. *Nucleic Acids Research* 43: D130-D137.

Neimark H. and Kirkpatrick B.C. 1993. Isolation and characterization of full-length chromosomes from non-culturable plant-pathogenic mycoplasma-like organisms. *Molecular Microbiology* 7:21-28.

Nejat N and Vadamalai G. 2013. Diagnostic techniques for detection of phytoplasma diseases: past and present. *Journal of Plant Diseases and Protection* 120(1): 16-25.

Neriya Y., Maejima K., Nijo T., Tomomitsu T., Yusa A., *et al.* 2014. Onion yellow phytoplasma P38 protein plays a role in adhesion to the hosts. *FEMS Microbiology Letters* 361(2): 115-122.

Nishigawa H., Miyata S.-I., Oshima K., Sawayanagi T., Komoto A., *et al.* 2001. *In planta* expression of a protein encoded by the extrachromosomal DNA of a phytoplasma and related to geminivirus replication proteins. *Microbiology* 147(2): 507-513.

Norden-Krichmar T.M., Allen A.E., Gaasterland T. and Hildebrand M. 2011. Characterization of the small RNA transcriptome of the diatom, *Thalassiosira pseudonana*. *PLoS ONE* 6(8): e22870-e22870.

Orlovskis Z., Canale M., Haryono M., Lopes J., Kuo C.-H., *et al.* 2017. A few sequence polymorphisms among isolates of Maize bushy stunt phytoplasma associate with organ proliferation symptoms of infected maize plants. *Annals of Botany* 119(5): 869-884.

Orlovskis Z., Canale M., Haryono M., Lopes J., Kuo C.-H., *et al.* 2017. A few sequence polymorphisms among isolates of Maize bushy stunt phytoplasma associate with organ proliferation symptoms of infected maize plants. *Annals of Botany* 119(5): 869-884.

Oshima K., Ishii Y., Kakizawa S., Sugawara K., Neriya Y., *et al.* 2011. Dramatic transcriptional changes in an intracellular parasite enable host switching between plant and insect. *PLoS ONE* 6: e23242.

Oshima K., Kakizawa S., Nishigawa H., Jung H.-Y., Wei W., *et al.* 2004. Reductive evolution suggested from the complete genome sequence of a plant-pathogenic phytoplasma. *Nature Genetics* 36(1): 27-29.

Oshima K., Maejima K. and Namba S. 2013. Genomic and evolutionary aspects of phytoplasmas. *Frontiers in Microbiology* 4: 230.

Oshima K., Shiomi T., Kuboyama T., Sawayanagi T., Nishigawa H., *et al.* 2001. Isolation and characterization of derivative lines of the onion yellows phytoplasma that do not cause stunting or phloem hyperplasia. *Phytopathology* 91(11): 1024-1029.

Pant B.D., Buhtz A., Kehr J. and Scheible W.-R. 2008. MicroRNA399 is a long-distance signal for the regulation of plant phosphate homeostasis. *The Plant journal* 53(5): 731-738.

Park M.Y., Wu G., Gonzalez-Sulser A., Vaucheret H. and Poethig R.S. 2005. Nuclear processing and export of microRNAs in *Arabidopsis*. *Proceedings of the National Academy of Sciences of the United States of America* 102(10): 3691.

Park W., Li J., Song R., Messing J. and Chen X. 2002. CARPEL FACTORY, a Dicer homolog, and HEN1, a novel protein, act in microRNA metabolism in *Arabidopsis thaliana*. *Current Biology* 12(17): 1484-1495.

Pasquinelli A.E., Reinhart B.J., Slack F., Martindale M.Q., Kuroda M.I., *et al.* 2000. Conservation of the sequence and temporal expression of *let-7* heterochronic regulatory RNA. *Nature* 408: 86.

Pastore M., Piccirillo P., Simeone A.M., Tian J., Paltrinieri S., *et al.* 2001. Transmission by patch grafting of ESFY phytoplasma to apricot (*Prunus armeniaca* L.) and Japanese plum (*Prunus salicina* lindl). *Acta horticulturae* 550: 339-344.

Přibylová J. and Spak J. 2013. Dodder transmission of phytoplasmas. *Methods in Molecular Biology* 938: 41-46.

Quaglino F., Kube M., Jawhari M., Abou-Jawdah Y., Siewert C., *et al.* 2015. '*Candidatus* Phytoplasma phoenicium' associated with almond witches'-broom disease: from draft genome to genetic diversity among strain populations. *BMC Microbiology* 15(1): 148.

Razin S., Yogev D. and Naot Y. 1998. Molecular biology and pathology of mycoplasmas. *Microbiology and Molecular Biology Review* 62: 1094-1156.

Reyes J.L. and Chua N.-H. 2007. ABA induction of miR159 controls transcript levels of two MYB factors during *Arabidopsis* seed germination. *The Plant Journal* 49(4): 592-606.

Rhoades M.W., Reinhart B.J., Lim L.P., Burge C.B., Bartel B., *et al.* 2002. Prediction of plant microRNA targets. *Cell* 110(4): 513-520.

Robinson M.D. and Oshlack A. 2010. A scaling normalization method for differential expression analysis of RNA-seq data. *Genome Biology* 11(3): R25.

Robinson M.D., Mccarthy D.J. and Smyth G.K. 2010. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* 26(1): 139-140.

Rueda A., Barturen G., Lebrón R., Gómez-Martín C., Alganza Á., *et al.* 2015. sRNAtoolbox: an integrated collection of small RNA research tools. *Nucleic Acids Research* 43(W1): W467-W473.

Ruiz-Ferrer V. and Voinnet O. 2009. Roles of plant small RNAs in biotic stress responses. *Annual Review of Plant Biology* 60(1): 485-510.

Saccardo F., Martini M., Palmano S., Ermacora P., Scortichini M., *et al.* 2012. Genome drafts of four phytoplasma strains of the ribosomal group 16SrIII. *Microbiology* 158(11): 2805-2814.

Santana M., Ionescu M.S., Vertes A., Longin R., Kunst F., *et al.* 1994. Bacillus subtilis F0F1 ATPase: DNA sequence of the atp operon and characterization of *atp* mutants. *Journal of Bacteriology* 176: 6802-6811.

Sarkar D., Maji R.K., Dey S., Sarkar A., Ghosh Z., *et al.* 2017. Integrated miRNA and mRNA expression profiling reveals the response regulators of a susceptible tomato cultivar to early blight disease. *DNA research : an international journal for rapid publication of reports on genes and genomes* 24(3): 235-250.

Schroeder A., Mueller O., Stocker S., Salowsky R., Leiber M., *et al.* 2006. The RIN: an RNA integrity number for assigning integrity values to RNA measurements. *BMC Molecular Biology* 7: 3-3.

Seddas A., Meignoz R., Daire X. and Boudon-Padieu E. 1996. Generation and characterization of monoclonal antibodies to Flavescence dorée phytoplasma: Serological relationships and differences in electroblot immunoassay profiles of Flavescence dorée and elm yellows phytoplasmas. *European Journal of Plant Pathology* 102(8): 757-764.

Seemüller E., Marcone C., Lauer U., Ragozzino A. and Göschl M. 1998. Current status of molecular classification of the phytoplasmas. *Journal of Plant Pathology* 80: 3-26.

Seemüller E., Schneider B., Maurer R., Ahrens, U., Daire, X., *et al.* 1994. Phylogenetic classification of phytopathogenic mollicutes by sequence analysis of 16S ribosomal DNA. *International Journal of Systematic Bacteriology* 44: 440-446.

Shivaprasad P.V., Chen H.-M., Patel K., Bond D.M., Santos B.a.C.M., *et al.* 2012. A microRNA superfamily regulates nucleotide binding site-leucine-rich repeats and other mRNAs. *The Plant Cell* 24(3): 859-874.

Sims D., Sudbery I., Ilott N.E., Heger A. and Ponting C.P. 2014. Sequencing depth and coverage: key considerations in genomic analyses. *Nature Reviews Genetics* 15: 121.

Sinha R.C. and Benhamou N. 1983. Detection of mycoplasmalike organisms antigens from aster yellows-diseased plants by two serological procedures. *Phytopathology* 73: 1199-1202.

Soneson C. and Delorenzi M. 2013. A comparison of methods for differential expression analysis of RNA-seq data. *BMC Bioinformatics* 14: 91-91.

Sparks M.E., Bottner-Parker K.D., Gundersen-Rindal D.E. and Lee I.-M. 2018. Draft genome sequence of the New Jersey aster yellows strain of '*Candidatus* Phytoplasma asteris'. *PLoS ONE* 13(2): e0192379.

Spies D., Renz P.F., Beyer T.A. and Ciaudo C. 2017. Comparative analysis of differential gene expression tools for RNA sequencing time course data. *Briefings in Bioinformatics*: bbx115-bbx115.

Spinas N.L. 2013. *The efficacy of the antimicrobial peptides D4E1, VvAMP-1 and Snakin1 against the grapevine pathogen aster yellows phytoplasma.* Master's degree thesis, Department of Genetics, Stellenbosch University. http://scholar.sun.ac.za/handle/10019.1/80066

Stocks M.B., Moxon S., Mapleson D., Woolfenden H.C., Mohorianu I., *et al.* 2012. The UEA sRNA workbench: a suite of tools for analysing and visualizing next generation sequencing microRNA and small RNA datasets. *Bioinformatics* 28(15): 2059-2061.

Sugawara K., Honma Y., Komatsu K., Himeno M., Oshima K., *et al.* 2013. The alteration of plant morphology by small peptides released from the proteolytic processing of the bacterial peptide TENGU. *Plant Physiology* 162(4): 2005.

Sugio A., Maclean A.M., Kingdom H.N., Grieve V.M., Manimekalai R., *et al.* 2011. Diverse targets of phytoplasma effectors: from plant development to defense against insects. *Annual Review of Phytopathology* 49(1): 175-195.

Sunkar R. and Zhu J.-K. 2004. Novel and stress-regulated microRNAs and other small RNAs from *Arabidopsis*. *The Plant cell* 16(8): 2001-2019.

Sunkar R., Kapoor A. and Zhu J.-K. 2006. Posttranscriptional induction of two Cu/Zn superoxide dismutase genes in *Arabidopsis* is mediated by downregulation of miR398 and important for oxidative stress tolerance. *The Plant cell* 18(8): 2051-2065.

Sunkar R., Li Y.-F. and Jagadeeswaran G. 2012. Functions of microRNAs in plant stress responses. *Trends in Plant Science* 17(4): 196-203.

Szczesniak M.W. and Makalowska I. 2014. miRNEST 2.0: a database of plant and animal microRNAs. *Nucleic Acids Research* 42: D74-D77.

Tam S., Tsao M.-S. and Mcpherson J.D. 2015. Optimization of miRNA-seq data preprocessing. *Briefings in bioinformatics* 16(6): 950-963.

Tanne E., Boudon-Padieu E., Clair D., Davidovich M., Melamed S., *et al.* 2001. Detection of phytoplasma by polymerase chain reaction of insect feeding medium and its use in determining vectoring ability. *Phytopathology* 91(8): 741-746.

Tao Y., Xie Z., Chen W., Glazebrook J., Chang H.-S., *et al.* 2003. Quantitative nature of *Arabidopsis* responses during compatible and incompatible interactions with the bacterial pathogen *Pseudomonas syringae*. *The Plant cell* 15(2): 317-330.

Tarazona S., Furió-Tarí P., Turrà D., Pietro A.D., Nueda M.J., *et al.* 2015. Data quality aware analysis of differential expression in RNA-seq with NOISeq R/Bioc package. *Nucleic Acids Research* 43(21): e140-e140.

Toruño T.Y., Seruga Musić M., Simi S., Nicolaisen M. and Hogenhout S.A. 2010. Phytoplasma PMU1 exists as linear chromosomal and circular extrachromosomal elements and has enhanced expression in insect vectors compared with plant hosts. *Molecular Microbiology* 77(6): 1406-1415.

Toth K.F., Harrison N. and Sears B.B. 1994. Phylogenetic relationships among members of the class Mollicutes deduced from rps3 gene sequences. *International Journal of Systematic and Evolutionary Microbiology* 44: 119-124.

Tran-Nguyen L.T.T., Kube M., Schneider B., Reinhardt R. and Gibb K.S. 2008. Comparative genome analysis of "*Candidatus* Phytoplasma australiense" (subgroup *tuf*-Australia I; *rp*-A) and "*Ca.* phytoplasma asteris" strains OY-M and AY-WB. *Journal of Bacteriology* 190(11): 3979.

Trapnell C., Hendrickson D.G., Sauvageau M., Goff L., Rinn J.L., *et al.* 2012. Differential analysis of gene regulation at transcript resolution with RNA-seq. *Nature Biotechnology* 31: 46.

Trapnell C., Roberts A., Goff L., Pertea G., Kim D., *et al.* 2012. Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and Cufflinks. *Nature Protocols* 7(3): 562-578.

Trapnell C., Williams B.A., Pertea G., Mortazavi A., Kwan G., *et al.* 2010. Transcript assembly and abundance estimation from RNA-Seq reveals thousands of new transcripts and switching among isoforms. *Nature Biotechnology* 28(5): 511-515.

Trindade I., Capitão C., Dalmay T., Fevereiro M.P. and Santos D.M.D. 2010. miR398 and miR408 are up-regulated in response to water deficit in *Medicago truncatula*. *Planta* 231(3): 705-716.

Válóczi A., Hornyik C., Varga N., Burgyán J., Kauppinen S., *et al.* 2004. Sensitive and specific detection of microRNAs by northern blot analysis using LNA-modified oligonucleotide probes. *Nucleic Acids Research* 32(22): e175-e175.

Van Dijk E.L., Auger H., Jaszczyszyn Y. and Thermes C. 2014. Ten years of next-generation sequencing technology. *Trends in Genetics* 30(9): 418-426.

Várallyay É., Burgyán J. and Havelda Z. 2008. MicroRNA detection by northern blotting using locked nucleic acid probes. *Nature Protocols* 3: 190.

Varkonyi-Gasic E., Wu R., Wood M., Walton E.F. and Hellens R.P. 2007. Protocol: a highly sensitive RT-PCR method for detection and quantification of microRNAs. *Plant Methods* 3(1): 12.

Veneziano D., Nigita G. and Ferro A. 2015. Computational approaches for the analysis of ncRNA through deep sequencing techniques. *Frontiers in bioengineering and biotechnology* 3: 77-77.

Voinnet O. 2009. Origin, biogenesis, and activity of plant microRNAs. *Cell* 136(4): 669-687.

Wang L., Feng Z., Wang X., Wang X. and Zhang X. 2010. DEGseq: an R package for identifying differentially expressed genes from RNA-seq data. *Bioinformatics* 26(1): 136-138.

Wang Q., Guo Y., Wang N., Li Y., Chen W., *et al.* 2014. Identification of a conserved core genome with group-specific genes from comparative genomics of ten different *Candidatus* Phytoplasma strains. *Journal of Phytopathology* 162(10): 650-659.

Wang W.-C., Lin F.-M., Chang W.-C., Lin K.-Y., Huang H.-D., *et al.* 2009. miRExpress: analyzing high-throughput sequencing data for profiling microRNA expression. *BMC Bioinformatics* 10: 328-328.

Wang X., Zhang J., Li F., Gu J., He T., *et al.* 2005. MicroRNA identification based on sequence and structure alignment. *Bioinformatics* 21(18): 3610-3614.

Wei W., Davis R.E., Jomantiene R. and Zhao Y. 2008. Ancient, recurrent phage attacks and recombination shaped dynamic sequence-variable mosaics at the root of phytoplasma genome evolution. *Proceedings of the National Academy of Sciences* 105(33): 11827.

Weiberg A. and Jin H. 2015. Small RNAs - the secret agents in the plant-pathogen interactions. *Current Opinion in Plant Biology* 26: 87-94.

Weintraub P.G. and Beanland L. 2006. Insect vectors of phytoplasmas. *Annual Review of Entomology* 51: 91-111.

Weisburg W.G., Tully J.G., Rose D.L., Petzel J.P., Oyazu H., *et al.* 1989. A phylogenetic analysis of the mycoplasmas: basis for their classifi cation. *Journal of Bacteriology* 171: 6455-6467.

Woese C.R. 1987. Bacterial evolution. *Microbiological Reviews* 51: 221-271.

Wu T.D. and Nacu S. 2010. Fast and SNP-tolerant detection of complex variants and splicing in short reads. *Bioinformatics* 26(7): 873-881.

Xie Z., Allen E., Fahlgren N., Calamar A., Givan S.A., *et al.* 2005. Expression of *Arabidopsis MIRNA* genes. *Plant Physiology* 138(4): 2145.

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 Biology* 10: 123-123.

Yamasaki H., Hayashi M., Fukazawa M., Kobayashi Y. and Shikanai T. 2009. SQUAMOSA promoter binding protein-like7 is a central regulator for copper homeostasis in *Arabidopsis*. *The Plant cell* 21(1): 347-361.

Yang J., Zhang F., Li J., Chen J.-P. and Zhang H.-M. 2016. Integrative analysis of the microRNAome and transcriptome illuminates the response of susceptible rice plants to rice stripe virus. *PLoS ONE* 11(1): e0146946-e0146946.

Yang X. and Li L. 2011. miRDeep-P: a computational tool for analyzing the microRNA transcriptome in plants. *Bioinformatics* 27(18): 2614-2615.

Yang Z., Ebright Y.W., Yu B. and Chen X. 2006. HEN1 recognizes 21-24 nt small RNA duplexes and deposits a methyl group onto the 2' OH of the 3' terminal nucleotide. *Nucleic Acids Research* 34(2): 667-675.

Yu B., Bi L., Zheng B., Ji L., Chevalier D., *et al.* 2008. The FHA domain proteins DAWDLE in *Arabidopsis* and SNIP1 in humans act in small RNA biogenesis. *Proceedings of the National Academy of Sciences of the United States of America* 105(29): 10073-10078.

Zeng Q.-Y., Yang C.-Y., Ma Q.-B., Li X.-P., Dong W.-W., *et al.* 2012. Identification of wild soybean miRNAs and their target genes responsive to aluminum stress. *BMC Plant Biology* 12: 182-182.

Zhai J., Jeong D.-H., De Paoli E., Park S., Rosen B.D., *et al.* 2011. MicroRNAs as master regulators of the plant NB-LRR defense gene family via the production of phased, trans-acting siRNAs. *Genes & Development* 25(23): 2540-2553.

Zhang B. and Wang Q. 2014. MicroRNA-based biotechnology for plant improvement. *Journal of Cellular Physiology* 230(1): 1-15.

Zhang B., Pan X., Cobb G.P. and Anderson T.A. 2006. Plant microRNA: A small regulatory molecule with big impact. *Developmental Biology* 289(1): 3-16.

Zhang J., Xu Y., Huan Q. and Chong K. 2009. Deep sequencing of *Brachypodium* small RNAs at the global genome level identifies microRNAs involved in cold stress response. *BMC Genomics* 10: 449-449.

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 Molecular Biology* 75(1-2): 93-105.

Zhang Y. 2005. miRU: an automated plant miRNA target prediction server. *Nucleic Acids Research* 33: W701-W704.

Zhang Z., Yu J., Li D., Zhang Z., Liu F., *et al.* 2010. PMRD: plant microRNA database. *Nucleic Acids Research* 38: D806-D813.

Zhao J.-P., Jiang X.-L., Zhang B.-Y. and Su X.-H. 2012. Involvement of microRNA-mediated gene expression regulation in the pathological development of stem canker disease in *Populus trichocarpa*. *PLoS ONE* 7(9): e44968.

Zhao Y., Wei W., Lee I.-M., Shao J., Suo X., *et al.* 2009. Construction of an interactive online phytoplasma classification tool, *i*PhyClassifier, and its application in analysis of the peach X-disease phytoplasma group (16SrIII). *International Journal of Systematic and Evolutionary Microbiology* 59(10): 2582-2593.

Zheng Y., Li Y.-F., Sunkar R. and Zhang W. 2012. SeqTar: an effective method for identifying microRNA guided cleavage sites from degradome of polyadenylated transcripts in plants. *Nucleic Acids Research* 40(4): e28-e28.

Zhou L., Liu Y., Liu Z., Kong D., Duan M., *et al.* 2010. Genome-wide identification and analysis of drought-responsive microRNAs in *Oryza sativa*. *Journal of Experimental Botany* 61(15): 4157-4168.

Zhou X., Wang G. and Zhang W. 2007. UV-B responsive microRNA genes in *Arabidopsis thaliana*. *Molecular Systems Biology* 3: 103-103.

Zhu Q.-H., Fan L., Liu Y., Xu H., Llewellyn D., *et al.* 2014. miR482 regulation of NBS-LRR defense genes during fungal pathogen infection in cotton. *PLoS ONE* 8(12): e84390.

Chapter 3

The use of high-throughput small RNA sequencing reveals differentially expressed microRNAs in response to aster yellows phytoplasma-infection in *Vitis vinifera* cv. 'Chardonnay'

3.1 Abstract

Phytoplasmas are cell wall-less plant pathogenic bacteria responsible for major crop losses throughout the world. In grapevine they cause grapevine yellows, a detrimental disease associated with a variety of symptoms. The high economic impact of this disease has sparked considerable interest among researchers to understand molecular mechanisms related to pathogenesis. Increasing evidence exist that a class of small non-coding endogenous RNAs known as microRNAs (miRNAs), play an important role in post-transcriptional gene regulation during plant development and responses to biotic and abiotic stresses. Thus, we aimed to dissect complex high-throughput small RNA sequencing data for the genome-wide identification of known and novel differentially expressed miRNAs, using read libraries constructed from healthy and phytoplasma-infected Chardonnay leaf material. Furthermore, we utilised computational resources to predict putative miRNA targets to explore the involvement of possible pathogen response pathways. We identified multiple known miRNA sequence variants (isomiRs), likely generated through post-transcriptional modifications. Sequences of 13 known, canonical miRNAs were shown to be differentially expressed. A total of 175 novel miRNA precursor sequences, each derived from a unique genomic location, were predicted, of which 23 were differentially expressed. A homology search revealed that some of these novel miRNAs shared high sequence similarity with conserved miRNAs from other plant species, as well as known grapevine miRNAs. The relative expression of randomly selected known and novel miRNAs was determined with real-time RT-qPCR analysis, thereby validating the trend of expression seen in the normalised small RNA sequencing read count data. Among the putative miRNA targets, we identified genes involved in plant morphology, hormone signalling, nutrient homeostasis, as well as plant stress. Our results may assist in understanding the role that miRNA pathways play during plant pathogenesis, and may be crucial in understanding disease symptom development in aster yellows phytoplasma-infected grapevines.

3.2 Introduction

Phytoplasmas are known to infect hundreds of plant species worldwide and are responsible for devastating yield losses of many economically important crops, fruit trees, and ornamental plants (Lee *et al.*, 2000). They are obligate cell wall-less bacterial pathogens (class Mollicutes), and rely on plants and homopterous phloem-sucking insects for biological dispersal. In plants, they are mainly restricted to the phloem tissue where they can move and multiply through the sieve tube elements (Hogenhout *et al.*, 2008).

The aster yellows (AY) phytoplasma group (16SrI, subgroup A and B) represents the most diverse and widespread phytoplasma group and is also known as '*Candidatus* Phytoplasma asteris' (Lee, 2004). AY phytoplasma-infection can cause a severe disease in grapevine (*Vitis vinifera* L.), known as grapevine yellows (GY). Phytoplasma-like symptoms have been observed in South African vineyards since 2006, and were later shown to be caused by AY phytoplasma (16SrI-B) (Engelbrecht *et al.*, 2010). Transmission experiments conducted on vineyards in the vicinity of Vredendal (Western Cape) suggested that *Mgenia fuscovaria* (Hemiptera: Cicadellidae) is a vector of AY phytoplasma in South Africa (Krüger *et al.*, 2011). GY disease incidence in the same region was monitored for different cultivars (Chenin blanc, Shiraz, Chardonnay, Cabernet Franc, Sauvignon blanc, Pinotage and Colombar), and revealed that Chardonnay is especially susceptible, based on a GY increase from 0.5% to 7.5% in two years in a single vineyard (Carstens *et al.*, 2011). Typical symptoms caused by GY disease include discolouration and necrosis of leaf veins and laminae, downward curling of leaves, abnormal leaf shape and size, incomplete lignification, stunting and necrosis of shoots, flower abortion and berry withering. These symptoms eventually lead to reduced plant vitality and fruit yield that may hold devastating consequences for the wine and table grape industries (Lee *et al.*, 2000, Belli *et al.*, 2010). Currently, the only available control strategies include early eradication of infected crops, early eradication of infected source plants (weed control), and chemical control of vectors through regular insecticide treatments (Maixner, 2006)

V. vinifera is one of the most important fruit and/or beverage crops in the world and, like all land plants, grapevines have to develop various mechanisms at a physiological and molecular level in order to cope with their ever-changing environment. Significant progress has been made to understand plant-pathogen interactions and the multiple gene regulatory mechanisms they invoke during plant defence responses. The recent successful, axenic cultivation of phytoplasmas (Contaldo *et al.*, 2012) will allow direct *in planta* investigation of molecular interactions postulated to exist between phytoplasmas and their plant and insect vectors. In addition, high-throughput

transcriptome analysis of next-generation sequencing (NGS) and microarray data, as well as proteomics, have served as valuable approaches for gaining new insights into physiological, biochemical and molecular mechanisms underlying phytoplasma disease symptom development in grapevine and other plant species (Hren *et al.*, 2009, Albertazzi *et al.*, 2009; Margaria *et al.*, 2010; Mou *et al.*, 2013; Margaria *et al.*, 2013; Monavarfeshani *et al.*, 2013; Liu *et al.*, 2014).

Increasing evidence has shown that a class of small non-coding endogenous RNAs known as microRNAs (miRNAs), play a major role in post-transcriptional gene regulation during plant development and plant responses to biotic and abiotic stresses (Ruiz-Ferrer and Voinnet, 2009; Sunkar *et al.*, 2012). Mature miRNAs are typically 19 to 24 nt in length and originate from miRNA (*MIR*) genes that are transcribed by RNA Polymerase II. These transcripts, known as primary miRNAs (pri-miRNA), form imperfect fold-back hairpins that are cleaved by RNase III-like Dicer 1 (DCL1) to produce miRNA precursors (pre-miRNA). Each pre-miRNA contains one or more short intermediate complementary miRNA/miRNA* duplexes. These duplexes are then cleaved by DCL1 from the stem region and processed inside the nucleus to be exported to the cytoplasm where the leading miRNA is incorporated into the RNA-induced silencing complex (RISC). When associated with the RISC, guided binding of the miRNA to its complementary target mRNA(s) or non-coding trans-acting siRNA (*TAS*) transcript(s) occurs. This facilitates either translational inhibition or degradation of target mRNA(s), or slicing of *TAS* transcripts that lead to generation of trans-acting siRNAs (tasiRNAs). Target degradation occurs through endonucleolytic cleavage by the RISC core protein Argonaute 1 (AGO1) (Allen *et al.*, 2005; Vaucheret, 2006; Budak and Akpinar, 2015).

It has been suggested that the miRNA pathway contributes to pathogen-associated molecular pattern (PAMP)-triggered immunity (PTI), which refers to a basal defence response upon recognition of certain pathogenic elements, such as flagellin (Padmanabhan *et al.*, 2009). The bacterial PAMP peptide flg22 causes induced expression of the *Arabidopsis* miR393, which was the first miRNA identified to play a role in plant PTI. Overexpression of miR393 caused down-regulation of auxin receptor mRNAs, including *transport inhibitor response 1* (*TIR1*), through degradation, which caused increased resistance to virulent *Pseudomonas syringae* pv. *tomato* (*Pst*) DC3000 (Navarro *et al.*, 2006).

