“Deciphering the inter-kingdom RNA exchanges in the *Alternaria*-Tomato Pathosystem”

Christopher Norman

A degree submitted in fulfilment of the Universities requirements for degree of Doctor of Philosophy.

2022

Abstract

Fungal infections make up most plant crop losses in the agricultural industry. This contrasts with mammals, in which the major threat to our existence is viruses and bacteria with fungi making up only a small minority of the infections we suffer. Due to this there has been a large investment in the prevention of fungal infection, mainly using chemical fungicides. However, chemical fungicides have many disadvantages, most notably the runoff resulting in damage to wildlife and other forms of ecological impact, occurrence of resistant pathogen isolates and health risks for consumers and farmers using them. This has gradually led to public rejection of chemical fungicides as shown by the growth of the organic food industry. The growing technology of RNA silencing shows great potential in alleviating this problem. Host induced gene silencing (HIGS), naturally exists in plants as a means of protection from pathogens and has already become cited as a potential new resistance engineering method in crops. Either by spraying the plant directly with RNA molecules or inducing the RNA production in the plant using gene delivery vectors like *Agrobacterium* and viruses. The more we understand these interactions the more we can perfect this technology. Here the *A. alternata*-tomato pathosystem will be used to evaluate the role of sRNA in infection and defence of the plant and evaluate genes that may have value as a target for sRNA. In addition, this project will look at the potential of certain genes as targets for targeted RNA silencing to protect tomato from *Alternaria alternata*. Specifically, the Chitin synthase genes (CHS) and the effect silencing RNAs has on these genes has and whether they inhibit fungal infection. The methods used in this project focused on two areas, determining if CHS is a viable target for spray induced silencing through treating spores with small RNA solutions and determining any effect on germination rates. These experiments found limited effect on germination however did discover a significant effect on germination tube size and spore production, with tube size increasing significantly on treatment with sRNA and spore production occuring earlier and in larger quantities. sRNA produced by the plant during infection was also evaluated using sRNA sequencing, to identify novel miRNA. My conclusion for this project are that CHS is likely a poor candidate for sRNA silencing however with the number of miRNA identified to show upregulation during infection it can be concluded that natural miRNA play a role in infection and many of these miRNA could be the targets of future research to improve the defence response of the tomato plant or to develop new SIGS or HIGS systems.

Table of contents

List of figures

List of tables

Abbreviations

bp Base pair
CABI CAB International
°C Degree Celsius
cDNA Complementary DNA
CDS Coding sequence
Cp Cross point
Cq Quantification cycle
CHS Chitin synthase genes
DNA Deoxyribonucleic acid
ds sRNA double stranded sRNA
disiRNAs Dicer-independent small non-coding RNAs
ETI Effector-Triggered Immunity
FPKM Fragments per kilobase per million
gDNA Genomic DNA
HR Hypersensitive Response
HIGS Host induced gene silencing
ITS internal transcribed spacer region
kb Kilobase
LTR Long terminal repeat
M Molar
mg Milligrams
min Minute
ml Millilitre
mm Millimetre
mM Millimolar
mRNA Messenger RNA
R gene Resistance gene
NaCl Sodium chlorate
NaOH Sodium hydroxide
NATs Natural antisense transcripts
NB-LRR Nucleotide-binding leucine rich repeat
ncRNA non-coding RNA
ng Nanogram
nt Nucleotide
ORF open reading frame
PAMPs pathogen-associated molecular patterns
piRNA Piwi-interacting RNA
PCR Polymerase chain reaction
PDA Potato dextrose agar
qPCR Quantitative polymerase chain reaction
RIN RNA integrity number
RPM Revolutions per minute
RNA Ribonucleic acid
RPKM Reads per kilobase per million
RT Room temperature
RT-PCR Reverse transcription polymerase chain reaction
RNA-seq High-throughput RNA sequencing
RNAi RNA interference
snRNAs small nuclear RNAs
snoRNAs small nucleolar RNAs
miRNA micro RNA
RISC RNA induced silencing complex
RNP ribonuclear proteins
tsRNA tRNA derived sRNA
s Second
sRNA small-RNA
srRNA small ribosomal derived sRNA
scaRNAs small Cajal body specific RNAs
siRNAS small interfering RNAs
SIGS Spray induced gene silencing
Tm Melting temperature or annealing temperature
TPM Transcripts per million
TIR toll/interleukin-1 receptor
TAS Plant Trans-acting siRNAs
µg Micrograms
µl Microliter
µm Micrometre
V Volts
v/v Volume / Volume
w/v Weight/ Volume

Declaration of the authors rights

I declare that the thesis here presented is my sole own work and generated by myself during my research work.
I confirm that the work was done wholly while in candidature for the PhD degree at this University and have not been submitted to any other qualification or institution.
Where any published work of others was used, it was clearly attributed or stated in the written work. When join contributions of other authors were made, acknowledgement was expressed.

Acknowledgements

Acknowledgements

Throughout this long and arduous journey I have been supported by many wonderful and talented people. All of which have been invaluable for making this thesis possible.

I would like to thank the university as well as the research school who have supported my research and made finishing the thesis possible.

I would like to thank the BSPP who provided covid relief funding to me when the extension needed for covid threatened to derail the project, their funding made continuing to completion possible. They also provided many opportunities to attend and present at conferences.

I want to give a big thank you to my supervisor Prof. Mahmut Tör, who took on the difficult and arduous task of mentoring me which I offer my full and heartfelt thanks for. Mahmut is an amazing supervisor and I hope he has many more successful students in the future.

I also extend my recognition to the supervisory team, Dr Yiguo Hong, whose expertise in sRNA has been invaluable especially in the beginning and Dr David Studholme whose knowledge of bioinformatics has helped greatly in a field I struggle with.

I'd like to thank those in the labs who have supported me including Dr Catherine Jimenez Quiros for all her support and teaching and Dr Emeka Okechukwu for all the knowledge and guidance he has given me. I want to continue to tank all those in the SSE, there are too many of you to name but you have all supported me in one way or another.

And finally, I wish to thank my family who without which I would not have been able to fund or complete this thesis, my Auntie Shirley for the many times she helped growing tomato plants, and my wonderful wife Jessica Norman for helping me to proofread and edit this thesis. Without them I would have never come this far.

Chapter 1

General Introduction

1.1 Introduction

Fungal pathogens play a significant role in crop losses and consumer health making up most plant diseases world-wide [Archana Jain et al,. 2019]. Understanding causes of crop yield loss is not only important in combating the increased need for food for an increasing human population but also important to investors in a multi-billion-dollar industry. Designing new strategies, new fungicides, resistance genes or other strategies to reduce fungal growth is essential to the survival of this industry. Every year, current resistant plant lines lose their protection as pathogens adapt to the resistance, and fungicides used can lose their effectiveness. This has led to an arms race between the pathogen and the host plant [Archana Jain et al,. 2019], and the plant breeders and researchers have continuously developed new fungicides and searched for new resistance genes. Major losses to crop yield are due to insect pests and plant diseases; fungal diseases make up 70-80% of plant diseases. This is due to there being far more plant pathogenic fungi than pathogenic bacteria, virus, or parasites as well as their ability to destroy their hosts.

Of the losses inflicted on crops world-wide, the five most important are wheat, rice, maize, potatoes, and soybean. It has been calculated that if these losses alone could be mitigated, up to 600 million people could be fed annually [Davies CR,. 2021]. It has also been suggested that if there were a scenario where in the same year all these crops were to face large crop losses, 62% of the world's population would be without food. This potential scenario becomes even more likely due to modern cultivation methods. These methods of plant breeding and cultivation have vastly increased crop yield but, in many cases, to the expense of crop resistance to disease. However, this increase has been necessary to compensate for growing populations and development of resistance.

1.2 Plant diseases

There are over 19000 known plant diseases in commercial crops world-wide, many of these plant diseases can remain inactive in both living and dead plant material until the conditions are correct for it to cause disease [Archana Jain et al,. 2019]. Pathogenic fungi and bacteria can cause a variety of diseases, such as leaf spot, rust, wilt, scab, canker, mildew, and dieback. Each of the different types of plant disease have different methods for control. The best control method for fungals leaf spot for example is to keep the leaves dry to prevent excess moisture and humidity promoting fungal growth. To avoid this irrigation is applied directly to the soil and roots, preventing the spores from being spread by water droplets. In addition, fungal spores that cause leaf spot often need water to germinate [R. Barkai-Golan et al., 2001] these spores then access the leaf by opening such as the stomata. Disease results in the yellowing of the leaves and necrosis, spreading through tissue as dark brown patches until the leaves die out. Blight is a disease caused by bacterial pathogens and so has a completely different process of infection and spread. Bacteria like *E.amylovora* penetrate the plant via opening like the stomata and wounds in the tissue, leading to localised necrosis, without colonising adjacent tissue [Piqué N et al,. 2015]. Fusarium Wilt disease is a serious vascular disease in crops especially potatoes and tomatoes. The difference between fusarium wilt and leaf spot because it spreads through the soil and plant debris. This results in leaves losing structure and rigidity, turning light green and yellow until they eventually lose integrity and die. Leaf rust is a fungal disease that predominantly targets beans, tomatoes and roses, these diseases first appear as white lesions. Over time these lesions lead to raised lumps and eventually red patches of spores. Severe infestations can lead to death of the plant and deformation [Figueroa M et al,. 2018].

Late blight is an infection caused by *Phytophthora infestans* that begins with the older leaves and can be observed from green and grey spots on the leaf spots. After disease progression the spots become darker and a white mycelial mass forms on the lower area of the leaves. This becomes a systemic infection, affecting the entire plant. This disease does not survive winter, in contrast to the diseases called early blight. The Early blight diseases occur in potato and tomato, caused by *Alternaria solani* this begins as small brown lesions with rings surrounding them. As the disease progresses the leaves become yellow and die. This disease does overwinter in contaminated plants and can be spread by rain and insect vectors [Jindo K et al,. 2021]

Powdery mildew is a disease caused by a range of fungi [Glawe DA 2008].; these fungi often have mycotoxin contamination; mycotoxins are non-host n have a limited range of hosts. It can appear early in the growing year and begins by causing leaves to curl and expose their lower area. As the disease progresses the upper leaf becomes covered in white powder as the mycelial mass grows. This disease eventually results in death and necrosis of leaves and can prevent flowers from opening and being fertilised resulting in reduction in yield. Downy mildew is an oomycete pathogen, like fungi the pathogen is spread by spores and can cause disease by yellow patches on leaves. The spores are spread by rain and dew and can cause infected leaves to turn brown and die.

1.3 Alternaria alternata

A. alternata is predominantly an opportunistic pathogen which uses multiple different routes for infection. These routes include wounds, such as leaf or stem damage, and through the leaf stomata [Pinto et al,. 2017]. This allows the pathogen to remain in the plant system until the fruit ripens. The *Alternaria* genus are predominantly saprophytic fungi, meaning that they are predominantly necrotrophic breaking down and recycling dead plant material. However, some species have adapted to be an opportunistic plant pathogen capable of causing disease in a broad range of hosts.

A. alternata causes the development of circular black spots on many different vegetables and fruits. This can occur at both pre- and post-harvest stages. Either by causing losses in plant leaves through leaf blight pre-harvest, being retained in the fruit resulting in blight post-harvest or resulting in harmful toxin build up on harvests making them unsafe for consumption [Pinto VE et al,. 2017].

The *Alternaria spp* are present in many areas of the human environment. They are found in air conditioning systems, isolated from the backs of cockroaches, and most notably in leaf blight infection in crops. The USDA fungi host index includes over 4000 examples of the *A. alternata* and plant pathosystem, making it tenth in the number of potential hosts out of 2000 fungal genus [Bills et al., 1987]. *A. alternata* specifically has been observed to cause disease in over 100 plant species [Thomma et al., 2003] where it can cause loss of yields and mycotoxin contamination of crops. These usually occur on the leaves of the target plant; leaf spots are observable by the black lesions surrounded by yellow rings as observable in [figure 1.1]. Leaf

necrosis can reduce sale value of leafy vegetables, reduce photosynthetic ability of the plant, and hence reduce the crop productivity and growth. Fruit plants infected with the *Alternaria spp* can make plants unsellable, in cases of trees being grown for sale from nurseries, as well as making fruits inedible due to mycotoxin contamination.

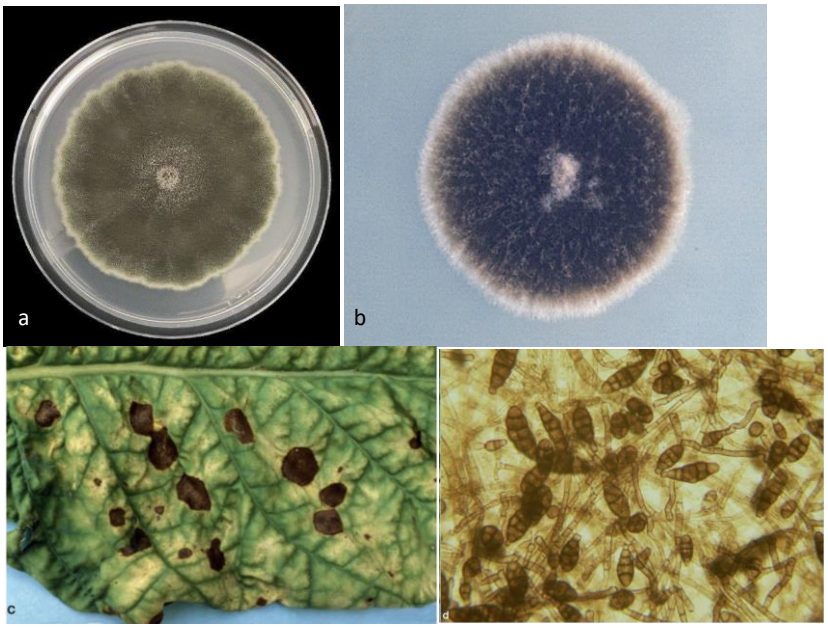


Figure 1.1: *A.alternata images:* *A.alternata* grows well on PDA Agar, beginning a light grey colour with white mycelia (figure 1a). Over time and in the correct conditions the mycelia create spore producing bodies that cause the culture to turn black (figure 1b). *A.alternata* appears as black lesions with a yellowing around them when infecting plant material (figure 1c). Fungal spores appear as dark green, segmented ovals larger at one end (figure 1d).

With Red Delicious apples grown in Argentina shown to have up to 50% have fungal lesions [Pavicich MA et al,. 2020]. These post-harvest fungal diseases are attributed to many fungi within the Alternaria spp, and the diseases can occur both in the field and in storage. With crops often being transported long distances before distribution, improper storage can lead to a perfect environment for fungal growth. The serious consequence of this is contamination by mycotoxins. Mycotoxins are toxins which are not host specific and which are produced by fungi during pathogenesis. There have been over 30 identified in the Alternaria spp [Robiglio

and Lopez et al., 1995]. In the tomato plant, this can be an increased issue due to their prevalence in crushed and canned produce, but it is also present in intact produce. These mycotoxins are associated with many human disease risks, such as having been linked to certain cancers [Marchese S et al,. 2018].

There has been a strong correlation between the spread of *Alternaria* caused leaf blight and the use of susceptible cultivars [Li et al., 2013]. However, it has been observed that leaf blight has had little effect on crops of apple and pear trees in Europe, possibly due to existing populations of resistant varieties. However, there have been examples of *A. alternata* causing leaf blight in Asia, specifically the Asian pear *Pyrus pyrifolia*, as well as by *A. Mali* in apple varieties in Italy. The UK has taken actions against these potential food contaminations by rejecting material that may contain pathogenic *Alternaria* spores, which has a negative effect on buyers and distributors. There has also been a correlation observed between red mite infestation and *Alternaria* leaf blight; trees with both have been recorded to suffer up to 60% loss in leaf matter. Demonstrating that the effects of the *Alternaria* are heightened by the presence of the European red mites. This is possibly due to increased and uncontrollable spread of the spores or increased mechanical damage to leaves, both of which lead to increased damage due to fungal infection.

Like all fungi, *A. alternata* needs specific conditions to grow and sporulate. These conditions are important considerations for pathology, as not only do we need to know these conditions for fungal growth, but also to understand the conditions of infection. The *Alternaria* spp needs an environment which is moist and warm in order to grow and produce spores, because of this it has been observed to thrive in humid climates [Peter W et al,. 1987].

The Alternaria are of scientific interest not just because of its effect on crops but also due to its allergenic qualities, in part due to the environment it thrives in. The airborne spores of *Alternaria spp* can be found in high numbers in any outdoor environment during summer months [Godfrey A. et al,. 2020]. These spores, in combination with other mould spores such as the *Ulocladium spp* and *Fusarium spp*, can also occur indoors due to increased humidity. However, there is little precedence in literature or examples of *A. alternata* being isolated from indoor building materials. In most cases where *Alternaria spp* spores are found, it's from the outdoors or in ventilation systems. Indoor cases of *A. alternata* can often be misidentified from *ulocladium chartarum*, which is a common indoor mould, growing on wallpaper, plants, and fibres.

Alternaria life cycle is demonstrated in [figure 1.2] under the correct conditions Alternaria mycelium produces conidiophores, long fungal growths that support chains of conidia spores. The conditions the spores are grown in, wet, windy, and warm, this results in the spore release and dispersal. Dispersed airborne spores land on new leaves and the presence of moisture and a flat surface induces germination and the creation of appressoria to breach the cell wall and infect cell leaves. This infection and mycelial growth on the plant leaves results in the expression of host specific toxins resulting in large scale colonisation and death of the leaves and the plant. These dead leaves then remain infected on the soil surface until conditions result in conidiophore production and continuation of the cycle [Kuang-Ren Chung., 2012]

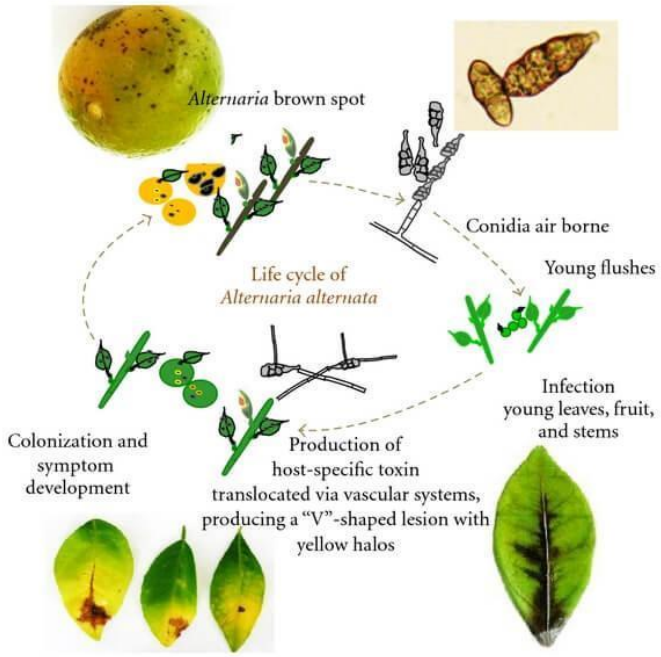


Figure 1.2 : the life cycle of *Alternaria alternata* [Kuang-Ren Chung,. 2012]
The *Alternaria alternata* life cycle begins when spores are carried by rain and wind onto a viable host, inducing disease, colonisation, and production of conidiophores and new spores to continue the cycle.

1.4 Economic importance

There are several reasons for choosing the *A. alternata*/tomato pathosystem. It can grow on potato-dextrose media; it is easy to isolate, genetically manipulate, and easily transformed. It

has a sequenced genome readily available and access to isolates in the laboratory. In addition, the tomato plant is one of the most widely cultivated commercial plants in the world [Kimura S, Sinha N. 2008]. The second most consumed plant after the potato, it can be grown in a range of temperatures, under cover when the outside temperature is not sufficient, with china being the world's largest producer at over 61 million tonnes per year ["Crops". FAOSTAT,. 2016] such as a long shelf life and the ability to be transported long distances to shops and markets. However, it is susceptible to diseases such as *A.alternata*. The *Alternaria* diseases are at their most dangerous post-harvest, they produce mycotoxins that can cause spoilage of fruit and loss of yields.

Mycotoxins are an incredibly important aspect of plant pathology and the prevention of plant disease induced crop yield loss. Mycotoxins consist of secondary metabolites that are produced by fungi which can cause disease in humans who consume them, as well as harming animals that consume them in the case of plants grown for fodder. Important mycotoxins include Aflatoxin, ergot alkaloids, zearalenone and citrinin [Bennett JW, Klich M,. 2003]

Mycotoxins can have a complex definition, all mycotoxins are natural products of filamentous fungi, low molecular weight secondary metabolites. These metabolites can show chemical and toxicological diversity, showing overlapping toxicity to plants and microorganisms. This can also include some antibiotics, such as patulin. Disease caused by mycotoxins is referred to as mycotoxicosis, and a toxin is only considered so if dangerous in low concentrations unlike metabolites like ethanol which is only dangerous in high concentrations [Bennett JW, Klich M,. 2003]

From the perspective of plant pathology there are many mycotoxins that also have the role of phytotoxins, toxins which are deadly to the host plant and play a role in pathogenicity. Many of the secondary metabolites produced by fungi can be virulence factors and cause disease in humans that consume them. Many mycotoxins produced by *Fusarium* are also important in plant disease progression [Bennett JW, Klich M,. 2003] this contrasts with the role of mycotoxins in vertebrate hoists where they play little role in disease progression [Bennett JW, Klich M,. 2003] an example of this is *Aspergillus fumigatus* which results in the disease aspergillosis, Aspergillus produces gliotoxins which inhibit T-cell activation and limit macrophage phagocytosis, both essential components for the human immune system. Despite this being a mycotoxin, it is not a pathogenicity factor in Aspergillus, and is not produced in

large quantities during human infection [Kwon-Chung, Kyung J, and Janyce A Sugui,. 2009] this is likely due to the optimum temperature for mycotoxin production being 20-30 degrees, making them far more suited for production as pathogenicity factors in plants than in humans where the fungi are forced to grow at a much higher temperature [Magan, Medina & Aldred, D et al,. 2011]

There have been studies of the 70 plus phytotoxins produced by the *Alternaria* sp which can range from non-host specific and host specific effects. Those of which are toxic for humans and animals are classed as mycotoxins. These include Alternariol, Tentoxin and Stemphyltoxin III, whose effects have been extensively studied [Zwickel T et al,. 2016]. There are many issues associated with removing produce contamination, treatments often focus on killing microorganisms that can leave the toxins they produce behind. This increases the need for effective disease prevention and the use of fungicides in agriculture.

1.5 Combatting plant disease

This knowledge can be used to reduce fungal infection in crops using alternative cultivation methods. The main means of infection comes through the leaves, entering through the stomata. This can partly be prevented by monitoring during high-risk months such as April to June, avoiding overhead irrigation, as well as treatments of fungicide during April [Jindo K et al,. 2021] . Areas with a known history of infection are likely to still have spores present in the soil and so are under greater risk.

Due to issues caused by mycotoxin contamination, it is often more efficient to prevent disease from occurring rather than treating a disease after observation of symptoms. Once a plant has developed leaf blight, it is highly likely, especially if the lesions are on the fruit or vegetable itself, that the crop may be contaminated. This has driven the agricultural strategy for defence against *Alternaria* through chemical pesticides, physical measures, biological control, and integrated approaches.

The environment the fungi grows best in is a key factor in preventing fungal growth but is also a major factor in shelf life of tomatoes. It is essential to maintain a temperature during storage of below 7 degrees Celsius and limit the duration of that storage to less than 10 days. This limits *Alternaria* growth and mycotoxin production. It has been observed while studying *Alternaria* in wheat that radiation exposure decreases colony diameter of *A. alternata* and

significantly reduced mycotoxin concentration [Braghini et al. 2009]. This study looked at the effect of gamma radiation on *A. alternata* in sunflower seeds which also greatly reduced mycotoxin production.

Alternaria spp can still be present in diseased plant matter over winter, due to the plant spores being able to survive cold temperatures and still being able to recolonise plant tissue upon their preferred conditions returning [Rotem, J. 1990]. Therefore, an important preventative strategy is to remove dead plant material and dispose of it correctly. This includes not using it in compost due to the likely hood of spores contaminating future crops. The best method is to burn the diseased material and remove it from the site. It is also possible for the soil to be contaminated with *Alternaria* spores, which also survive over winter to affect the next season's growth. This means that areas that have had *Alternaria* contamination should be subject to additional crop rotation. Due to the fungi's species specificity, it is unlikely that an unrelated crop will be infected by the same spores that affected the last crop. This adds to one of many reasons that make crop rotation essential to modern farming practices [Abuley IK, et al,. 2018] . In addition, pruning plants effectively can be essential, as non-sanitary cuts done with dirty tools that have not been sterilised can result in infection. Failing to do so can increase fungal spread by transmitting spores from infected plants to new areas.

There have been many approaches to the prevention and management of crop disease. These include agrochemical, biological and selective breeding for resistant genes [Jindo K et al,. 2021].. One of the most efficient and effective methods is to combine resistance breeding with a holistic method of crop disease management. However, this has resulted in a biological arms race, with new resistant and virulent strains emerging to take advantage of the reduced competition. Many researchers in the field believe that traditional plant breeding methods are insufficient to cope with the threat posed by new fungal strains and adaptation.

Biotechnology opens many new possibilities to combat the threat of pathogen resistance and adaptation, providing novel methods to create disease resistant crops. These can be bred to be resistant to multiple pathogens, as well as being environmentally non-damaging. Molecular research into the biology pathways of plant-pathogen interactions is key to this approach [Fonseca JP et al,. 2019]

1.5.1 R- genes

The development of resistance genes (*R*-genes) in plants is a response to plant pathogens, and those pathogens' response in adapting to those *R*-genes, is a major part of the study of plant/fungal pathogenesis. *R*-genes are diverse in sequences and protein domains and are essential for the effective defence of plants from disease.

A summary of the plant defence mechanism can be made as the sequence of steps leading to the plant undergoing a Hypersensitive Response (HR) in infected tissue [Goodman, R.N. and Novacky, A.J,. 1994]. HR is defined by localised and rapid cell death at the location of the pathogen growth. This localised cell death results in a suppression of the disease process. HR can be present in response to bacterial, Fungal and Ooomycete infections and is only effective against biotrophic infection rather than necrotrophic [Balint-Kurti P et al,. 2019].

Defence requires several steps, first is the detection of pathogens by known recognition mechanisms such as cell surface pathogen-associated molecular patterns (PAMPs) detection. This is counteracted by pathogen produced effector molecules adapted to suppress PAMP triggered immunity, which is in turn countered by the host mechanism of Effector-Triggered Immunity (ETI). These compose of most R- genes, R- genes encode for a variety of proteins, of which most can recognise pathogen effectors. These avirulence proteins are classified based on their recognition region. Those protein recognition regions include the Nucleotide-binding leucine rich repeat (NB-LRR) and the extracellular LRR proteins.The NB_LRR containing class of avirulence proteins is the most common class of these protiens and can also contain an Amino acid coiled-coil (CC) and toll/interleukin-1 receptor (TIR) domains [Belkhadir, Subramaniam, & Dangl, 2004]. One of the first examples of a transgenic R gene is the gene encoding the TIR-NB_LRR protein that results in a resistant phenotype against tobacco mosaic virus.

Several studies have investigated the role of R- genes in the Tomato/*A. alternata* pathosystem. This is predominantly done by looking at expressions under pathogen invasion [Tsuda and Somssich, 2015]. Reactive oxygen species scavenging compounds such as catalase, peroxidase and phenylalanine-ammonia-lyase, superoxide dismutase and total phenolic contents are all used by plants as defence against fungal disease [Hyun MW et al,.2011] as well as genes *S1WRKYs* and 7 PRs and genes encoding β-1,3 glucanase and chitinases shown to play a role in *A. alternata* resistance [Moghaddam et al., 2019]. Several factors in the host can determine both susceptibility and resistance to infection.

1.5.2 Fungicides

Fungicides make up a significant part of the agriculture industry. Strategies for applying fungicide are usually by seed application or direct crop application. Many of the seed applied fungicides are systemic; this means that they are taken up by the plant and transported to the tissue where they perform their anti-fungal activity. These seed-based fungicides are most effective in combating soil-borne pathogens. Fungicides are used extensively; however, fungal pathogens have adapted to these fungicides and produce resistant fungal strains. Fungal adaptation to fungicides is a complicated process, as fungi can vary in genome size, variability, and method of infection. Due to this, some fungi are far more prone to development of resistance than others and can result in resistant strains in as little as a few years of exposure [Avenot H. F et al,.2007]. This process has driven the industry in a constant arms race with fungal resistance with researchers working to find new fungicides and novel treatment strategies. This hasn't been limited to novel fungicides, resistant strains, or biological control. Research has been directed into variation of quantity and frequency of application of fungicide or combining fungicides to combat resistance [Van Den Bosch F et al,. 2014].

In some ways, this shows similarity to research into antibiotic resistance in bacteria in human systems with research motivated by finding new ways to evade this resistance [Hahn M et al,. 2014] There are two main drivers in the research of fungicides: the effectiveness (including cost, evading resistance, and preventing infection) and the environmental damage caused by the fungicide. The damage caused by fungicides has been a topic in public consciousness for many years and drives the organic food industry that avoids chemical fungicides in favour of more natural methods. As well as driving research into environmental damage and the cause of it, it also looks at alternatives to chemical fungicides that can be just as effective. Fungicide contamination is not a sole issue of the agricultural industry; much of that pollution can come from urban areas.

1.5.3 Resistosome formation

There have been modern advancements that have progressed the field of plant resistance, plant innate immunity can be induced by the NLR family of proteins [Burdett H et al,.2019].

Examples include ZAR1, which exists in an active state referred to as a Resistosome. This suggests a NLR function speciously not understood. The NLR protein family is an essential protein for plant and animal innate immunity. NLR's are defined by a central nucleotide binding domain (NBD) as well as a C-terminal leucine rich repeat domain, in plants they contain coiled-coil (CC) or Toll/interleukin-1 receptor/resistance protein (TIR). These are signalling domains that induce a defence response without other protiens. Resistance genes often encode NLR's that detect and induce resistance in the presence of Effectors.

ZAR1 is an example of such a R gene encoded receptor [Gong Z et al,.2022], encoded by HOPZ-activated resistance 1. ZAR1 is a capable of binding and detecting effector caused modifications of kinases modified by bacterial effectors. The 3D structure of inactive ZAR1:RKs1 complex has been studied as well as the ZAR1:RKS1:PBL^{UMP} activated complex which is the resistosome. ZAR1(ADP):RKS1 is the inactive state that exists passively within plant cells, when PBL2 is subject to Uridylation by an effector such as AvrAC this yields PBL2^{UMP} which is now able to binding to RKS1 which is able to bind ATP to induce a conformational change which forms the active resistosome which goes on to trigger cellular cascades that lead to cell death and the rest of the innate immune response.

1.5.4 PTI and ETI cross communication

Plant have clearly developed a two staged system of innate immunity; this allows the plants to detect and deal with a range of pathogens. The first stage of defence is initiated by recognition of PAMPS or similar damage associated molecular patterns (DAMPS) this is done by the cell surface receptors PRRs leading to PTI. Many virulence factors such as effectors have adapted to suppress PTI resulting in triggering effector triggered immunity (ETI) regnition by effectors by NLRs resulting in the process of ETI being triggered. Models have been proposed on how the two systems communicate and interact. Systems such as the 'zig-zag' model [Keller H et al,. 2016]but a detailed analysis of their interaction is still poorly understood.

PTI and ETI trigger two different pathways by two different classes of receptors [Yuan M et al,. 2021], however their downstream pathways often overlap. For example MAPK and ROS processes are connected to both pathways . Research in Arabidopsis showed the role of mutual initiation of both pathways by activating intracellular receptors, showing that activation of ETI and PTI is done through increasing the abundance of protien kinases and NADPH, and showed

that both extracellular and intracellular receptors are needed to prime the immune system against *Pseudomonas syringae.*

1.6 role of small RNA

Plant-pathogen interaction is defined by the transfer of and reaction to effector molecules sent from both the pathogen and the plant in response to infection. These interactions are well documented and understood [Dodds et al., 2010]. However, it has only recently been discovered that small RNAs (sRNAs) play a part in this exchange of signals and effectors [Weiberg et al., 2015]. sRNAs are a common tool in higher eukaryotes as a system of gene regulation and defence against various forms of pathogen infection [Catalanotto et al., 2016]. They are also used in prokaryotic pathogenicity; s RNA has a role in the family Vibrionaceae virulence factors [Pérez-Reytor et al., 2016] and in biofilm production in *Pseudomonas aeruginosa* [Falcone M et al,. 2018]. By producing and exchanging trans-kingdom sRNAs, plant pathogens can hijack the host machinery to its own ends, leading to pathogen beneficial gene silencing [Weiberg et al., 2013]. In the case of *B. cinerea*, this is done by the binding and silencing the *AGO1* gene in *Arabidopsis.*

Many papers have researched how these different forms of RNA are distinguished from each other, their roles in silencing, what they silence, and their biogenesis. All these aspects are essential to understanding the overall role of sRNA in an organism and how RNA interference (RNAi) can be used to benefit us in research or agriculture. sRNA is a broad term referring to all RNA below a certain size, cited as between 15 and 300 bp in length by most papers and reviews, including [Watson et al,. 2019] and [Storz,. 2002]. This term encompasses all these RNAs regardless of function and all come within the definition of non-coding RNAs, with no known s RNA being capable of expressing proteins. These can then be subdivided into roles and categories, as shown in [Figure 1.3] with those s RNAs then being able to be separated between those that have their role in the RNAi pathway or other forms of gene silencing or regulation and those with other roles. RNAi has taken much of the academic world by storm, with countless papers looking into the role of RNAi in biological systems [Fei Q et al,. 2013] and [Chen et al,. 2013] and looking into the sRNA interactions between organisms and identifying their role in infection.