The availability of two draft *V. vinifera* cv. 'Pinot noir' genome sequences obtained from NGS projects (Jaillon *et al.*, 2007; Velasco *et al.*, 2007) has enabled rapid discovery of miRNAs that further supports efforts to explore small RNA (sRNA)-based regulatory networks in grapevine. The use of computational analyses of high-throughput sequencing and microarray data, followed by

experimental validation, have been used to identify highly conserved miRNAs, some of which play important roles in grapevine development (Mica *et al.*, 2010; Pantaleo *et al.*, 2010). To date, 186 mature grapevine miRNA sequences from 47 different miRNA families have been deposited in miRBase v21 (Kozomara and Griffiths-Jones, 2011).

This study is the first to utilise a bioinformatics pipeline to dissect complex high-throughput sRNA sequencing (sRNA-seq) data in order to identify miRNAs that are differentially expressed in *V. vinifera* cv. 'Chardonnay' in response to AY phytoplasma-infection. Furthermore, we used computational resources for the *in silico* prediction and annotation of putative miRNA targets to explore the involvement of possible pathogen response pathways. Understanding sRNA-mediated gene regulation is crucial to expanding our knowledge of gene regulatory pathways involved in different stress-regulated physiological processes. Our results provide insight into miRNA-mediated pathogenesis in *V. vinifera* and may shed light on disease control strategies for molecular breeding in the future.

3.3 Materials and Methods

3.3.1 Plant material

We visually selected and tagged 50 symptomatic and 50 asymptomatic *V. vinifera* cv. 'Chardonnay' plants in a 7-year-old vineyard in the Olifants River Valley (Western Cape) (Figure 3.1). The vineyard was part of a high disease incidence area mapped by the Agricultural Product Inspection Services (APIS) of the Department of Agriculture, Forestry and Fisheries (DAFF). Permission was granted by the owner to conduct the study on his farm, Daltana. During the peak summer season, whole leaf material, including the blade and petiole, were collected from each plant, immediately flash frozen in liquid nitrogen, transported on dry ice and stored at -80°C until use. RNA was extracted using a modified CTAB method (White *et al.*, 2008), while genomic DNA was extracted using a NucleoSpin® Plant II kit (Macherey-Nagel; Düren, Germany). Phytoplasma infection was confirmed by a nested-PCR procedure, specifically amplifying a region of the phytoplasma 16S rDNA. The first PCR round was performed using a universal primer pair R16mF2/mR1, followed by a second PCR with the R16F2n/R2 primer pair (Gundersen and Lee, 1996). Afterwards samples were screened for the most prevalent grapevine viruses, including Grapevine leafroll-associated virus 3 (GLRaV-3), Grapevine virus A (GVA), Grapevine virus E (GVE), and Grapevine rupestris stempitting-associated virus (GRSPaV), using two-step RT-PCR assays. Primer sequences for virus screening were obtained from previous publications (File S1). Results from these diagnostics were

used to select material, free from these viruses, from three AY phytoplasma-infected, and three healthy plants for further experiments.

Figure 3.1: *Vitis vinifera* cv. 'Chardonnay' with asymptomatic leaves (A), and leaves showing typical aster yellows (AY) disease symptoms (B).

3.3.2 Total RNA extraction and sRNA-seq

Large-scale RNA extractions were carried out on one gram of plant material for each of the six experimental plants using PureLink® Plant RNA Reagent (Thermo Fisher Scientific, Waltham, Massachusetts, United States), according to the manufacturer's protocol, with an additional phenol-chloroform extraction step when further purification was necessary. Total RNA was quantified on a NanoDrop ND-1000, while RNA integrity was assessed using a Plant RNA Nano Assay using an Agilent 2100 Bioanalyzer. Ten micrograms of total RNA from each plant were sent to Fasteris SA (Plan-les-Ouates, Switzerland) for sRNA-seq. The six sRNA libraries were constructed using the TruSeq® Small RNA Library Prep Kit protocol (Illumina, San Diego, California, USA), followed by sRNA-seq (single-end; 1 x 50 bp) on an Illumina HiSeq2000 platform (https://support.illumina.com).

3.3.3 sRNA bioinformatic analysis

After sRNA-seq, high-quality, adapter-trimmed sequence data was received from the service provider in Illumina-fastq format. FastQC (www.bioinformatics.babraham.ac.uk/projects/fastqc/) was used as a tool to visualise different quality control measurements. In order to confirm RT-PCR results of the virus screening, we produced *de novo* assemblies with the 18 to 26 nt sRNA reads of each sample, using Velvet v1.1 (Zerbino and Birney, 2008). The resulting contigs were compared against the NCBI database using nucleotide BLAST (Altschul *et al.*, 1990).

The unique (non-redundant) 18 to 26 nt sequences with accompanying copy numbers, across all six libraries (representing the six biological samples), were submitted to miRanalyzer (Hackenberg *et al.*, 2011) (http://bioinfo5.ugr.es/miRanalyzer/miRanalyzer.php) for known miRNA analysis, allowing one mismatch. All reads that mapped to other non-coding RNAs (ncRNAs) in RFam (http://www.sanger.ac.uk/science/tools/rfam) and RepBase (http://www.girinst.org/repbase/) were removed, and the remaining reads were mapped against the canonical grapevine miRNA (vvi-miRNA) sequences deposited in miRBase v21. Mapped read counts for libraries obtained from the phytoplasma-infected group were compared to those from the healthy (control) group using the DESeq v2 package for differential expression analysis (Anders and Huber, 2010) (http://bioconductor.org/packages/release/bioc/html/DESeq.html).

For novel miRNA predictions, sRNA library files of the 18 to 26 nt reads, from all six libraries, were grouped into a single file that served as input for sRNAbench v0.9 (Barturen *et al.*, 2014) (http://bioinfo5.ugr.es/srnatoolbox/srnabench), and Shortstack v0.4.1 (Axtell, 2013), using the default parameters of the respective packages. sRNAbench was also used for the discovery of sequence variants of known miRNAs, also known as miRNA isoforms (isomiRs). The *V. vinifera* (cv. 'Pinot noir'; PN40024) 12x coverage genome assembly (http://www.genoscope.cns.fr/externe/GenomeBrowser/Vitis/) served as the reference sequence to which the sRNA reads were mapped (Jaillon *et al.*, 2007). Importantly, primary criteria described by Meyers *et al.* (2008), for duplex-forming precursors (pre-miRNAs) are used by both programs. These include that (1) the miRNA and miRNA* are derived from opposite arms within the stem region to form a duplex with two 3'-nucleotide overhangs; (2) extensive base-pairing exist between the miRNA and the other arm of the hairpin, which includes the miRNA*; and (3) asymmetric bulges are minimal in size and frequency, especially within the miRNA/miRNA* duplex.

The Unified Nucleic Acid Folding (UNAFold) software was used to calculate the minimum folding free energy (MFE; ΔG) of novel pre-miRNA sequences (Markham and Zuker, 2008)

(http://mfold.rna.albany.edu/). In an effort to find more comprehensive evidence that miRNAs differ from other RNAs, Zhang *et al.* (2006) described a statistical method incorporating pre-miRNA folding free energies, base pairing, nucleotide composition, and other characteristics. This method was defined by two criteria known as the adjusted minimum folding free energy (AMFE) and the minimal folding free energy index (MFEI). The AMFE and MFEI were calculated using the following equations:

$$\text{AMFE} = \frac{MFE}{Length\ of\ precursor\ (nt)} \times 100$$

$$\text{MFEI} = \frac{AMFE}{\%GC\ content\ of\ precursor}$$

Precursor sequences were analysed in RNAfold to view their stem-loop secondary structures (Gruber *et al.*, 2008). Novel mature miRNA sequences were compared against the miRBase v21 database using BLASTn v2.2.29+ (Altschul *et al.*, 1990; Camacho *et al.*, 2009; http://www.ncbi.nlm.nih.gov/books/NBK1763/) for the identification of miRNA homologs. Only the top BLAST results, with an identity of ≥90%, zero gaps and not more than two mismatches (over a seed region of 18 nt), were regarded as homologs. For each resulting BLAST hit, we compared the associated precursor sequence against miRBase with the miRBase BLASTn tool, using less stringent parameters, to identify homologous pre-miRNA sequences.

The number of sRNA reads that aligned to novel mature miRNA sequences present in all six libraries were obtained with Bowtie v1.0.1 (Langmead *et al.*, 2009), and a customised shell script. The resulting count data were analysed in DESeq v2 to obtain differentially expressed novel miRNAs. Only $\log_2$-fold changes with an adjusted *p*-value of ≤ 0.05 were considered significant.

3.3.4 Validation of miRNA expression by real-time RT-qPCR

Stem-loop reverse transcription quantitative PCR (RT-qPCR) assays were performed according to the methods of Chen *et al.* (2005) to validate the DESeq differential expression results. High-quality total RNA was prepared as described above. For each miRNA a 20 µl reverse transcription reaction was prepared containing 100 U of Superscript III reverse transcriptase (Invitrogen, Carlsbad, CA, USA), 20 U of RiboLock RNase inhibitor (Thermo Scientific, Waltham, Massachusetts, United States), 4 µl first-strand buffer (5x), 5 mM DTT, 500 nM dNTPs and 1 µl miRNA-specific stem-loop RT primer (10 µM) and 1.2 µg total RNA. Cycling conditions were as follows: 30 min at

16°C, 60 cycles at 30°C for 30 s, 42°C for 30 s, and 50°C for 1 s, heat inactivation for 5 min at 85°C, and cooling at 4°C. qPCR was performed using the Universal ProbeLibrary (UPL) probe assay with UPL probe #21 (Roche Diagnostics, Basel, Switzerland). Each 10 µl reaction mixture was prepared in triplicate and contained 1 µl cDNA, 5 µl FastStart TaqMan® Probe Master (2x) (Roche Diagnostics, Basel, Switzerland), 0.5 µl miRNA-specific forward primer (10 µM), 0.5 µl universal reverse primer (10 µM), 0.1 µl UPL probe (10 µM), and nuclease-free water. A control reaction, without cDNA template, was included for each miRNA. Based on previous results from geNorm analysis (qBase^{PLUS} v2.0, Biogazelle, Ghent, Belgium) (Hellemans *et al.*, 2007), miR167a was chosen as internal control to normalise miRNA expression levels (data not shown). PCR amplification was performed in an Applied Biosystems 7900HT Fast Real-Time PCR System**,** in which the baseline and threshold cycles (C_t) were automatically determined with SDS v2.3 software. Cycling conditions were as follows: 95°C for 5 min, 45 cycles at 95°C for 10 s and 60°C for 1 min. Relative miRNA expression analysis was performed using qBase^{PLUS} v2.0 software (Biogazelle, Ghent, Belgium).

3.3.5 miRNA target prediction and functional annotation

Potential targets of differentially expressed miRNAs were predicted using the psRNAtarget analysis server (Dai and Zhao, 2011; http://plantgrn.noble.org/psRNATarget/), with default parameters which included a threshold cut-off of 3.0 for low false-positive prediction, a complementarity scoring length of 20 bp, and the energy required for target accessibility equal to 25 kcal/mole. The collection of annotated transcript sequences of the *V. vinifera* (PN40024) 12x assembly was used for the miRNA target search (http://www.genoscope.cns.fr/externe/GenomeBrowser/Vitis/). Predicted targets for both the known and novel differentially expressed miRNAs were functionally annotated using Blast2GO v2.2.7 (Conesa and Götz *et al.*, 2008). This was done by using NCBI BLASTx to find homologous sequences, a mapping step to retrieve gene ontology (GO) terms associated with BLAST hits (http://geneontology.org/page/go-database), and assigning functional attributes to each query sequence in terms of biological processes, cellular components and molecular functions, in a species-independent manner. Afterwards a combined graph was generated using a GO sequence similarity level of 3 and an annotation cut-off value of 7.

3.4 Results and Discussion

3.4.1 Plant material

According to the diagnostic PCR screening results (data not shown), 19 out of the 50 plants that were visually tagged as 'healthy' were AY phytoplasma-positive, while 32 out of the 50 plants that were visually tagged as phytoplasma-infected were confirmed positive for AY phytoplasma. The remaining 31 'healthy' (no phytoplasma-infection) and 32 phytoplasma-infected candidate plants were subjected to further virus screening. All plants that tested positive, following the virus-screening, were eliminated from the study. BLAST results for the *de novo* assembled contigs also confirmed the absence of any prevalent grapevine viruses (data not shown). Our final test groups consisted of three phytoplasma and virus-free Chardonnay plants for the control group (h55, h85, h89), and three AY phytoplasma-infected, but virus-free, Chardonnay plants for our experimental group (p73, p93, p99).

3.4.2 sRNA-seq

To investigate miRNA expression profiles in response to GY disease, individual sRNA libraries were constructed from RNA extracted from pooled leaf material of the six plants. High-quality, adapter-trimmed reads were generated from the respective sRNA-seq libraries and the number of reads are displayed in Table 3.1.

Table 3.1: Summary of total small RNA reads.

	Small RNA library type	
Total high-quality reads	**Healthy**	**AY**
p73	N/A	10,893,265
p93	N/A	10,476,093
p99	N/A	10,511,436
h55	10,878,402	N/A
h85	12,424,487	N/A
h89	11,510,533	N/A
All	34,813,422	31,880,794
18-26 nt	26,474,279	24,314,330
18-26 nt: unique	6,388,422	5,726,632
18-26 nt: mapped	22,515,584	20,782,176

H: Health (control) sample group

AY: AY phytoplasma-infected sample group

N/A: Not applicable

Analysis of the size distribution of sRNA sequences in the 18 to 26 nt range showed the most abundant sequences to be between 21 and 24 nt in length, with sizes 21 nt and 24 nt as the major

classes (Figure 3.2). These results were consistent with those of other grapevine cultivars, as well as *Arabidopsis*, *Citrus trifoliate*, *Oryza sativa*, *Eugenia uniflora*, and *Glycine max* (Rajagopalan *et al.*, 2006; Zhu *et al.*, 2008; Pantaleo *et al.*, 2010; Song *et al.*, 2010; Song *et al.*, 2011; Guzman *et al.*, 2012). The library generated from the phytoplasma-infected samples indicated that 21 nt sRNAs were more abundant (34.2%) than those in the library obtained from the healthy plant samples (29.7%). A similar profile was observed for Mexican lime infected with '*Candidatus* Phytoplasma aurantifolia' (Ehya *et al.*, 2013). The 24 nt sRNAs, however, were more abundant in the library from the healthy plant samples (33.2%) compared to the library from the phytoplasma-infected samples (30.7%). This observation points to differences in complexity between the two pools of sRNAs that may infer an underlying miRNA-mediated regulatory response triggered by biotic stress. The unique (non-redundant) 21 nt reads were also more abundant in the phytoplasma-infected samples. Their length is characteristic of canonical miRNAs, and they possessed a high reads/unique reads ratio (Figure 3.2), reflecting their regulatory impact and abundance in plants. The 24 nt reads, which are predominantly repeat-associated siRNAs (rasiRNAs), exhibited the highest sequence diversity, consistent with the origin of this size class (Figure 3.2B) (Lelandais-Brière *et al.*, 2010).

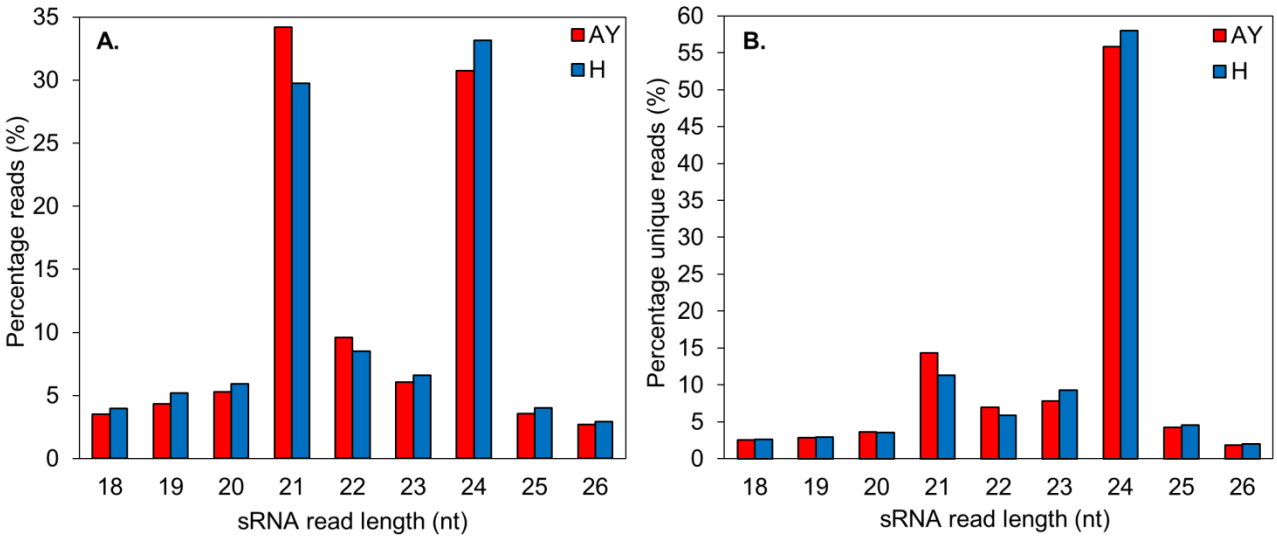

Figure 3.2: (A) The size distribution of the total 18 to 26 nt sRNA reads in the healthy (H) and AY phytoplasma-infected (AY) libraries. (B) The size distribution of the total 18 to 26 nt unique sRNA reads in the healthy (H) and AY phytoplasma-infected (AY) libraries.

3.4.3 Identification of known miRNAs and their sequence variants

We used sRNAbench v0.9 to detect both the canonical vvi-miRNA sequences (from miRBase), and all isomiRs present in the pooled phytoplasma-infected and healthy (control) read data, respectively.

The different sequences were classified and are presented in a simple table output (File S2). IsomiRs are defined as different sequence variants of known miRNAs that may arise from post-transcriptional modifications and alternative processing (Ebhardt *et al.*, 2009). IsomiR types included those reads having non-template additional nucleotides (where the read sequence starts and ends at the same position as the canonical sequence in the pre-microRNA, but shows sequence variation), "flush fitting" length variants (where the read sequence always starts or ends at the same position as the canonical sequence but a terminal trimming or extension is evident), and multiple length variants (where the read sequence does not coincide with either the 3' or 5' terminal nucleotides of the canonical sequence). Those reads that contained the same 5' terminal nucleotides as the canonical vvi-miR166b sequence, but showed divergence of length in their 3' terminal extension, as a result of alternative DCL1 cleavage, were the dominantly expressed isomiRs in the 'healthy libraries' (File S2). In the case of the 'AY phytoplasma-infected libraries', those reads that contained the same 3' terminal nucleotides as the canonical vvi-miR166e-5p sequence, but showed divergence of length in their 5' terminal extension, as a result of alternative DCL1 cleavage, were the most dominantly expressed isomiRs (File S2).The mechanism by which miRNA heterogeneity arises has been extensively reviewed. Different findings have suggested that multiple isomiRs that arose from a single miRNA locus are not randomly generated artefacts, but rather generated *in vivo* through biological relevant processes. Consequently, such sequence variations may drastically alter miRNA association with their targets, and also influence miRNA stability during Argonaute (AGO)-RISC loading (Chugh and Dittmer, 2012; Ameres and Zamore, 2013; Guo and Chen, 2014).

The vvi-miR166 family showed the highest levels of expression, but had no significant difference in terms of the total normalised read counts between the two different library types (File S2). Vvi-miR166b and its isomiRs constituted ~40% of the total normalised read counts in both library types. This high level of vvi-miR166 expression was also seen in a previous study where it was the most dominantly expressed miRNA family in all assayed grapevine tissues (Pantaleo *et al.*, 2010). A degradome sequencing approach revealed that vvi-miR166b regulates a Class III homeodomain leucine zipper (HDZIP-III) transcription factor which is involved in secondary cell wall biosynthesis (Velasco *et al.*, 2007; Carra *et al.*, 2009; Du and Wang, 2015). Direct evidence from the identification and analysis of corresponding activation tagged mutants has implicated the regulatory involvement of miR165/166 in leaf and vascular morphogenesis (Kim *et al.*, 2005; Sun, 2012).

3.4.4 Differential expression analysis of known miRNAs

Comparative profiling, with DESeq v2, between the healthy (control) and AY phytoplasma-infected samples was used to determine the differential expression of known miRNAs in the AY phytoplasma-infected material. Based on false discovery rate (FDR) for multiple testing, we encountered seven significantly differentially expressed known vvi-miRNA families that had $\log_2$-fold changes with adjusted *p*-values (q) ≤ 0.05. (Table 3.2; Figure 3.3).

Table 3.2: List of significantly differentially expressed known vvi-miRNAs.

Kown miRNA	Sequence (5'-3')	Length (nt)	Avg of normalised read counts†		DESeq results (H vs AY)		
			H	AY	$\log_2$FC	*p*-value	Adj. *p*-value (*q*-value)
vvi-miR156b,c,d¶	UGACAGAAGAGAGUGAGCAC	20	29.29	11.06	-1.18	0.0011	0.0102
vvi-miR159c	UUUGGAUUGAAGGGAGCUCUA	21	3869.83	8860.93	1.15	3.89E-05	0.0007
vvi-miR399g	UGCCAAAGGAGAUUUGCCCCU	21	463.24	103.94	-2.02	3.89E-05	0.0007
vvi-miR171a,c,d,I,j	UGAUUGAGCCGUGCCAAUAUC	21	20.56	37.71	0.87	0.0006	0.0071
vvi-miR172d	UGAGAAUCUUGAUGAUGCUGCAU	23	243.48	736.45	1.42	0.0007	0.0075
vvi-miR160c,d,e¶	UGCCUGGCUCCCUGUAUGCCA	21	14.23	30.9	1.00	0.0060	0.0477
vvi-miR2950-5p¶	UUCCAUCUCUUGCACACUGGA	21	22.15	69.15	1.59	3.26E-10	2.35E-08

H: Healthy sample group

AY: AY phytoplasma-infected sample group

¶Validated using real-time RT-qPCR

†Average of reads per million mapped reads (RPM) between three biological replicates

An additional nine known miRNAs from seven families, had $\log_2$-fold changes with significant *p*-values ($p \leq 0.05$), which indicate they may be of biological importance (File S3). A total of eight miRNA families, *viz.* vvi-miR159c, vvi-miR160c-e, vvi-miR171acdij, vvi-miR172d, vvi-miR2950-5p, vvi-miR319bcef, vvi-miR3627-5p, and vvi-miR395a-m, were up-regulated, and five, *viz.* vvi-miR156bcd, vvi-miR3629(a-3p, b-3p, c-5p), vvi-miR3638-5p, vvi-miR399aheg, vvi-miR479, were down-regulated (Table 3.2; File S3). The differential expression of conserved miRNA families (vvi-miR156, miR159, vvi-miR160, vvi-miR171, vvi-miR172, vvi-miR319), known to be involved in different aspects of plant development (Sunkar and Jagadeeswaran, 2012), make these potential candidates that play a role in the interactions leading to symptoms associated with GY.

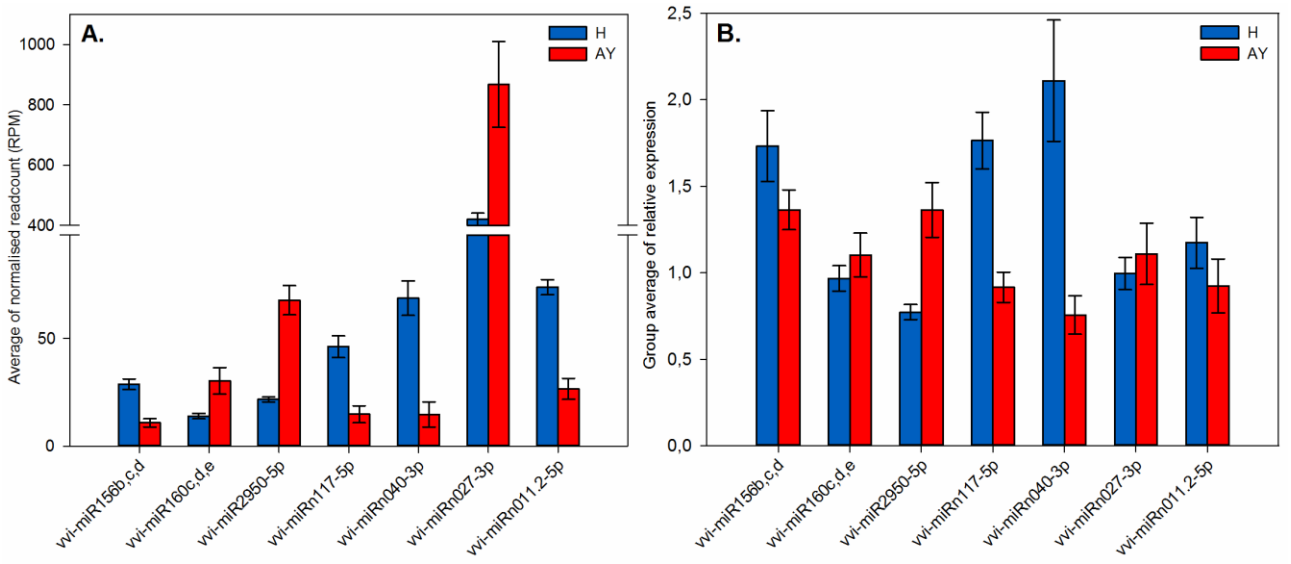

Figure 3.3: Bar charts displaying profiles of differentially expressed vvi-miRNAs ($q \leq 0.05$) in healthy (H) and AY phytoplasma-infected (AY) samples that were further validated. Vertical bars indicate the standard error (SE) of the mean. (A) Average normalised read counts of vvi-miRNAs. Group averages were given in terms of the average of reads per million mapped reads (RPM) for three biological replicates. (B) Relative expression analysis with real-time RT-qPCR, confirming expression profiles of vvi-miRNAs. Each bar represents the average of three biological replicates with three technical replicates.

3.4.5 Novel miRNA prediction and differential expression analysis

The pooled sRNA reads from all six libraries served as input for sRNAbench v0.9 and Shortstack v0.4.1 for predicting novel miRNAs. These sRNA sequences were aligned to the *V. vinifera* (PN40024) 12x assembled genome sequence to identify loci that may harbour potential pre-miRNA sequences, based on secondary structure and read distribution. Known *V. vinifera* pre-miRNA chromosomal locations found in miRBase v21 were flagged during each analysis to obtain unique precursor sequences that did not match these loci. Secondary fold structures were viewed using RNAfold and all miRNA precursors displayed appropriate stem-loop hairpin secondary structures (data not shown). Based on structural criteria described by Meyers *et al.* (2008), these miRNAs can be regarded as authentic candidates that adhere to biogenesis and expression criteria for confident miRNA annotation.

In total, 175 novel pre-miRNA sequences were predicted, each derived from a unique genome location (File S4). Three of the pre-miRNAs were predicted with both prediction pipelines. We also identified multiple pre-miRNAs that produce mature miRNAs with similar sequences, e.g. vvi-miRn024a to vvi-miRn024c. These miRNAs can be considered members of the same miRNA family (File S4). Likewise, vvi-miRn019a to vvi-miRn019g represents a larger family of duplicated miRNA paralogs with identical precursor and mature miRNA sequences (File S4).