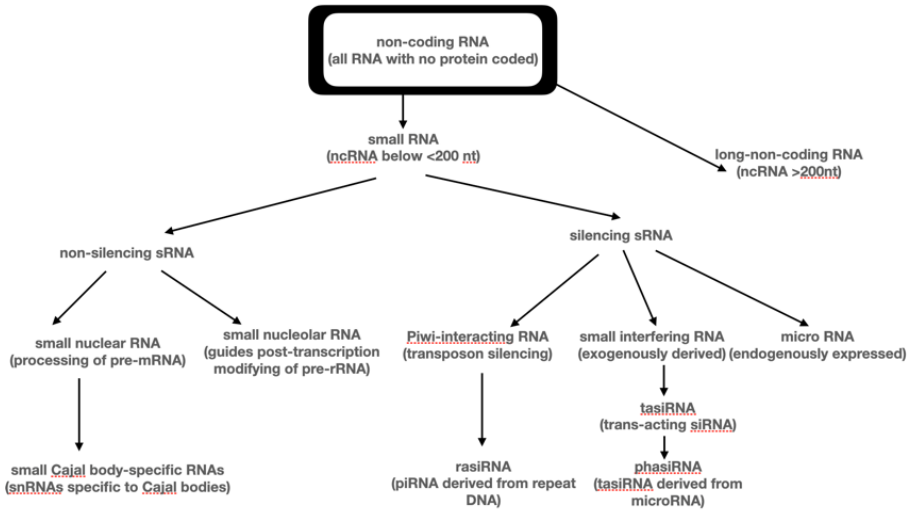


Figure 1.3: sRNA classification:

sRNA are all grouped into the non-coding RNA but are distinguished from long non-coding RNAs by size (sRNAs being smaller than 200nt) [Langenberger D et al,. 2010] It is important to understand that not all small RNAs perform gene silencing with a variety of functions for s RNAs occurring within the nucleus. These consist of the small nuclear RNAs (snRNAs) and the small nucleolar RNAs (snoRNAs) [Kufel J, Grzechnik P. 2018] which are distinguished by their location and role, with snoRNAs guiding post-transcriptional modification of rRNA and snRNAs in the processing of mRNA, these having their subcategory of Cajal body specific RNAs [Meier UT et al,. 2017] that only perform their role in the Cajal bodies. The rest fall under the broader category of silencing RNAs of the naturally occurring silencing sRNA these can be separated into three broader categories, miRNA which is expressed endogenously by specific miRNA, form a distinctive mRNA stem-loop structure and function as a gene regulator by associating to the miRNA-protein complex [Cai Y et al,. 2009] Piwi-interacting RNA are sRNAs that predominately silence transposable elements post-transcription through forming a complex with Argonaute with their sub-group of rasiRNA which are piRNA that specifically silence repeat elements. Small interfering RNA refers to all small RNA that perform silencing using the RNA induced silencing complex (RISC) usually doubles stranded small RNAs, these include the subgroups of tasiRNAs and phasiRNAs which are small RNAs in plants that act to silence post-transcriptionally and are generated by miRNA degradation of mRNA. sRNA was first discovered in animals in the nematode in 1992 [Lee et al,. 1993] and later in plants in 1999 [Chen et al,. 2010] in which a microRNA called miRNA171 accumulated in tissues and targeted the mRNA of scarecrow-like (SCL) transcription factors. siRNAs are often endogenous, capable of repressing expression either post-transcriptionally by degrading mRNA or by blocking mRNA translation and by modifying histones to prevent DNA transcription. All sRNA observed form complexes with proteins to perform their role, these complexes are called ribonuclear proteins (RNPs). Biogenesis of sRNA can vary wildly based on the type of sRNA and the organism of origin (for example differences between

phasiRNA in plants and piRNA in animals for example) when produced by expression from a sRNA gene locus it is done through DNA polymerase ll or lll and their overall copy number per cell can vary from 100000 to 10,000,000 per cell.

1.7 RNA Interference Pathway

The RNA interference pathway is the main source of sRNA mediated gene silencing [Muhammad T et al,. 2019]done specifically through miRNA and siRNA. RNAi is understood to be an effective method of silencing genes in eukaryotes, used in natural systems as defence from hostile genes and transposons or for gene regulation. It is also used as an artificial method to induce gene silencing in eukaryotes by introducing sRNAs designed to silence specific genes. The process depends on the sRNA guiding the silencing and the means of inducing this silencing is the RNP called the RNA induced silencing complex (RISC). The RISC is how the RNAi pathway silencing is performed, acting as the protein instrument to silencing using the sRNA as the guide.

The components of the RISC are often conserved [Zhang Y.et al,. 2013]. They contain multiple proteins with many variations between organisms and specific gene regulation pathways. Protein families are known to play a role include Argonaute, an essential component to the complex. The Argonaute proteins play the role of RNA cleavage within the complex; once the Argonaute is associated with its sRNA guide through its interaction with the PAZ binding domain within the complex, it is guided to the target RNA and RNA cleavage is catalysed by the PIWI domain which resembles a Rnase in its structure and function.

1.8 Small Interfering RNA

With the sRNA being differentiated into multiple types [Kong et al,. 2022] it is important to understand which perform what roles. Many of the most well studied and characterised of the sRNA, siRNA are those sRNAs capable of gene silencing through degradation of the mRNA or blocking ribosomes from performing mRNA translation. They are derived from degrading larger dsRNA into smaller dsRNA, but with notable structural motifs such as 3' overhangs. These work through perfect RNA-RNA hybridisation with target mRNA at any point within the sequence which leads to degradation by the RISC.

Dicer caused degradation of the dsRNA leads to siRNA formed from the short dsRNA, this siRNA is then incorporated into the RISC which uses the siRNA as a guide to direct silencing. Once incorporated, the siRNA is unwound to a single-stranded state and the less thermodynamically stable is degraded. Once this RNA has bound to a complementary mRNA the complex induces RNA cleavage, the cleaved RNA now likely to be degraded due to now lacking the prerequisite structures to be processed as a mRNA. siRNAs are distinguishable from miRNA due to their predominance in post-transcriptional silencing over blocking translation [Carthew et al,.2009], the fact they are derived from dsRNA cleavage and not endogenous gene expression, their lack of stem-loop structure and requirement for more specific total complementarity with the target.

1.9 Micro RNAs

miRNAs are also a RNAi inducing RNA, produced by specific miRNA genes. They are single-stranded and found in plants, animals, viruses, and fungi, all of which are observed to be involved in RNA silencing and gene regulation. These RNAs function due to their sequence complementarity between them and their target, forming a ribonucleoprotein complex with Argonaute family proteins they can silence the gene through multiple mechanisms, cleavage of the mRNA, shortening of the polyA tail [Carthew RW et al,. 2009] and blocking the translation of the mRNA through preventing ribosomal binding. Another distinctive aspect of miRNA is their structure, when compared to the siRNA, miRNA fold onto themselves to form a hairpin structure while siRNA form dsRNAs. For the major miRNA pathways in Plants, miRNA is classified into different groups based on their expression, function, and conservation. There is miRNA which are highly conserved such as the miRNA used by Angiosperms, these miRNA are only conserved in a particular species of plant [Jaiswal et al,. 2019].

Of the conserved miRNAs, 20 miRNA families share evolutionary origins [Cuperus et al,. 2011], with 9 originating from eukaryotes with two from liverwort. 12 families are derived from Spermatophytes and 1 from Gymnosperms. Some miRNA families are restricted to specific lineages, such as miR1432, miR538 from monocots and miR403 being restricted to eudicots [Chen.C et al,. 2018]. Conserved miRNA often shares their target gene with multiple different plant species, regulating conserved key pathways such as plant growth and development. Examples like this can be seen in the miRNA family in regulating the transition

between the vegetative and reproductive growth cycles through downregulation of SPL like genes [Liu et al,. 2017]. While the less conserved miRNA is more restricted in their target range, showing species or lineage specificity in their target genes [Wang et al,. 2020] which have been observed to be involved functions which are specific to their plant group. An example of this is defence against viral infection in rice [Baldrich P,. 2019]. Identification of 36 species-specific miRNA in Australian Nicotiana benthamiana species of Tobacco, specifically miR399, miR827 and miR398 [Huen et al,. 2018], these miRNAs showed variation in their cis elements. miR399 and miR827 showed that they had diverged from their primary conserved miRNA role in phosphate starvation pathways that had been demonstrated in *Arabidopsis*.

miRNA is produced by a process of biogenesis from a miRNA locus by Pol ll which produces a precursor miRNA; this is what has the hairpin structure that can be used to identify the miRNA from another sRNA [Xiao et al,. 2011]. DCL1 converts the hairpin into a double-stranded miRNA with overhangs at the 3'terminal end. These are methylated by HEN1. The miRNA complementary strand is loaded into the AGO1 part of the RISC to perform its function. This process is working by the same asymmetry rule phenomena that processes siRNA; the two strands of the pre-miRNA are not equally viable. This makes sense, as both strands cannot have complementarity to the same mRNA; instead, the stability of the 5' end determines which of the two strands will be incorporated and which one will be degraded. This results in an inherent asymmetry present in the biogenesis of the precursor miRNA to allow control of which strand is incorporated during complex formation [Hutvagner et al,. 2005]. Following this, HST transportation the RISC is transported to the cytoplasm where it can perform the needed silencing. The degradation of mRNA by miRNA can also lead to new sRNA called phasiRNA being produced. PhasiRNA is produced in 21-24 nt by DCL4 and DCL5 in cases where only one miRNA has targeted a transcript. This forms the final stage of the RISC where the miRNA is used as a guide using sequence complementarity to the target gene to induce gene silencing. In some cases, multiple miRNAs can target the same sequence (two-hit mode). Sequential cleaves of the upstream RNA is performed by DCL4 from the 3' to the 5' end. In cases where two or more have targeted the mRNA, the target sites are cleaved by miRNA-AG07, with the fragments converted into double-stranded RNA by RDR6 for DCL4 to cleave to convert into phasiRNA [Zhao et al,. 2020]. miRNAs are some of the most well researched and functionally important of the s RNAs especially in plants [Borges and Martienssen 2015], and their biogenesis has been a subject of many studies due to its complex

process. miRNAs are widely conserved in plants [Zheng et al,. 2015]. the miRNA gene once transcribed results in a 5' cap and 3' polyadenylation the same as a mRNA.

1.10 Phase-induced RNA

phasiRNA were named due to being phased secondary small interfering RNAs, which make up a significant group of s RNA in plant species. Some of these phasiRNA are known to target complementary mRNA as a form of post-transcriptional silencing s RNA, specifically a tasiRNAs (trans-acting small interfering RNA) from a long non-coding RNA locus present in the TAS genes [Peragine et al,. 2004]. tasiRNAs form a specialised sub-group of the phasiRNA and are distinct in that they only function to perform post-transcriptional regulation of gene expression. PhasiRNA biogenesis was briefly described when talking about mRNA function and shown in [figure 1.4], as the mRNA that leads to phasiRNA production are called phasiRNA triggers [Liu et al,. 2020]. In the case of TAS1/2, its biogenesis is induced by miR828, which, along with miR159 and miR858, could silence a range of 81 MYB genes in apple [Xia et al,. 2012]. 10 of the 19 MYB targeted by miR828 resulted in siRNA biogenesis predominantly from the 3' end of the mRNA and led to the generation of 100 sequence unique sRNAs

phasiRNA were confirmed to target 70 different genes through post-transcriptional regulation. An alternative method of phasiRNA generation is referred to as the two-hit model [Deng et al,. 2018], in which two separate miRNAs target the same region of the genome resulting in two triggering miRNAs. This was observed in the petunia genome. Studies into the distribution of phasiRNA have shown that their expression patterns can be tissue specific. It was demonstrated that the predominant form of si RNA have been observed in the anther tissue of Angiosperm were phasiRNA [Xia et al,.2019], and analysis of the sRNA sequences of the maize crop showed that phasiRNA expression can vary based on cell development.

There were phasiRNA that mapped to intergenic regions [Komiya R et al,. 2017].; 21-nt phasiRNA from 463 loci are selectively expressed after germinal and somatic cell specification and then reduce in expression after while the 24nt versions which are derived from 176 known loci and are instead expressed and accumulate during meiosis and continue to be present when the cells specialise to anther somatic cells. Each of these types of phasiRNA show independent

regulation depending on the phase and specialisation of the cells. PhasiRNA can be separated into two distinct types based on sequence length, 21nt and 24nt, with the 21nt known to be active before the 24-nt in various stages of cell specification. The 24-nt are associated stronger with meiosis. However, there is little research to suggest that they work by differing mechanism; both act post-transcriptionally by sequence-specific recognition of mRNA sequence. However, there observed differences in the loci for each and areas they have in common.

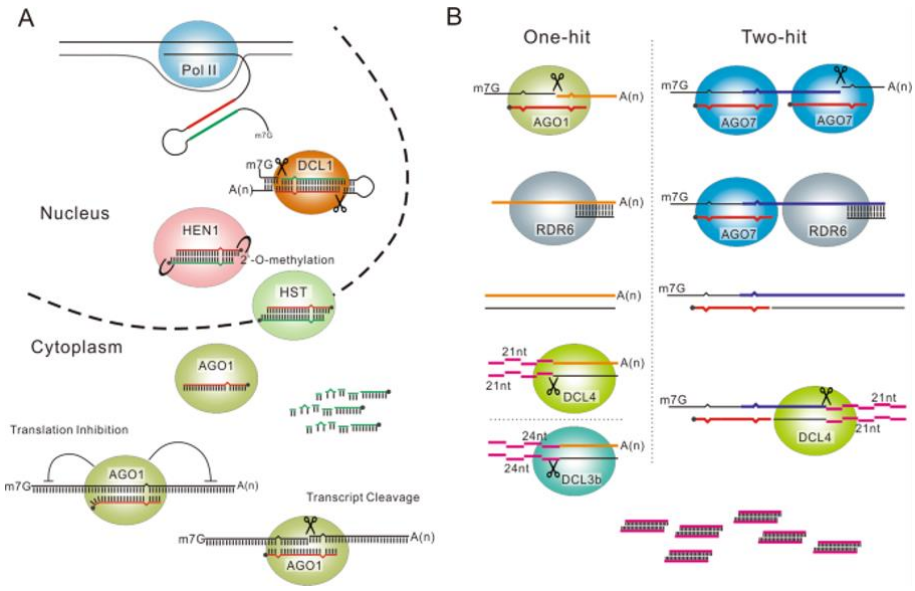


Figure 1.4: Biogenesis of miRNA and phasiRNA **from [**Chen et al,. 2018] :

This diagram presents the process by which both miRNA and phasiRNA are generated; A) Biogenesis of miRNA, the miRNA locus is converted into RNA by Pol ll which produces the precursor miRNA, this is then processed by DCL1 to be methylated by HEN1 at the 3' end. HST transports the mature miRNA to the cytoplasm where it is loaded into AGO1, and the non-target strand is degraded. B) the two methods of phasiRNA biogenesis, the one hit model, the dsRNA is created by RDR6 after cleavage by miRNA and converted into the correct length phasiRNA by DCL4 or 5. For the two-hit model, a sequence with two miRNA target regions is targeted by miRNA-AGO7 with the fragments converted into dsRNA and then into phasiRNA by DCL4.

1.11 Major phasiRNA pathways

As a unique component of plant sRNA interactions, phasiRNA have been extensively studied and many pathways uncovered. Many of these pathways are known to be conserved highlighting their importance to the plant's survival; PHAS genes capable of producing phasiRNA have been identified in *Arabidopsis*, named the TAS genes, all these identified regions are non-coding regions like TAS1 through TAS4 while some like TAS1 and 2 and its respective miRNA is unique to Arabidopsis. This does not mean that phasiRNA exclusively originate from non-coding genes; there have been observed examples of protein coding PHAS genes. Examples include phasiRNA for nucleotide-binding leucine-rich proteins, the repression of these genes was observed to be unusual especially due to the genes high level of genetic redundancy [Zhai et al,.2011], and production of phasiRNA from PRR in Angiosperm [Xia R et al,. 2019]. It's likely in many cases that the use of phasiRNA is a form of positive regulation of the silencing activity, in which breakdown of the mRNA transcripts leads to the production of more phasiRNA which leads to further degradation of other transcripts. This process has been observed in genes silenced by siRNA, which originated from those genes being first degraded by DICER [Dukowic-Schulze et al,. 2016]. Piwi-interacting RNA (piRNA) are the most abundant of the non-coding sRNA present in animal cells; they are distinguished by a lack of conservation, larger size than other s RNAs (26–31 nucleotides) and the fact they are only found in animals [Houwing et al,. 2007]. They form protein complexes specifically involved with epigenetic modification and post-transcriptional silencing mostly for the repression of transposable elements, viral RNA and repeat genes [Aravin et al,. 2001], although they can be involved in some cases of gene regulation in the germline [Klattenhoff et al,. 2007]. This creates a form of RNA derived immunity to transposons, in which expression of a damaging transposable element can be dealt with.

They are distinct from miRNA and siRNA due to their radically different biogenesis and their function independent of DICER. Biogenesis is done via two main pathways; this is the primary pathway and the secondary pathway. The primary pathway works by piRNA clusters; these exist in pericentromeric regions near the centromere of chromosomes as well as some in

telomeric regions. Both can be sources for piRNA. These can produce both single and double-stranded piRNA. The secondary pathway is referred to as "ping Pong" replication, in which the silencing of transposons by the piRNA can lead to the production of more complimentary piRNA. The silencing complex works like the RISC but independent of DICER and mediated by Ago3, AuB and Piwi proteins. The piRNA directed cleavage leads to the to the creation of new piRNAs with the necessary sequence and RNA length to be used for further silencing, leading to a potential positive feedback loop feeding silencing. At no point has there been any observations of the involvement of DICER or double-stranded RNA. Also, it has been observed that this complex is capable of guiding epigenetic modifications [Yin et al,. 2007].

1.12 Non-RNA Interference sRNA

As shown in the literature, there are many classes of sRNA that are not directly involved in the RNAi pathway [Borges, F., Martienssen, R. 2015] including small nucleolar RNAs (snoRNA), small nuclear RNA (snRNA), tRNA derived sRNA (tsRNA) small ribosomal derived sRNA (srRNA). All of these have important functions and not all directly with gene regulation. Small nuclear RNA is found in the eukaryotic organisms that perform DNA splicing to remove exon regions from mRNA; these are usually 150nt, making them among the larger sRNA and like other sRNAs function through an RNA protein complex. In this case, small ribonuclear proteins which are themselves complexes of several different proteins and RNA, such as the U1-6 splicosomal RNA. These snRNAs [Hanley BA, Schuler MA.1991] are distinguished from other small RNAs by their high uridine contact and the fact they accumulate in the nucleus. snoRNAs differ from snRNAs in terms of structure and function; they are involved in chemical modification of other RNAs, usually ribosomal RNAs and transfer RNAs as well as in the biogenesis of ribosomes.

These can be distinguished structurally by the size: 17-19nt and greater than 27nt are often associated with methylation, while the 20-24 nt snoRNAs are more commonly found to be involved with pseudouridylation. Within the class of snoRNA come the sub-class called the small Cajal body specific RNAs (scaRNAs), which are found in the Cajal body nuclear organelle and play a role in creating snRNPs for interaction with snRNAs and mediating methylation of the same splicosomal RNAs involved with those snRNAs [Adachi et al,. 2007]. Small rDNA derived RNAs are s RNAs expressed from ribosomal genes, usually between

18-30nt. These exist specifically to bind to Argonaute, but without driving or seeming to guide the RNA silencing complex [Wei et al,. 2013], distinguishing them from siRNA and miRNA that are also derived from ribosomal genes. tsRNA are s RNAs expressed from tRNA usually by a biogenesis pathway involving the cleaving or processing of existing tRNA. They have been observed in prokaryotes and eukaryotes. These RNAs are highly conserved and divided into 2 subclasses: tsRNA, which are the products of cleaved tRNA producing 28-36nt sRNA, and those tsRNA produced by nucleases found in bacteria, yeast and by angiogenin in human genomes. These are distinguished from tRFs by them being 14-22nt fragments of tRNA, which is cleaved by DICER or by a RNase.

A greater understanding of the cross-kingdom RNA transfer between plants and pathogens could lead to greater advances in pathogen treatment, protection from pathogens and in the field of RNA silencing as a pesticide. Modified plants can increase the production of the pathogen interfering RNAs or target the sRNA originating from the pathogen

1.13 sRNA in plant-pathogen interactions.

Plant and pathogen interactions have mostly been understood in terms of the interactions between effectors and the plants innate immunity. This research has focussed on the interactions between proteins at the plant-pathogen level and how these pathways have impacted plant resistance. sRNAs have been recognised as players in the interactions between plant and pathogen, first as regulators in host-bacterial interactions [Katiyar-Agarwal S,. 2010] and later in fungi and plant interactions. Both plant and animal pathogens have similarities in that they have adapted virulence or effector proteins to counter the immune systems of their host; many of these effectors have been discovered in fungal pathogens and oomycetes thanks to whole genome sequencing and studying the secretome.

These studies show that sRNAs can act as effectors in plant-pathogen interactions, making them readily useful for study in plant pathology. By silencing host genes these pathogens can improve infection and so by understanding them we can better inhibit infection. In the case of *B. cinerea*, the RNAi pathway of the host plant was used to mediate the silencing. As has been shown in this paper on the roles of s RNA and miRNA, they are often used for regulation of genes by plants. Other pathogens like *Verticillium dahlia* have also shown to use host mechanism dependent silencing to enhance pathogenicity [Ellendorff U et al,. 2009].

There have been various types of sRNA identified in plants and pathogens capable of performing host induced gene silencing. This includes multiple different types of parasites and pathogen not just related to fungi infection. These RNA can result in upregulation or down regulation of genes involved in plant defence. Most of the sRNA studied are not trans-kingdom silencing RNAs, but instead are sRNAs that are produced by the plant regulate plant genes involved in immunity. These genes can also be silenced to regulate immunity, such as susceptibility factors, defence regulators and auxin receptors [Si-Ammour A et al,.2011]. Many of these targeted genes are known resistance genes, and often the sRNA originates from the same gene that it targets. However, many of the studies plant sRNAs have no target transcript within the fungal genome; these are likely candidates to be transkingdoms sRNAs by a process known as inter-kingdom translocation.

The discovery of bidirectional sRNA transmission is both relatively recent and limited to a few fungi pathogens. *B. cinerea* showed the fungi originating sRNA were transferred to the plant during infection and functioned similar to effectors, repressing host immunity and silencing host genes [Weiburg et al,. 2013] and in *Verticillium dahlia* showing plants, sRNA transmitted to the fungi during infection [Zhang et al., 2016]; both fungal pathogens have a large host range and a limited number of host resistance genes found in agricultural cultivars.

Due to the complexity of fungal and plant genomes, observing and confirming the roles of silencing suppressors could be very difficult. An example of this is in the *P. sojae*, which has multiple effectors that suppress RNAi in the host [Qiao et al,. 2015]. However, alone it is ambiguous from this, whether that is because the RNAi pathway poses a threat to the fungi or because the regulation it performs is essential for plant defence against pathogens. It is well known that upon infection, plants use sRNA to moderate gene expression and sacrifice plant growth for defence.

One observation in *V. dahliae* is that it uptakes large amounts of plant derived sRNAs on infection [Zhang et al., 2016b]. This may be in part due to its location during infection; the movement of bidirectional sRNA is mediated by the plasmodesmata and when transported over long distances is mediated by vasculatures [Chitwood and Timmermans, 2010]. In addition, phloem-mediated silencing has been shown to be more effective than cell to cell [Liang et al., 2012]. It is likely that both means of sRNA movement between plant and pathogen are needed

for trans-kingdom RNA silencing; however, how the host sRNA is delivered to the fungal pathogen is far less understood. We know that they require the sRNA to be both mobile and stable, and most hypotheses of this transfer are that it is done by extracellular vesicle transport. It has been observed that bacterial infection can increase production of extracellular vesicles [Rutter and Innes, 2017], including observations of the EV's containing the prerequisite proteins for transport and trafficking across membranes, which may indicate a role in RNA trafficking.

The silencing capability of these mobile sRNAs depends on multiple different factors, these include RNA motifs including the 3' untranslated regions which can be recognised by RNA binding protiens and influence long distance mobility, [Leblanc et al., 2012]. Mobiles sRNA are also potentially more likely to be single stranded due to a biases in the sRNA found in phloem sap [Buhtz et al., 2008].

There is a correlation with vascular fungal pathogens and observations of sRNA transfer, possibly due to the fungal pathogen being more likely to encounter mobile sRNA during infection process than non-vascular fungi. There have been observations of plant miRNAs in *V. dahlae* hyphae, which indicates a role of plant miRNA in immunity through methods not just involving endogenous gene regulation.

1.14 Aim and Objectives

The overall aim of this research is to determine What are sRNA interactions between *A. alternata* and Tomato, A pathosystem that has not been explored for sRNA interactions, whether cross kindom silencing or involvement in plant defence. This needs to be done in order to identify new miRNAs and determine whether sRNA can be a viable control method for *A.alternata* leaf spot disease and reduce the role of fungicides in Agriculture.

To achieve these aims, specific objectives were as follows.

1. To assess the interaction between *A.alternata* and tomato to understand the basic pathology of the interaction to from a background of knowledge.
2. Establish a method for evaluating sRNAs targeting Chitin synthase and to determine if Chitin synthase targeting sRNAs are effective at inhibiting spore germination.
3. To sequence the sRNA of Tomato, *Alternaria alternata* and infected tomato tissue to determine novel miRNA involved in infection and sRNA of interest that may be involved in infection.
4. To sequence the *A.alternata* genome for my chosen strain carry out Bioinformatic analysis to validate sRNA results in comparison to the reference genome.

Chapter 2

MATERIALS AND METHODS

2.1 Materials

2.1.1 Plant materials and growth conditions

Tomato plants were grown from the commercially available Money-maker variety, acquired from local gardening supply stores. Seeds were germinated in trays with soil soaked with water under 12-hour light and 12-hour dark conditions for 14 days before being transferred into individual pots for completion of growth until harvest. Tomato leaves were harvested 6 weeks after transfer in which time leaves will have grown to at least 5 centimetres in width and 10 centre metres in length to optimise available leaf surface area for lesion development and available biomass for RNA and DNA extraction, and in enough quantity for biological replicas. Seeds were routinely sowed and grown to replace harvested tomato plants.

2.1.2 *A. alternata* strains

Four strains of *Alternaria* were acquired, A001 and Fera 15968 strains of *A. alternata* were obtained from Dr Andrew Armitage at the NIAB EMR institute, East Malling, strain LSHB SM-O430 from CABI (CAB International), and a sample of *A. brassicicola* from the samples available at Worcester university.

2.1.3 Cultivation of fungal isolates

A. alternata was cultivated by guidelines given in [Armitage., 2013]. *A. alternata* samples were transferred to PDA plates, which were prepared by dissolving 39 grams commercial PDA powder to 1 litre distilled water and autoclaving. After 7 days under temperature of 25°C and a 12-hour light and 12-hour dark conditions, plates are tested for spore production, and are flooded with sterile distilled water. Spore suspension was then collected and passed through sterile gauze and number of spores were measured in a haemocytometer.

The haemocytometer measurement was carried out with a standard method, using a pipette, 100 µL of spore suspensions were dropped to a haemocytometer, filling both chambers underneath the coverslip, allowing the spore suspension to be drawn out by capillary action. Using a light microscope, grid lines of the haemocytometer was examined.

Using a hand tally counter, the distinctive *Alternaria* spores were counted in one set of 16 squares, only spores wholly within the squares were counted to keep the method consistent between multiple counts. The average cell count was taken from each of the sets of 16 squares and multiplied by 10,000 (10^4) to achieve the final count. This was often needed to be diluted or concentrated to achieve the desired number. This was achieved by either adding more sterile distilled water to dilute or by centrifuging them in a 50ml tube with the spore suspension at 2000 rpm to pellet the spores, before resuspending in lower volume of liquid and re-counting. Spores were observed to confirm correct shape and to confirm no contamination.

2.2 Methods

2.2.1 Fungal assays:

2.2.2 Determining spore viability

Spores needed to be evaluated for their viability and this was carried out on cellophane strips, 1.5 cm in length, which were taken from plain transparent cellophane, this was autoclaved in prepared distilled water. After the autoclaving was complete, the cellophane strips were placed onto a PDA medium in Petri dishes and under a laminar airflow. Ten microliter spore solution were placed on cellophane and spore germination was observed after 24 hours incubation under light microscope. Observation of spore tubules with deformities, discoloration and abnormal growth would not be considered viable for germination assay

2.2.3 Optimising spore production

Spore production needed to be optimised to ensure reliable spore harvesting for germination assays. Multiple techniques were assayed for reliable conidia production.

Agar plugs from purified fungus was transferred to PDA plates and incubated at 25˚C under a 12-hour light and dark conditions. Similar conditions were tested but varied, incubation in the absence of the light and dark conditions and with up to 5 plugs per PDA plate to increase scarcity of available nutrients. In addition, the following techniques from [Carvalho et al., 2008] were evaluated.

The mycelium stress technique: This was carried out by the use of a needle being pressed into the colonies at 4 mm deep. The Petri dish was returned to incubate at the same conditions and after 24h, spore production was evaluated. Spore evaluation was also carried out in the same way as shown in 2.2.1. All plates were flooded with sterile distilled water and observed under light microscope using a Haemocytometer.

Cold shock technique: Fungal cultures were produced using the standard method and were transferred to a refrigerator in the absence of light. After this, the petri dishes were then taken to be induced to the mycelium stress technique, described above and transferred to another incubator and kept for 24h at 25°C with a 12 h light and dark conditions before being evaluated.

In addition to techniques of increasing spore production, the medium was also tested. Of the available media four were tested, PDA, low nutrient PDA, V8 media and sporulation media. Pad media was made using standard procedure, low nutrient PDA was made using the instructions provided by Andrew Armitage used to optimise spore production in his thesis, 15 grams of agar powder and 3.9 grams of potato dextrose agar was dissolved into 1 litre of distilled water and autoclaved for 20 minutes. This was allowed to cool completely before being poured into agar dishes, sporulation media was made due to instructions from research done in [S. Chaurasia, R. Chand and A. K. Joshi,. 1998] in which spore production was optimised by the addition of 20g per litre of calcium carbonate to normal PDA media a d addition of 2ml of sterile water to the agar surface for increased humidity and v8 media made with 1 gram potato extract, 5 grams of glucose, 150ml V8 vegetable juice and (Campbell south Cc.) and 3 grams $CaCO_3$ and 20 gr per litre of agar.

2.2.4 DNA extraction of fungal samples

Fungal samples in petri dishes are taken to a class 2 safety cabinet where fungal mycelia are scraped into a petri dish which has been cooled with liquid nitrogen and ground while gradually covering the mycelia with Liquid nitrogen. The ground mycelium powder is

transferred to Eppendorf tubes for extraction using DNA extraction kit Isolate 11(Bioline). The extracted DNA is stored at -20.

2.2.5 ITS sequencing

All the isolates were subjected to PCR verification (Thermal Cycler, Applied Biosystems) using the universal ITS 4 and 5 primers. Bio Mix™ Red (Bio line) was used and the PCR regime employed were 96˚C for 5 min, 94˚C for 30 sec, 56˚C for 30 sec and 72˚C for 1 min for 30 cycles followed by 10 min of extension at 72˚C. The PCR products were run in a 1% agarose gel to determine if the PCR was successful. PCR products were then cleaned using a Plasmid mini prep kit and sent for Sanger sequencing at Eurofin Genomics. BLASTN analysis against the NCBI databases were performed with the sequence results obtained. If the sample came back as not *Alternaria* the batch would be rejected as non-viable.

2.3 Plant inoculation

Spores from each of the available isolates were collected and were diluted to 5×10^4 /mL and drop inoculated on tomato leaves. These were incubated at 28 °C at an 8h light and 16h dark conditions for at least 7 days to observe the generation of *A. alternata* lesions. 2 droplets were placed on the right side of each leaf with 2 drops of sterile distilled water as the control placed on the opposite side of each leaf.

2.3.1 Lesion measurement

After 7 days the pictures were taken using a camera using a stand. Each experiment was done using 3 biological replicates and each experiment was repeated at least three times. All the pictures were taken with the camera (Canon EOS-350D) this camera was set at the same height, focus and resolution by every batch of experiments was kept the same to remove any possible discrepancies when using ImageJ software to quantify lesions.

2.3.2 Image J analysis

Every batch of treatment's pictures were saved in individual files to acquire the data of the fungal growth and the batches of the pictures were evaluated using ImageJ. The scale of the image was set using a 10 mm measuring ruler and applied to all samples. The area of fungal

growth was selected and measured. This tool was used and exported before the values were analysed using statistical analysis.

2.3.3 Statistical analysis of lesion measurements

Statistical analysis of the pathology data was performed to determine if there was any significant difference ($\alpha = 0.05$) between the lesion size induced by the different strains of *Alternaria* and between the fungi and the control. Because the experiments were established with multiple variables, a Two-way Anova was performed using SPSS 25.0 to determine the effect of the differences between the strains on infectivity in tomato, *A. alternata* Cabi, Fera, A001 and *A. brassicola*. In addition to this, a one-way Anova was performed to evaluate statistical differences in each group between the infected samples and its control, as this was separate for each sample and only containing two samples a one-way Anova was more appropriate. In cases where samples are significantly different a post-hoc test was performed.

2.4 Molecular genetics techniques.

2.4.1 Agarose gel electrophoresis

All DNA, RNA and PCR products were run on gels to verify success in extraction or amplification. The method for gel production varied depending on whether it is to be used for DNA or RNA; the DNA products were mixed with 5ul of DNA loading buffer (Gel Loading Dye, Purple (6X), NEB B7024) and run on 1.5 % agarose gels, gels made by mixing the amount of agarose powder in a W to volume ratio to 1x T.A.E buffer and heating briefly in a microwave to dissolve, allowed to cool before adding Safe view (NBSBio) in order to visualise DNA and RNA under UV. These were then run-on different settings based on the product. DNA and PCR products were run on 75 volts for 30 minuets, DNA fragments were visualised using a transilluminator (BioSpectrum® 310 IS, UVP LLC).

Since RNA is less stable at high temperatures, a separate gel electrophoresis machine with a lower heat generation was used at a lower voltage, running for 50 volts for 45 minuets. In addition, agarose gels were mixed with 200ul of common household Bleach (6% sodium hypochlorite) which degrades RNases and denatures RNA secondary structure [Aranda Et al 2012] and improves RNA visualisation.

2.4.2 DNA isolation

Genomic DNA isolation was performed using the ISOLATE II Genomic DNA Kit (Bioline), following the instructions of the manufacturer. Fresh samples from 24h grown of *A. alternata* cultures were used to do the extraction. To assess the quality of the extraction, 5 ul of gDNA samples were carefully mixed with 5ul DNA loading buffer (Gel Loading Dye, Purple (6X), NEB B7024) and run in a 1% agarose gel as shown in 2.3.1 and a picture of the gel was taken. The quality and concentration of the gDNA was measured with Nanodrop 2000c (ThermoScientific).

2.4.3 Touch down PCR

Touchdown PCR was performed to determine viability of primers and compensate for primers with significantly different annealing temperatures interfering with gene expression comparison. Initial heating at 95 ˚C for 5 minutes, then 10 cycles touchdown of 95 ˚C for 30 sec, annealing temperature of 65 °C, decreasing 1 degree every cycle to 56, and extension at 72 for 30 seconds. After 10 cycles of touchdown, a further 25 cycles of 95 °C for 30 seconds, 60 °C for 30 seconds and 72 °C for 30 seconds and a final extension at 72 °C for 5 minutes were carried out.

2.4.4 RNA extraction

Total RNA for RNA-sequencing and generating the s RNA library was generated through an optimised Trizol method. Samples ground in liquid nitrogen were dissolved in 1ml Trizol Sigma (Gillingham, UK) with 10ul beta-mercaptethanol or 40mM DTT. These were vortexed and centrifuged briefly to remove large sediment. The removed liquid was then mixed with 150 ul chloroform and span at 10000 rpm for 15 minutes, the clear top layer was extracted, and the process were repeated 2 more times. The final extracted upper layer was added to 1ml Isopropanol and chilled at 4 °C overnight, before being centrifuged for 10 minutes at 10000 rpm. The pellet was then cleaned three times with ice cold 70% ethanol before being air dried and resuspended in sterile DEPC water and tested under a Nanodrop 2000c (ThermoScientific). For the optimal RNA readings at 260 /280 and 260 /230 values were recorded. The RNA was treated with DNase Ambion® TURBO DNA-free™ according to manufacturer's instructions, samples were inactivated with the provided inactivation reagent and centrifuged for 5 minutes

at maximum speed to remove the inactivation reagent. The aliquoted samples were either used immediately or stored at – 80˚C.