Pre-vvi-miRn027, predicted with Shortstack, may serve as an example of a large precursor that could give rise to two different miRNA duplexes since the sRNAbench-predicted pre-vvi-miRn136 falls within its location (Figure 3.4; File S4). Precursor sequences ranged from 54 nt to 742 nt while mature miRNA sequences ranged from 20 nt to 25 nt in length, the majority being 21 nt. Most mature miRNA sequences started with an uracil at the first position, corroborating data described by Baumberger and Baulcombe (2005) that showed a preferential association of the AGO1 protein with sRNAs containing a 5'-terminal uracil. This may indicate an important characteristic for miRNA biogenesis through recognition of miRNA duplexes by RISC.

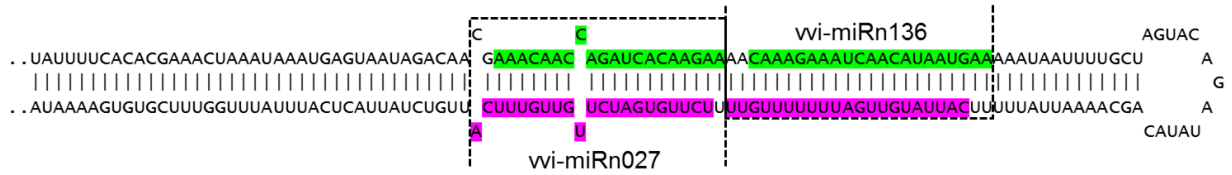

Figure 3.4: An example of a novel pre-miRNA hairpin structure that may give rise to two different miRNA duplexes (see File S4). The sequences highlighted in green and magenta represents the 5' and 3' mature miRNA sequences, respectively.

Zhang *et al.* (2006), implemented a criterion to better distinguish miRNAs from other sRNAs, known as MFEI which incorporates MFE, sequence length and GC content. The Unified Nucleic Acid Folding (UNAFold) software was used to calculate the MFE. MFE for predicted novel pre-miRNAs ranged from -13.2 kcal/mol to -428.9 kcal/mol and the MFEI ranged from -0.47 to -2.51. The majority (>94%) of the novel Shortstack-predicted pre-miRNAs possessed MFEI-values in accordance with the expected value (≥-0.85), while only 39% of sRNAbench-predicted pre-miRNAs had strong negative MFEI-values of more than -0.85 (File S4).

Each novel pre-miRNA sequence, as well as its most abundant mature miRNA sequence, was subjected to a homology-based search against miRBase using BLASTn. Results indicated that some of the newly identified pre-miRNAs have either precursor and/or mature sequences homologous to conserved miRNAs from other plant species, including grapevine miRNAs. Vvi-miRn025a and vvi-miRn025b, for example, share high homology with vvi-miR171 (File S4). We identified 17 novel vvi-miRNAs belonging to 13 miRNA families that can be classified as newly-identified members of known miRNAs, based on homology to known plant miRNAs for both precursor and mature sequences (File S4).

Differential expression analysis revealed that 10 novel miRNAs were significantly up-regulated, while 13 novel miRNAs were significantly down-regulated in sRNA libraries prepared from AY phytoplasma-infected leaf material (Table 3). We used a Perl script provided by Shen *et al.* (2012) to generate hairpin structures of these differentially expressed novel miRNAs (Figure S1). These images demonstrated complementary 5' and 3' mature miRNA sequences, each within a duplex that was possibly DCL1-derived.

Table 3.3: List of significantly differentially expressed novel vvi-miRNAs.

Novel miRNA	Sequence (5'-3')	Length (nt)	Avg of normalised read counts†		DESeq results (H vs AY)		
			H	AY	log_2FC	*p*-value	Adj. *p*-value (*q*-value)
vvi-miRn010.2-3p (miR529_new)	GCUGUACCCUCUCUCUUCCCC	21	8.71	2.20	-1.58	3.88E-07	5.39E-05
vvi-miRn025b/n025a-3p (miR171_new)	UGAUUGAGCCGUGCCAAUAUC	21	20.56	37.71	0.93	4.40E-07	5.39E-05
vvi-miRn011.2-5p¶ (miR391_new)	AGGAGAGAUGACGCCGUCGCC	21	75.28	27.16	-1.23	9.60E-07	7.84E-05
vvi-miRn133-5p	AGACUGGUAGAAAGAUUUAUA	21	19.36	2.73	-1.78	7.27E-06	4.45E-04
vvi-miRn140-3p	UCACCUUGUUGAGUGCCCGGU	21	6.48	1.59	-1.50	1.29E-05	6.31E-04
vvi-miRn040-3p¶	UGGGUUCAAAGUAGACAAUAUUUA	24	70.10	14.94	-1.57	2.09E-05	8.53E-04
vvi-miRn131-3p (miR399_new)	UGCCAAAGGAGAUUUGCCCCG	21	2.73	0.53	-1.56	4.22E-05	1.48E-03
vvi-miRn117-5p¶	UGGACCCUCAUGACUUUAAAAUGC	24	47.07	15.09	-1.29	6.07E-05	1.86E-03
vvi-miRn139-3p	GGGGGCUGACCUGUUGAAGAG	21	21.50	8.60	-1.04	0.0002	0.0045
vvi-miRn150-5p	UUUUUCAUGGUCUGAUUGAGC	21	15.97	36.25	1.11	0.0002	0.0045
vvi-miRn022b-5p (miR1446_new)	UCUGAACUCUCUCCCUCAUUGGC	23	0.76	2.45	1.35	0.0002	0.0045
vvi-miRn008.1-3p (miR169_new)	AGGCAGUCACCUUGGCUAACU	21	3.72	1.17	-1.22	0.0004	0.0081
vvi-miRn147-5p	UGGUGAACCAAAUAACUCUGG	21	33.29	63.81	0.93	0.0009	0.0174
vvi-miRn027-3p¶	UCUUGUGAUCUUGUUGUUUCA	21	420.78	867.56	0.99	0.0010	0.0174
vvi-miRn115-3p	AGGAAUGUGCUUCUUGGCAUA	21	6.45	1.84	-1.19	0.0016	0.0261
vvi-miRn070-3p	UAAGGACUAAAUUGGUAGACC	21	1.92	4.03	0.97	0.0022	0.0334
vvi-miRn089-5p	UACACAUGUAGUGCCAUCAUAUGA	24	53.03	16.67	-1.13	0.0025	0.0365
vvi-miRn007.1-3p	UGAUAUUAGCAGCUGAGAACA	21	7.19	3.71	-0.76	0.0032	0.0386
vvi-miRn003-5p	UUACACAGAGAGAUGACGGUGG	22	24.78	53.83	0.98	0.0031	0.0386
vvi-miRn051-5p	AGAGACCACCUAGUCCUGUUAAGA	24	31.20	19.81	-0.52	0.0029	0.0386
vvi-miRn129-5p	UUUUGGAACUAGAGUGCUUGC	21	1.34	2.89	0.98	0.0035	0.0410
vvi-miRn137 -5p	CAACAAUCUAAAUGAAACAUAGA	23	3.40	6.61	0.90	0.0043	0.0478
vvi-miRn022a-5p (miR1446_new)	UCUGAACUCUCUCCCUCAUGGC	22	8.36	15.35	0.85	0.0047	0.0497

H: Healthy sample group
AY: AY phytoplasma-infected sample group
¶Validated using real-time RT-qPCR
†Average of reads per million mapped reads (RPM) between three biological replicates
Known miRNA homologs are given in brackets

3.4.6 Validation of miRNA expression profiles by real-time RT-qPCR

A stem-loop RT-qPCR assay was applied to verify the results for the miRNA differential expression analysis. Primers sequences are listed in File S5. The relative expression of seven significantly differentially expressed miRNAs, three known (*viz.* miR156bcd, miR160cde, and

miR2950-5p) and four novel (*viz.* miRn011.2-5p, miRn040-3p, miRn117-5p, and miRn027-3p), was measured in healthy and phytoplasma-infected leaves using real-time RT-qPCR analysis (Figure 3.3B). We were also able to validate the expression of less significant known miRNAs ($q \leq 0.15$) (*viz.* miR319e, miR399e, and miR479) (File S3). This suggests that modulation of these miRNAs may hold biological importance. Although the non-conserved miRNAs (*viz.* vvi-miR479 and vvi-miR2950), and certain novel miRNAs were present at low levels, they were detected using real-time RT-qPCR. The trend of expression obtained from the RT-qPCR analysis was consistent with the average normalised read abundance observed in the sRNA-seq data (Figure 3.5; File S3). Since the expression of the novel miRNA candidates were confirmed using real-time RT-qPCR they can be tentatively classified as authentic miRNAs. The use of stable and robust degradome data, however, will provide us with more concrete evidence to confirm these results (German *et al.*, 2008).

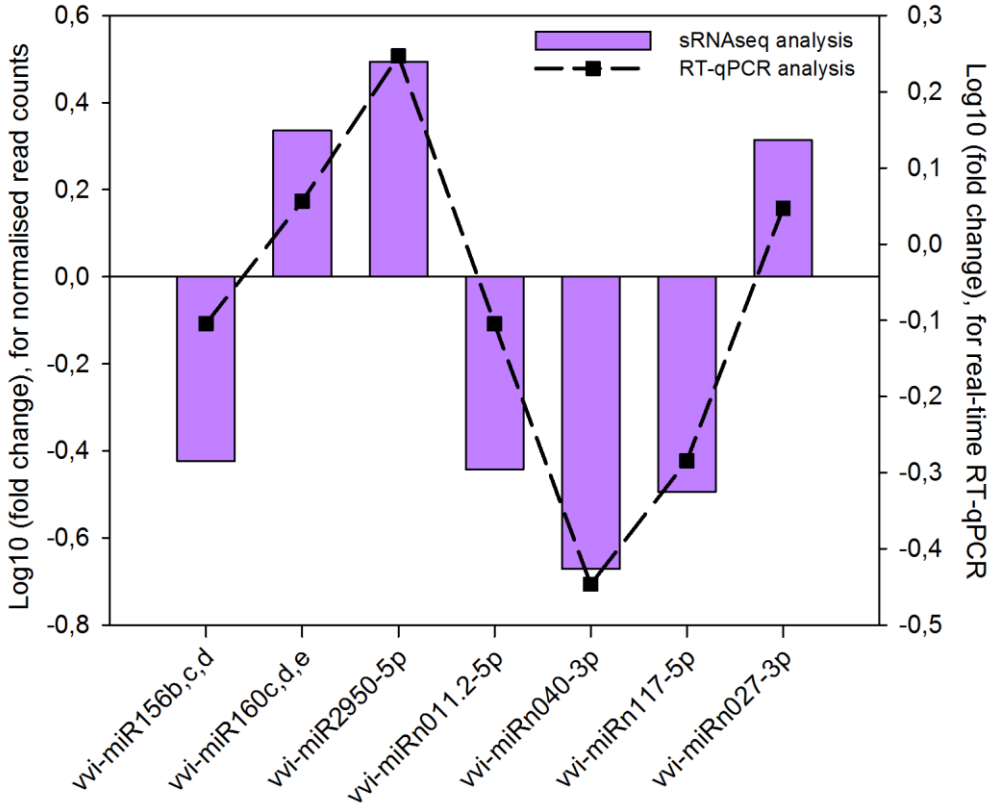

Figure 3.5: Correlation graph comparing average normalised read counts with real-time RT-qPCR results, thereby confirming vvi-miRNA expression patterns.

3.4.7 Identification of putative targets for differentially expressed known miRNAs

Over the past decade, increasing evidence have demonstrated how miRNAs can play an important role in modulating gene expression during plant-microorganism interactions (Khraiwesh *et al.*, 2012). It is important to consider miRNA target identification and validation in order to elucidate

the biological functions of miRNAs. Multiple 'Pinot noir' target mRNAs have been identified for known miRNAs using a high-throughput degradome sequencing approach (Pantaleo *et al.*, 2010). To gain further insight into the function of the differentially expressed miRNAs found in this study, we performed a complementary-based search with psRNAtarget to search for putative target-binding sites found in grapevine mRNAs. We adopted strict parameters, which provided perfect or near-perfect complementarity between a miRNA and its target, suggesting DCL1-cleavage or translational inhibition of miRNA-targeted mRNAs (Schwab *et al.*, 2005; Carrington and Ambros, 2003) (File S6).

In order to obtain a holistic view of biological pathways possibly influenced by miRNA-mediated regulation, *in silico* predicted targets for both the differentially expressed known and novel miRNAs were functionally annotated using Blast2GO v2.2.7. After GO analysis we found 71 functionally annotated putative targets for 15 of the known miRNAs and 54 functionally annotated putative targets for 17 of the novel miRNAs. For some of the targets, however, functional attributes could not be assigned, using default parameters within Blast2GO. Detailed annotation results are provided in File S6. A combined graph was generated and depicted different categories in which the targets grouped in terms of biological processes (Figure 3.6).

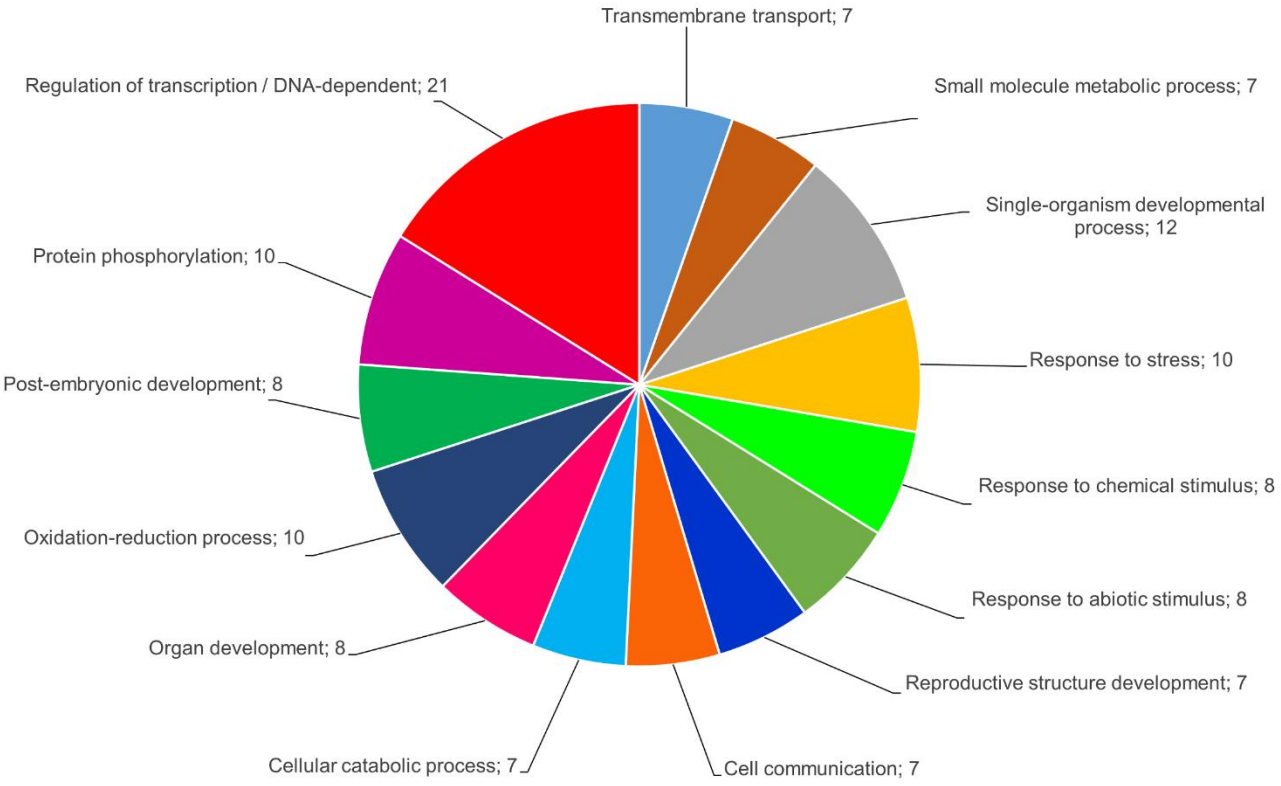

Figure 3.6: A combined graph depicting the main categories of putative vvi-miRNA targets grouped in terms of biological processes (GO level 3; annotation cut-off = 7.0).

There were 14 categories of which the four major processes included transcriptional regulation, developmental processes, response to stress, and metabolic processes that included phosphoregulation and oxidoreductase activity. This suggested that the differentially expressed miRNAs are involved in a broad range of physiological functions. Putative targets of conserved miRNA families, such as miR156, miR159, miR171 and miR399, identified in this study, correspond to targets found in numerous plant species, including several grapevine cultivars, while the predicted functions of these targets were also similar with previous findings (Pantaleo *et al.*, 2010; Mica *et al.*, 2010; Wang *et al.*, 2011a; Wang *et al.*, 2014; Belli Kullan *et al.*, 2015; Du and Wang, 2015) (File S6). Recent studies have revealed that several of these miRNA targets share common roles in the crosstalk between signalling pathways modulated by both biotic and abiotic stresses (Curaba *et al.*, 2014). Our results revealed that some of these target genes encode transcription factors, including squamosa-promoter binding protein (SPB)-box, MYB, NAC-domain, Scarecrow-like/GRAS-domain, AP2, HDZIP-III and bHLH transcription factors, previously reported for grapevine and other plant species (Kantar *et al.*, 2011; Gonzalez-Ibeas *et al.*, 2011; Colaiacovo *et al.*, 2010; Ozhuner *et al.*, 2013).

Phytoplasma-infection may cause miRNA-mediated changes in plant morphology and architecture

Plant morphological changes can be attributed to changes in the expression of certain transcription factors, as well as regulatory changes at a post-transcriptional and epigenetic level. The miR156/157 family, which is highly conserved in plants, can target numerous members of the SBP-box genes in *V. vinifera*. Evidence has shown that changes in the expression levels of these genes play a role in phase transition and reproductive development (Hou *et al.*, 2013; Huijser and Schmid, 2011). Studies on *Arabidopsis* and rice showed that cleavage of squamosa-promoter binding-like (SPL) proteins, due to miR156 overexpression, give rise to plants that are smaller, show delayed flowering and loss of apical dominance, initiate growth of more leaves with shorter plastochrons (in *Arabidopsis*) and causes reduced panicle size (in rice) (Xie *et al.*, 2006; Schwarz *et al.*, 2008). Likewise, miR156-overexpression in poplar (*Populus* spp.) caused an increase in leaf size and leaf initiation rate, and reduced apical dominance (Wang *et al.*, 2011b). The modification of leaf morphology due to regulation of SBP-box transcripts by miR156 overexpression was demonstrated in phytoplasma-infected Mexican lime trees, mulberry, and red date (Ehya *et al.*, 2013; Gai *et al.*, 2014; Shao *et al.*, 2016). Expression analysis in this study, however, revealed a significant decrease in abundance of certain vvi-miR156 members in the phytoplasma-infected samples, which cannot be explained at this point.

It has been shown that down-regulation of miR156 results in an increase in SPLs that promote juvenile to adult phase transition and flowering through activation of miR172 and MADS box genes in *Arabidopsis* (Wang *et al.*, 2008; Wu *et al.*, 2009). The *Arabidopsis AtSPL9*, can positively regulate the expression of miR172, demonstrating the presence of a miR156-AtSPL9-miR172 regulatory cascade (Chen *et al.*, 2010). It was proposed that the miR156-SPL-miR172 regulatory pathway was activated in mulberry in response to phytoplasma infection (Gai *et al.*, 2014). Higher levels of miR172 associated with viral pathogenesis in tomato leaf curl disease and grapevine leafroll disease have also been reported (Naqvi *et al.*, 2010; Alabi *et al.*, 2012). The APETALA2/Ethylene-responsive transcription factor (*AP2/ERF*)-like mRNA was identified as a possible target of vvi-miR172 in our study. The interaction between miR172 and *AP2/ERF*-like targets is well conserved and is known to be involved in transitions between developmental stages, regulating flowering time and specifying floral organ identity (Chen, 2004; Zhu and Helliwell, 2011). Differential expression of vvi-miR156 and vvi-miR172, leading to restricted phase transition, may lead to symptoms associated with GY such as abnormal leaf shape and size, as well as downward curling of leaves and flower abortion. It was suggested that expression changes of miR156 and miR172 may lead to development of green leaf-like structures instead of flowers, also referred to as phyllody, as well as flower sterility in phytoplasma-infected red date and mulberry (Gai *et al.*, 2014; Shao *et al.*, 2016).

Levels of vvi-miR159 and vvi-miR319 were also significantly higher in the AY phytoplasma-infected leaves, and *in silico* analysis predicted that they may target a *GAMYB*-like mRNA and a *R2R3-MYB* mRNA. Recent studies identified their association with plant disease during fungal and bacterial infection in *Arabidopsis* and *Populus trichocarpa*, respectively (Zhang *et al.*, 2011; Zhao *et al.*, 2012), and have experimentally validated their targets as being mRNAs encoding MYB transcription factors (Du *et al.*, 2012). MYB genes constitutes a large and widespread gene family in plants (estimated at 279 members in grapevine, of which 108 belong to the R2R3 family), and are involved in a variety of plant-specific functions including primary and secondary metabolism, cell fate and differentiation, developmental processes and responses to biotic and abiotic stresses (Dubos *et al.*, 2010; Galbiati *et al.*, 2011). Consequently, altered expression of miR159 and miR319 may also contribute to deformation of grapevine leaves (Naqvi *et al.*, 2010; Shao *et al.*, 2016).

Differential miRNA expression may lead to modulated auxin signalling

Disease symptoms caused by certain pathogens have also been described as a result of interference with plant hormone signalling that lead to the disturbance of plant defence responses. Phytoplasma diseases have been classified as ‘auxonic diseases’ which refers to possible interactions with the

auxin balance of the host (Musetti, 2010). Auxin, an important phytohormone, regulates many plant developmental processes, and its influence during pathogen resistance responses has been described (Kazan and Manners, 2009). A substantial increase of indole-3-acetic acid (IAA) was observed in phytoplasma-infected Mexican lime trees, possibly indicating susceptibility to the pathogen (Ehya *et al.*, 2013). Certain proteins, known as virulence effectors, are secreted by pathogens during infection and are known to modulate hormone and signalling pathways by altering gene transcription levels. AY-WB effectors, SAP11 and TENGU, are known to be unloaded from the phloem sieve cells to the target cell nuclei where they interact and destabilise certain transcription factors, resulting in severe changes in leaf morphology and increased susceptibility to phytoplasma insect vectors (Palatnik *et al.*, 2007; Hoshi *et al.*, 2009; Sugio *et al.*, 2011). Microarray analysis of transgenic *Arabidopsis* lines overexpressing TENGU demonstrated regulation of several auxin responsive genes and auxin efflux carrier genes. SAP11 destabilises Teosinte branched1/Cincinnata/Proliferating cell factor (TCP) transcription factors 1 and 2, known to be regulated by miR319 in *Arabidopsis*, resulting in the suppression of Jasmonate (JA) production that create favourable conditions for insect vector proliferation (Palatnik *et al.*, 2007; Hoshi *et al.*, 2009; Sugio *et al.*, 2011).

A group of miRNAs can promote plant defence responses by coordinate regulation of hormone signalling pathways in response to pathogen attack. Among them, miR160, miR167, miR390 and miR393 contribute to PTI by regulating the expression of genes encoding different auxin response factors (ARFs) and auxin receptors involved in auxin signalling, thereby promoting inhibition of pathogen growth (Zhang *et al.*, 2011). miR393 expression, induced by bacterial elicitor flg22, was the first shown to be implicated in the repression of auxin receptor genes in *Arabidopsis* (Navarro *et al.*, 2006). Our results showed that vvi-miR160, which may target *ARF* mRNAs, was significantly up-regulated in the AY phytoplasma-infected leaves. ARF transcription factors are known to regulate auxin-inducible genes by binding to elements in their auxin-responsive promoters to either activate or repress transcription (Hagen and Guilfoyle, 2002). Other instances where miR160 accumulated during biotic stress response were demonstrated in clubroot-infected *Brassica napus* root (Verma *et al.*, 2014), powdery mildew infection in wheat (Xin *et al.*, 2010), and phytoplasma-infected mulberry (Gai *et al.*, 2014).

Phytoplasma-responsive miRNA expression may play a role in nutrient homeostasis

The AY phytoplasma chromosome is extremely reduced and lacks many essential genes related to amino acid and fatty acid biosynthesis, the tricarboxylic acid cycle and oxidative phosphorylation. This suggested that phytoplasmas have evolved as intracellular parasites in nutrient-rich host

environments and therefore possess multiple transporter genes in order to assimilate important mineral nutrients for their survival (Oshima *et al.*, 2004). Several plant miRNAs have been reported for their role in nutrient homeostasis in response to deficiencies of phosphate, nitrogen, sulphur, and copper (Shriram *et al.*, 2016). A few of these, including vvi-miR395 and vvi-miR399, were differentially expressed in the present study, possibly in response to AY phytoplasma-infection. miR395 is known to target members of the *ATP-sulphurylase* (*ATPS*) gene family and a low-affinity sulphate transporter gene *SULTR2;1*, both crucial for regulating sulphate homeostasis in *Arabidopsis* (Matthewman *et al.*, 2012) (File S6). The induction of miR395 levels leads to sulphate accumulation in the leaves due to increased translocation from the roots (Liang *et al.*, 2010; Kawashima *et al.*, 2011). In this study miR395 was up-regulated in the AY phytoplasma-infected leaves, and may contribute to favourable conditions for pathogen growth. Alternatively, sulphur starvation can cause physiological imbalances, impaired plant growth, and reduced plasticity against environmental changes and pathogen attack (Jagadeeswaran *et al.*, 2014).

The role of miR399 in the maintenance of phosphate homeostasis has been well characterised. It is involved in the regulation and allocation of inorganic phosphate (Pi) from the roots to the shoots as well as remobilisation from old to young leaves (Chiou *et al.*, 2006; Pant *et al.*, 2008). miR399 positively regulates Pi uptake and translocation by down-regulating *PHO2*, which encodes a ubiquitin-conjugating E2 enzyme, UBC24 (Chiou *et al.*, 2006; Hsieh *et al.*, 2009). *PHO2*, on the other hand, acts as a negative regulator by suppressing these activities when external Pi is ample, thereby preventing phosphate toxicity. Our results revealed significant down-regulation of vvi-miR399 in AY phytoplasma-infected leaves. Interestingly, lower levels of miR399 was also found in phytoplasma-infected material of Mexican lime trees, mulberry, and red date (Ehya *et al.*, 2013; Gai *et al.*, 2014; Shao *et al.*, 2016). An adequate supply of Pi is required for optimal growth and reproduction due to its involvement in essential plant functions, including energy transfer, photosynthesis, enzyme regulation, metabolite transport and nucleic acid synthesis. Therefore, the down-regulation of miR399 may cause suppression of Pi uptake, further contributing to disease symptom development.

3.4.8 Identification of putative targets for differentially expressed novel miRNAs

In addition to the targets of known miRNA, we also predicted possible targets for the in silico predicted novel miRNAs that were significantly differentially expressed in the AY phytoplasma-infected leaves (File S6). Some of these target mRNAs encode certain transcription factors, such as Scarecrow-like/GRAS-domain protein, TPR-like protein, MADS-box protein, bHLH-like protein, and a NAC-domain protein. We also identified targets that encode proteins involved in hydrolase

activity, e.g. ARM repeat superfamily isoform 2-like protein, beta-fructofuranosidase, glucan endo-1,3-beta-glucosidase, and a calcineurin-like metallo-phosphoesterase. Receptor-like kinase (RLK) proteins that are involved in signal transduction, such as a G-type lectin S-receptor-like serine/threonine-protein kinase, a leucine-rich repeat (LRR) receptor EXS-like kinase, a disease resistance At3g14460-like protein, a RLK HSL1-like protein, and a LRR receptor-like serine/threonine At4g08850-like kinase were also identified.