2.4.5 cDNA production

cDNA synthesis was carried out in PCR tubes using the following mixture: 1 µl of RNase inhibitor, 4.0µl of 5x cDNA Synthesis Mix (PCR biosystems) and 1µl of 20x RTase (PCR biosystems) were added to 5 µl of total RNA, and PCR grade dH_2O was added up to 20µl final volume. The mixture for each RNA sample was incubated at 42°C for 30 minutes followed with an incubation at 85°C for 10 minutes to denature RTase

2.4.6 Quantitative Real Time PCR of *Chitin synthase genes* (*CHS*)

A. alternata spores were extracted from the plates flooded with sterile distilled water. The spore solution was viewed under microscope to confirm the confluent presence of spores. Spores were separated into 11 200ul spore solutions and 10ul of sRNA solution was added and incubated for 1 hour. The spores were then spun in a centrifuge at 10000 rpm to pellet the spores. Samples of the spores' post sRNA treatment were taken for RNA extraction. TRIzol™ protocol (Themofisher, #15596026) this was used according to manufacturer's instruction. The RNA is treated with DNase 1 (New England Biolabs; Ipswich, MA, USA; M0303) as according to the manufacturers protocol to account for any residual gDNA. RNA was stored at -80 degrees to take for qPCR analysis, with the sample separated into 10ul aliquots to reduce risk of degradation. Quality of the samples was measured with a Nanodrop 2000c (Thermofisher; Wilmington, Del, USA) with samples preferably seen to be >1.8 in 260/280 and 260/230 absorbance ratios. This indicates the purity and quality of samples. PCR was done with the primers for ACTIN and the relevant primers for the chitin synthase targeted in that sample. In addition to validate the samples to be free of DNA, each aliquot was tested with a standard PCR as explained in [2.6.3] using ITS primers ITS1/2 (5'–TCCGTAGGTGAACCTGCGG -3', 5'–GCTGCGTTCTTCATCGATGC -3')

Quantitative Real Time PCR (qRT-PCR) was carried out using SensiFAST™ SYBR® No-ROX OneStep Kit (Bioline). The first goal was to compare *CHS* expression in fungal mycelia to a housekeeping gene, *Actin*. Five µl of SensiFAST™ SYBR® No-ROX OneStep mix, 2 µl of template, 0.5 µl of each primer, 0.05 µl of reverse transcriptase, 0.002 µl RNA-inhibitor and DEPC water were added to give a final reaction volume of 20 µl. qPCR reactions were carried

out in 96-well plates using a Roche Light Cycler Real-Time PCR System. The mycelial cDNA was evaluated for expression of each of the 11 *CHS* genes. Actin expression was used for normalisation, and negative controls of primer and reagent mix with no template (no template control) were also included as a negative control.

2.4.7 Calculating the relative quantification using 2–ΔΔCt method

To measure the relative expression of the 11 *CHS* genes in relation to the house keeping gene Actin, a qPCR was done using SensiFAST cDNA synthesis kit following the manufacturer instructions. RNA samples were stored on ice while the master mix was prepared over an ice block using SensiFAST™ SYBR No-ROX One Step mix (2x) (5µl), primers 10mM (0.5µl), template (2µl), Ribo Safe RNase inhibitor (0.02µl), Reverse transcriptase (0.01µl) and DEPC-H2O for a total volume of 10µl. The samples were transferred to a 96 well plate, sealed and placed in the LightCycler 480 II (Roche Molecular Systems, Inc.) each sample was prepared with three replications in addition with the negative control with distilled water instead of the sample and the house keeping gene in which the actin primers were used instead of the chitin synthase primers. The light cycler was then run as shown. Calculating relative expression to calculate the relative expression the Cp values were recorded from the light cycler and the 2–ΔΔCt (Livak and Schmittgen, 2001) method was used to determine the relative quantification of the gene expression taking into account the background expression using the actin gene. To do this the cross point (Cp) from the absolute quantification were placed in excel and used to create the normalised values. Using the change in expression from the sample C removing the actin C resulting in the (ΔCt). This allows the calculation of the fold change using the control sample normalised (ΔΔCt) this resulted in both an ΔCt and ΔΔCt plot of the 11 chitin synthase genes showing the change in expression comparing the treatment samples (sRNA targeting the tested CHS gene) and the control with no sRNA treatment.

2.4.8 qPCR statistical analysis

qPCR was performed in replicates and a Two-way Anova was performed to compare to each other as described in Section 2.5.3.

2.5 Bioinformatics

2.5.1 Total Genome sequence

Genomic DNA (gDNA) were extracted from seven-day mycelial cultures by using ISOLATE II Genomic DNA Kit (Bioline). Extracted DNA was run on Agarose Gel to determine the quality of the sample. The quality and concentration of the gDNA was measured with Nanodrop 2000c (ThermoScientific). gDNA were then sent to BGI for sequencing.

2.5.2 Primers used for qPCR

Primers were generated using Primer3 and the target region was chosen on Geneious, using domains and open reading frames determined using InterProScan [Figure 2.1] . Primers were tested for effectiveness in DNA first using the touchdown PCR method (Section 2.3.3). These were run on an 1% agarose gel before being used for qPCR.

Table 2.1 Primers used for the amplification of the CHS genes

Gene ID	Primer orientation	primer sequence
Actin	R	CACTGCCGAGCGAGAAAT
ACTIN	F	CAAGACGGAAGGCTGGAA
OAG19177	R	GAGCGCGGTGTAGAAGTTG
OAG19177	F	GCACAGTCGACCTTGAGTTG
OAG19173	R	TAATGACGACCTTGGGAGCG
OAG19173	F	GATTTCCTCCAGTGGTTCGC
OAG21670	R	CGACCGTGTGTTCTCCAGAT
OAG21670	F	TCTCTACCGCAGCGACCC
OAG18953	R	GGGGTATTGTGTTGGCCTGT
OAG18953	F	CCAGGAGCATGACGACGAAG
OAG25348	R	TTCACCCTTGCTGCTGACTC
OAG25348	F	CGGCGCCAACCAAGAATATC
OAG23547	R	CCTCTTCGTACGCTCCTTCA
OAG23547	F	GTCAAGTTGATTCAAGGCTCTGT
OAG20922	R	AATGGCGCTGGTTGAACTTG
OAG20922	F	GAGCGTCCGTCAAGTAGGTC
OAG13919	R	TCCATAGGCTCCTTGTCCCA
OAG13919	F	AAGGCATCGAGGACCACAAG
OAG23516	R	ACGACGTGATCTCCTTTCGG
OAG23516	F	CTATGCGCTCTTGTGGGTCA
OAG24007	R	CGAAGCTGAACTGTGATGGGA
OAG24007	F	CCAGGGACGGATTGAAAGATTC

2.5.3 Generation of small interfering RNAs (siRNAS)

CHS gene sequences acquired from the Ensembl Fungi database, and protein domains and ORFs were found using InterProScan (Figure 2.1). These genes were then compared against the total genome sequence acquired from BGI using BLASTN, to confirm their presence and that there were no differences in our strain (Table 3). sisRNAs were generated to target each gene, optimised to be 30 nucleotides (30-nt) long and located at the 5' region of the gene (Table 2). Three siRNAs were designed for each gene. A single siRNA was chosen for each gene, these were then blasted into the genome of our sample to confirm that there were no SNP differences that may interfere with sRNA effectiveness (Table 4) and to determine there is no off-target site within the genome. These sequences were ordered from Sigma (Gillingham, UK) pre-prepared as double stranded siRNA. Genomic data and BLAST

Table 2.2: Sequences of siRNAs targeting *CHS* genes in *A. alternata*.

siRNA name	Gene location	sRNA sense strand	sRNA Anti-sense strand
OAG13919 sRNA 1	*1009 - 1038*	GCAGAAAGGCATCGAGGACCACAAGCAGTG	CGTCTTTCCGTAGCTCCTGGTGTTCGTCAC
OAG13919 sRNA 2	*1069 - 1098*	CGGTCCAGGCATGACGCAGACCACTGGCGA	GCCAGGTCCGTACTGCGTCTGGTGACCGCT
OAG13919 sRNA 3	*1146 - 1179*	TGAAGGATGCATACATGGCCTGGGACAAGG	ACTTCCTACGTATGTACCGGACCCTGTTCC
OAG18953 sRNA 1	*1148 - 1177*	GAACCCTACCGGCGCCCACTCAGACCCTTT	CTTGGGATGGCCGCGGGTGAGTCTGGGAAA
OAG18953 sRNA 2	*1215 - 1244*	AGCTTGACCGAGTCGTATGCCACCGACAGG	TCGAACTGGCTCAGCATACGGTGGCTGTCC
OAG18953 sRNA 3	*1249 - 1280*	CACAATACCCCGGCCAAGCTCACGATGGCT	GTGTTATGGGGCCGGTTCGAGTGCTACCGA
OAG19173 sRNA 1	*1067 - 1096*	CGTTTTGGTCTTTCGCCGGTCTGTTGACTG	GCAAAACCAGAAAGCGGCCAGACAACTGAC

OAG19173 sRNA 2	*1100 - 1129*	CGTGGCTGTTCTGTGTTTACCAGCTTGTCA	CGTGGCTGTTCTGTGTTTACCAGCTTGTCA
OAG19173 sRNA 3	*1156 - 1185*	ATCAAATCTGTTCTCGACGAGTCATCCGCT	TAGTTTAGACAAGAGCTGCTCAGTAGGCGA
OAG19177 sRNA 1	*971-1000*	TTTGCCCGGACATCGACCATCTGGGACGAC	AAACGGGCCTGTAGCTGGTAGACCCTGCTG
OAG19177 sRNA 2	*1004 - 1033*	TGCAGTTGACACACGGTACCACCTCATCTA	ACGTCAACTGTGTGCCATGGTGGAGTAGAT
OAG19177 sRNA 3	*1039 - 1068*	TGCAGTTGACACACGGTACCACCTCATCTA	TGCAGTTGACACACGGTACCACCTCATCTA
OAG19178 sRNA 1	*1007 - 1036*	CCTGTCCAAGAAGATCGAACACGCCAACTT	GGACAGGTTCTTCTAGCTTGTGCGGTTGAA
OAG19178 sRNA 2	*1047 - 1076*	GCCTTCACCACTACCAAGTCCCTCACTACC	CGGAAGTGGTGATGGTTCAGGGAGTGATGG
OAG19178 sRNA 3	*1122 - 1151*	ACCACTACCGCAAACAACGCCACTCTTCTG	TGGTGATGGCGTTTGTTGCGGTGAGAAGAC
OAG20922 sRNA 1	*411-440*	GAGAAGCCAGAGAACTTTGTCTACCTGCTA	CTCTTCGGTCTCTTGAAACAGATGGACGAT
OAG20922 sRNA 2	*535-564*	TGATCTGCGACGGCCGAGTCAGCAGCAAGG	ACTAGACGCTGCCGGCTCAGTCGTCGTTCC
OAG20922 sRNA 3	*757-786*	TTCTTATTCGATCGTTCATGCGCAAGTTCA	AAGAATAAGCTAGCAAGTACGCGTTCAAGT
OAG21670 sRNA 1	*940-969*	TCGCCGCCTCGACGGGGCGAGCGGCGACAT	AGCGGCGGAGCTGCCCCGCTCGCCGCTGTA
OAG21670 sRNA 2	*1040 - 1069*	CCAATAGCAAGAGCGGCACCCAGGCCTACT	GGTTATCGTTCTCGCCGTGGGTCCGGATGA
OAG21670 sRNA 3	*1001 - 1030*	CTATGGAGCGCGATCCCATAATGAGAGGCC	GATACCTCGCGCTAGGGTATTACTCTCCGG
OAG23516 sRNA 1	*842-871*	CATTCAGACCGAGTGCTATGCGCGGACGGT	GTAAGTCTGGCTCACGATACGCGCCTGCCA
OAG23516 sRNA 2	*926-955*	ATACCCTCATCTTCGAGACGGCCAACGACT	TATGGGAGTAGAAGCTCTGCCGGTTGCTGA

OAG23516 sRNA 3	*883-912*	GGCCATGGTCATGACAGTCGTCATGATTAT	CCGGTACCAGTACTGTCAGCAGTACTAATA
OAG23547 sRNA 1	*1060-1098*	AACAAGAAGAGGCCGCTCAACACCTTCTCA	TTGTTCTTCTCCGGCGAGTTGTGGAAGAGT
OAG23547 sRNA 2	*1094-1123*	CTCACTTATCCCTACCAACGACTACTCACG	GAGTGAATAGGGATGGTTGCTGATGAGTGC
OAG23547 sRNA 3	*1160-1189*	TACGGAAACCTACAGTGCAATACAGCCCCC	ATGCCTTTGGATGTCACGTTATGTCGGGGG
OAG24007 sRNA 1	*1056-1085*	AAAAAACACTCCAAGGATACCCTACGTGTC	TTTTTTGTGAGGTTCCTATGGGATGCACAG
OAG24007 sRNA 2	*1123-1154*	TGCGGTGTTACCCGAGTCGCTCCACATTGC	ACGCCACAATGGGCTCAGCGAGGTGTAACG
OAG24007 sRNA 3	*1091-1120*	AAGACCGAGTGGACGATCATTCTCCAGACA	TTCTGGCTCACCTGCTAGTAAGAGGTCTGT
OAG25348 sRNA 1	*357-386*	GTCATTATCTTCTTCATCATAAGACGATTG	AAGACCGAGTGGACGATCATTCTCCAGACA
OAG25348 sRNA 2	*483-512*	GATTCGTTGGCCGAGCAGCAGTTATTAGAT	CTAAGCAACCGGCTCGTCGTCAATAATCTA
OAG25348 sRNA 3	*549-578*	CGAGTCAGCAGCAAGGGTGAACCCAAAACA	GCTCAGTCGTCGTTCCCACTTGGGTTTTGT

Table 2.3 CHS genes were aligned against the Whole genome sequence of our A.alternata strain

Gene ID	e Value	Bit score	query cover	hit start	Hit end	gene name	
OAG13919	0	5602.01	100.00%	30204	33236	NW_017306225.1	Alternaria alternata strain SRC1lrK2f
OAG18953	0	8909.36	100.00%	465438	470261	NW_017306203.1	Alternaria alternata strain SRC1lrK2f
OAG19173	0	7147.65	100.00%	41104	44973	NW_017306202.1	Alternaria alternata strain SRC1lrK2f
OAG19177	0	13415.2	100.00%	55761	48498	NW_017306202.1	Alternaria alternata strain SRC1lrK2f
OAG19178	0	12621.1	100.00%	55846	62679	NW_017306202.1	Alternaria alternata strain SRC1lrK2f
OAG20922	0	5554	100.00%	1105315	1108321	NW_017306199.1	Alternaria alternata strain SRC1lrK2f
OAG25348	0	3338.02	100.00%	1105915	1107721	NW_017306199.1	Alternaria alternata strain SRC1lrK2f
OAG21670	0	10200.2	100.00%	701795	707317	NW_017306197.1	Alternaria alternata strain SRC1lrK2f
OAG23516	0	5007.39	100.00%	1468849	1471559	NW_017306194.1	Alternaria alternata strain SRC1lrK2f
OAG23547	0	8556.65	100.00%	17945	22577	NW_017306193.1	Alternaria alternata strain SRC1lrK2f
OAG24007	0	16151.9	100.00%	1120982	1112237	NW_017306193.1	Alternaria alternata strain SRC1lrK2f

CHS gene a	E value	Bit score	grade	Hit start	hit end	Name in sequence	Description
OAG13919 sRNA 1	3.19E-09	56.5198	75.00%	19104	19133	NW_017306193.1	Alternaria alternata strain SRC1lrK2f
OAG18953 sRNA 1	3.19E-09	56.5198	75.00%	1119927	1119898	NW_017306225.1	Alternaria alternata strain SRC1lrK2f
OAG19173 sRNA 2	3.19E-09	56.5198	75.00%	1469690	1469719	NW_017306202.1	Alternaria alternata strain SRC1lrK2f
OAG19177 sRNA 1	3.19E-09	56.5198	75.00%	702734	702763	NW_017306202.1	Alternaria alternata strain SRC1lrK2f
OAG19178 sRNA 3	3.19E-09	56.5198	75.00%	1106671	1106700	NW_017306202.1	Alternaria alternata strain SRC1lrK2f
OAG20922 sRNA 3	3.19E-09	56.5198	75.00%	1106397	1106426	NW_017306203.1	Alternaria alternata strain SRC1lrK2f
OAG21670 sRNA 1	3.19E-09	56.5198	75.00%	54791	54762	NW_017306197.1	Alternaria alternata strain SRC1lrK2f
OAG23516 sRNA 1	3.19E-09	56.5198	75.00%	56967	56996	NW_017306199.1	Alternaria alternata strain SRC1lrK2f
OAG23547 sRNA 3	3.19E-09	56.5198	75.00%	42203	42232	NW_017306199.1	Alternaria alternata strain SRC1lrK2f
OAG24007 sRNA 1	3.19E-09	56.5198	75.00%	466585	466614	NW_017306193.1	Alternaria alternata strain SRC1lrK2f
OAG25348 sRNA 1	3.19E-09	56.5198	75.00%	31212	31241	NW_017306194.1	Alternaria alternata strain SRC1lrK2f

2.6 Germination Assay

2.6.1 sRNA treatment

The *A. alternata* spores were first collected from their PDA plates, spore samples were then washed in sterile distilled water and the spore concentration was taken with a haemocytometer. This concentration was adjusted to 5×10^4 /ml using the haemocytometer. The sRNA solution was added to the spore suspension at a final concentration of 5, 10, 15 and 20 µM. PDA medium was sterilized by autoclaving at 15 psi and 121 ℃ for 15 min.

Approximately 20 mL of the medium was aliquoted into each of the sterile Petri dishes in a laminar airflow unit. 10 µL spore suspension was placed on each segment of cellophane. Plates were incubated with 12 h light and dark regime at 20 °C. Spores were then examined using a light microscope and 24 h after incubation. Following this the germinated spores were counted using a Leica microscope and camera. In addition, subsequent experiments were made to determine the effect of sRNA on the length of germination tubes. Each siRNA treatment was replicated 3 times, from each replicate 10 spores were measured using the 40x Leica microscope um scale. The average of these 10 spores were taken for each replicate, and then the average taken from these three replicates. This average was used to calculate the standard error.

2.6.2 Statistical analysis of spore germination assays

Statistical analysis using a two-way Anova was performed on the spore germination data to evaluate the means of more than two variables for statistical significance. This was done using SPSS using the same method as in Section 2.5.3.

2.7 Sequencing small RNAs.

Sample collection was performed by taking sections of *A.alternata* mycelium from a Petri dish. Once RNA was extracted using the methods shown in [2.4.4] the RNA needed to be quantified and its quality measured. Measuring RNA degradation was done by running he samples on 1% agarose gels the RNA quality was meased using the NanoPhotometer® spectrophotometer (IMPLEN, CA, USA) and the degradation measured using the RNA Nano 6000 Assay Kit of the Agilent Bioanalyzer 2100 system (Agilent Technologies, CA, USA) carried out by Novogene(Novogene Experimental Department) .

The next step was to prepare small RNA libraries, the sequencing process was carried out by Novogene (Novogene Experimental Department). The total quantity of RNA was 3 μg per sample. The samples were taken as the input material and sequencing was undertaken using NEBNext® Multiplex Small RNA Library Prep Set for Illumina® (NEB, USA.) this was done following manufacturer's recommendations. The NEB 3' SR Adaptor was then ligated to 3' end of extracted s RNAs. Once the ligation reaction was complete, the RT Primer was hybridized to the of 3' SR Adaptor. The now double stranded RNA molecule was now ready for the next step. The 5´end adapter was then ligated to the 5´ends of the s RNA. This allowed for cDNA synthesis, cDNA was synthesized using M-MuLV Reverse Transcriptase.

A PCR amplification reaction was performed using the LongAmp Taq 2X Master Mix, SR Primer for Illumina, and an index primer. The PCR products were then purified on a 8% polyacrylamide gel at a voltage of 100 for 80 min. the DNA fragments which were between 140 and 160 bp which corresponds to the length of the s RNA including the adaptors were recovered from the solution and dissolved in 8 μL elution buffer.

The quality of the library was then tested using Agilent Bioanalyzer 2100 system making use of DNA High Sensitivity Chips. The index needed to be clustered,a method used to explore the groups of RNA and their relationships, this was done using a cBot Cluster Generation System

using TruSeq SR Cluster Kit v3-cBot-HS (Illumina) following the manufacturer's guidlines. Once cluster generation was performed the library samples were taken for genetic sequencing on an Illumina platform programmed for samples 50bp in size.

2.7.1 Data analysis (Novogene Gene Regulation Department)

The samples were first processed for quality control, this was done by analysing the raw sequence reads through a custom python program. For the next step the clean sRNA reads were separated from other fragments by separating those with a poly-N, with 5' adapter contaminants, without 3' adapter, containing ploy A or T or G or C. this and in addition low quality reads were removed from the raw data. While this was being performed the Q20, Q30, and GC-content of the raw data was measured. A range of clean sRNA reads was taken for analysis, the tagged s RNAs were mapped using Bowtie using a setting without mismatch to use for expression analysis and understand their distribution across the genome

The Known mirRNA were determined using the miRBase2 database as a reference, modified software mirdeep2 and sRNA-tools-cli were used to determine the miRNA sequences and to generate the secondary structures. Following this Custom scripts (Novogene) were used to obtain the miRNA read counts as well as Bp biases on the miRNA first position with certain length and on each position of all identified miRNA respectively. The sRNA that originated from protein-coding genes that corresponded to protiens or other known s RNA such as snoRNA were mapped to RepeatMasker and the Rfam database.

2.7.2 Novel miRNA prediction

The hairpin structure of novel miRNA can be used to predict and learn valuable information. Using this method, the precursor can be used to predict novel miRNA from raw data. This was done by NOVOGENE using a method based on miREvo software [Wen et al., 2012] and mirdeep2 [Friedlander et al., 201]. This methodology allowed the prediction of novel miRNA based on the secondary structure. These structures include the Dicer site, this was done using Custom scripts (Novogene) and were used to learn the read counts and sequences of identified miRNA in addition to their bp bias on the first position of the miRNA depending on the length of that miRNA.

2.7.3 Bioinformatic analysis of sRNA data

Bioinformatic data from the sRNA sequencing resulted in total sRNA samples in three replicates from the three different materials: plant alone, fungi alone and infected tissue. These sRNA libraries were uploaded into Geneious for analysis.

2.7.4 sRNA mapping

Total sRNA for the samples from the infected, fungi and plant samples were mapped to the reference genomes for both *A. alternata* and the tomato to identify sRNA mapped to a particular gene and the read counts for each gene mapped to. This was quantified and compared between the infected and the fungi alone when mapped to the fungal genome and for the plant alone and the infected when mapped against the plant genome. sRNAs identified to map to the reference genome were recorded and compiled into a list. These sRNAs were then mapped to the opposite genome to look for similar RNA that matches to the proposed target genome.

The read numbers for each unique sRNA were normalised based on the total sRNA mapped to the *A. alternata* or tomato genome, or *A. alternata* infected sRNA libraries. The sRNA must follow the following criteria 1) it must be present in both the infected and the Host/Target libraries; 2) its normalised read count must be higher than 100 once mapped to the respective genome; 3) the normalised reads must be higher in the infected than in either the tomato or Fungal libraries; and 4) it must have predicated targets in the plant for fungi genomes.

Target gene prediction was performed using TAPIR1.1 with the following requirements, no gap within the alignments between the sRNA and target, the 10^{th} nucleotide must match the target, a maximum of 1 mismatch or two was permitted from position 2-12. A maximum of 2 continuous mismatches were permitted and a score of 4.5 was used as the cut off for acceptance. If the sRNA meets all these criteria, it would be accepted as a possible sRNA of interest. In addition, sRNA was grouped together if their 5' end and 3' end were within 3 nucleotides on the genomic loci.

Chapter 3

Pathology and Morphology

3.1.1 Causes of plant disease

Disease is defined as a malfunctioning of the organism; in most cases plant disease is the result of the influence of an infectious organism. This can include mycoplasma, viruses, parasites and fungi, these causes can vary from plant to plant as for which diseases are most dangerous and prevalent. For the ornamental plants the most common diseases are abiotic diseases, defined as diseases resulting from non-biological factors such as drought nutrient loss, overwatering, pollution, mineral toxicity, and light damage [Robert Schutzki, 2015]. Usually, the symptoms of the disease can be used to diagnose the cause of disease and determine biotic from abiotic disease. Leaf disease is often associated with spots and lesions, appearance of spores and fruiting bodies. Other observations such as cankers, defined as unusual lesions on the stem or branches that can interfere with water transport and movement of valuable plant nutrients in the phloem, and additionally isolating and identifying the biological agent at fault. Many fungi especially can be identified by their spore or mycelial structure under a microscope or at least narrowed down to a particular genus or family [Doehlemann G et al 2016].

Plant disease is not understood in isolation. Plant disease is a result of a combination of factors, including pathogen, environmental factors, and the hosts response to these factors [Gullino, M.L., Garibaldi, A. 2018] Due to this a weak or unaggressive pathogen could easily cause a significant disease so the environment is conductive to this or if the host is particularly susceptible. Often it is the very conditions that favour and encourage plant growth that also favour disease. Increased water use, fertilisation, warmth, and humidity can all increase plant growth but also provide a fertile breeding ground for infection.

3.1.2 Pathology of *A.alternata*

The morphology of the *Alternaria* genus has been well studied and understood for many years, this has been set out in [Simmons 2007] there are 276 taxa of the Alternaria genus, many of these morphological groups do not correlate with genetic variability within the genus [Andrew et al., 2009]. Morphological identification is the process of classifying Alternaria strains within the genus by physical differences, observations in spore shape and other factors there is much disputation as to whether these morphological differences qualify as different species or not. *A.alternata* morphology is defined by spore formation groups, with the species within the genus being separated into six different spore groups, which were first advocated for Alternaria as a morphology in 1993 [Simmons and Roberts 1993] this process of identification began after hundreds of Alternaria isolates were extracted from leaves, fruit, and fungal lesions from a variety of plant hosts. The six sporulation groups were identified based on physical differences in these isolates.

The next major identifier of *Alternaria* isolates is host specificity, within the six sporulation groups alternaria isolates can be separated based on host-specificity [Takashi Tsuge et al,. 2013] Two of the sporulation groups are identified as being pear host-specific, sporulation groups 1-2 later identified as *Alternaria gaisen*. *Alternaria* species are defined in these papers based on morphology, defining them as distinctive Alternaria which grow in solid culture, with distinctive sporulation features, and microscopically distinguishable characteristics and not based on evolutionary relationships between.

Some of these are major understood *Alternaria* species, *A. alternata* spores are observed to be 40mm in diameter at seven-fourteen days of culture, chains consisting of 4-6 conidia of short conidiophores, detective sporulation patterns based on a single strand of the canephore as the centre of a branch of apical clusters of branching spore chains, containing up to 20 conidia.

The proposed Host specific Toxin (HST) must be capable of reproducing in isolation the philological effects of the infecting pathogen and lastly that the physiological changes caused by the proposed HST in the host must result in colonisation of the host by the pathogen

ensuring that the effects are not slimily a by-product of the pathogens presence but plays a direct role in colonisation of place cells with fungal mycelia [Takashi Tsuge et al,. 2013].

3.1.3 HST identification in *Alternaria alternata*

The characterisation for toxins in the *Alternaria* were originally done on the bases of their morphology and the identification of mycotoxins [Simmons,. 1967] this obviously focussed heavily on the mycotoxins which spoiled fruit production and caused crop loss. It was found that even though some species were morphologically similar these different fungi strains produced different toxins that conferred on them different host specificity. These could be found in Apple, strawberry, Tomato, pear, and citrus. These were only after a long period of study identified to have a range of hosts resulting in them being identified as HSTs [Friesen TL et al, 2008]. These in some cases have contradicted the morphological descriptions, however, can still be identified as the same species by sequencing of the internal transcribed spacer region (ITS).

Of the species in the *Alternaria spp* eleven species possess HSTs that meet these criteria [Armitage., 2013], extracts of these HSTs have been shown to lose these effects when treated with proteinase k or heat treated to denature the host, demonstrating that's the toxins are protiens. On identification and study these HSTs were shown to be like polyketides and worked with rough lemon toxin (ACRT). These six polyketides like structured HSTs have been extensively studied, all strains are within the *Alternaria alternata* species, they show structural similarities with each other, but are shown to be unique for each pathotype, with pear, strawberry HST and tangerine showing the most similarity. The apple pathotype however does not possess this similar region, instead having a cyclic nature.

HSTs are produced by germinating spores and allows the spores to recognise a host. These protiens bind to the receptors on host cells and this leads to a repression of host cellular defences making mycelial growth easier and able to penetrate the plant cells. However, it has been observed that in resistant strains these HST instead result in an induction in resistance response, while they facilitate necrotrophic infection in susceptible plants it instead results in a defence response in the insensitive plants. This interaction leads to the host-specificity we

see in plant pathogenic fungi where plants from different species can be unaffected by a pathogenic fungus even in optimal growth conditions [Friesen TL et al,. 2008]

3.1.4 DNA Barcoding and identification of *A.alternata*

DNA barcoding is a methodology of identifying a species by sequencing a region of the genome. Usually, a region with little sequence conservation between species to ensure genetic diversity that can be observed. This allows researchers to use a qualitative method of definitively identifying a fungal isolate without relying on morphological differences. These areas are usually between 500-1000 base pairs in length and are amplified by broad range PCR primers. An example of this in animals is the cytochrome oxidase subunit which has also been used in fungi [B.E.Deagle et al,. 2014] however has fallen less in favour in compared to ITS region barcoding. ITS region is an area of the ribosomal DNA which is a non-coding region not including the 18s and 28s rDNA. The locus itself is present in all members of the fungal kingdom and demonstrates the high level between species variance needed for it to be an effect bar-coding locus. Because of this ITS has become one of the most common diagnostic methods for species identification [Bellemain, E. et al,. 2010].

Despite this the procedure still has some flaws, in the case of some genera of fungi with many species the method can be unable to distinguish between genetically similar species [Kipling W Will et al,. 2004]. As a method for identifying species within the Alternaria genius however, ITS has been very effective. The method is capable of easily and quickly identifying members of the different *Alternaria* species even those which would normally have very similar morphological characteristics. This would include differentiating between species with the same spore group such as *A. infectoria* and *A. alternata*, which both have small hard to distinguish conidial spores [Birgitte Andersen and Ulf Thrane,. 1996]. There however cases of morphological district groups that do not show differences in the ITS locus, allowing them to be potentially classed as variances within a species. There are other regions that have been considered and may show greater levels of between species genetic variation. Examples include LSU region and the SSU region both have been used in other research to distinguish *Alternaria* isolates as a member of a particular species [Rasime Demirel ,. 2016]

3.2 AIMS

This project began with multiple available cultivars of *A. alternata* and other trains of the *Alternaria* genus, this chapter began with learning which of these available *Alternaria* strains were the most well suited for further analysis. The chosen strains would need to be easily cultivated, produce spores easily without lightly processes, and infect the tomato leaves of the money-maker tomato with visible quantifiable lesions. To do this pathology work had to be done to cultivate, produce spores and evaluate our strains most effectively

Specific aims were:

1. Determine which of the available *A. alternata* strains is most suitable for downstream analysis by evaluating theist pathogenicity against tomato.

2. Develop a methodology for cultivation and sporulation of the chosen Alternaria strains so spores could be easily acquired and tested for downstream analysis.

3.3 Results

3.3.1 Determining *A. alternata* spore viability

It was observed early in the process of studying *A.alternata* spores that not all spore extractions were successful. Some either only produced a solution of mycelial debris and others producing spores that upon later being used for germination assays would give unusual results and sub optimal control germinations. To avoid this, we began doing preliminary germination assays on extracted spores [Figure 3.1] . Extracting from a small region of the plate and observing under a microscope. I found in some cases spores could appear misshapen and darker, or perfectly normal but upon incubating for 24 hours to induce germination, would produce no germination tubes like those shown in [Figure 3.1d]. Extracting and testing before using the spores for downstream applications allowed me to verify the viability of spores before use in germination assays, lesion measurements or RNA extraction.

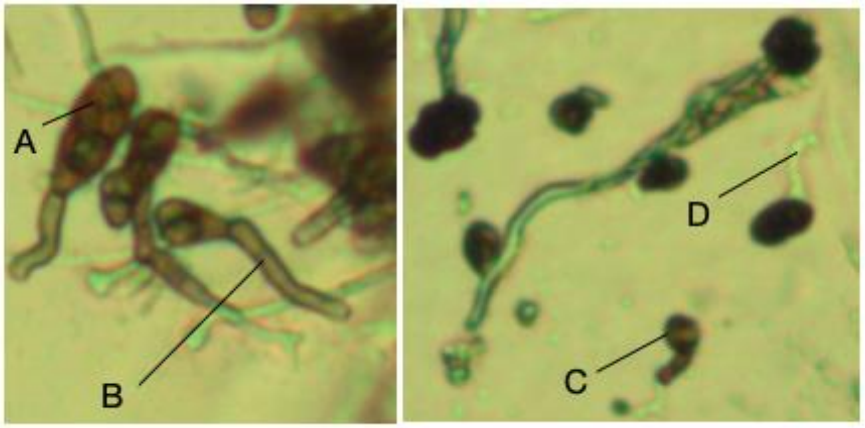


Figure 3.1**:** Determining spore viability using light microscope:

Spore extraction was observed under light microscope to determine spore viability, its likelihood to germinate and be a viable spore solution for downstream applications. All extractions carried out were observed for characteristics that determined viability and potentially indicated non-viability as shown in these two images by the labelled spores. a) example of healthy spore 24h after incubation showing the archetypical 'hand grenade'

like shape, clearly defined segments and green colouration; b) example of a healthy spore tube distinguished by length, at least twice spore conidia length after 24 hours, green colour and visible segments; c) example of a non-viable spore conidia, often darker colouration, spherical shaped and no visible segmentation; and d) example of a non-viable spore germination tube, observed to be short, translucent and with no visible segmentation.

3.3.2 Inoculation of plant materials.

One of the first steps in achieving our aims was establishing a strain of *A. alternata* that would be suitable for all downstream applications. To do this, spores from four available *Alternaria* isolates were extracted and diluted to 3 concentrations, 2.5×10^5, 1.25×10^5 and 0.625×10^5 and these were drop inoculated onto the left side of 3 tomato leaves with the opposite side drop inoculated with the same quantity of sterile distilled water as a control. The observed lesions were measured and compared to determine the best strain for further applications. Each individual strain was tested with the Cabi strain shown in [figure 3.2] AOO1 shown in [figure 3.3] Fera strain tested and shown in [figure 3.4] and the brassica strain tested in [figure 43.5] the overall results plotted in [Figure 3.6]. these results show the relative lesion sizes of the 4 isolates, demonstrating that the CABI LSHB SM-O430 strain had an average lesion diameter of near twice that of the other *A.alternata* strains we tested and almost no lesion development from *A. brassicicola*, which was expected as it is not a known tomato infecting strain.