Some signal transduction proteins are surface-located, transmembrane receptor molecules that are activated by external stimuli, such as plant hormones and pathogens. These, in turn, are sequentially transmitted to initiate complex downstream signalling pathways that induce PAMP -and effector-triggered immunity (PTI and ETI) and/or hypersensitive cell death resistance responses. The majority of these innate immune receptors are proteins that contain a nucleotide-binding site (NBS) and leucine-rich repeats (LRR) that are encoded by resistance (*R*) genes (Rafiqi *et al.*, 2009). Another versatile function of certain miRNAs is targeting diverse members of NBS-LRRs which are then processed by RNA-dependent RNA polymerase 6 (RDR6) to dsRNA and then cleaved by DCL4 to produce phased, secondary siRNAs (phasiRNAs) (Fei *et al.*, 2013). This was demonstrated in resistant *Solanum lycopersicum* (tomato) and *Gossypium raimondii* (cotton) where the miR482-mediated silencing cascade was suppressed in pathogen-infected plants so that certain NBS-LRRs were up-regulated to confer resistance (Shivaprasad *et al.*, 2012; Zhu *et al.*, 2014). Interestingly, vvi-miRn027, which may target a disease resistance mRNA (*GSVIVG01027229001*), was severely increased in AY phytoplasma-infected leaves in comparison to the other novel miRNAs that might target RLK mRNAs. This miRNA and its putative target may serve as potential candidates in transient expression studies to investigate an underlying defence response to AY phytoplasma.

Furthermore, most of the other novel miRNAs were expressed at a lower abundance than that of conserved miRNAs and are likely to be grapevine-specific miRNAs, which may be classified into non-conserved miRNAs. It can be suspected that they are likely candidates involved in developmental, metabolic and transmembrane transport processes as proposed by the gene ontology results, but it would require additional experimental approaches to address these hypotheses.

3.5 Conclusions

In summary, our study employed different computational tools to provide the first report on the identification of differentially expressed miRNAs in grapevine leaves infected with AY phytoplasma. In addition to known vvi-miRNAs, we detected a large group of putative novel

miRNAs by utilizing two different analysis pipelines. Some of the novel miRNAs shared a high degree of homology with other known plant miRNAs, and were therefore classified as newly-identified members of existing miRNA families. Further experimentation concerning the regulation of their target mRNA(s), however, would be required to confirm this.

Differential expression analysis was done via comparative miRNA profiling between sRNA libraries constructed from healthy control plants and plants diagnosed with AY phytoplasma, respectively. Changes in the expression of various miRNAs were clearly observed in the diseased group, possibly modulated in response to biotic stress. The relative expression of certain known and novel miRNAs was determined with real-time RT-qPCR analysis, thereby demonstrating a similar trend in expression regarding the normalised sRNA read data. There is increasing evidence for the involvement of miRNAs in plant-microorganism interaction and how they mediate gene expression related to pathogenesis.

In order to identify potential miRNA targets, we applied a simple complementary-based, *in silico* approach with psRNAtarget. This method relies on perfect or near-perfect complementarity of plant miRNAs with their target(s), known to facilitate gene regulation through mRNA cleavage or translational inhibition. To further validate grapevine-specific miRNAs and the mRNAs they target would require the use of stable and robust degradome sequencing data that would assist in the elucidation of different modes of regulation in a tissue-specific and developmental stage-specific manner. Target mRNAs regulated by translational inhibition, however, would be undetectable in degradome data. Furthermore, high-throughput gene expression profiling techniques such as microarray-hybridisation analysis and RNAseq/transcriptome analysis would allow us to observe expression levels of miRNAs and their anti-correlated target mRNAs.

The miRNA expression patterns observed in AY phytoplasma-infected grapevine leaves, followed by putative miRNA target description and annotation, led us to believe that our results were compatible with evidence of perturbations found in other pathogen-infected plants. Putative miRNA target predictions indicated the involvement of miRNA pathways that may influence plant development and morphology either directly or by auxin imbalance. We also identified targets involved in nutrient homeostasis, as well as a few important novel miRNA targets involved in signal transduction, which may hold the key to activating pathogen-resistance pathways in grapevine. Taken together, our findings suggest some hypothetical associations between miRNAs and certain physiological changes that may be crucial in understanding disease symptom development in AY phytoplasma-infected grapevines. Further investigations of these miRNA-

mediated pathways may shed new light on the roles and mechanisms of miRNAs in plant pathogenesis.

3.6 Supporting information

Figure S1: Hairpin structures of differentially expressed novel vvi-miRNAs.
(Document available online: https://doi.org/10.1371/journal.pone.0182629.s001)

File S1: Virus RT-PCR primer list.
(Document available online: https://doi.org/10.1371/journal.pone.0182629.s002)

File S2: IsomiR summary for each of the mature vvi-miRNAs detected in the healthy and AY phytoplasma-infected libraries.
(Document available online: https://doi.org/10.1371/journal.pone.0182629.s003)

File S3: Additional differentially expressed known miRNAs.
(Document available online: https://doi.org/10.1371/journal.pone.0182629.s004)

File S4: Putative novel vvi-miRNAs.
(Document available online: https://doi.org/10.1371/journal.pone.0182629.s005)

File S5: miRNA real-time RT-qPCR primer list.
(Document available online: https://doi.org/10.1371/journal.pone.0182629.s006)

File S6: psRNATarget AND GO analysis results.
(Document available online: https://doi.org/10.1371/journal.pone.0182629.s007)

3.7 Acknowledgements

This work is based upon the research supported by the National Research Foundation (NRF) of South Africa. The grant holder (UID 78073) acknowledges that opinions, findings and conclusions or recommendations expressed in any publication generated by the NRF-supported research are those of the author(s) and the NRF accepts no liability whatsoever in this regard. Furthermore, the authors are also very appreciative towards the Technology Innovation Agency (TIA) for financial support for the duration of this study. We would also like to express our gratitude to Anelda van der Walt, Marike Visser and Beatrix Coetzee (Department of Genetics, Stellenbosch University), as well as Charl Möller (IT, Stellenbosch University), for their generous contributions and/or technical assistance regarding the computational work.

3.8 References

Alabi O.J., Zheng Y., Jagadeeswaran G., Sunkar R. and Naidu R.A. 2012. High-throughput sequence analysis of small RNAs in grapevine (*Vitis vinifera* L.) affected by grapevine leafroll disease. *Molecular Plant Pathology* 13(9): 1060-1076.

Albertazzi G., Milc J., Caffagni A., Francia E., Roncaglia E., *et al.* 2009. Gene expression in grapevine cultivars in response to Bois Noir phytoplasma infection. *Plant Science* 176(6): 792-804.

Allen E., Xie Z., Gustafson A.M. and Carrington J.C. 2005. microRNA-directed phasing during trans-acting siRNA biogenesis in plants. *Cell* 121(2): 207-221.

Altschul S.F., Gish W., Miller W., Myers E.W. and Lipman D.J. 1990. Basic local alignment search tool. *Journal of Molecular Biology* 215(3): 403-410.

Ameres S.L. and Zamore P.D. 2013. Diversifying microRNA sequence and function. *Nature Reviews Molecular Cell Biology* 14(8): 475-488.

Anders S. and Huber W. 2010. Differential expression analysis for sequence count data. *Genome Biology* 11(10): R106.

Axtell M.J. 2013. ShortStack: Comprehensive annotation and quantification of small RNA genes. *RNA* 19(6): 740-751.

Barturen G., Rueda A., Hamberg M., Alganza A., Lebron R., *et al.* 2014. sRNAbench: profiling of small RNAs and its sequence variants in single or multi-species high-throughput experiments. *Methods in Next Generation Sequencing* 1(1): 21-31.

Baumberger N. and Baulcombe D.C. 2005. *Arabidopsis* ARGONAUTE1 is an RNA Slicer that selectively recruits microRNAs and short interfering RNAs. *Proceedings of the National Academy of Sciences* 102(33): 11928-11933.

Belli G., Bianco P.A. and Conti M. 2010. Grapevine yellows in Italy: past, present and future. *Journal of Plant Pathology* 92(2): 303-326.

Belli Kullan J., Lopes Paim Pinto D., Bertolini E., Fasoli M., Zenoni S., *et al.* 2015. miRVine: a microRNA expression atlas of grapevine based on small RNA sequencing. *BMC Genomics* 16(1): 1-23.

Budak H. and Akpinar B.A. 2015. Plant miRNAs: biogenesis, organization and origins. *Functional & Integrative Genomics* 15(5): 523-531.

Camacho C., Coulouris G., Avagyan V., Ma N., Papadopoulos J., *et al.* 2009. BLAST+: architecture and applications. *BMC Bioinformatics* 10(1): 421.

Carra A., Mica E., Gambino G., Pindo M., Moser C., *et al.* 2009. Cloning and characterization of small non-coding RNAs from grape. *The Plant Journal* 59(5): 750-763.

Carrington J.C. and Ambros V. 2003. Role of MicroRNAs in Plant and Animal Development. *Science* 301(5631): 336-338.

Carstens R., Petersen Y., Stephan D. and Burger J.T. 2011. Current status of Aster yellows disease in infected vineyards in the Vredendal grape producing area of South Africa. *Phytopathogenic Mollicutes* 1(2): 83-85.

Chen C., Ridzon D.A., Broomer A.J., Zhou Z., Lee D.H., *et al.* 2005. Real-time quantification of microRNAs by stem-loop RT-PCR. *Nucleic Acids Research* 33(20): e179-e179.

Chen X. 2004. A microRNA as a translational repressor of APETALA2 in *Arabidopsis* flower development. *Science* 303(5666): 2022-2025.

Chen X., Zhang Z., Liu D., Zhang K., Li A., *et al.* 2010. SQUAMOSA Promoter-Binding Protein-Like Transcription Factors: Star Players for Plant Growth and Development. *Journal of Integrative Plant Biology* 52(11): 946-951.

Chiou T.-J., Aung K., Lin S.-I., Wu C.-C., Chiang S.-F., *et al.* 2006. Regulation of Phosphate Homeostasis by MicroRNA in *Arabidopsis*. *The Plant Cell* 18(2): 412-421.

Chugh P. and Dittmer D.P. 2012. Potential pitfalls in microRNA profiling. *Wiley Interdiscip Reviews: RNA* 3(5): 601-616.

Colaiacovo M., Subacchi A., Bagnaresi P., Lamontanara A., Cattivelli L., *et al.* 2010. A computational-based update on microRNAs and their targets in barley (*Hordeum vulgare* L.). *BMC Genomics* 11(1): 595.

Conesa A. and Götz S. 2008. Blast2GO: A comprehensive suite for functional analysis in plant genomics. *International Journal of Plant Genomics* 2008: 619832.

Contaldo N., Bertaccini A., Paltrinieri S., Windsor H.M. and Windsor G.D. 2012. Axenic culture of plant pathogenic phytoplasmas. *Phytopathologia Mediterranea* 51(3): 607-617.

Curaba J., Singh M.B. and Bhalla P.L. 2014. miRNAs in the crosstalk between phytohormone signalling pathways. *Journal of Experimental Botany* 65(6): 1425–1438.

Dai X. and Zhao P.X. 2011. psRNATarget: a plant small RNA target analysis server. *Nucleic Acids Research* 39: W155-W159.

Du H., Feng B.-R., Yang S.-S., Huang Y.-B. and Tang Y.-X. 2012. The R2R3-MYB transcription factor gene family in maize. *PLoS ONE* 7(6): e37463.

Du Q. and Wang H. 2015. The role of HD-ZIP III transcription factors and miR165/166 in vascular development and secondary cell wall formation. *Plant Signaling & Behavior* 10(10): e1078955.

Dubos C., Stracke R., Grotewold E., Weisshaar B., Martin C., *et al.* 2010. MYB transcription factors in *Arabidopsis*. *Trends in Plant Science* 15(10): 573-581.

Ebhardt H.A., Tsang H.H., Dai D.C., Liu Y., Bostan B., *et al.* 2009. Meta-analysis of small RNA-sequencing errors reveals ubiquitous post-transcriptional RNA modifications. *Nucleic Acids Research* 37(8): 2461-2470.

Ehya F., Monavarfeshani A., Mohseni Fard E., Karimi Farsad L., Khayam Nekouei M., *et al.* 2013. Phytoplasma-responsive microRNAs modulate hormonal, nutritional, and stress signalling pathways in Mexican lime trees. *PLoS ONE* 8(6): e66372.

Engelbrecht M., Joubert J. and Burger J.T. 2010. First report of aster yellows phytoplasma in grapevines in South Africa. *Plant Disease* 94(3): 373-373.

Fei Q., Xia R. and Meyers B.C. 2013. Phased, secondary, small interfering RNAs in posttranscriptional regulatory networks. *Plant Cell* 25(7): 2400-2415.

Gai Y.-P., Li Y.-Q., Guo F.-Y., Yuan C.-Z., Mo Y.-Y., *et al.* 2014. Analysis of phytoplasma-responsive sRNAs provide insight into the pathogenic mechanisms of mulberry yellow dwarf disease. *Scientific Reports* 4: 5378.

Galbiati M., Matus J.T., Francia P., Rusconi F., Cañón P., *et al.* 2011. The grapevine guard cell-related VvMYB60 transcription factor is involved in the regulation of stomatal activity and is differentially expressed in response to ABA and osmotic stress. *BMC Plant Biology* 11(1): 1-15.

German M.A., 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 Biotech* 26(8): 941-946.

Gonzalez-Ibeas D., Blanca J., Donaire L., Saladié M., Mascarell-Creus A., *et al.* 2011. Analysis of the melon (*Cucumis melo*) small RNAome by high-throughput pyrosequencing. *BMC Genomics* 12(1): 393.

Gruber A.R., Lorenz R., Bernhart S.H., Neubock R. and Hofacker I.L. 2008. The Vienna RNA websuite. *Nucleic Acids Research* 36: W70-W74.

Gundersen D.E. and Lee I.M. 1996. Ultrasensitive detection of phytoplasmas by nested-PCR assays using two universal primer pairs. *Phytopathologia Mediterranea* 35(3): 144-151.

Guo L. and Chen F. 2014. A challenge for miRNA: multiple isomiRs in miRNAomics. *Gene* 544(1): 1-7.

Guzman F., Almerão M.P., Körbes A.P., Loss-Morais G. and Margis R. 2012. Identification of microRNAs from *Eugenia uniflora* by high-throughput sequencing and bioinformatics analysis. *PLoS ONE* 7(11): e49811.

Hackenberg M., Rodriguez-Ezpeleta N. and Aransay A.M. 2011. miRanalyzer: an update on the detection and analysis of microRNAs in high-throughput sequencing experiments. *Nucleic Acids Research* 39: W132-W138.

Hagen G. and Guilfoyle T. 2002. Auxin-responsive gene expression: genes, promoters and regulatory factors. *Plant Molecular Biology* 49(3): 373-385.

Hellemans J., Mortier G., De Paepe A., Speleman F. and Vandesompele J. 2007. qBase relative quantification framework and software for management and automated analysis of real-time quantitative PCR data. *Genome Biology* 8: R19.

Hogenhout S.A., Oshima K., Ammar E.-D., Kakizawa S., Kingdom H.N., *et al.* 2008. Phytoplasmas: bacteria that manipulate plants and insects. *Molecular Plant Pathology* 9(4): 403-423.

Hoshi A., Oshima K., Kakizawa S., Ishii Y., Ozeki J., *et al.* 2009. A unique virulence factor for proliferation and dwarfism in plants identified from a phytopathogenic bacterium. *Proceedings of the National Academy of Sciences of the United States of America* 106(15): 6416-6421.

Hou H., Li J., Gao M., Singer S.D., Wang H., *et al.* 2013. Genomic organization, phylogenetic comparison and differential expression of the SBP-box family genes in grape. *PLoS ONE* 8(3): e59358.

Hren M., Nikolić P., Rotter A., Blejec A., Terrier N., *et al.* 2009. 'Bois noir' phytoplasma induces significant reprogramming of the leaf transcriptome in the field grown grapevine. *BMC Genomics* 10(1): 460.

Hsieh L.-C., Lin S.-I., Shih A.C.-C., Chen J.-W., Lin W.-Y., *et al.* 2009. Uncovering small RNA-mediated responses to phosphate deficiency in *Arabidopsis* by deep sequencing. *Plant Physiology* 151(4): 2120.

Huijser P. and Schmid M. 2011. The control of developmental phase transitions in plants. *Development* 138(19): 4117-4129.

Jagadeeswaran G., Saini A. and Sunkar R. 2009. Biotic and abiotic stress down-regulate miR398 expression in *Arabidopsis*. *Planta* 229(4): 1009-1014.

Jaillon O., Aury J.-M., Noel B., Policriti A., Clepet C., *et al.* 2007. The grapevine genome sequence suggests ancestral hexaploidization in major angiosperm phyla. *Nature* 449(7161): 463-467.

Kantar M., Lucas S.J. and Budak H. 2011. miRNA expression patterns of *Triticum dicoccoides* in response to shock drought stress. *Planta* 233(3): 471-484.

Kawashima C.G., Matthewman C.A., Huang S., Lee B.-R., Yoshimoto N., *et al.* 2011. Interplay of SLIM1 and miR395 in the regulation of sulfate assimilation in *Arabidopsis*. *The Plant Journal* 66(5): 863-876.

Kazan K. and Manners J.M. 2009. Linking development to defense: auxin in plant pathogen interactions. *Trends in Plant Science* 14(7): 373-382.

Khraiwesh B., Zhu J.-K. and Zhu J. 2012. Role of miRNAs and siRNAs in biotic and abiotic stress responses of plants. *Biochimica et Biophysica Acta (BBA) - Gene Regulatory Mechanisms* 1819(2): 137-148.

Kim J., Jung J.-H., Reyes J.L., Kim Y.-S., Kim S.-Y., *et al.* 2005. microRNA-directed cleavage of ATHB15 mRNA regulates vascular development in *Arabidopsis* inflorescence stems. *The Plant Journal* 42(1): 84-94.

Kozomara A. and Griffiths-Jones S. 2011. miRBase: integrating microRNA annotation and deep-sequencing data. *Nucleic Acids Research* 39: D152-D157.

Krüger K., De Klerk A., Douglas-Smit N., Joubert J., Pietersen G., *et al.* 2011. Aster yellows phytoplasma in grapevines: identification of vectors in South Africa. *Bulletin of Insectology* 64: S137-S1387.

Langmead B., Trapnell C., Pop M. and Salzberg S.L. 2009. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biology* 10(3): R25.

Lee I.M. 2004. '*Candidatus* Phytoplasma asteris', a novel phytoplasma taxon associated with aster yellows and related diseases. *International Journal of Systematic and Evolutionary Microbiology* 54(4): 1037-1048.

Lee I.-M., Davis R.E. and Gundersen-Rindal D.E. 2000. Phytoplasma: Phytopathogenic Mollicutes 1. *Annual Review of Microbiology* 54(1): 221-255.

Lelandais-Brière C., Sorin C., Declerck M., Benslimane A., Crespi M., *et al.* 2010. Small RNA diversity in plants and its impact in development. *Current Genomics* 11(1): 14-23.

Liang G., Yang F. and Yu D. 2010. MicroRNA395 mediates regulation of sulfate accumulation and allocation in *Arabidopsis thaliana*. *The Plant Journal* 62(6): 1046-1057.

Liu L.Y.D., Tseng H.I., Lin C.P., Lin Y.Y., Huang Y.H., *et al.* 2014. High-throughput transcriptome analysis of the leafy flower transition of *Catharanthus roseus* induced by peanut witches'-broom phytoplasma infection. *Plant and Cell Physiology* 55(5): 942-957.

Maixner M. Grapevine yellows - current developments and unsolved questions. 15th meeting of the international council for the study of virus and virus-like diseases of the grapevine (ICVG), 2006. Stellenbosch, South Africa. 86-88.

Margaria P., Abbà S. and Palmano S. 2013. Novel aspects of grapevine response to phytoplasma infection investigated by a proteomic and phospho-proteomic approach with data integration into functional networks. *BMC Genomics* 14(1): 38.

Margaria P. and Palmano S. 2010. Response of the *Vitis vinifera* L. cv. 'Nebbiolo' proteome to Flavescence dorée phytoplasma infection. *Proteomics* 11(2): 212-224.

Markham N.R. and Zuker M. 2008. UNAFold. *Bioinformatics.* Springer Science + Business Media.

Matthewman C.A., Kawashima C.G., Húska D., Csorba T., Dalmay T., *et al.* 2012. miR395 is a general component of the sulfate assimilation regulatory network in *Arabidopsis*. *FEBS Letters* 586(19): 3242-3248.

Meyers B.C., Axtell M.J., Bartel B., Bartel D.P., Baulcombe D., *et al.* 2008. Criteria for annotation of plant microRNAs. *THE PLANT CELL ONLINE* 20(12): 3186-3190.

Mica E., Piccolo V., Delledonne M., Ferrarini A., Pezzotti M., *et al.* 2010. Correction: High throughput approaches reveal splicing of primary microRNA transcripts and tissue specific expression of mature microRNAs in *Vitis vinifera*. *BMC Genomics* 11(1): 109.

Monavarfeshani A., Mirzaei M., Sarhadi E., Amirkhani A., Khayam Nekouei M., *et al.* 2013. Shotgun proteomic analysis of the Mexican lime tree infected with '*Candidatus* Phytoplasma aurantifolia'. *Journal of Proteome Research* 12(2): 785-795.

Mou H.-Q., Lu J., Zhu S.-F., Lin C.-L., Tian G.-Z., *et al.* 2013. Transcriptomic analysis of Paulownia infected by Paulownia witches'-broom phytoplasma. *PLoS ONE* 8(10): e77217.

Musetti R. 2010. Biochemical changes in plants infected by phytoplasmas. *In: PHYTOPLASMAS: Genomes, Plant Hosts and Vectors.* WEINTRAUB, P. G. (ed.). 138-146.

Naqvi A.R., Haq Q.M. and Mukherjee S.K. 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. *Virology Journal* 7(1): 1-16.

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(5772): 436-439.

Oshima K., Kakizawa S., Nishigawa H., Jung H.-Y., Wei W., *et al.* 2004. Reductive evolution suggested from the complete genome sequence of a plant-pathogenic phytoplasma. *Nature Genetics* 36(1): 27-29.

Ozhuner E., Eldem V., Ipek A., Okay S., Sakcali S., *et al.* 2013. Boron stress responsive microRNAs and their targets in barley. *PLoS ONE* 8(3): e59543.

Padmanabhan C., Zhang X. and Jin H. 2009. Host small RNAs are big contributors to plant innate immunity. *Current Opinion in Plant Biology* 12(4): 465-472.

Palatnik J.F., Wollmann H., Schommer C., Schwab R., Boisbouvier J., *et al.* 2007. Sequence and expression differences underlie functional specialization of *Arabidopsis* microRNAs miR159 and miR319. *Developmental Cell* 13(1): 115-125.

Pant B.D., Buhtz A., Kehr J. and Scheible W.-R. 2008. MicroRNA399 is a long-distance signal for the regulation of plant phosphate homeostasis. *The Plant Journal* 53(5): 731-738.

Pantaleo V., Szittya G., Moxon S., Miozzi L., Moulton V., *et al.* 2010. Identification of grapevine microRNAs and their targets using high-throughput sequencing and degradome analysis. *The Plant Journal*: no-no.

Rafiqi M., Bernoux M., Ellis J.G. and Dodds P.N. 2009. In the trenches of plant pathogen recognition: Role of NB-LRR proteins. *Seminars in Cell and Developmental Biology* 20: 1017-1024.

Rajagopalan R., Vaucheret H., Trejo J. and Bartel D.P. 2006. A diverse and evolutionarily fluid set of microRNAs in *Arabidopsis thaliana*. *Genes & Development* 20(24): 3407-3425.

Ruiz-Ferrer V. and Voinnet O. 2009. Roles of plant small RNAs in biotic stress responses. *Annual Review of Plant Biology* 60(1): 485-510.

Schwab R., Palatnik J.F., Riester M., Schommer C., Schmid M., *et al.* 2005. Specific effects of microRNAs on the plant transcriptome. *Developmental Cell* 8(4): 517-527.

Schwarz S., Grande A.V., Bujdoso N., Saedler H. and Huijser P. 2008. The microRNA regulated SBP-box genes SPL9 and SPL15 control shoot maturation in *Arabidopsis*. *Plant Molecular Biology* 67(1): 183-195.

Shao F., Zhang Q., Liu H., Lu S. and Qiu D. 2016. Genome-wide identification and analysis of microRNAs involved in witches'-broom phytoplasma response in *Ziziphus jujuba*. *PLoS ONE* 11(11): e0166099.

Shen W., Chen M., Wei G. and Li Y. 2012. MicroRNA prediction using a fixed-order Markov model based on the secondary structure pattern. *PLoS ONE* 7(10): e48236.

Shivaprasad P.V., Chen H.-M., Patel K., Bond D.M., Santos B.a.C.M., *et al.* 2012. A microRNA superfamily regulates nucleotide binding site-leucine-rich repeats and other mRNAs. *The Plant Cell* 24(3): 859-874.

Shriram V., Kumar V., Devarumath R.M., Khare T.S. and Wani S.H. 2016. MicroRNAs as potential targets for abiotic stress tolerance in plants. *Frontiers in Plant Science* 7: 817.

Song C., Wang C., Zhang C., Korir N., Yu H., *et al.* 2010. Deep sequencing discovery of novel and conserved microRNAs in trifoliate orange (*Citrus trifoliata*). *BMC Genomics* 11(1): 431.

Song Q.-X., Liu Y.-F., Hu X.-Y., Zhang W.-K., Ma B., *et al.* 2011. Identification of miRNAs and their target genes in developing soybean seeds by deep sequencing. *BMC Plant Biology* 11(1): 5.

Sugio A., Maclean A.M., Kingdom H.N., Grieve V.M., Manimekalai R., *et al.* 2011. Diverse targets of phytoplasma effectors: from plant development to defense against insects. *Annual Review of Phytopathology* 49: 175-195.

Sun G. 2012. MicroRNAs and their diverse functions in plants. *Plant Molecular Biology* 80(1): 17-36.

Sunkar R., Li Y.-F. and Jagadeeswaran G. 2012. Functions of microRNAs in plant stress responses. *Trends in Plant Science* 17(4): 196-203.

Vaucheret H. 2006. Post-transcriptional small RNA pathways in plants: mechanisms and regulations. *Genes & Development* 20(7): 759-771.

Velasco R., Zharkikh A., Troggio M., Cartwright D.A., Cestaro A., *et al.* 2007. A high quality draft consensus sequence of the genome of a heterozygous grapevine variety. *PLoS ONE* 2(12): e1326.

Verma S.S., Rahman M.H., Deyholos M.K., Basu U. and Kav N.N.V. 2014. Differential expression of miRNAs in *Brassica napus* root following infection with *Plasmodiophora brassicae*. *PLoS ONE* 9(1): e86648.

Wang C., Leng X., Zhang Y., Kayesh E. and Sun X. 2014. Transcriptome-wide analysis of dynamic variations in regulation modes of grapevine microRNAs on their target genes during grapevine development. *Plant Molecular Biology* 84 (3): 269-285.

Wang C., Wang X., Kibet N.K., Song C., Zhang C., *et al.* 2011a. Deep sequencing of grapevine flower and berry short RNA library for discovery of novel microRNAs and validation of

precise sequences of grapevine microRNAs deposited in miRBase. *Physiologia Plantarum* 143(1): 64-81.

Wang J.-W., Park M., Wang L.-J., Koo Y., Chen X.-Y., *et al.* 2011b. miRNA control of vegetative phase change in trees. *PLoS Genetics* 7(2): e1002012.

Wang J.-W., Schwab R., Czech B., Mica E. and Weigel D. 2008. Dual effects of miR156-targeted SPL genes and CYP78A5/KLUH on plastochron length and organ size in *Arabidopsis thaliana*. *The Plant Cell* 20(5): 1231-1243.

White E.J., Venter M., Hiten N.F. and Burger J.T. 2008. Modified Cetyltrimethylammonium bromide method improves robustness and versatility: The benchmark for plant RNA extraction. *Biotechnology Journal* 3(11): 1424-1428.

Wu G., Park M.Y., Conway S.R., Wang J.-W., Weigel D., *et al.* 2009. The sequential action of miR156 and miR172 regulates developmental timing in *Arabidopsis*. *Cell* 138(4): 750-759.

Xie K., Wu C. and Xiong L. 2006. Genomic organization, differential expression, and interaction of SQUAMOSA promoter-binding-like transcription factors and microRNA156 in rice. *Plant Physiology* 142(1): 280-293.

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 Biology* 10(1): 1-11.

Zerbino D.R. and Birney E. 2008. Velvet: Algorithms for de novo short read assembly using de Bruijn graphs. *Genome Research* 18(5): 821-829.

Zhang B.H., Pan X.P., Cox S.B., Cobb G.P. and Anderson T.A. 2006. Evidence that miRNAs are different from other RNAs. *Cellular and Molecular Life Sciences* 63(2): 246-254.

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 Molecular Biology* 75(1-2): 93-105.

Zhao J.-P., Jiang X.-L., Zhang B.-Y. and Su X.-H. 2012. Involvement of microRNA-mediated gene expression regulation in the pathological development of stem canker disease in *Populus trichocarpa*. *PLoS ONE* 7(9): e44968.

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 Research* 18(9): 1456-1465.

Zhu Q.-H., Fan L., Liu Y., Xu H., Llewellyn D., *et al.* 2014. miR482 regulation of *NBS-LRR* defense genes during fungal pathogen infection in cotton. *PLoS ONE* 8(12): e84390.

Zhu Q.-H. and Helliwell C.A. 2011. Regulation of flowering time and floral patterning by miR172. *Journal of Experimental Botany* 62(2): 487-495.

Chapter 4

High-throughput mRNA transcriptome sequencing of aster yellows phytoplasma-infected *Vitis vinifera* cv. 'Chardonnay'

4.1 Introduction

The aster yellows (AY) phytoplasma group '*Candidatus* Phytoplasma asteris' (16SrI) is a diverse and widespread phytoplasma group (Lee *et al.*, 2000; Lee, 2004), associated with a detrimental disease in grapevine known as grapevine yellows (GY). GY symptoms were first observed in South African vineyards in 2006, and were shown to be caused by AY phytoplasma, group 16SrI-B (Engelbrecht *et al.*, 2010). During a recent GY incidence study, different cultivars (Chenin Blanc, Shiraz, Chardonnay, Cabernet Franc, Sauvignon Blanc, Pinotage and Colombar), were monitored and revealed that Chardonnay is especially susceptible to AY phytoplasma infection (Carstens, 2014). Cultivars showing the highest mean cumulative disease incidences over four years were Pinotage (10.87%), Chenin Blanc (32.31%) and Chardonnay (37.77%). Such an infection rate may infer Chardonnay vineyards to be 100% infected with AY phytoplasma after 10 years which could have ruinous consequences for wine production in the future (Carstens, 2014).

V. vinifera is one of the most important fruit and/or beverage crops in the world and, like all land plants, grapevines have to develop various mechanisms at a physiological and molecular level in order to cope with their ever-changing environment. Significant progress has been made to understand plant-pathogen interactions and the multiple gene regulatory mechanisms they invoke during plant defence responses. Recently, different groups of phytoplasmas were successfully cultivated on complex media which will contribute to studies on host susceptibility and will help design effective GY control measures (Contaldo *et al.*, 2016). In addition, RNA-seq, microarrays, high-resolution mass spectrometry and high-performance liquid chromatography with a diode array detector (HPLC-DAD), have served as valuable approaches for comprehensive analysis of massive gene, transcript and proteomic datasets. These studies have added valuable insights into the physiological, biochemical and molecular mechanisms underlying phytoplasma disease symptom development in grapevine and other plant species (Hren *et al.*, 2009, Albertazzi *et al.*, 2009; Liu *et al.*, 2013; Margaria *et al.*, 2013; Monavarfeshani *et al.*, 2013; Mou *et al.*, 2013; Margaria *et al.*, 2014; Abbà *et al.*, 2014; Luge *et al.*, 2014; Fan *et al.*, 2017; Wang *et al.*, 2018).

The aim of this study was to investigate responses induced by AY phytoplasma in *V. vinifera* cv. 'Chardonnay' at the gene expression level. We extracted leaf RNA from both healthy and AY-

infected plants, which was subjected to mRNA-seq on the Illumina HiSeq 2000 platform. Millions of high quality library reads were obtained and analysed to identify differentially expressed genes (DEGs). We assigned functional annotations to the DEGs using gene ontology (GO) terms. This enabled the identification of gene classes involved in important processes, such as stress response, signal transduction, carbohydrate metabolism, transcriptional regulation, hormone signalling, photosynthesis, and cell development. Our results may help understand the basis of host susceptibility to phytoplasma diseases and may assist in the development of control strategies to prevent further spread of phytoplasma-infection.

4.2 Materials and Methods

4.2.1 Plant material and phytoplasma detection

Visual selection of symptomatic and asymptomatic *V. vinifera* cv. 'Chardonnay' plants in a vineyard in the Olifants River Valley (Western Cape), was described in Chapter 3 (section 3.3.1) (Figure 4.1). Leaf material were collected from each plant, immediately flash frozen in liquid nitrogen, transported on dry ice and stored at -80°C until use. Small-scale RNA extractions were carried out using a modified CTAB method (White *et al.*, 2008), while genomic DNA was extracted using a NucleoSpin® Plant II kit (Macherey-Nagel; Düren, Germany).

Figure 4.1: *Vitis vinifera* cv. 'Chardonnay' with asymptomatic leaves (A), and leaves showing typical grapevine yellows (GY) disease symptoms (B).

Phytoplasma infection was confirmed by a nested-PCR procedure, specifically amplifying a region of the phytoplasma 16S rDNA. The first PCR round was performed using a universal primer pair R16mF2/mR1, followed by a second PCR with the R16F2n/R2 primer pair (Gundersen and Lee, 1996). Additionally, samples were screened for the most prevalent grapevine viruses, including Grapevine leafroll-associated virus 3 (GLRaV-3), Grapevine virus A (GVA), Grapevine virus E (GVE), and Grapevine rupestris stempitting-associated virus (GRSPaV), using two-step RT-PCR assays. Primer sequences for virus screening were obtained from previous publications (Table 4.1). Results from these diagnostics were used to select material, free from these viruses, from three AY phytoplasma-infected, and three healthy plants for further experiments.

Table 4.1: List of primers used in RT-PCR assays for virus-screening.

Target Virus	Primer Pair	Sequence (5'-3')	Amplicon size (bp)	Reference
GVA	GVA-P-F-7038	AGGTCCACGTTTGCTAAG	236	MacKenzie, 1997
	GVA-P-R-7273	CATCGTCTGAGGTTTCTACTA		
GVE	GVE-1-For	AATGGAGTCAAAAGCGATCC	991	Coetzee *et al.*, 2010
	GVE-Rev	GTAGGGTCAATCAACCAACA		
GLRaV-3	LR3.HRM4.F	TAATCGGAGGTTTAGGTTCC	226	Bester *et al.*, 2012
	LR3.HRM4.R	GTCGGTTCGTTAACAACAC		
GRSPaV	StempitCP-F	ACTTTCAAAGACGGTGGACATGAG	523	Noach, 2010
	StempitCP-R	AGCCATAGCTTGTCTGAGCACTTG		

4.2.2 Total RNA extraction and mRNA-seq

The large-scale method for extracting leaf total RNA was described in Section 3.3.2. A NanoDrop ND-1000 spectrophotometer was used for both quantification and purity analysis, while RNA integrity was assessed using a Plant RNA Nano Assay using an Agilent 2100 Bioanalyzer. Leaf total RNA from each plant was sent to Fasteris SA (Plan-les-Ouates, Switzerland) for transcriptome sequencing. Following ribosomal RNA depletion, six cDNA libraries were constructed using a TruSeq Stranded RNA Library Prep kit (Illumina, San Diego, California, USA), followed by quality assessment on an Agilent 2100 Bioanalyzer using a High Sensitivity DNA Assay. Libraries with acceptable quality were subjected to single-channel multiplexing and paired-end mRNA-seq (2 x 100 nt) on an Illumina HiSeq2000 platform (https://support.illumina.com).

4.2.3 Gene expression analysis and functional annotation

Raw mRNA-seq data was received from the service provider in Illumina-fastq format. FastQC was used to visualise certain quality control measurements (www.bioinformatics.babraham.ac.uk/projects/fastqc/). FASTX-Toolkit was used for adaptor-trimming and quality filtering. It was specified that ≥80% of bases must have Phred quality scores of at least 30 or more, while the rest were trimmed (https://hannonlab.cshl.edu/fastx_toolkit/). Following pre-processing, we used the Tuxedo suite for analysing RNA-seq data. The remaining (clean) reads were aligned to the annotated *V. vinifera* cv. 'Pinot noir' 12x coverage genome assembly, release 12X.v0 (PN40024; http://www.genoscope.cns.fr/externe/GenomeBrowser/Vitis/), using TopHat v2.0.7 (Trapnell *et al.*, 2009), with default parameters. The resulting read alignment file, in BAM format, obtained for each of the six libraries, was used as input for the Cufflinks package v1.3.0 (Trapnell *et al.*, 2010; Trapnell *et al.*, 2012). Cuffdiff was used to obtain normalised gene expression values in fragments per kilobase of transcript per million mapped reads (FPKM), and to calculate differential gene expression between the three AY phytoplasma-infected and three healthy (control) samples. In order to correct for false discovery rate (FDR) during multiple testing, $\log_2$-fold changes of DEGs with adjusted *p*-values (*q*) of ≤ 0.05 were considered significant (Benjamini and Hochberg, 1995).

Significant DEG sequences were subjected to functional annotation using Blast2GO v4.1.9 (Conesa and Götz, 2008). This was done by comparing each query sequence against the NCBI non-redundant protein database using BLASTx to find homologous sequences, mapping BLAST results against the gene ontology (GO) database (http://geneontology.org/page/go-database), and assigning functional attributes to each sequence in terms of biological process, cellular component and molecular function, using GO terms. InterProScan (Finn *et al.*, 2017), was also used to complement the BLAST-based annotations by searching for domain/motif information in a sequence-wise manner and then linking the results to the existing annotations. In cases where limited GO information was available, a manual search was performed in the UniProt Knowledgebase (The UniProt Consortium, 2018) to assign a putative function to the gene product in question.

To predict whether any resulting DEGs were potential targets of the differentially expressed miRNAs (identified in Chapter 3), we utilised the psRNAtarget analysis server (Dai and Zhao, 2011; http://plantgrn.noble.org/psRNATarget/). Default parameters were used, which included a threshold cut-off of 3.0 for low false-positive prediction, a complementarity scoring length of 20 bp, and the energy required for target accessibility equal to 25 kcal/mole.

4.2.4 Validation of gene expression by RT-qPCR

In order to validate the mRNA-seq data analysis results, we selected four significant DEGs ($q \leq 0.05$) for quantification using reverse transcription quantitative PCR (RT-qPCR). Primer sequences were generated with Oligo Explorer v1.1.2 (http://www.genelink.com) and submitted for custom oligo synthesis (Table 4.2). High-quality total RNA was prepared as described above. Each 20 µl reverse transcription reaction was prepared with a High-Capacity cDNA Reverse Transcription Kit (Thermo Scientific, Waltham, Massachusetts, United States), and contained 50 U MultiScribe® Reverse Transcriptase, 1x RT buffer, 4 mM dNTPs, 1x random primers, 1 µg total RNA, and nuclease-free water. Cycling conditions were as follows: 10 min at 25°C, 2 hours at 37°C, 5 min at 85°C and cooling at 4°C. Each 10 µl qPCR reaction mixture was prepared in triplicate and contained 1x Power SYBR® Green PCR master mix (Thermo Scientific, Waltham, Massachusetts, United States), 125 nM forward primer, 125 nM reverse primer, 1 µl cDNA, and nuclease-free water. Control reactions, each without cDNA template, were included for each target gene.

Actin, glyceraldehyde 3-phosphate dehydrogenase (*GAPDH*), and elongation factor 1-alpha (*EF1-a*) were chosen as internal control genes to normalise gene expression data, and their stable expression levels were confirmed with geNorm analysis (qBase^{PLUS} v2.0, Biogazelle, Ghent, Belgium; Hellemans *et al.*, 2007) (data not shown). We designed *Actin*, *GAPDH* and *EF1-a* primers based on sequences used by Gutha *et al.* (2010) (Table 4.2). PCR amplification was performed in an Applied Biosystems 7900HT Fast Real-Time PCR System, in which the baseline and threshold cycles (Ct) were automatically determined with SDS v2.3 software. Cycling conditions were as follows: 50°C for 2 min, 95°C for 10 min, 40 cycles at 95°C for 10 s and 60°C for 1 min, followed by melt curve analysis. Melt curve analysis was used to ensure that a single PCR product was obtained. Relative gene expression analysis using the ΔΔCt-method was performed using qBase^{PLUS} v2.0 software (http://www.qbaseplus.com; Biogazelle, Ghent, Belgium).

Table 4.2: List of primers used in the RT-qPCR assays for relative gene expression analysis.

Target gene	Primer ID	Sequence (5'-3')	Tm (°C)	Amplicon size (bp)
Pathogenesis-related protein 1 (*PR-1*) (Genbank: XM_002273752)	vvPR1-F	TTGTGGGTGGGGGAGAAG TC	62	167
	vvPR1-R	CGTGGATCGTAGTTGCATGTGA	62	
Thaumatin-like protein (Genbank: AF003007)	vvPR5-F2	TTCACCACCACCCTCCCA	58	141
	vvPR5-R2	AAGTCTCCGTCCGCCACC	60	
Homeobox leucine zipper (Genbank: XM_002271487)	vvHLZ-F	TGACCAGGAAGAAGAGCAGGAACAA	65	143
	vvHLZ-R	CTGTCTCGGCTGCATCCCAA	62	
Peroxidase N1-like protein (Genbank: XM_002269136)	vvPeroxidase-F2	AGGCACTAGGGTTGGCTTCT	60	95
	vvPeroxidase-R2	TGGCTGGATTGGACTGGA	56	
Actin (Genbank: GU585869)	vvActin-F	CTTGCATCCCTCAGCACCTT	60	82
	vvActin-R	TCCTGTGGACAATGGATGGA	58	
Glyceraldehyde 3-phosphate dehydrogenase (*GAPDH*) (Genbank: GU585870)	vvGAPDH-F	TTCTCGTTGAGGGCTATTCCA	59	70
	vvGAPDH-R	CCACAGACTTCATCGGTGACA	61	
Elongation factor 1-alpha (*EF1-a*) (Genbank: GU585871)	vvEF1a-F	GAACTGGGTGCTTGATAGGC	60	164
	vvEF1a-R	AACCAAAATATCCGGAGTAAAAGA	58	

4.3 Results and Discussion

4.3.1 Plant material and phytoplasma detection

Diagnostic PCR screening results were discussed in section 3.4.1. The healthy control plants (h55, h85, h89) consisted of three phytoplasma and virus-free Chardonnay plants, while the three AY phytoplasma-infected, but virus-free, plants were our experimental group (p73, p93, p99). These samples were the same as those used in the small RNA-seq (sRNA-seq) experiment (Chapter 3; Appendix 1.1).

4.3.2 mRNA-seq analysis

Large-scale RNA preparations delivered good quality total RNA with acceptable RIN values of > 7 (Appendix 4.1). Following RNA-seq of the six libraries, each with an average insert size of 200 ± 40 nt, we obtained a total of 178,039,188 paired-end reads (Table 4.3). Following pre-processing, quality evaluation revealed that the reads possessed an acceptable mean Phred quality score (Q score) of >26 across the length of all reads. Figure 4.2 given as an example of the read quality data across all bases for sample h55. All quality-trimmed reads were 50 bp in length with a mean Q score of ~36 (Appendix 4.2). An average of ~85% of quality-trimmed paired-end reads mapped successfully against the *V. vinifera* (PN40024) 12x assembly (Table 4.3), which contains a total of 26,346 annotated genes.

Table 4.3: Summary of total number of mRNA-seq reads.

Sample type	Sample ID	Reads in pairs	Mapped reads	% mapped reads
AY phytoplasma-infected	p73	26,408,222	22,540,966	85%
	p93	33,043,064	28,621,470	87%
	p99	24,971,712	21,525,712	86%
Healthy	h55	25,819,410	21,789,577	84%
	h89	27,103,134	23,311,453	86%
	h85	40,693,646	34,795,339	86%

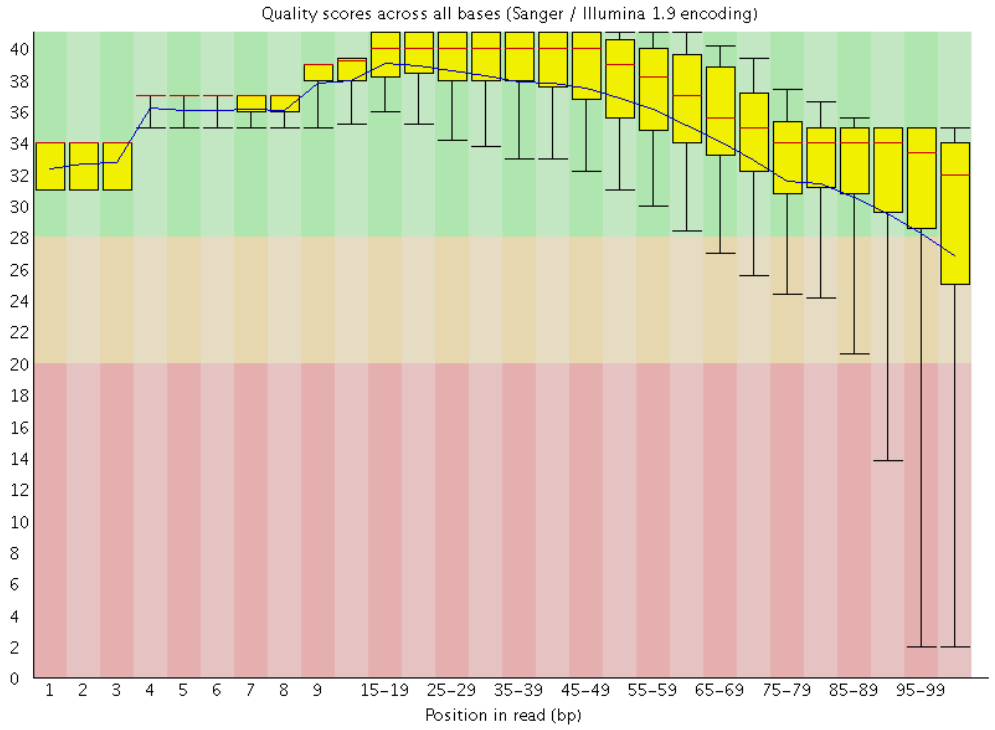

Figure 4.2: BoxWhisker plot representing the range of quality values across all bases at each position in the FastQ file of sample h55. Red line: median value; yellow box: inter-quartile range (25-75%); upper and lower whiskers: 10% and 90% points; blue line: mean quality.

After read mapping, the Cufflinks workflow was used to test for significant differences in transcript levels, using normalised read counts (FPKM), between the biological replicates of the experimental (diseased) group and those of the healthy (control) group. Results showed that 175 genes were differentially expressed in the AY phytoplasma-infected leaf material ($q \leq 0.05$). Of these, 119 genes were up-regulated, while the remaining 56 genes were down-regulated (Appendix 4.3).

We utilised psRNAtarget to perform reverse complementary matching between the previously-identified miRNAs (Chapter 3) and the predicted DEGs in order to search for potential miRNA targets. Unexpectedly, there were no potential anti-correlated targets found for the differentially

expressed miRNAs. In corroboration, none of the putative targets predicted for the miRNAs in Chapter 3 were among the DEGs found in the mRNA-seq transcriptome data. Genotypic sequence bias between the Pinot noir genome assembly and the Chardonnay transcriptome could be a reason for this. As a result, alternative transcripts may have been predicted as possible miRNA targets. We can also speculate that because leaf material was used to perform this experiment, the hypothetical miRNA/mRNA interactions simply did not occur in those tissues. Another possibility may be that miRNA/mRNA interactions were too subtle to measure significant changes in anti-correlated transcript levels. Therefore, the DEGs found in the RNA-seq transcriptome data may have been the result of different regulatory mechanisms.

4.3.3 Functional annotation of DEGs

Blast2GO analysis was used to gain further insight into the function of the DEGs found in this study since the molecular mechanisms involved during AY phytoplasma-infection of Chardonnay are still largely unknown. Overall, GO terms were assigned to 157 of the 175 DEGs in the AY phytoplasma-infected plants (Appendix 4.3). Functional assignments were distributed among 31 categories. Of the total annotated DEGs, the major GO categories included gene products involved in ion- and cyclic/heterocyclic compound binding (molecular functions: GO:1901363; GO:0043167), gene products involved in organic compound-, cellular- and primary metabolic processes (biological processes: GO:0071704; GO:0044237; GO:0044238), gene products involved in processes taking place intracellularly (GO:0005622; GO:0044424), and gene products that are intrinsic components of the cell membrane and plastids (cellular component: GO:0031224; GO:0043229) (Figure 4.3).

The UniProt Knowledgebase (UniProtKB) and Blast2GO were used to assign general putative functions to the DEGs which were summarised in Figure 4.4. Interestingly, most up-regulated genes were associated with plastid and/or cell wall development, plant-pathogen interactions possibly related to stress response, as well as transcriptional regulation. The majority of down-regulated genes were involved in plastid and/or cell wall development, transcriptional regulation, oxidative stress and photosynthesis. Involvement of these DEGs in stress-related pathways and gene regulation was of particular interest for this study.

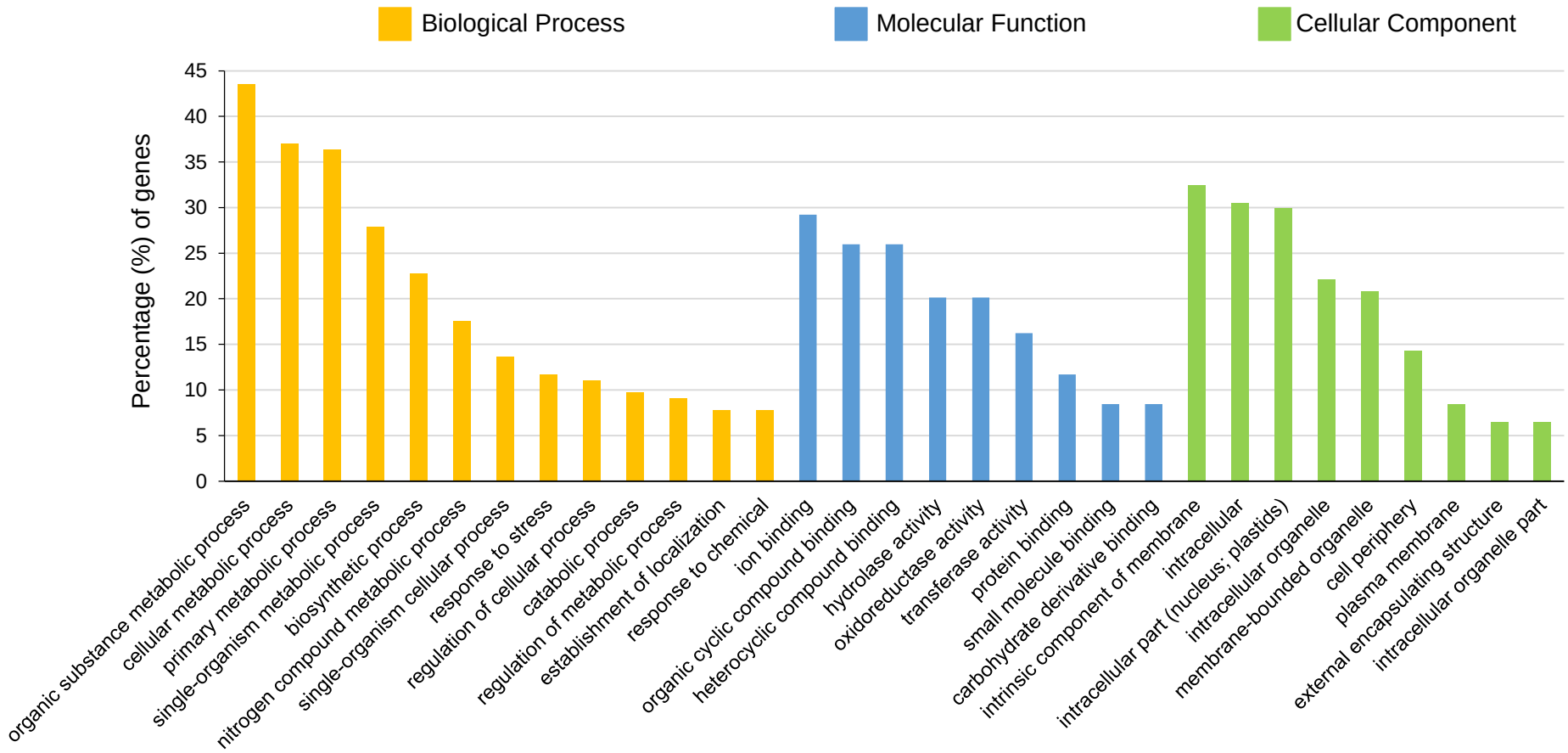

Figure 4.3: Gene Ontology (GO) classification of DEGs in AY phytoplasma-infected *V. vinifera* cv. 'Chardonnay'. Genes were annotated in three categories: biological process, molecular function and cellular component.