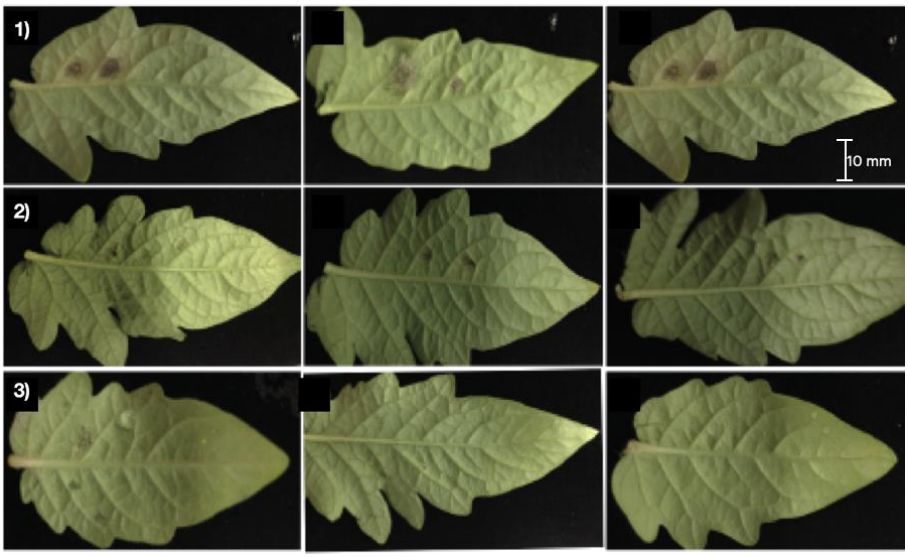


Figure 3.2: Lesion development on tomato leaves infected with *A. alternata* CABI LSHB SM-0430 isolate:

For this experiment 9 tomato leaves were each drop inoculated twice on the left side with a serial dilution of *A. alternata* spores with a control inoculation of sterile distilled water on the opposite side of the leaf. 1) show three replicates of tomato leaves inoculated with a spore concentration of 2.5 $x10^5$, these show the largest of all the shown lesions. 2) show the results of inoculation with a spore concentration of $1.25x10^5$, showing less than half the lesion size of the higher concentration samples. 3) show the results of the lowest concentration $0.625x10^5$ showing little to no visible lesions. I was able to determine from the optimum concentration of spore inoculation for visible lesion formation and by comparing to the other available strains that the *A. alternata* would produce significantly larger lesions compared to the other strains

The first isolate tested was the Cabi isolate, this tested showed a marked difference between the higher and lower concentrations and showed clear difference between the sample and the control on the left side of the leaf.

Figure 3.3*: A. alternata* A001 strain lesion size by spore concentration:

For this experiment 9 Tomato leaves were each drop inoculated twice on the right side with a serial delusion of *A.alternata* A001 spores with a control inoculation of sterile distilled water on the opposite side of the leaf. 1) show three replicates of tomato leaves inoculated with a spore concentration of 2.5 $x10^5$, these show the largest of all the shown lesions. 2) show the results of inoculation with a spore concentration of $1.25x10^5$, showing small barely visible lesions. 3) show the results of the lowest concentration $0.625x10^5$ showing no visible lesions and no difference when compared to the negative control. I was able to determine from this that the Aoo1 strain was likely less infective than the CABI strain, however this would need to be verified by quantifying the lesions size by measuring in ImageJ

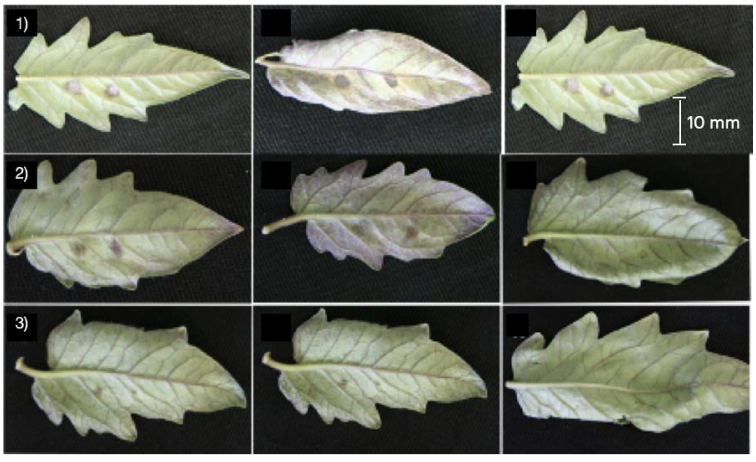


Figure 3.4 *: A. alternata* Fera 15968 strain lesion size by spore concentration:

For this experiment 9 Tomato leaves were each drop inoculated twice on the right side with a serial delusion of *A.alternata* Fera 15968 spores with a control inoculation of two drops sterile distilled water on the opposite side of the leaf. 1) show three replicates of tomato leaves inoculated with a spore concentration of 2.5 $x10^5$, these show the largest of all the shown lesions. 2) show the results of inoculation with a spore concentration of $1.25x10^5$, showing small but clearly visible lesions. 3) show the results of the lowest concentration $0.625x10^5$ showing very small visible lesions. I determined form this that the fera was likely a more viable candidate for our *A.alternata* strain that the A001 based on lesion size at all concentrations, showing some visible lesions at low concentrations while the A001 showed none, and would need to be measured in ImageJ to compare to the CABI strain.

Figure 3.5: *A. brassicola* lesion size by spore concentration:

For this experiment 9 Tomato leaves were each drop inoculated twice on the right side with a serial delusion of *A. brassicicola* spores with a control inoculation of two drops sterile distilled water on the opposite side of the leaf. 1) show three replicates of tomato leaves inoculated with a spore concentration of 2.5 $x10^5$, 2) show the results of inoculation with a spore concentration of 1.25x105 and 3) show the results of the lowest concentration $0.625x10^5$ all these concentrations show no visible lesions and no difference from the negative control. I determined from this that the *A. brassicicola* was not a viable candidate for sequencing or further infection assays as it clearly had no infectivity in the tomato plant.

The *A.brassicola* sample was expected to not give a lesion as the tomato leaves are not its specific plant target. But analyses were performed to give a control to show that lesion size was significantly different between the Alternaria virulent strains and the non-virulent strains.

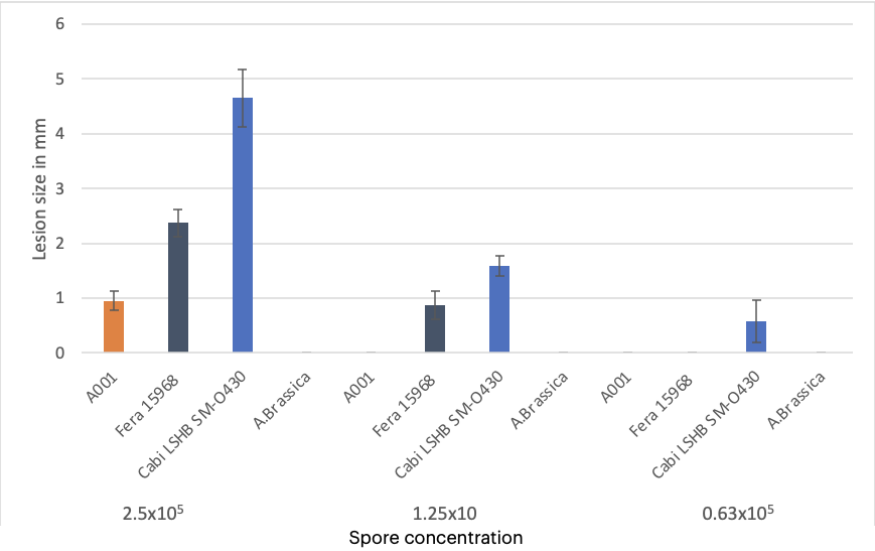


Figure 3.6 :The lesion size for each tested *Alternaria* isolate. This data shows that of the three *A.alternata* strains we tested, with 9 replicate sample size for each concentration the CABI LSHB S M-0430 strain showed consistently significantly higher lesion size compared to the A001 and Fera 15968 at all dilutions. The CABI strain showed over twice the lesion size of the other *A. alternata* strains at $2.5x10^5$ concentration, significantly greater lesion size is $1.25x10^5$ and the only strain to show lesions at the lowest concentration of $0.63x10^5$. The data also demonstrates that the *A. brassicicola* variety of fungus has no pathogenic effect on the tomato producing no lesions at any concentration.

3.3.3 Optimisation of spore production

The first step in optimisation in spore production was to see in our strains of *A. alternaria* how different growth conditions affected spore production and which method would produce the most consistent and reliable spores for downstream applications. This was carried out by evaluating 4 different techniques based on literature in which different methods have shown satisfactory results in inducing *A. alternata* conidia production in the quantity we needed, which is usually more than 2.5 $x10^5$ spores. These methods included standard incubation

under a light and dark photoperiod of 12 hours and incubation at 25˚C, which we found did not produce the numbers of spores we needed.

Mycelial stress technique, which consisted of creating multiple 4mm incisions into the mycelial growth on the pad of media to induce physical stress to induce sporulation was also ineffective for our fungal strain, despite having been shown to be effective by others [Carvalho et al., 2008]. The same can be said for several other methods we used, including cold shock technique by which physical stress was combined with incubating at lower temperatures of 4˚C but was found not to increase the total spore production. In addition, multiple sporulation media were evaluated, including sporulation media containing 20 g l -1 calcium carbonate, which has been shown to increase spore production [Romelia et al, 2000] but showed little positive increase for our strain and low nutrient media consisting of PDA made up to less than 10% of the nutrient concentration of standard PDA [Armitage., 2013], which also showed little improvement these were tested and shown in graph form in [figure 3.7].

At the end of this process the method that showed the best results was the PDA oversaturation method showed in [Figure 3.7 and 3.8]. PDA plates made using the standard concentration were inoculated with multiple plugs of *A.alternata* mycelia incubated with no photoperiod, this was done to create a confluent and nutrient deficient media plate faster and in doing so inducing sporulation. This was found to be highly successful, producing confluent *A. alternata* plates producing high concentration of spore solutions consistently within 2 weeks of inoculation.

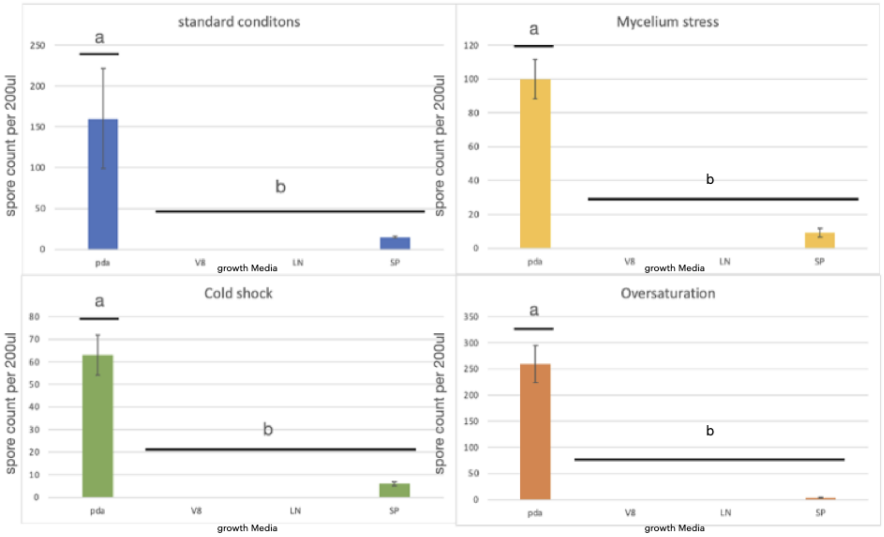


Figure 3.7: PDA demonstrates increased spore production:

Spore count taken from each of the four different conditions for each of the 4 growth mediums. these graphs show that in each growth condition showed PDA as far more likely produce spores and at higher concentrations than the others. 5 replicates were used for each condition for each media type and statistical signifigance was determined using 2-way anova and Post Hoc Turkey test, with group b shown to not be statistically significant (p = .056), with group a statistically significant (p = .00001)

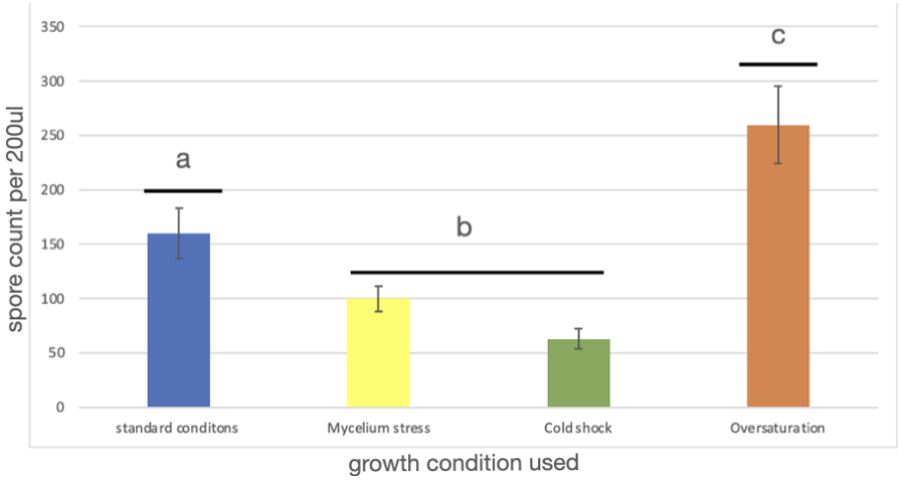


Figure 3.8: Oversaturation is the most effective condition within the PDA:

All 4 growth conditions are plotted on the x axis against spore count on the y axis, this showed that among the conditions tested in the PDA agar the oversaturation method was statistically significantly more productive than the others followed by standard conditions and the mycelium stress method and cold shock both having little difference as shown using a 2-way Anova test and Post Hoc Turkey test, sample group a and c showing stasitically signifigant difference to all other groups (p=0.0001) and b containing groups not significantly different (p=0.052)

3.4 Discussion

3.4.1 *A. alternata pathology* impacted by strain and cultivation method.

The first pathology study was performed to look at the available strains and how they infected the tomato leaves. This study looked at 4 strains from the Alternaria Genus: *A. alternata* CABI LSHB SM-O430, *A. alternata* FERA ex Potato, *A. alternata* A001 ex Potato and *A. brassicicola*. In these results we can see a clear difference in lesion size between the different fungi, the CABI strain showed significantly larger lesion size at all concentrations compared to both the FERA and A001 strains of *A. alternata* and *A. brassicicola* showed zero growth at all concentrations as seen by the measurements of lesion sizes [Figure 3.6] .

This difference can be attributed to host-specificity, the difference between the four strains lies in which host they have adapted to infect, in respect to the FERA and A001 both had been extracted from different plant tissues, potato and parsnip respectively, resulting in lower pathogenicity towards tomato, and the *A. brassicicola* which is species of *A. alternata* adapted to infect the *Brassica* genus of plants. A fungus cannot infect all plant species, as they are limited in the range of plant they can infect by host-specificity. The mechanisms of which have been extensively studied, different fungi strains encode different virulence (*Avr*) proteins, host specific toxins that determine host-specificity [Petit-Houdenot et al,.2012].

In *A. alternata* host specificity can be identified by looking at *conditionally dispensable Chromosomes (CDC),* these regions of the fungi genome have been identified as containing many of the genes determining fungi host-specificity [Armitage., 2013] For the tomato pathotype a single gene has been identified in *Allergic protein (ALT*) synthesis specifically *ALT1* [Akagi et al., 2009] and sequencing of the tomato phenotype has shown that there are contigs

that do not align with the *A. brassicicola* genome consisting of up to 200 genes. This could very well be an explanation for the stark differences in infectivity between the tomato pathotype and the *A. brassicicola*.

As for the difference between the A001, Fera strains and the CABI strain this can be best explained by looking into the loss of pathogenicity through subculturing. It is well known that long periods of subculturing can lead to loss of virulence [M.A. Ansari and T.M. Butt 2011] but the underlying mechanism are less understood. I would hypothesis based on these studies that there is a connection between loss of virulence due to continued subculturing *in vitro* and sub-culturing on and re-extracting from a different host. This is based on findings by [Breen et al,. 2016] showing a link between differentially methylated genes/genomic regions associated with virulence upon analysing changes to the pathogen's genome during culture and after re-incubation on plant tissue. Virulence declined with timed and could be recovered upon inoculation on *A. thaliana* and whole-genome sequencing showed changes in methylation during culture but no significant genetic changes. This role of methylation is seen in *A. alternata* and other ascomycetes [He et al, 2020]. It is possible that if changes in methylation can occur due to cultivation *in vitro* due to lack of host-specific factors, then the same can occur if cultivated on a different host for an extended period. This can lead to reduced virulence to that specific host, as we see in the FERA and AOO1 strains.