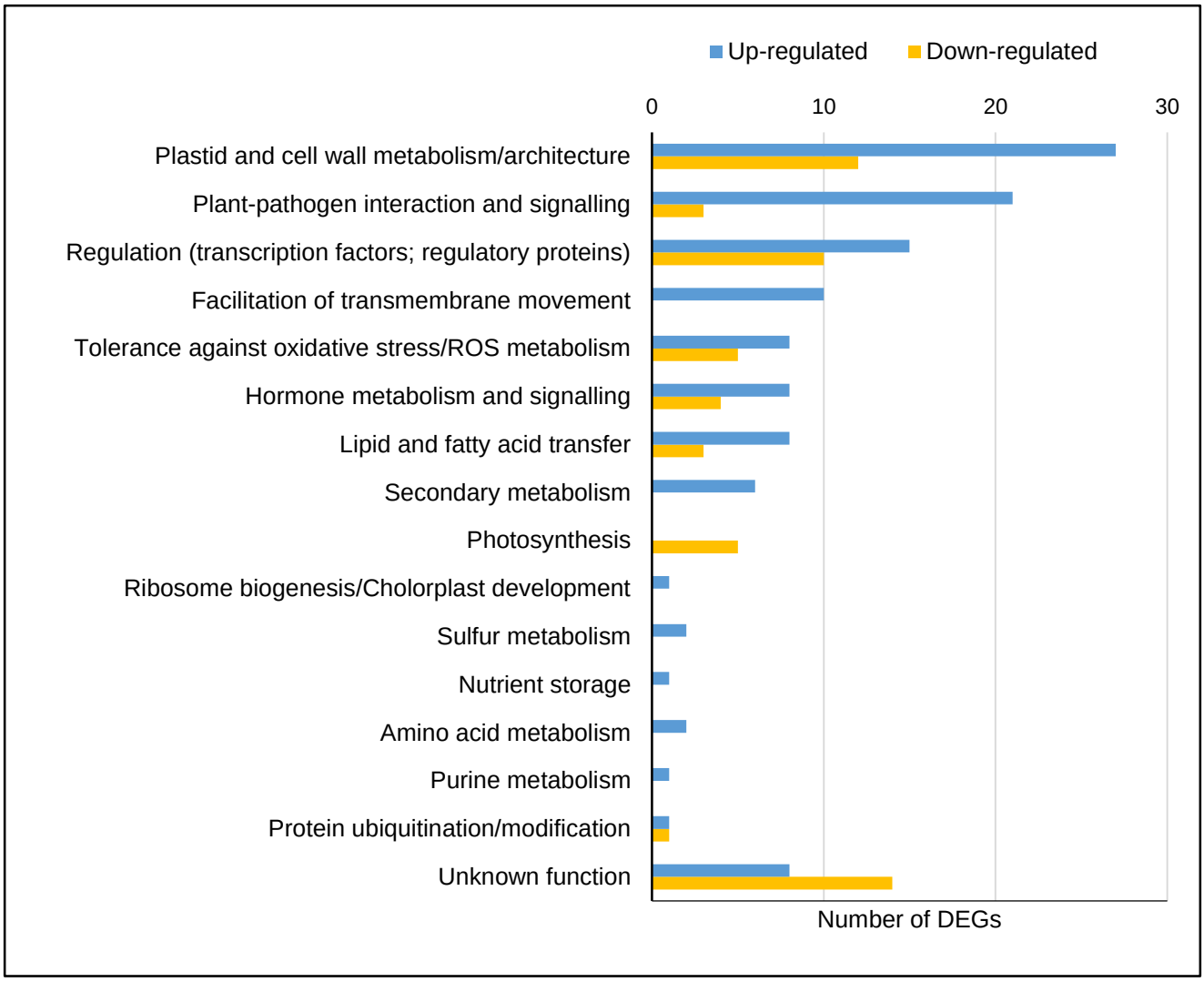

Figure 4.4: Simplified functional classification of the number of DEGs discovered in the AY phytoplasma-infected plant material.

4.3.3.1 DEGs involved in plant-pathogen interaction and signalling

Our analyses revealed DEGs involved in innate immunity and pathogen defence response, including genes encoding pathogenesis-related (PR) proteins, receptor-like kinases (RLKs), and leucine-rich repeat receptor-like kinases (LRR-RLKs) (Appendix 4.3; Figure 4.4).

Among the seven RLKs and LRR-RLKs, five were up-regulated and two down-regulated in the AY phytoplasma-infected grapevines (Appendix 4.3). Protein kinases are a large group of membrane-bound receptor molecules that play an essential role in the detection of various stimuli and, in turn, generate signals via protein phosphorylation to induce a relevant defence response (Romeis, 2001; Ho, 2015). They control various processes, including plant growth and morphology, disease resistance, homeostatic mechanisms in response to abiotic stimuli, and activation of hormone

signalling (Osakabe *et al.*, 2013; Ho, 2015). Previous studies have identified several protein kinase genes that were differentially expressed in response to phytoplasma-infection, including mitogen-activated protein kinases (MAPK) and LRR receptor kinase brassinosteroid insensitive 1 (BAK1) (Mardi *et al.*, 2015; Fan *et al.*, 2015).

RLKs, which are surface-located transmembrane receptors, provide the first line of defence against invading pathogens by recognising pathogen-associated molecular patterns (PAMPS), such as flagella, liposaccharides and chitin (Figure 4.5). Intracellular proteins, such as translation elongation factor thermo unstable (EF-Tu) and cold shock proteins (CSPs) can also act as PAMPS (Sugio *et al.*, 2011). Molecular sensing of PAMPS can activate different kinase signalling pathways which result in the induction of PAMP-triggered immunity (PTI) (Rafiqi *et al.*, 2009). Phytoplasmas are devoid of an outer cell wall and flagella, therefore lacking these PAMPs, but do have genes encoding CSPs and the EF-Tu, which may induce PTI (Sugio *et al.*, 2011).

Interestingly, among the modulated LRR-RLK genes, two were resistance genes (*R* genes), *viz. GSVIVT01006610001* (up-regulated) and *GSVIVT01023557001* (down-regulated) (Appendix 4.3). They encode specialised forms of intracellular receptors known as R proteins, most of which contain a nucleotide binding (NB) domain and leucine-rich repeats (LRR) (NB-LRR proteins) (Rafiqi *et al.*, 2009). R proteins provide a second layer of defence known as effector-triggered immunity (ETI) against PTI-evading microbes (Figure 4.5).

ETI involves sensing of pathogen effector proteins which strongly induce basal defence through the actions of PR proteins. If these molecules do not succeed, programmed cell death (PCD) is activated. These reactions are collectively known as the hypersensitive response (HR) (Dangl *et al.*, 1996) (Figure 4.5). It is believed that although PTI and ETI networks involve different receptor kinases and signalling pathways, they can interconnect to inhibit pathogen growth (Truman *et al.*, 2006). To date, only four phytoplasma-derived effector proteins *viz.* TENGU and PHYL1 from '*Ca.* Phytoplasma asteris' OY wild type strain (OY-W), as well as SAP11 and SAP54 from '*Ca.* Phytoplasma asteris' AY-WB, and have been functionally characterised (Minato *et al.*, 2014). They can disrupt host developmental processes through interaction with transcription factors and modulation of phytohormones (auxin and JA), that in addition to regulating plant development, also have fundamental roles in plant defence signalling (Sugio *et al.*, 2011; Minato *et al.*, 2014; Orlovskis and Hogenhout, 2016).

Many plant species display the accumulation of PR proteins in response to pathogens, suggesting a common ancestral role of these proteins during biotic stress conditions (Figure 4.5). We identified a number of PR genes, most of which were up-regulated (Appendix 4.3). Although the function of

PR-1 proteins in grapevine is still unclear, *PR-1* genes were shown to be up-regulated in response to SAR elicitors (e.g. SA), and during infection by *Botrytis cinerea* and *Plasmopara viticola* (Repka, 2001). A uniquely duplicated *PR-1* gene, *VvPR1b1*, from a *Vitis* interspecific hybrid were shown to confer resistance to *Pseudomonas syringae* pv. *tabaci* during over-expression in transgenic tobacco (Li *et al*., 2011). In a previous study, genes encoding PR-1 and PR-2 (β-1,3-glucanase) proteins were also up-regulated in Chardonnay in response to Bois noir (BN) phytoplasma-infection (Hren *et al.*,2009), as well as the red grape cultivar 'Barbera' in response to Flavescence dorée (FD) infection (Gambino *et al.*, 2013). PR-2 and PR-3/4 (chitinase) deploy antimicrobial activity through hydrolysis of fungal cell walls, causing cell lysis and inhibition of fungal growth. The resulting release of oligosaccharides can act as elicitors that further induce downstream defence mechanisms (Ebrahim *et al.*, 2011; Enoki and Suzuki, 2016).

Up-regulation of PR-2-type glucanase genes in this study may have caused callose degradation (Figure 4.5). This was demonstrated in other studies and is believed to clear the way for spread of phytoplasmas through the phloem (Albertazzi *et al.*, 2009; Landi and Romanazzi, 2011; Gambino *et al.*, 2013; Paolacci *et al.*, 2017). Abscisic acid (ABA) treatment of *Arabidopsis* was shown to promote callose deposition through transcriptional repression of *PR-2* (Oide *et al.*, 2013). Callose deposition (by means of callose synthase) and degradation occurs frequently at regions of the cell wall surrounding the plasmodesmata, where it is generally believed to regulate the transport of molecules through the symplast in response to biotic and/or abiotic stresses (Chen and Kim, 2009).

The expression of *PR-5* can be induced by a number of biotic and abiotic signals (Fagoaga *et al.*, 2001; Kumar *et al.*, 2015). PR-5 proteins, including thaumatin-like proteins and osmotin, are believed to enhance fungal membrane permeability that causes osmotic rupture of the fungal plasma membrane (Stintzi *et al.*, 1991). Genes encoding thaumatin were up-regulated in our study, which correspond to other reports where members of the *PR-5* class were up-regulated in response to phytoplasma-infection in *Chrysanthemum coronarium* (Zhong and Shen, 2004), *Malus domestica* (Giorno *et al.*, 2013), and symptomatic FD- and BN phytoplasma-infected grapevines (Hren *et al.*, 2009a; Landi and Romanazzi, 2011; Gambino *et al.*, 2013; Albertazzi *et al.*, 2015).

ETI results in the activation of jasmonic acid/ethylene (JA/ET) and salicylic acid (SA) biosynthesis in chloroplasts that trigger the expression of PR genes, as well as genes related to systemic acquired resistance (SAR) (Balakireva and Zamyatnin, 2018) (Figure 4.5). Previous studies showed that expression of *PR-1*, *PR-2* and *PR-5* gene families were modulated by exogenous application of SA, whereas exogenous JA mainly induced the expression of *PR-3*, *PR-4* and *PR-6* (protease inhibitor) genes (Hamiduzzaman *et al.*, 2005; Belhadj *et al.*, 2006; Chong *et al.*, 2008; Le Henanff *et al.*, 2009). The induction of SA-dependent SAR in leaves of BN phytoplasma-infected tomatoes and

grapevines has also been suggested (Ahmad *et al.*, 2015; Dermastia *et al.*, 2015). *VvPR-2* and *VvPR-5* genes were up-regulated together with increased levels of SA in FD phytoplasma-infected grapevine (Prezelj *et al.*, 2016), which may support findings that these genes are co-ordinately regulated by SA and act as molecular markers for SA-dependent SAR signalling (Frías etal. 2013).

Interestingly, three paralogs encoding salicylate carboxymethyltransferases were up-regulated in our study (Appendix 4.3). SA carboxymethyltransferase catalyses the conversion of SA to a volatile methyl ester known as methyl salicylate (MeSA) (Figure 4.5). MeSA is a mobile signal molecule that is transported from the site of infection to distal parts through the plasmodesmata and phloem. SAR in uninfected areas can be accomplished once MeSA is converted back to SA (Gao *et al.*, 2015). MeSA is also hypothesized to have an accessory role in SAR signalling by acting as an airborne signal that triggers defence responses in uninfected plants (Seskar *et al.*, 1998). Additional research is required to explore the involvement SA- and JA-mediated defence that may co-ordinately regulate *PR* genes in response to AY phytoplasma-infection in Chardonnay.

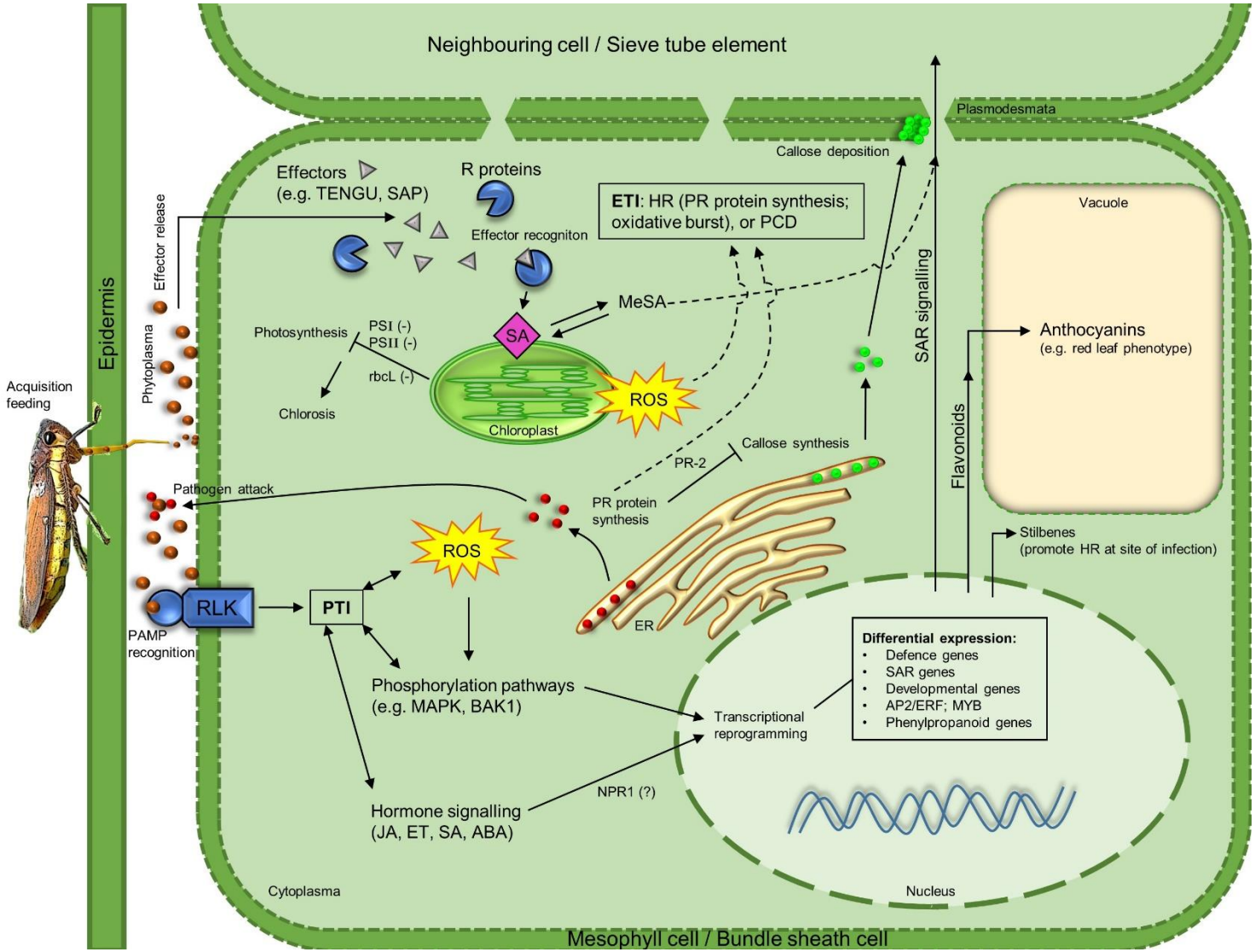

Figure 4.5: A schematic diagram presenting a model for putative host response pathways in a Chardonnay leaf upon AY phytoplasma-infection. The DEGs identified in this study could be classified according to their roles in either PAMP-triggered immunity (PTI) and/or effector-triggered immunity (ETI). Abbreviations: PAMP, pathogen-associated molecular patterns; RLK, receptor-like protein kinase; R protein, resistance protein; ROS, reactive oxygen species; PR, pathogenesis-related; MAPK, mitogen-activated protein kinases; BAK1, brassinosteroid insensitive-associated receptor kinase 1; PCD, programmed cell death; HR, hypersensitive response; ABA, abscisic acid; JA/ET, jasmonic acid/ethylene; SA, salicylic acid; SAR, systemic acquired resistance; MeSA, methyl salicylate; PSI, photosystem I; PSII, photosystem II; rbcL, ribulose 1,5-bisphosphate carboxylase/oxygenase large subunit; ER, endoplasmic reticulum; NPR1, Nonexpressor of pathogenesis-related genes 1.

4.3.3.2 DEGs involved in ROS metabolism

HR is often accompanied by an oxidative burst, causing accumulation of reactive oxygen species (ROS) such as superoxide radicals, hydrogen peroxide (H_2O_2), and hydroxyl radicals. ROS production is a common feature during activation of both ETI and PTI and has been observed in plant-pathogen reactions involving avirulent bacteria, fungi, and viruses (Apel and Hirt, 2004; Camejo *et al.*, 2016) (Figure 4.5). ROS contribute to plant defence by killing pathogens either directly (in the case of hydroxyl radicals), or limiting pathogen entry through the formation of physical barriers by reinforcing cell walls via lignin and glycoprotein cross-linking (Bradley *et al.*, 1992; Huckelhoven, 2007), or by a signalling function that mediates defence gene activation (Torres, 2010).

Genes involved in the production and neutralisation of ROS, such as glutathione S-transferase (GST), polyamine oxidase, polyphenol oxidase (PPO), peroxidases (POX), and superoxide dismutase (SOD), were differentially expressed in this study (Appendix 4.3). Antioxidant/scavenger enzymes such as POX and SOD were down-regulated which may lead to increased ROS levels (Mittler, 2004). In turn, higher ROS levels may activate phosphorylation cascades such as the MAPK signalling pathway to induce the expression of *PR* genes (Mou *et al.*, 2003, Ho, 2015) (Figure 4.5). Several host cultivars, such as Manzoni Bianco, can exhibit a strong defence response against phytoplasma-infection to increase its tolerance to the pathogen. In a previous study Manzoni Bianco displayed a strong down-regulation of genes encoding POX and GST, which may support the notion that down-regulation of ROS-scavengers may influence a response to phytoplasma-infection (Albertazzi *et al.*, 2009). In another study, tolerant grapevine plants had distinctly lower levels of scavenger enzymes in their leaf tissue compared to healthy and infected plants. This may be associated with long-term, sustained and tissue-specific accumulation of H_2O_2, which reduces titre or prevents further infection by FD phytoplasma (Musetti *et al.*, 2007).

ROS, in association with SA, was proposed to mediate the induction of SAR in *Arabidopsis*. This involves the accumulation of the SA which induces defences gene expression via activation of Nonexpressor of pathogenesis-related genes 1 (NPR1), the master regulator of SA-dependent transcriptional responses (Figure 4.5). Once NPR1 is transported to the nucleus it can interact with TGA transcription factors to induce the expression of *PR* genes (Fan and Dong, 2002; Weigel *et al.*, 2005). The expression of *PR-2* and *PR-5* genes is co-ordinately regulated by SA (Frías *et al.*, 2013), and it was suggested that BN phytoplasma induces SA-dependent SAR in leaves of infected grapevines and tomatoes (Ahmad *et al.*, 2015; Dermastia *et al.*, 2015).

A *Chlorophyllase-1*-like gene (*GSVIVT01001207001*) was significantly down-regulated in the AY-phytoplasma-infected grapevines (Appendix 4.3). Certain forms of biotic and abiotic stress, as well as excess light, can damage plant tissues and may cause the release of chlorophyll from the thylakoid membranes (Karpinski *et al.*, 2003). This results in the need for chlorophyll degradation in order to avoid cellular damage by their photodynamic action (Takamiya *et al.*, 2000). Failure to degrade chlorophyll through the action of chlorophyllase 1 can increase the amount of ROS produced (Figure 4.5). This may cause an over-toxifying effect that could override the detoxification capacity of the antioxidant systems, leading to either organelle damage, cell death or SA-dependent SAR (Kariola *et al.*, 2005). Decreased chlorophyll catabolism was shown to result in the accumulation of photosensitive porphyrin rings, causing lesion development and severe oxidative stress in both tobacco and *Arabidopsis* (Matile and Hörtensteiner, 1999; Mock *et al.*, 1999; Mach *et al.*, 2001). Therefore, defects in the degradation of chlorophyll may contribute to the augmentation of GY disease symptoms through discolouration and necrosis of leaf veins and laminae, as well as downward curling of leaves.

Expression analysis also revealed a two-fold up-regulation of two *polyphenol oxidase* (*PPO*) genes upon AY phytoplasma-infection in Chardonnay (Appendix 4.3). PPO catalyses the oxidation of phenolic compounds to free radicals, which in turn can react with oxygen and proteins to form ROS and typical brown-pigmented complexes that create unfavourable conditions for pathogen growth (Ngadze *et al.*, 2012). Different studies have suggested that PPO plays a role in pathogenesis during bacterial, fungal and phytoplasma infection (Li and Steffens, 2002; Zafari and Niknam, 2011; Ngadze *et al.*, 2012; Nasir *et al.*, 2017).

4.3.3.3 DEGs involved in secondary metabolism

Our findings also revealed differentially expressed genes involved in plant secondary metabolism during biotic stress, i.e. *Stilbene synthase* (*STS*), *Trans-resveratrol di-O-methyltransferase* (*ROMT*), *Flavanone 3-hydroxylase* (*F3H*), each being significantly induced in the AY phytoplasma-infected grapevines (Appendix 4.3; Figure 4.4).

The *STS* gene family has been well-characterised in grapevine and encode the key enzyme responsible for the biosynthesis of phytoalexins (such as resveratrol and stilbenes), which can accumulate around a site of infection (as part of the HR) to help limit the spread of invading pathogens (Armijo *et al.*, 2016; Hasan and Bae, 2017) (Figure 4.5). The accumulation of stilbenes in grapevine tissues in response to biotic stresses have been documented for a range of different pathogens, including powdery mildew (*Erysiphe necator*), downy mildew (*Plasmopara viticola*), gray mold (*Botrytis cinerea*), and *Aspergillus carbonarius* (Vannozzi *et al.*, 2012). *STS* genes were

previously up-regulated in BN-diseased grapevine plants, which corroborate our data. This was believed to promote stilbene accumulation which possibly contributed to symptom development (Paolacci *et al.*, 2017).

Trans-resveratrol di-O-methyltransferase can catalyse the conversion of resveratrol to pterostilbene *in vitro* and *in planta* (Schmidlin *et al.*, 2008). Pterostilbene was shown to also possess antifungal activity against various grapevine pathogens, and can inhibit fungal growth five to ten times more effectively than resveratrol (Jeandet *et al.*, 2002).

Flavanone 3-hydroxylase (F3H) is involved in the biosynthesis of dihydroflavonols which serve as precursors for different flavonoid compounds, including flavonols, anthocyanins and condensed tannins (Winkel-Shirley, 2011). Flavonoids are known for their role as antioxidants, as well as antimicrobial, pigmentation and/or UV-B protective compounds (Petrussa *et al.*, 2013). A few studies have described changes in flavonoid biosynthetic gene expression upon phytoplasma-infection, where *F3H*, as well as other flavonoid biosynthetic genes, were up-regulated in diseased plants (Hren *et al.*, 2009; Margaria *et al.*, 2014; Prezelj *et al.*, 2016). Therefore, we suggest the possible involvement of flavonoid biosynthesis in response to AY phytoplasma-infection in Chardonnay (Figure 4.5).

Unlike genes that play a role in host resistance, certain genes known as susceptibility genes (S genes) are required for successful pathogen infection and spread (Zaidi *et al.*, 2018). A paralog of *Downy mildew resistance 6* (*DMR6*), called *DMR6-like oxygenase 2* (or *DLO2*) (Appendix 4.3), was up-regulated in this study and may increase the level of host susceptibility to AY phytoplasma. *DMR6* was previously characterised by Van Damme *et al.* (2008), and was shown to be activated during infection with *Hyaloperonospora arabidopsidis*, the causal agent of downy mildew in *Arabidopsis thaliana*. *DMR6* belongs to a multigene family encoding 2-oxoglutarate/Fe(II)-dependent dioxygenases (2-ODDs). 2-ODDs are involved in both primary and secondary metabolism during the biosynthesis of signalling molecules (e.g. gibberellin, ethylene and SA), as well as flavonoids and alkaloids (Farrow and Facchini, 2014). It has been shown that *dmr6* mutant lines associated with SA homeostasis resulted in plants with resistance to certain bacteria (*Pseudomonas* spp.) and oomycetes (Zeilmaker *et al.*, 2015). As with *DMR6*-overexpression, overexpression of *DLO1* and *DLO2* restored susceptibility of the resistant *dmr6* mutant to downy mildew. This indicated that all three proteins can act as suppressors of immunity (Zeilmaker *et al.*, 2015).

New breeding techniques combined with the advent of gene editing tools, particularly the CRISPR/Cas9 system, has provided crop varieties with either improved yield and/or resistance to

certain biotic and abiotic stresses (Zaidi *et al.*, 2018). Recently, the CRISPR/Cas9 system was employed to mutate a tomato *DMR6* orthologue in order to inactivate this gene. The resulting tomato plants demonstrated disease resistance against a variety of pathogens without affecting growth and development (De Toledo Thomazella *et al.*, 2016). Giacomelli *et al.* (2017) are attempting to produce mildew-resistant grapevines by utilizing the CRISPR/Cas9 technology to knock out susceptibility genes, such as *DMR6* and *MLO* (*Mildew resistance locus*). The CRISPR method has been used to successfully engineer economically important plant species, such as wheat and rice, with biotic stress resistance (Zaidi *et al.*, 2018). Therefore, we can speculate that susceptibility genes may serve as potential targets for genome editing tools in order to engineer grapevines with resistance to a number of diseases, including GY.

4.3.3.4 DEGs involved in photosynthesis and carbohydrate metabolism

Our expression analysis results also displayed down-regulated genes involved in photosynthesis (Figure 4.4). According to Bilgin *et al.* (2010), biotic stress is believed to cause a universal down-regulation of photosynthesis-related genes as an adaptive response by "investing resources in immediate defence needs without debilitating near term losses in photosynthetic capacity".

Previous studies revealed a significant breakdown in the photosynthesis chain, mainly due to repression of photosystem II (PSII) activity in phytoplasma-infected leaves of grapevine (Bertamini and Nedunchezhian, 2001; Bertamini *et al.*, 2002b; Albertazzi *et al.*, 2009; Hren *et al.*, 2009; Margaria and Palmano, 2011; Margaria *et al.*, 2013; Prezelj *et al.*, 2016) (Figure 4.5). The analysis of chlorophyll fluorescence indicated a significant reduction in PSII activity, particularly at the electron donor site, and was supported by a pronounced loss of several thylakoid polypeptides, as well as a decrease in total chlorophyll and carotenoids in phytoplasma-infected leaves of apple and grapevine (Bertamini *et al.*, 2002a; Bertamini *et al.*, 2002b). In other studies, transcripts encoding proteins responsible for photosystem I (PSI) activity were strongly down-regulated. Such cases were also observed in Chardonnay infected with BN phytoplasma (Albertazzi *et al.*, 2009), paulownia trees infected with Paulownia Witches' Broom (PaWB) phytoplasma (Mou *et al.*, 2013), and jujube trees infected with Jujube Witches' Broom (JWB) phytoplasma (Wang *et al.*, 2018). In agreement with these findings, we identified three significantly down-regulated genes in the AY phytoplasma-infected Chardonnay plants, encoding a PSI reaction center subunit II protein (psaD), a PSII protein D1 (psbA), and a PSII repair protein PSB27, respectively (Appendix 4.3) (Figure 4.5).

Transcriptome analysis also revealed the significant repression of a gene encoding the large subunit of ribulose 1,5-bisphosphate carboxylase/oxygenase (RuBisCO; *rbcL*) (Appendix 4.3) (Figure 4.5). This is consistent with a marked decrease in soluble proteins and RuBisCO proteins in phytoplasma-infected apple and grapevine leaves following SDS-PAGE analysis (Bertamini *et al.*, 2002a; Bertamini *et al.*, 2002b). RuBisCO catalyses the carboxylation of ribulose 1,5-bisphosphate in the Calvin cycle (light-independent photosynthesis reactions), during which atmospheric CO_2 is used to produce 3-phosphoglycerate molecules. 3-Phosphoglycerate serves as precursor for consecutive steps where ATP and NADPH, produced during the light-dependent photosynthetic reactions, are consumed to produce triose phosphate as net product. These triose phosphate molecules, as well as those formed during gluconeogenesis, ultimately serves as intermediates for the production of carbohydrates such as sugars and polysaccharides (starch, glycogen, cellulose, pectin and chitin) (Kruger, 1998; MacDonald and Buchanan, 1998).

Our findings may support the hypothesis that phytoplasma-infection can cause serious inhibition of the whole photosynthesis chain, leading to chlorosis and rapid leaf senescence (Figures 4.5). This may strongly contribute to GY symptoms seen in the field and could have detrimental effects on plant vitality.

4.3.3.5 DEGs involved in transcriptional regulation

Many of the abovementioned biological processes may be the result of upstream events involving regulatory genes such as transcription factors. DEGs involved in transcriptional regulation of cellular and metabolic processes were also modulated in response to AY phytoplasma (Figure 4.4). These DEGs encode proteins with DNA-binding transcription factor (TF) activity, including MYB, AP2/ERF-like, zink finger domain, homeodomain leucine zipper (HD-ZIP) and basic-helix-loop-helix (bHLH) transcription factors (TFs), as well as genes encoding histone H4 proteins (Appendix 4.3). The products of these regulatory genes are known to play important roles in plant stress responses that involve a complex interplay of activation and repression. Recent studies have suggested that signal transduction, upon recognition of MAMPS and PAMPS, is strongly dictated by regulatory networks involving transcription factors and associated co-factors (Tsuda and Somssich, 2015).

The majority of transcription factor genes, encoding MYB, AP2/ERF, bHLH, and HD-ZIP TFs, were up-regulated in this study. *MYB* genes were shown to be down-regulated in BN phytoplasma-infected Chardonnay, while up-regulated in a less susceptible cultivar, Manzoni Bianco (Hren *et al.*, 2009; Albertazzi *et al.*, 2009). MYB TFs are known to be master regulators of secondary cell wall

formation where they play a key role in the regulation of genes involved in biosynthesis of lignin, cellulose and hemicellulose (Ko *et al.*, 2014), as well as regulation of genes involved in phenylpropanoid metabolism (Liu *et al.*, 2015). Therefore, incomplete lignification of shoots in a susceptible cultivar such as Chardonnay is commonly associated with the repression of *MYB*, compared to tolerant cultivars where cell wall reinforcement is induced to limit the spread of infection (Hren *et al.*, 2009; Albertazzi *et al.*, 2009). Two grapevine R2R3-MYB-type TFs, *viz.* MYB14 and MYB15, were demonstrated to specifically activate the promoters of *STS* genes in response to biotic and abiotic stresses (See section 4.3.3.3). The subsequent increase in STS expression leads to the accumulation of glycosylated stilbenes *in planta* (Höll *et al.*, 2013), which can facilitate antimicrobial activity.

AP2/ERF TFs are plant-specific TFs that participate in the regulation of genes involved in disease resistance pathways, in particular those related to JA and ethylene signalling, as well as PR genes (Tsuda and Somssich, 2015; Licausi *et al.*, 2013). It was demonstrated that JA and ethylene signalling mediates plant defence against necrotrophic pathogens, such as the bacterial pathogen *Pectobacterium carotovorum* (Kwenda *et al.*, 2016), and fungal pathogens such as *Alternaria brassicicola*, *Botrytis cinerea*, and *Fusarium oxysporum*. JA signalling has also been shown to provide resistance against certain biotrophic and hemibiotrophic pathogens such as *Meloidogyne graminicola* and *Xanthomonas oryzae* in rice (Zhang *et al.*, 2017), and resistance to phytoplasma in Chinese jujube (Liu *et al.*, 2016). We identified two *AP2/ERF* genes that were significantly up-regulated in our study. This could imply that *AP2/ERF* expression was induced in the AY phytoplasma-infected Chardonnay to elicit a possible defence response (Figure 4.5).

Three *HD-Zip* genes were up-regulated in this study, including two *ATHB-12*-like genes and an *AtHB1*-like gene. In *Arabidopsis* they were shown to play a tissue-specific developmental role following analyses of constitutive-expressed and knock-down mutant phenotypes (Hur *et al.*, 2015; Capella *et al.*, 2015). Four *bHLH* genes were also significantly up-regulated in the AY phytoplasma-infected grapevines. In plants these genes are usually expressed in a tissue-specific and developmental stage-dependent manner. They are predominantly involved in the regulation of flowering time and anthocyanin biosynthesis (Vimolmangkang *et al.*, 2013; Wu *et al.*, 2015; Li *et al.*, 2017). The biological functions of most members of this gene family are still largely unknown and few have been associated with a role in plant immunity (Rahaie *et al.*, 2013).

4.3.4 Validation of differential gene expression using RT-qPCR

In order to evaluate technical and biological variation in transcript levels, following mRNA-seq analysis, RT-qPCR analysis was used to determine the relative expression of four DEGs, which were possibly involved in host response to AY phytoplasma-infection. These genes encode a thaumatin-like protein (*PR-5*; *GSVIVT01019848001*), a pathogenesis-related 1 protein (*PR-1*; *GSVIVT01037005001*), a peroxidase N1-like protein (*GSVIVT01029774001*), and a homeodomain leucine zipper HTHB-12 (*GSVIVT01019655001*), respectively. Results confirmed their up-regulation (Figure 4.6). In addition, linear regression analysis demonstrated good correlation between the RT-qPCR results and the $\log_2$-ratios determined with the mRNA-seq data, with a goodness of fit (R^2) of ~0.96 (Figure 4.7). These result were used to confirm the trend of expression of significant DEGs.

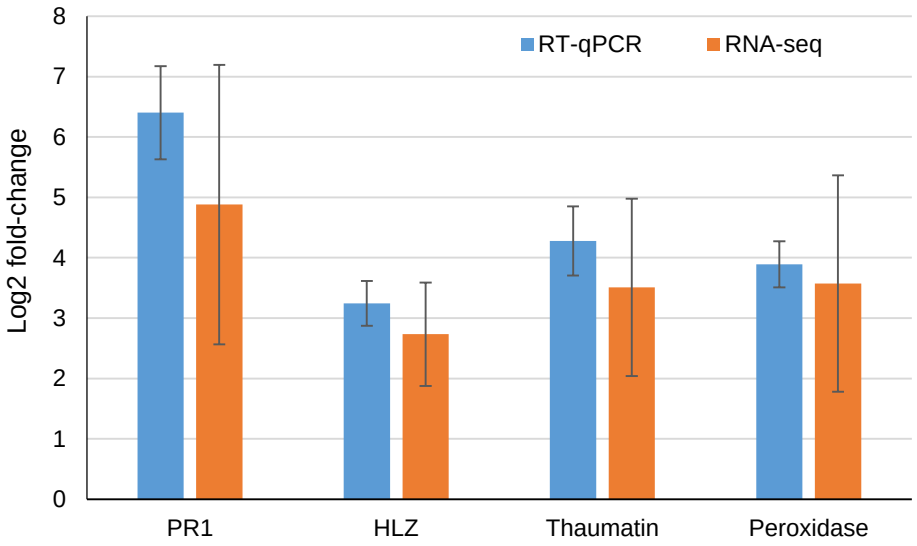

Figure 4.6: RT-qPCR confirmation of differential expression of four DEGs, *viz. Thaumatin-like protein, pathogenesis-related 1 protein* (*PR1*), *Peroxidase N1-like protein*, and a *Homeodomain-leucine zipper HTHB-12* (*HLZ*). The $\log_2$ fold-change in expression of each gene determined with real-time RT-qPCR analysis is plotted for comparison with $\log_2$ ratios obtained using FPKM values in the mRNA-seq data. Vertical bars indicate the standard error (SE).

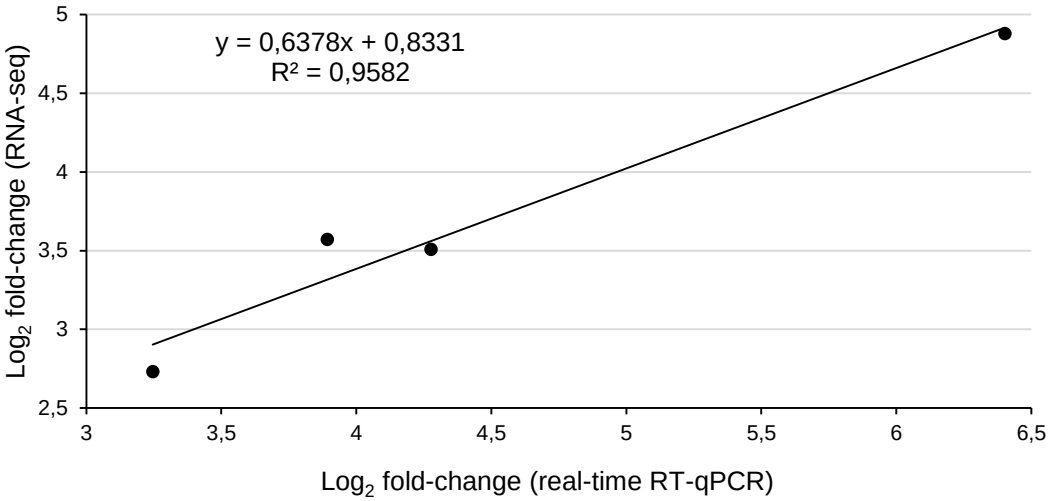

Figure 4.7: Comparison between gene expression ratios reported for mRNA-seq and RT-qPCR analysis.

4.4 Conclusions

In summary, we used RNA-seq to sequence the mRNA transcriptome of *V. vinifera* cv. 'Chardonnay' infected with AY phytoplasma, the causal agent of GY disease in South African vineyards. The aim was to explore changes in gene expression in order to understand the host response pathways induced by the disease. Therefore, we sequenced the leaf transcriptomes from three healthy (control) plants and three infected (experimental) plants. Following acquisition of 178,039,188 paired-end reads, ~86% was successfully mapped to the annotated reference genome. The remaining unmapped reads may be a result of genotypic reference bias introduced by the Pinot noir reference genome available at the time. DEGs were obtained by comparing transcript abundances between healthy and AY phytoplasma-infected plants. GO analysis was used to assign functional annotations to the DEGs in an attempt to understand their involvement in plant-pathogen interactions.

The major DEGs were involved in plastid and cell wall metabolism/architecture, signalling, innate immunity, pathogen defence, secondary metabolism and photosynthesis. DEGs involved in plastid and cell wall metabolism/architecture may potentially be correlated with the severe symptom development seen in susceptible cultivars such as Chardonnay. Interestingly, our results revealed significant up-regulation of multiple pathogenesis-related (*PR*) genes and salicylate carboxymethyltransferase paralogs involved in SA-dependent SAR signalling. The identification of DEGs involved in ROS and chloroplast breakdown was also a strong indication of HR in the infected plant material. We also identified a susceptibility gene which could be targeted using a genome editing approach to acquire a higher level of tolerance or complete disease resistance.

This study may shed new light on phytoplasma pathogenicity and may help unravel the basis of host susceptibility to GY disease. Furthermore, our results may also set some grounds for developing control strategies to prevent further spread of phytoplasma-infection, and contribute to studies aiming to establish grapevine varieties with resistance to AY phytoplasma.

4.5 Supporting information

Appendix 4.1: Bioanalyzer high sensitivity DNA assay results for leaf total RNA quality (File attached separately)

Appendix 4.2: FASTQC results for quality-filtered paired-end reads (File attached separately)

Appendix 4.3: Functional annotations of DEGs (File attached separately)

4.6 References

Abbà S., Galetto L., Carle P., Carrère S., Delledonne M., *et al.* 2014. RNA-Seq profile of flavescence dorée phytoplasma in grapevine. *BMC Genomics* 15(1): 1088.

Ahmad J.N., Renaudin J. and Eveillard S. 2015. Molecular study of the effect of exogenous phytohormones application in stolbur phytoplasma infected tomatoes on disease development. *Phytopathogenic Mollicutes* 5: 121–122.

Albertazzi G., Milc J., Caffagni A., Francia E., Roncaglia E., *et al.* 2009. Gene expression in grapevine cultivars in response to Bois Noir phytoplasma infection. *Plant Science* 176(6): 792-804.

Apel K. and Hirt H. 2004. Reactive oxygen species: metabolism, oxidative stress, and signal transduction. *Annual Review of Plant Biology* 55(1): 373-399.

Armijo G., Schlechter R., Agurto M., Muñoz D., Nuñez C., *et al.* 2016. Grapevine pathogenic microorganisms: understanding infection strategies and host response scenarios. *Frontiers in Plant Science* 7: 382.

Balakireva A.V. and Zamyatnin A.A. 2018. Indispensable role of proteases in plant innate immunity. *International Journal of Molecular Sciences* 19(2): 629.

Belhadj A., Saigne C., Telef N., Cluzet S., Bouscaut J., *et al.* 2006. Methyl jasmonate induces defense responses in grapevine and triggers protection against *Erysiphe necator*. *Journal of Agricultural and Food Chemistry* 54(24): 9119-9125.

Belli G., Bianco P.A. and Conti M. 2010. Grapevine yellows in Italy: past, present and future. *Journal of Plant Pathology* 92(2): 303-326.

Benjamini Y.H., Y 1995. Controlling the false discovery rate: A practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society* 57(1): 289-300.

Bertamini M., Grando M.S., Muthuchelian K. and Nedunchezhian N. 2002a. Effect of phytoplasmal infection on photosystem II efficiency and thylakoid membrane protein changes in field grown apple (*Malus pumila*) leaves. *Physiological and Molecular Plant Pathology* 61(6): 349-356.

Bertamini M. and Nedunchezhian N. 2001. Effects of phytoplasma [stolbur-subgroup (Bois noir-BN)] on photosynthetic pigments, saccharides, ribulose 1,5-biphosphate carboxylase, nitrate and nitrite reductases, and photosynthetic activities in field-grown grapevine (*Vitis vinifera* L. cv Chardonnay) leaves. *Photosynthetica* 39: 119-122.

Bertamini M., Nedunchezhian N., Tomasi F. and Grando M.S. 2002b. Phytoplasma [Stolbur-subgroup (Bois Noir-BN)] infection inhibits photosynthetic pigments, ribulose-1,5-bisphosphate carboxylase and photosynthetic activities in field grown grapevine (*Vitis vinifera* L. cv. Chardonnay) leaves. *Physiological and Molecular Plant Pathology* 61(6): 357-366.

Bester R., Jooste A.E.C., Maree H.J. and Burger J.T. 2012. Real-time RT-PCR high-resolution melting curve analysis and multiplex RT-PCR to detect and differentiate grapevine leafroll-associated virus 3 variant groups I, II, III and VI. *Virology Journal* 9: 219-219.

Bilgin D.D., Zavala J.A., Zhu J.I.N., Clough S.J., Ort D.R., *et al.* 2010. Biotic stress globally downregulates photosynthesis genes. *Plant, Cell & Environment* 33(10): 1597-1613.

Bradley D.J., Kjellbom P. and Lamb C.J. 1992. Elicitor- and wound-induced oxidative cross-linking of a proline-rich plant cell wall protein: A novel, rapid defense response. *Cell* 70(1): 21-30.

Camejo D., Guzmán-Cedeño Á. and Moreno A. 2016. Reactive oxygen species, essential molecules, during plant–pathogen interactions. *Plant Physiology and Biochemistry* 103: 10-23.

Capella M., Ribone P.A., Arce A.L. and Chan R.L. 2015. *Arabidopsis thaliana* HomeoBox 1 (AtHB1), a Homedomain-Leucine Zipper I (HD-Zip I) transcription factor, is regulated by PHYTOCHROME-INTERACTING FACTOR 1 to promote hypocotyl elongation. *New Phytologist* 207(3): 669-682.

Carstens R. 2014. *The incidence and distribution of grapevine yellows disease in South African vineyards.* Master's degree thesis, Department of Genetics, Stellenbosch University. http://scholar.sun.ac.za/handle/10019.1/86683

Chen X.-Y. and Kim J.-Y. 2009. Callose synthesis in higher plants. *Plant Signaling & Behavior* 4(6): 489-492.

Chung K., Igari K., Uchida N. and Tasaka M. 2008. New perspectives on plant defense responses through modulation of developmental pathways. *Molecules and Cells* 26(2): 107-12.

Coetzee B. 2010. *A metagenomic approach using next-generation sequencing for viral profiling of a vineyard and genetic characterization of grapevine virus E.* Master's degree thesis, Department of Genetics, Stellenbosch University. http://scholar.sun.ac.za/handle/10019.1/5186

Conesa A. and Götz S. 2008. Blast2GO: A comprehensive suite for functional analysis in plant genomics. *International Journal of Plant Genomics* 2008: 619832.

Contaldo N., Satta E., Zambon Y., Paltrinieri S., Bertaccini A. 2016. Development and evaluation of different complex media for phytoplasma isolation and growth. *Journal of Microbiological Methods* 127: 105-110.

Dai X. and Zhao P.X. 2011. psRNATarget: a plant small RNA target analysis server. *Nucleic Acids Research* 39: W155-W159.

De Toledo Thomazella D.P., Brail Q., Dahlbeck D. and Staskawicz B.J. 2016. CRISPR-Cas9 mediated mutagenesis of a DMR6 ortholog in tomato confers broad-spectrum disease resistance. *bioRxiv* 064824: 1-23.

Dermastia M., Nikolić P., Chersicola M. and Gruden K. 2015. Transcriptional profiling in infected and recovered grapevine plant responses to '*Candidatus* Phytoplasma solani'. *Phytopathogenic Mollicutes* 5: 123–124.

Ebrahim S., Usha K. and Singh B. 2011. Pathogenesis related (PR) proteins in plant defense mechanism. *In: Science against microbial pathogens: communicating current research and technological advances.* Méndez-Vilas, A. (ed.). pp.1043-1054.

Engelbrecht M., Joubert J. and Burger J.T. 2010. First report of aster yellows phytoplasma in grapevines in South Africa. *Plant Disease* 94(3): 373-373.

Enoki S. and Suzuki S. 2016. Pathogenesis-related proteins in grape. *In: Grape and Wine Biotechnology*. Morata, A. & Loira, I. (eds.). IntechOpen. pp.43-58.

https://www.intechopen.com/books/grape-and-wine-biotechnology/pathogenesis-related-proteins-in-grape

Fagoaga C., Rodrigo I., Conejero V., Hinarejos C., Tuset J.J., *et al.* 2001. Increased tolerance to *Phytophthora citrophthora* in transgenic orange plants constitutively expressing a tomato pathogenesis related protein PR-5. *Molecular Breeding* 7(2): 175-185.

Fan G., Niu S., Xu T., Deng M., Zhao Z., *et al.* 2015. Plant–pathogen interaction-related microRNAs and their targets provide indicators of phytoplasma infection in *Paulownia tomentosa* × *Paulownia fortunei*. *PLoS ONE* 10(10): e0140590.

Farrow S.C. and Facchini P.J. 2014. Functional diversity of 2-oxoglutarate/Fe(II)-dependent dioxygenases in plant metabolism. *Frontiers in Plant Science* 5: 524.

Fan W. and Dong X. 2002. *In vivo* interaction between NPR1 and transcription factor TGA2 leads to salicylic acid-mediated gene activation in *Arabidopsis*. *The Plant Cell* 14(6): 1377-1389.

Fan X.-P., Liu W., Qiao Y.-S., Shang Y.-J., Wang G.-P., *et al.* 2017. Comparative transcriptome analysis of *Ziziphus jujuba* infected by jujube witches' broom phytoplasmas. *Scientia Horticulturae* 226: 50-58.

Finn R.D., Attwood T.K., Babbitt P.C., Bateman A., Bork P., *et al.* 2017. InterPro in 2017 - beyond protein family and domain annotations. *Nucleic Acids Research* 45: D190-D199.

Frías M., González C. and Brito N. 2011. BcSpl1, a cerato-platanin family protein, contributes to *Botrytis cinerea* virulence and elicits the hypersensitive response in the host. *New Phytologist* 192(2): 483-495.

Gambino G., Boccacci P., Margaria P., Palmano S. and Gribaudo I. 2013. Hydrogen peroxide accumulation and transcriptional changes in grapevines recovered from flavescence dorée disease. *Phytopathology* 103(8): 776-784.

Gao Q.-M., Zhu S., Kachroo P. and Kachroo A. 2015. Signal regulators of systemic acquired resistance. *Frontiers in Plant Science* 6: 228.

Giacomelli L., Zeilmaker T., Malnoy M., Velasco R., Viola R., *et al.* 2017. Generation of mildew-resistant grapevine clones via genome editing. 20th GiESCO International Meeting: "Sustainable viticulture and winemaking in climate change scenarios", Mendoza, Argentina, 5-10 November 2017, p.965-969.

Giorno F., Guerriero G., Biagetti M., Ciccotti A.M. and Baric S. 2013. Gene expression and biochemical changes of carbohydrate metabolism in in vitro micro-propagated apple plantlets infected by '*Candidatus* Phytoplasma mali'. *Plant Physiology and Biochemistry* 70: 311-317.

Gundersen D.E. and Lee I.M. 1996. Ultrasensitive detection of phytoplasmas by nested-PCR assays using two universal primer pairs. *Phytopathologia Mediterranea* 35(3): 144-151.

Gutha L.R., Casassa L.F., Harbertson J.F. and Naidu R.A. 2010. Modulation of flavonoid biosynthetic pathway genes and anthocyanins due to virus infection in grapevine (*Vitis vinifera* L.) leaves. *BMC Plant Biology* 10: 187-187.

Hamiduzzaman M.M., Jakab G., Barnavon L., Neuhaus J.-M. and Mauch-Mani B. 2005. β-Aminobutyric acid-induced resistance against downy mildew in grapevine acts through the potentiation of callose formation and jasmonic acid signaling. *Molecular Plant-Microbe Interactions* 18(8): 819-829.

Hasan D.-M. and Bae H. 2017. An overview of stress-induced resveratrol synthesis in grapes: perspectives for resveratrol-enriched grape products. *Molecules* 22: 294.

Hellemans J., Mortier G., De Paepe A., Speleman F. and Vandesompele J. 2007. qBase relative quantification framework and software for management and automated analysis of real-time quantitative PCR data. *Genome Biology* 8: R19.

Ho H.L. 2015. Functional roles of plant protein kinases in signal transduction pathways during abiotic and biotic Stress. *Journal of Biodiversity, Bioprospecting and Development* 2(2): 147.

Höll J., Vannozzi A., Czemmel S., D'onofrio C., Walker A.R., *et al.* 2013. The R2R3-MYB transcription factors MYB14 and MYB15 regulate stilbene biosynthesis in *Vitis vinifera*. *The Plant Cell* 25(10): 4135-4149.

Hren M., Nikolić P., Rotter A., Blejec A., Terrier N., *et al.* 2009. 'Bois noir' phytoplasma induces significant reprogramming of the leaf transcriptome in the field grown grapevine. *BMC Genomics* 10(1): 460.

Hückelhoven R. 2007. Cell wall-associated mechanisms of disease resistance and susceptibility. *Annual Review of Phytopathology* 45(1): 101-127.

Hur Y.-S., Um J.-H., Kim S., Kim K., Park H.-J., *et al.* 2014. *Arabidopsis thaliana* homeobox 12 (ATHB12), a homeodomain-leucine zipper protein, regulates leaf growth by promoting cell expansion and endoreduplication. *New Phytologist* 205(1): 316-328.

Jeandet P., Douillet-Breuil A.-C., Bessis R., Debord S., Sbaghi M., *et al.* 2002. Phytoalexins from the Vitaceae: biosynthesis, phytoalexin gene expression in transgenic plants, antifungal activity, and metabolism. *Journal of Agricultural and Food Chemistry* 50(10): 2731-2741.

Jones J.D.G. and Dangl J.L. 2006. The plant immune system. *Nature* 444: 323.

Kariola T., Brader G., Li J. and Palva E.T. 2005. Chlorophyllase 1, a damage control enzyme, affects the balance between defense pathways in plants. *The Plant Cell* 17(1): 282-294.

Karpinski S., Gabrys H., Mateo A., Karpinska B. and Mullineaux P.M. 2003. Light perception in plant disease defence signalling. *Current Opinion in Plant Biology* 6(4): 390-396.

Ko S.-S., Li M.-J., Sun-Ben Ku M., Ho Y.-C., Lin Y.-J., *et al.* 2014. The bHLH142 transcription factor coordinates with TDR1 to modulate the expression of *EAT1* and regulate pollen development in rice. *The Plant Cell* 26(6): 2486.

Kruger N.J. 1998. Carbohydrate synthesis and degradation. *In: Plant Metabolism.* Dennis, T., Layzell, D., Lefebvre, D. & Turpin, D. (eds.). Addison Wesley Longman Ltd. pp.83-104.

Kumar S.A., Kumari P.H., Kumar G.S., Mohanalatha C. and Kishor P.B.K. 2015. Osmotin: a plant sentinel and a possible agonist of mammalian adiponectin. *Frontiers in Plant Science* 6: 163.

Kwenda S., Motlolometsi T.V., Birch P.R.J. and Moleleki L.N. 2016. RNA-seq profiling reveals defense responses in a tolerant potato cultivar to stem infection by *Pectobacterium carotovorum* ssp. Brasiliense. *Frontiers in Plant Science* 7: 1905.

Landi L. and Romanazzi G. 2011. Seasonal variation of defense-related gene expression in leaves from Bois noir affected and recovered grapevines. *Journal of Agricultural and Food Chemistry* 59(12): 6628-6637.

Le Henanff G., Heitz T., Mestre P., Mutterer J., Walter B., *et al.* 2009. Characterization of *Vitis vinifera* NPR1 homologs involved in the regulation of Pathogenesis-Related gene expression. *BMC Plant Biology* 9: 54-54.

Lee I.M. 2004. '*Candidatus* Phytoplasma asteris', a novel phytoplasma taxon associated with aster yellows and related diseases. *International Journal of Systematic and Evolutionary Microbiology* 54(4): 1037-1048.

Lee I.-M., Davis R.E. and Gundersen-Rindal D.E. 2000. Phytoplasma: Phytopathogenic Mollicutes 1. *Annual Review of Microbiology* 54(1): 221-255.

Li C., Qiu J., Ding L., Huang M., Huang S., *et al.* 2017. Anthocyanin biosynthesis regulation of DhMYB2 and DhbHLH1 in Dendrobium hybrids petals. *Plant Physiology and Biochemistry* 112: 335-345.

Li L. and Steffens J.C. 2002. Overexpression of polyphenol oxidase in transgenic tomato plants results in enhanced bacterial disease resistance. *Planta* 215(2): 239-247.

Li Z.T., Dhekney S.A. and Gray D.J. 2011. PR-1 gene family of grapevine: a uniquely duplicated PR-1 gene from a *Vitis* interspecific hybrid confers high level resistance to bacterial disease in transgenic tobacco. *Plant Cell Reports* 30(1): 1-11.

Licausi F., Ohme-Takagi M. and Perata P. 2013. APETALA2/Ethylene Responsive Factor (AP2/ERF) transcription factors: mediators of stress responses and developmental programs. *New Phytologist* 199(3): 639-649.

Liu J., Osbourn A. and Ma P. 2015. MYB transcription factors as regulators of phenylpropanoid metabolism in plants. *Molecular Plant* 8(5): 689-708.

Liu L.Y.D., Tseng H.I., Lin C.P., Lin Y.Y., Huang Y.H., *et al.* 2014. High-throughput transcriptome analysis of the leafy flower transition of *Catharanthus roseus* induced by peanut witches'-broom phytoplasma infection. *Plant and Cell Physiology* 55(5): 942-957.

Liu R., Dong Y., Fan G., Zhao Z., Deng M., *et al.* 2013. Discovery of genes related to witches broom disease in *Paulownia tomentosa* × *Paulownia fortunei* by a *de novo* assembled transcriptome. *PLoS ONE* 8(11): e80238.

Liu Z., Zhao J. and Liu M. 2016. *Photosynthetic responses to phytoplasma infection in Chinese jujube*.

Luge T., Kube M., Freiwald A., Meierhofer D., Seemüller E., *et al.* 2014. Transcriptomics assisted proteomic analysis of *Nicotiana occidentalis* infected by *Candidatus* Phytoplasma mali strain AT. *Proteomics* 14(16): 1882-1889.

Macdonald F.D. and Buchanan B.B. 1998. The reductive pentose phosphate pathway and its regulation. *In: Plant Metabolism.* Dennis, T., Layzell, D., Lefebvre, D. & Turpin, D. (eds.). Addison Wesley Longman. pp.299-313.

Mach J.M., Castillo A.R., Hoogstraten R. and Greenberg J.T. 2001. The *Arabidopsis*-accelerated cell death gene ACD2 encodes red chlorophyll catabolite reductase and suppresses the spread of disease symptoms. *Proceedings of the National Academy of Sciences of the United States of America* 98(2): 771-776.

Mackenzie D.J. 1997. A standard protocol for the detection of viruses and viroids using a reverse transcription polymerase chain reaction technique. *The Canadian Food Inspection Agency*. Document CPHBTRT-PCR1.00.

Maixner M. Grapevine yellows - current developments and unsolved questions. 15th meeting of the international council for the study of virus and virus-like diseases of the grapevine (ICVG), 2006 Stellenbosch, South Africa. 86-88.

Mardi M., Karimi Farsad L., Gharechahi J. and Salekdeh G.H. 2015. In-depth transcriptome sequencing of Mexican lime trees Infected with *Candidatus* Phytoplasma aurantifolia. *PLoS ONE* 10(7): e0130425.

Margaria P., Abbà S. and Palmano S. 2013. Novel aspects of grapevine response to phytoplasma infection investigated by a proteomic and phospho-proteomic approach with data integration into functional networks. *BMC Genomics* 14(1): 38.

Margaria P., Ferrandino A., Caciagli P., Kedrina O., Schubert A., *et al.* 2014. Metabolic and transcript analysis of the flavonoid pathway in diseased and recovered Nebbiolo and Barbera grapevines (*Vitis vinifera* L.) following infection by Flavescence dorée phytoplasma. *Plant, Cell & Environment* 37(9): 2183-2200.

Matile P., Hörtensteiner S. and Thomas H. 1999. Chlorophyll degradation. *Annual Review of Plant Physiology and Plant Molecular Biology* 50(1): 67-95.

Minato N., Himeno M., Hoshi A., Maejima K., Komatsu K., *et al.* 2014. The phytoplasmal virulence factor TENGU causes plant sterility by downregulating of the jasmonic acid and auxin pathways. *Scientific Reports* 4: 7399.

Mittler R., Vanderauwera S., Gollery M. and Van Breusegem F. 2004. Reactive oxygen gene network of plants. *Trends in Plant Science* 9(10): 490-498.

Mock H.-P., Heller W., Molina A., Neubohn B., Sandermann H., *et al.* 1999. Expression of uroporphyrinogen decarboxylase or coproporphyrinogen oxidase antisense RNA in tobacco induces pathogen defense responses conferring increased resistance to tobacco mosaic virus. *Journal of Biological Chemistry* 274(7): 4231-4238.

Monavarfeshani A., Mirzaei M., Sarhadi E., Amirkhani A., Khayam Nekouei M., *et al.* 2013. Shotgun proteomic analysis of the Mexican lime tree infected with '*Candidatus* Phytoplasma aurantifolia'. *Journal of Proteome Research* 12(2): 785-795.

Mou H.-Q., Lu J., Zhu S.-F., Lin C.-L., Tian G.-Z., *et al.* 2013. Transcriptomic analysis of *Paulownia* infected by Paulownia witches'-broom phytoplasma. *PLoS ONE* 8(10): e77217.

Mou Z., Fan W. and Dong X. 2003. Inducers of plant systemic acquired resistance regulate NPR1 function through redox changes. *Cell* 113(7): 935-944.

Musetti R., Marabottini R., Badiani M., Martini M., Sanità Di Toppi L., *et al.* 2007. On the role of H_2O_2 in the recovery of grapevine (*Vitis vinifera* cv. Prosecco) from Flavescence dorée disease. *Functional Plant Biology* 34(8): 750-758.

Nasir F., Pervaiz Akhtar K., Hameed A., Yousaf S., Gulzar T., *et al.* 2017. Biochemical alterations in the leaves of different Desi and Kabuli type chickpea genotypes infected by phytoplasma. *Turkish Journal of Biochemistry.*

Ngadze E., Icishahayo D., Coutinho T.A. and Van Der Waals J.E. 2011. The role of polyphenol oxidase, peroxidase, phenylalanine ammonia lyase, chlorogenic acid and total soluble phenols in the resistance of potatoes to soft rot. *Plant Disease* 96(2): 186-192.

Noach L.C. 2010. *The molecular characterization of South African isolates of grapevine rupestris stem pitting-associated virus (GRSPaV).* Master's degree thesis, Department of Genetics, Stellenbosch University. http://scholar.sun.ac.za/handle/10019.1/5252

Oide S., Bejai S., Staal J., Guan N., Kaliff M., *et al.* 2013. A novel role of PR2 in abscisic acid (ABA) mediated, pathogen-induced callose deposition in *Arabidopsis thaliana*. *New Phytologist* 200(4): 1187-1199.

Orlovskis Z. and Hogenhout S.A. 2016. A bacterial parasite effector mediates insect vector attraction in host plants independently of developmental changes. *Frontiers in Plant Science* 7: 885.

Osakabe Y., Yamaguchi-Shinozaki K., Shinozaki K. and Tran L.-S.P. 2013. Sensing the environment: key roles of membrane-localized kinases in plant perception and response to abiotic stress. *Journal of Experimental Botany* 64(2): 445-458.

Paolacci A.R., Catarcione G., Ederli L., Zadra C., Pasqualini S., *et al.* 2017. Jasmonate-mediated defence responses, unlike salicylate-mediated responses, are involved in the recovery of grapevine from bois noir disease. *BMC Plant Biology* 17(1): 118.

Petrussa E., Braidot E., Zancani M., Peresson C., Bertolini A., *et al.* 2013. Plant flavonoids - biosynthesis, transport and involvement in stress responses. *International Journal of Molecular Sciences* 14(7): 14950-14973.

Prezelj N., Covington E., Roitsch T., Gruden K., Fragner L., *et al.* 2016. Metabolic consequences of infection of grapevine (*Vitis vinifera* L.) cv. "Modra frankinja" with Flavescence Dorée phytoplasma. *Frontiers in Plant Science* 7: 711.

Rafiqi M., Bernoux M., Ellis J.G. and Dodds P.N. 2009. In the trenches of plant pathogen recognition: Role of NB-LRR proteins. *Seminars in Cell and Developmental Biology* 20: 1017-1024.

Rahaie M., Xue G.-P. and Schenk P.M. 2013. The role of transcription factors in wheat under different abiotic stresses. *In: Abiotic Stress - Plant Responses and Applications in Agriculture.* Vahdati, K. & Leslie, C. (eds.). IntechOpen. pp.367-385.

Repka V. 2001. Elicitor-stimulated induction of defense mechanisms and defense gene activation in grapevine cell suspension cultures. *Biologia Plantarum* 44(4): 555-565.

Romeis T. 2001. Protein kinases in the plant defence response. *Current Opinion in Plant Biology* 4(5): 407-414.

Seskar M., Shulaev V. and Raskin I. 1998. Endogenous methyl salicylate in pathogen-inoculated tobacco plants. *Plant Physiology* 116(1): 387-392.

Stintzi A., Heitz T., Kauffmann S. and Fritig B. 1991. Identification of a basic pathogenesis-related, thaumatin-like protein of virus-infected tobacco as osmotin. *Physiological and Molecular Plant Pathology* 38(2): 137-146.

Sugio A., Maclean A.M., Kingdom H.N., Grieve V.M., Manimekalai R., *et al.* 2011. Diverse targets of phytoplasma effectors: from plant development to defense against insects. *Annual Review of Phytopathology* 49(1): 175-195.

Takamiya K.-I., Tsuchiya T. and Ohta H. 2000. Degradation pathway(s) of chlorophyll: what has gene cloning revealed? *Trends in Plant Science* 5(10): 426-431.

The UniRrot Consortium 2017. UniProt: the universal protein knowledgebase. *Nucleic Acids Research* 45(D1): D158-D169.

Torres M.A. 2010. ROS in biotic interactions. *Physiologia Plantarum* 138(4): 414-429.

Trapnell C., Pachter L. and Salzberg S.L. 2009. TopHat: discovering splice junctions with RNA-Seq. *Bioinformatics* 25(9): 1105-1111.

Trapnell C., Roberts A., Goff L., Pertea G., Kim D., *et al.* 2012. Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and Cufflinks. *Nature Protocols* 7(3): 562-578.

Trapnell C., Williams B.A., Pertea G., Mortazavi A., Kwan G., *et al.* 2010. Transcript assembly and abundance estimation from RNA-Seq reveals thousands of new transcripts and switching among isoforms. *Nature Biotechnology* 28(5): 511-515.

Truman W., Zabala M.T. and Grant M. 2006. Type III effectors orchestrate a complex interplay between transcriptional networks to modify basal defence responses during pathogenesis and resistance. *The Plant Journal* 46(1): 14-33.

Tsuda K. and Somssich I.E. 2015. Transcriptional networks in plant immunity. *New Phytologist* 206(3): 932-947.

Van Damme M., Huibers R.P., Elberse J. and Van Den Ackerveken G. 2008. *Arabidopsis DMR6* encodes a putative 2OG-Fe(II) oxygenase that is defense-associated but required for susceptibility to downy mildew. *The Plant Journal* 54(5): 785-793.

Vannozzi A., Dry I.B., Fasoli M., Zenoni S. and Lucchin M. 2012. Genome-wide analysis of the grapevine stilbene synthase multigenic family: genomic organization and expression profiles upon biotic and abiotic stresses. *BMC Plant Biology* 12: 130-130.

Vimolmangkang S., Han Y., Wei G. and Korban S.S. 2013. An apple MYB transcription factor, MdMYB3, is involved in regulation of anthocyanin biosynthesis and flower development. *BMC Plant Biology* 13: 176-176.

Wang J., Song L., Jiao Q., Yang S., Gao R., *et al.* 2018. Comparative genome analysis of jujube witches'-broom Phytoplasma, an obligate pathogen that causes jujube witches'-broom disease. *BMC Genomics* 19(1): 689.

Weigel R.R., Pfitzner U.M. and Gatz C. 2005. Interaction of NIMIN1 with NPR1 modulates PR gene expression in *Arabidopsis*. *The Plant Cell* 17(4): 1279-1291.

Winkel-Shirley B. 2001. Flavonoid biosynthesis. A colorful model for genetics, biochemistry, cell biology, and biotechnology. *Plant Physiology* 126(2): 485-493.

Wu Z.-G., Jiang W., Mantri N., Bao X.-Q., Chen S.-L., *et al.* 2015. Transciptome analysis reveals flavonoid biosynthesis regulation and simple sequence repeats in yam (Dioscorea alata L.) tubers. *BMC Genomics* 16(1): 346.

Zafari S., Niknam V., Musetti R. and Noorbakhsh S.N. 2012. Effect of phytoplasma infection on metabolite content and antioxidant enzyme activity in lime (*Citrus aurantifolia*). *Acta Physiologiae Plantarum* 34(2): 561-568.

Zaidi S.S.-E.-A., Mukhtar M.S. and Mansoor S. 2018. Genome Editing: Targeting Susceptibility Genes for Plant Disease Resistance. *Trends in Biotechnology* 36(9): 898-906.

Zeilmaker T., Ludwig N.R., Elberse J., Seidl M.F., Berke L., *et al.* 2014. DOWNY MILDEW RESISTANT 6 and DMR6-LIKE OXYGENASE 1 are partially redundant but distinct suppressors of immunity in *Arabidopsis*. *The Plant Journal* 81(2): 210-222.

Zhang L., Zhang F., Melotto M., Yao J. and He S.Y. 2017. Jasmonate signaling and manipulation by pathogens and insects. *Journal of Experimental Botany* 68(6): 1371-1385.

Zhong B.-X. and Shen Y.-W. 2004. Accumulation of pathogenesis-related type-5 like proteins in phytoplasma infected garland chrysanthemum *Chrysanthemum coronarium*. *Acta Biochimica et Biophysica Sinica* 36(11): 773–779.

Chapter 5

CONCLUDING REMARKS

Grapevine yellows (GY) is a detrimental disease of grapevine and responsible for devastating yield losses that pose a serious threat to the wine and table grape industries. Mycoplasma-like organisms known as phytoplasmas have been clearly described as the causal agents of the disease, despite the fact that Koch's postulates have not been fully met. Phytoplasmas colonise the nutrient-rich phloem tissue of infected plants where they can move and multiply through the sieve tube elements. GY is associated with an array of similar deleterious symptoms, depending on interference with the host physiology, but differs in aetiology and epidemiology. The principle step towards understanding host susceptibility and developing preventative control strategies would be grasping the complex molecular interactions between a phytoplasma and its host.

NGS approaches have been invaluable towards elucidating gene regulatory mechanisms and understanding pathogen response pathways. Therefore, the aim of this study was to utilise high-throughput transcriptome sequencing methods and a comparative profiling approach to identify differentially expressed miRNAs and mRNAs in AY phytoplasma-infected Chardonnay. In order to capture a true-life situation, healthy and infected leaf material were collected from field-grown plants within the same vineyard.

sRNA-seq and subsequent bioinformatics analysis were used to identify known vvi-miRNAs in Chardonnay leaf material on a genome-wide scale. Both canonical miRNAs, as well as known sequence variants (isomiRs) were identified, with isomiRs of vvi-miR166 being the dominantly expressed miRNAs in both the healthy and infected samples. Different findings suggest that multiple isomiRs can arise from a single hairpin precursor as a result of a biological relevant process. Previous studies revealed that miR166 regulate an HDZIP-III transcription factor involved in secondary cell wall biosynthesis (Du and Wang, 2015). This infers complex miRNA processing mechanisms involved during the regulation leaf and vascular morphogenesis. Certain candidates of known miRNA families were differentially expressed in the AY phytoplasma-infected samples. Some of these were previously characterised as plant developmental miRNAs, making them potential candidates involved in GY symptom development. On a different note, the pre-miR166 sequence may serve as the potential 'backbone' in an artificial miRNA (amiRNA) construct, because miR166 is highly conserved among different plants and mosses, and contains the necessary sequence parameters for sufficient target selection (Schwab *et al*., 2005). An amiRNA vector can

potentially be used in transient expression assays to either silence a virus or endogenous gene(s) of interest during functional studies (Sablok *et al*., 2011).

Two software packages were used to identify a large group of new pre-miRNA sequences, each derived from a previously uncharacterised genome location. Some of these sequences were homologues to known plant miRNAs, and were therefore classified into these known miRNA families. Such miRNA sequence data will expand the current miRNA knowledgebase and be used in future sRNA-based research on plant stress and development. The new miRNAs identified in this study were recently catalogued in the miRVIT database and used to analyse miRNA-mediated responses in Flavescence dorée infected grapevine (Chitarra *et al.*, 2018). Comparative sRNA profiling also revealed significant differential expression of some of the new miRNAs in the AY phytoplasma-infected leaf material. Functional characterisation of these vvi-miRNAs may increase knowledge on pathogenesis-related pathways in grapevine.

Possible miRNA targets identified in this study corresponded to targets found in numerous plant species, most of them encoding transcription factors. These targets are involved in major processes such as plant morphology and architecture, hormone signalling, nutrient homeostasis, hydrolase activity, and signal transduction. Some of our findings correspond to previous studies with similar miRNA targets. Therefore, hypothetical associations between miRNAs and physiological mechanisms were provided that may be directly correlated with GY disease symptom development. Interestingly, a transcript (*GSVIVG01027229001*) with a high degree of sequence similarity to as NBS-LRR disease resistance gene was identified as a potential target of a highly up-regulated novel miRNA (vvi-miRn027). Some NBS-LRR proteins are involved in pathogen detection and defence signalling (Rafiqi *et al.*, 2009). Assuming this gene is indeed a true target of vvi-miRn027, both can be ideal candidates in transient expression assays aiming to increase tolerance to GY disease. However, *in vitro* studies are required to confirm this.

mRNA-seq, followed by comparative gene profiling enabled the identification of differentially expressed genes (DEGs) involved in plastid and cell wall metabolism/architecture, signalling, innate immunity, pathogen defence, secondary metabolism and photosynthesis. DEGs involved in plastid and cell wall metabolism/architecture, ROS metabolism and photosynthesis may be involved in the development of GY symptoms, commonly seen in susceptible cultivars such as Chardonnay. A group of genes encoding receptor-like protein kinases (RLKs) were differentially expressed in the AY phytoplasma-infected leaves. Certain RLKs are frequently involved in pathogen recognition and subsequent activation of kinase signalling pathways (Rafiqi *et al.*, 2009). Although up-regulation of these genes may suggest their potential role in pathogen-associated molecular pattern

(PAMP)-triggered immunity (PTI), as a result of AY phytoplasma sensing, this mechanism remains to be investigated.

In agreement with similar studies on plant-pathogen interactions, different pathogenesis-related (*PR*) genes were up-regulated in response to AY phytoplasma-infection. PR protein expression suggested the existence of basal defence pathways as part of a hypersensitive response (HR), and further emphasised the ancestral role of these proteins during biotic stress. *PR* genes are usually up-regulated during a sharp increase of jasmonate (JA) and salicylic acid (SA) (Balakireva and Zamyatnin, 2018), suggesting the involvement of effector triggered immunity (ETI) response pathways during AY phytoplasma-infection. Moreover, the potential coordinate up-regulation of *PR* genes and salicylate carboxymethyltransferase paralogs also suggested a SA-mediated response, often present during systemic acquired resistance (SAR) signalling. Additional research would be required to explore the involvement of JA and SA-mediated defence responses upon AY phytoplasma-infection.

Plant defence responses during insect feeding have been well documented and add an additional level of complexity to the activation of phytohormone signalling pathways (Lazebnik *et al*., 2014). Evidence exists for the significant crosstalk between plant-insect or plant-pathogen interactions. Several transcriptome studies demonstrated a significant overlap in gene expression, including genes involved in defense hormone signaling, across PAMP and herbivory-associated interaction molecular pattern (HAMP) responses that activated pattern recognition receptors (PRRs) and PTI (Campos *et al*., 2014; Zhang *et al*., 2016). Phytophagous insects such as leafhoppers have developed certain strategies to manipulate the JA and SA pathways to gain easier access to plant nutrients (Kallenbach *et al*., 2012; Cowles *et al*., 2018). This includes vectoring and transmission of microorganisms, such as phytoplasmas, to manipulate plant defence signalling, and would add an extra layer of complexity to the interpretation of data on signalling pathways. The use of natural phytoplasma-infection of field-grown plants has the complication that the bacterium is introduced into the plant by the insect, so presumably the plant is interacting initially to both the presence of the insect and the presence of the bacterium. In an attempt to account for this, RNA was extracted from AY phytoplasma-infected material that were collected during the growing season when the highest number of positive diagnoses was obtained, as was shown in a previous study (Smyth, 2015). The use of a controlled transmission experiment to monitor temporal gene expression would greatly expand our knowledge on the intracate molecular responses to AY phytoplasma-infection.

Genes involved in the production and neutralisation of reactive oxygen species (ROS) were also differentially expressed in the AY phytoplasma-infected leaves. Manzoni Bianco is known as a highly phytoplasma-tolerant grapevine cultivar that displays a strong down-regulation of ROS-

scavengers (Albertazzi *et al.*, 2009). In the AY phytoplasma-infected Chardonnay leaves, however, the ROS-scavenger genes encoding polyphenol oxidase (PPO) and glutathione S-transferase (GST) were up-regulated more than two-fold. In order to reduce titre or prevent further phytoplasma-infection, sustained accumulation of ROS, such as H_2O_2, may be necessary (Musetti *et al.*, 2007). Therefore, transient and stable expression of ROS-scavenger genes may be used to engineer Chardonnay plants with enhanced tolerance to AY phytoplasma.

Among the genes involved in secondary metabolism, a paralog of *Downy mildew resistance 6* (*DMR6*), called *DMR6-like oxygenase 2* (or *DLO2*), was up-regulated in this study and may have a role in host susceptibility to AY phytoplasma. This may be an important discovery for future studies since the advent of gene editing tools, particularly the CRISPR/Cas9 system, can be used to potentially obstruct susceptibility (*S*) genes expression in order to engineer grapevines with resistance to a number of diseases, including GY. Previous studies have described these "*S* genes" as targets which can be knocked out to gain broad-spectrum and durable disease resistance (Zaidi *et al.*, 2018).

Prior to analysing the mRNA-seq data, a complementary-based *in silico* approach was used to identify putative targets for differentially expressed known and novel vvi-miRNAs. Unexpectedly, none of the DEGs identified in the mRNA-seq transcriptome data displayed anti-correlation to the differentially expressed miRNAs, even though miRNA expression was validated using RT-qPCR. We can only speculate on whether these miRNA/mRNA interactions were indeed leaf-specific, or whether they were too subtle to deliver significant changes in anti-correlated transcript levels. The level of precision of the miRNA target prediction software is also brought into question. Evidence suggest that many of these programs predict a large number of targets that are often biologically irrelevant or false positives (Karbiener *et al*., 2014; Pinzón *et al*., 2017; Seitz, 2017). Even though the number of validated miRNA targets is still proportionately small, many programs favour certain prediction algorithms while underrating others (Liu *et al*., 2014). In addition, some algorithms and databases rely heavily on sequence or structure criteria, physical binding properties, experimental validations, and even expression levels. It should be emphasised that not many programs exist that incorporate plant-specific parameters and that prediction criteria are mostly based on miRNA targets derived from a wide range of species, under different conditions.

The use of the Pinot noir mRNAs dataset may also have been sub-optimal for the purpose of identifying miRNA targets in this study, since some miRNA(s) may not guide cleavage of their complementary mRNA(s) under certain conditions, and other non-conserved miRNAs may even lack functional targets. Future studies would greatly benefit from the use of stable and robust

degradome sequencing data to predict miRNA targets, combined with RT-qPCR to validate their trend of expression.

sRNA-seq can provide a plethora of sequence data, not only for miRNAs but for many other sRNA species. These sRNAs can be involved in sRNA silencing pathways that play an important role in development and plant-pathogen interactions (Chen, 2012; Peláez and Sanchez, 2013). Therefore, the use of bioinformatics analysis to interrogate the remainder of the sRNA fraction of sequences may be used to identify other sRNAs potentially associated with AY phytoplasma-infection. Integrating such data may further elucidate molecular mechanisms involved during phytoplasma pathogenesis.

A similar study on domesticated cultivars or wild varieties of grapevine, displaying different degrees of susceptibility to AY phytoplasma, may potentially reveal a naturally occurring mechanism for GY resistance, and remains open for investigation. The development of an interactive "omics" resource, tailored specifically for *Vitis* spp., will provide a platform for the storage and analysis of high-throughput genome and transcriptome data. Including a database with experimentally validated expression profiles of different sRNAs and their targets, under different developmental and stress conditions, may provide clarity during functional characterisation of sRNAs.

This study presents the first report on the modulation of miRNAs and genes associated with AY phytoplasma-infection in Chardonnay. The use of high-throughput transcriptome sequencing and bioinformatics analysis provided valuable information on a number of miRNAs and genes that could potentially be linked to host susceptibility and symptom development in Chardonnay with GY disease. These results allow for greater biological insight into plant-pathogen reactions related to AY phytoplasma-infection, and may be utilised in preventative strategies to either mitigate the scale of infection or even engineer grapevines with resistance to the pathogen.

References

Albertazzi G., Milc J., Caffagni A., Francia E., Roncaglia E., *et al.* 2009. Gene expression in grapevine cultivars in response to Bois Noir phytoplasma infection. *Plant Science* 176(6): 792-804.

Balakireva A.V. and Zamyatnin A.A. 2018. Indispensable role of proteases in plant innate immunity. *International Journal of Molecular Sciences* 19(2): 629.

Campos M.L., Kang J.-H. and Howe G.A. 2014. Jasmonate-triggered plant immunity. *Journal of Chemical Ecology* 40(7): 657-675.

Chen X. 2012. Small RNAs in development - insights from plants. *Current Opinion in Genetics and Development* 22(4): 361-367.

Chitarra W., Pagliarani C., Abbà S., Boccacci P., Birello G., *et al.* 2018. miRVIT: a novel miRNA database and its application to uncover *Vitis* responses to Flavescence dorée infection. *Frontiers in Plant Science* 9: 1034-1034.

Cowles K.N., Groves R.L. and Barak J.D. 2018. Leafhopper-induced activation of the jasmonic acid response benefits *Salmonella enterica* in a flagellum-dependent manner. *Frontiers in Microbiology* 9(1987).

Du Q. and Wang H. 2015. The role of HD-ZIP III transcription factors and miR165/166 in vascular development and secondary cell wall formation. *Plant Signaling and Behavior* 10(10): e1078955.

Kallenbach M., Bonaventure G., Gilardoni P.A., Wissgott A. and Baldwin I.T. 2012. Empoasca leafhoppers attack wild tobacco plants in a jasmonate-dependent manner and identify jasmonate mutants in natural populations. *Proceedings of the National Academy of Sciences of the United States of America* 109(24): E1548-E1557.

Karbiener M., Glantschnig C. and Scheideler M. 2014. Hunting the needle in the haystack: a guide to obtain biologically meaningful microRNA targets. *International Journal of Molecular Sciences* 15(11): 20266-20289.

Lazebnik J., Frago E., Dicke M. and Van Loon J.J.A. 2014. Phytohormone mediation of interactions between herbivores and plant pathogens. *Journal of Chemical Ecology* 40(7): 730-741.

Liu B., Li J. and Cairns M.J. 2014. Identifying miRNAs, targets and functions. *Briefings in Bioinformatics* 15(1): 1-19.

Musetti R., Marabottini R., Badiani M., Martini M., Sanità Di Toppi L., *et al.* 2007. On the role of H_2O_2 in the recovery of grapevine (*Vitis vinifera* cv. Prosecco) from Flavescence dorée disease. *Functional Plant Biology* 34(8): 750-758.

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

Pinzón N., Li B., Martinez L., Sergeeva A., Presumey J., *et al.* 2017. microRNA target prediction programs predict many false positives. *Genome Research* 27(2): 234-245.

Rafiqi M., Bernoux M., Ellis J.G. and Dodds P.N. 2009. In the trenches of plant pathogen recognition: Role of NB-LRR proteins. *Seminars in Cell and Developmental Biology* 20: 1017-1024.

Sablok G., Pérez-Quintero Á.L., Hassan M., Tatarinova T.V. and López C. 2011. Artificial microRNAs (amiRNAs) engineering – On how microRNA-based silencing methods have affected current plant silencing research. *Biochemical and Biophysical Research Communications* 406(3): 315-319.

Schwab R., Ossowski S., Riester M., Warthmann N. and Weigel D. 2006. Highly specific gene silencing by artificial microRNAs in Arabidopsis. *The Plant Cell* 18(5): 1121-1133.

Seitz H. 2017. Issues in current microRNA target identification methods. *RNA Biology* 14(7): 831-834.

Smyth N. 2015. *The determination of the spatial and temporal distribution of aster yellows phytoplasma in grapevine.* Master's degree thesis, Department of Genetics, Stellenbosch University.

Zaidi S.S.-E.-A., Mukhtar M.S. and Mansoor S. 2018. Genome editing: targeting susceptibility genes for plant disease resistance. *Trends in Biotechnology* 36(9): 898-906.

Zhang L., Zhang F., Melotto M., Yao J. and He S.Y. 2017. Jasmonate signaling and manipulation by pathogens and insects. *Journal of Experimental Botany* 68(6): 1371-1385.