Among the three *Alternaria alternata* samples available for use, all of them showed some growth on tomato leaves. All of them showed adequate growth on pda media and showed significant spore production. This means that any of the conditions that could have inhibited spore growth on the host could not be accounted for, such as lack of infectivity due to over culturing in media [M.A. Ansari et al,.2011] lack of spore production or germination. However, the samples were all different in specific ways, the culture for the Fera strain was isolated from potato, and the A001 from parsnip. This is a common method in fungi culturing to re-gain an infective phenotype from a culture that has lost it from prolonged sub-culture. However, it is likely that the process of growing in media and then in a host different from the host that we needed had reduced infectivity. It was for this reason that the Cabi isolate was chosen and tested against the others. Despite this all isolates of *Alternaria alternata* did show growth and formation of the distinctive black spot lesions on the leaves. However, it clearly

seen from the results that the lesions were up to twice as large in the Cabi strain than the others. The *Alternaria brassica* was included as a negative control, as it was not expected for the non-tomato targeting group of the *Alternaria* to form a lesion. As was shown to be the case in [Figure 3.5] showing no lesion formation and an identical phenotype to the water only control.

Understanding this was a simple but important step in furthering the projects goals, for further experiments lesions will need to be created and extracted for total RNA for sequencing, in addition lesion size can be used as a determinant for the effectiveness of inhibiting a spore's ability to infect tomato leaves.

3.4.2 Genetic component of Alternaria pathology.

All the tested strains were evaluated for two main genetic components, the presence of the ITS region 4/5 region for *Alternaria alternata* as well as the presence of the Gene family that determines pathogenicity in tomato. This was chosen as a method sue to how pathogenicity works in *Alternaria* which was studied extensively in (Armitage., 2013) and used as a basis for my pathogenicity work. The pathotypes in *Alternaria alternata* are determined by the 6 main HSTs that determine host/pathogen specificity. These HSTs all contain a polyketide structure. All fungi that produce all 6 of these HSTs have been proven to be members of the *Alternaria alternata* species. However there has been shown structural differences between the 6 HSTs in different pathosystems. These protiens include AKT for the pear pathotype, AFT for strawberry, ACTT for Tangerine, AMT for apple and AATT for Tomato. These can then be subdivided further by additional groups by additional structural derivatives attached to the core protiens structure. In the case of the tomato isolate there has been a single gene identified responsible for production of AALT which is identified as ALT1. This gene was tested for in each strain by a PCR test and each of the three *Alternaria* strains each showed a positive band and the negative control *Alternaria brassicola* strain showing no amplification. The tomato pathotype has been sequenced and studied in [Hu et al 2012] and specifically studied the presence of contigs that are not present in A. brassicicola which makes the lack of presence of ALT1 not unlikely.

This was confirmed by the total lack of pathogenicity in the *A. brassicicola* strain [figure 3.5]. This does not mean however that tomato pathogenesis is a pure matter of only if it expresses ALT1 as was observed in the pathogenicity analysis lesion size varied between the different isolates of the alternaria pathotype. This is likely due to an array of genetic variables. There have been over 200 genes identified that could be present in the tomato *Alternaria alternata* pathotype that may impact the degree of pathogenicity to the target host.

This was the reason why identifying the expression of genes was not enough to confirm effective pathogenicity in the host. The three *Alternaria* strains each could express some of the genes needed to be part of the tomato pathotype but not all of genes for full effectiveness. If this study of the sRNA present in the pathosystem is to have value, then the relevance must as closely be attributable to the *Alternaria* strains found in field.

3.4.3 Spore production can be optimised by correct media and conditions.

There have been many studies on spore production and on maximising and optimising that spore production for use in pathogenic analysis. The goal is sallying the same, producing spores on demand for use in various experiments including germination, spore identification and for inoculating leaves. However, the fungi can be temperamental in spore production, plates that under seeming the same condition had previously produced spores could then grant nothing. Plates can under the wrong conditions can produce only mycelia, creating a paler white covered plate with much fungal material. This is acceptable for some goals such as RNA extraction from mycelia, in which spores are more difficult to extract from due to their melanin content [Dörrie J et al,.2006]. However, certain functions can only be done by spores, and so there needs to be a methodology put in place for your isolate to extract spore reliably and consistently. Spore production has been the subject of study in multipole papers, and these papers provided the conditions used in this experiment. Low nutrient media was demonstrated in [Armitage,. 2013] and was used as a standard practice for cultivating and extracting spores from *Alternaria*, the reasoning behind this media is that the fungi respond to nutrient availability. Providing the fungi with an abundance of nutrients results in more mycelia while reproducing nutrient stress can induce greater spore production. The standard

PDA had produced spores for my isolate in previous experiments, however it was inconsistent. It was my belief that it was less of a factor what media and more what conditions. With plates that had been stored at -4 for long periods of time after becoming confluent producing far more spores that plates that had been extracted fresh from the growth cabinet. Other papers expose on the subject, the sporulation media made with CaCO3 came from the paper [T.M.S. Rodrigues et al,. 2010] in which multiple methodologies and media types were tested for spore production in *Alternaria alternata* in addition to V8 media. V8 media I had expected to be successful due to the presence of tomato extract and this strain having been proven to grow well on tomato.

Observation for the various media and conditions turned out to demonstrate the effectiveness of standard PDA with the correct conditions. PDA media when allowed to go confluent, usually after 1-2 weeks incubation, and then stored at -4 for 4 days was found to be the most effective method of spore production, with other methods for artificially producing stress, such as mycelial stress technique, not having as near a large effect as cold shock.

Chapter 4

Chitin synthase sRNA Treatment

Chapter 4 Impact of chitin synthase on spore germination

4.1 Introduction

4.1.1 Background

Fungal cell wall comprises Chitin, β-1,3-glucan other components. Chitin is a long-chain polymer of N-acetylglucosamine, an amide derivative of glucose. The second most abundant polysaccharide in nature, it is a primary component of cell walls in fungi. All these components are needed for the growth of the cell, as well as ensuring the survival of the cell from possible conditions. Many of these are unique to fungal systems and not present in host cells. Biosynthesis of fungal cell components requires many protein enzymes. Two such proteins have been targets for known anti-fungal chemicals, i.e., echinocandins inhibit β-glucan synthase, and nikkomycins and polyoxins inhibit chitin synthase [Stevens, 2000]. Inhibitors of enzymes can be either competitive or non-competitive inhibitors depending on if they compete with the enzyme for binding with its target or work by a different mechanism. The echinocandins are non-competitive inhibitors and have multiple medically approved variants.

CHS family of proteins are responsible for the production and processing of chitin via chitin synthesis. Chitin varies in structure and function between dimorphic or filamentous fungi, and *CHS* genes can be grouped into 7 different categories [Lenardon et al.,2010].

Some dimorphic fungi like Saccharomyces *cerevisiae* encodes 3-4 chitin synthases and filamentous ascomycetes can have up to 10 chitin synthase genes. This does not suggest redundancy, but there is evidence that chitin synthase can be produced by CHS domains that are not normally involved in chitin synthase synthesis. Of the CHS domains, CHS 1 and 3 are known to be responsible for the septum formation while CHS 2 and CHS 8 are responsible for chitin synthase activity and can be upregulated by the presence of certain chitin synthase inhibitors [Grover et al., 2010] . Many CHS families are only known to have a variety of functions in filamentous fungi, predominantly CHS 3, 5, 6 and 7 with some exception in dimorphous fungi with a higher than normal quantity of chitin in the cell wall. Classes 5 and

7 chitin synthases are defined as containing the myosin head motor domain [Treitschke et al., 2010].

The classes of the chitin synthases can be further defined as follows and found in papers such as [Werner S et al,. 2007] and [Bowan AR et al,. 1992] Classes 4-7 CHSs contain a chitin synthase-2 (CS2) domain, while CHSs from Classes 1-3 have absent regions of the first portion of the CS2 domain. These are placed in these classes with a type 2 chitin synthase domain and in addition a chitin synthase N-terminal domain. These still possess the first conserved region, which contains the dominant region of chitin synthase 1. Class 1 chitin synthases have conserved regions in the terminal amino acid region, Classes 4,5 and 7 contain a domain, which is like cytochrome b5, Classes 5 and 7 contain the myosin motor domain, also known as the myosin head. Class 6 has a domain structure which is far simpler, it contains only a CS2 domain. These *CHS* genes can be further divided into multiple divisions: Division 1 including CHS 1-3, Division 2 containing CHS 4,5,7 and Division 3 containing CHS 6.

The chitin synthases and other cell wall structural components are known as a viable target for antifungal drugs. This is because they are essential for fungal growth and for their adaptation and survival to environmental factors. In addition to the fact that these components do not exist in the host cell. It is also the area that first interacts with the host during infection, giving it an essential role in plant-pathogen interaction [Latgé et al., 2017].

The process of chitin synthesis is complex and performed by a large range of enzymes including the chitin synthases [Merzendorfer, 2010]. Of the chitin synthases some of the most important are CHS 5 and CHS 8. Both have domains with chimeric structure containing the N terminal myosin motor like domain, this is fused to the chitin synthase domain. CHS5 nd CHS8 have been linked to the process of hyphae growth and the infection capability of filamentous fungi [Gandía et al., 2010]. There are many other studies on the individual roles of different CHS genes as well as studying the deleted mutant versions absent of these genes. Many of these genes have had roles found for each class of CHS, these roles can include growth, adaption to stress, cell wall damage and infection. For instance, *Penicillium dignatum* is one of the fungi studied for these genes and their function. Studies of the CHS genes in this fungus found that the CHS 3 is the most expressed and CHS 2 has the lowest expression [Gandía et al., 2010]. Many of the chitin synthase genes were induced during specific circumstances such as fruit infection which, in this case resulted in upregulation of CHS 1, 3,

5 and 7 with 2 being upregulated during conidia production. This study found that *CHS 7* silencing had the largest impact on the fungi reducing infection, sporulation, and growth. Silencing of the CHS 7 gene also resulted in induction in CHS 2 expression and some limited CHS 5 expression.

Chitin synthases are located primarily in the plasma membrane of the fungal cells and transfer of N-acetylglucosamine to the chitin chains as they grow and become longer. Chitin itself is an essential component of the fungal cell wall, a polymer of N-acetylglucosamine and one of the most abundant sources of mass in nature like cellulose [Merzendorfer, 2011], which is a well-known PAMP, a trigger of plant innate immune system. The different roles of the different CHS domains are an indication of family diversification over time. scCHS2 is predominately involved in the formation of septa and in the primary processes of cell division. Despite these roles in yeast, the most CHS is produced by CHS3 which is responsible for the ring of chitin produced at the site of cell budding.

Roles of CHSs can be even more complex in filamentous fungi, one example of this being studied is in *Magnaporthe oryzae* which showed that the CHS genes also played a role in infection. For example, mochs1 mutants in *M. orzae* showed a decreased ability to produce conidia and to form appressoria, a process necessary for fungal infection and penetration of plant tissue. The roles of the CHS have been extensively studied, the class 3 CHS genes were predominantly found in roles in animal and plant pathogenesis in multiple different fungal groups. Members of Class 3 and 4 were upregulated during infection [Kong et al., 2012]. It was also found that these CHS pathways were conserved in plant pathogenetic fungi as well as being involved in multiple different processes and that plant specificity determination sites also play a role in the diversification of CHS evolution. The evolution of the chitin synthase genes has been more extensively researched with more accuracy with the advent of better fungal genome sequencing. This was observed in comparison of two genomes of the *dothideomycetes* family [Li et al., 2016]. This paper indicates that powerful selective pressures, such as those present in pathogenetic fungi vs non-pathogenetic fungi of the same genus act to retain genes to form functional groups.

4.2 **Aims and objectives**

The aim of this chapter is to look in detail at a candidate gene target for small RNA silencing, the chitin synthase genes which play an essential role in fungal cell wall creation and cell replication. This is set to determine if the chitin synthase genes are viable targets for sRNA silencing for the purpose of controlling *A.alternata* infection through the following objectives.

1. Determine the chitin synthase genes for *A. alternata* and design novel sRNA targeting those genes
2. Synthesis the novel sRNA and test their effect on spore germination through observation under light microscope after treatment with sRNA.
3. Verify the effectiveness of sRNA silencing through qPCR assay of the chitin genes

4.3 Results

4.3.1 Selection of target genes for sRNA mediated gene silencing

The main cell wall components of the Ascomycetes fungi are chitin and β-glucans [Beauvais, 2018]. We have shown in previous research that targeting cell wall components can be an effective means to reduce infectivity of the oomycete *Hyaloperonospora arabidopsidis* [Bilir et al,, 2019]. Chitin synthase is a glycotransferase that plays an essential role in production of chitin, specifically catalysing the polymerase reaction producing N-acetyl-d-glucosamine, a monomeric unit of the chitin polymer [Merzendorfer, 2013]. Chitin synthase is highly conserved and essential for fungal survival and growth, making it a viable target for gene silencing to reduce fungal infectivity.

Determining this, my first step was to identify *A. alternata CHS* genes using the ensembled genome database. Compared to whole genome sequence of the CABI LSHB SM-O430 strain using BLAST, 11 candidate genes were identified and, these genes encode 11 proteins containing at least one CHS domain. There were no genetic differences between gene sequences on the Fungi database and the whole genome sequence of our strain. I checked all the designed RNAs had exactly matched the target genome sequences. I also tried to find which genes had sequence homology between the CHS domains to design sRNA that could target multiple CHS genes at once, potentially overcoming any issues from gene redundancy among *C*HSs that may prevent total CHS gene silencing.

InterProScan was used to check for the presence of CHS domains, determining the Interpro scan domain accession number for CHS to be IPR004834. Using this, I chose the database with the most comprehensive characterisation of the gene domains in all fungal genes. This was the domains characterised in Pfam and used these to determine the location of the protein coding domains and create a diagram depicting the domains. In addition, multiple amino acid alignments for the 11 CHSs were performed to compare the CHS domains this shown in [figure 4.1]. This was carried out to find homology between the different *CHS* genes to determine groups of the genes that could potentially be targeted by a single sRNA. This could potentially remove any compensation for the silenced genes function by the similar non-sRNA-targeted genes.

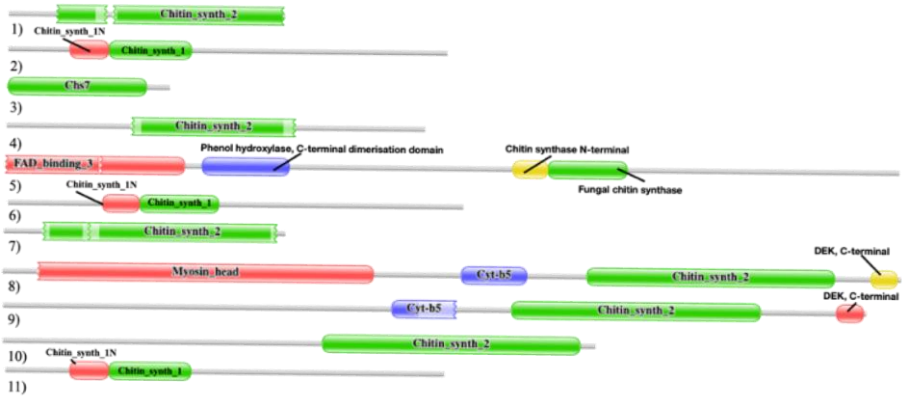


Figure 4.1 Diagram of the 11 CHSs and locations of relevant protein domains: After the 11 *CHS* genes were identified, Interpro scan was used to determine protein domains. 1) Domain structure for protein OAG 20922 derived from the *CC77DRAFT_1061482* gene; CHS2 domain (amino acid residues 121-563). 2) Domain structure for protein OAG1953 derived from the CC77DRAFT_1021806 gene; CHS N-terminal (amino acid residues 132-211) and Fungal CHS (amino acid residues 212-380). 3) Domain structure for protein OAG23516 derived from the CC77DRAFT_1092832 gene; CHS export chaperone domain (amino acid residues 2-284). 4) Domain structure for protein OAG19173 derived from the CC77DRAFT_1031975 gene, CHS domain (amino acid residues 44-804). 5) Domain structure for protein OAG24007 derived from the CC77DRAFT_929028 gene; FAD-binding domain (amino acid residues 1-367) Phenol hydroxylase, C-terminal dimerisation domain at position (amino acid residues 404-582) CHS N-terminal (amino acid residues 1042-1113) Fungal *Chitin synthase* (amino acid residues 1114-1276). 6) Domain structure for protein OAG25348 derived from the CC77DRAFT_1027903 gene; CHS N-terminal (amino acid residues 195-269), Fungal *Chitin synthase* domain (amino acid residues 270-432). 7) Domain structure for protein OAG13919 derived from the CC77DRAFT_1002156 gene; CHS domain (amino acid residues 95-549). 8) Domain structure for protein OAG19178 derived from the CC77DRAFT_1062346 gene, Myosin head motor domain (amino acid residues 77-760), Cytochrome b5-like home/steroid binding domain (amino acid residues 944-1076) CHS domain (amino acid residues 1202-1709) DEK, C-terminal (amino acid residues 1785-1838). 9) Domain structure for protein OAG19177 derived from the CC77DRAFT_1021349 gene; Cytochrome b5-like home/steroid

Commented [1]: Gene names italic

binding domain (amino acid residues 801-854) CHS domain (amino acid residues 1046-1554), DEK, C-terminal (amino acid residues 1713-1767). 10) Domain structure for protein OAG21670 derived from the CC77DRAFT_987140 gene; CHS domain (amino acid residues 659-1185) 11) Domain structure for protein OAG18953 derived from the CC77DRAFT_1021806 gene; CHS N-terminal (amino acid residues 132-211) Fungal *Chitin synthase* (amino acid residues 212-380).

Figure 4.2 Alignment of *A. alternata* CHS domains

DNA sequence of each chitin synthase domain were aligned and compared for reagies of genetic similarity to use for designing sRNAs to target multiple genes. CHS genes were found to align in groups with some genes having similarit and some having none.

In order to determine which of the genes may share similar roles or potential redundancy, the CHS domains were multiply aligned to look for protein similarity (Fig 4.2). Of these the proteins OAG19178, OAG19177 and OAG21670 aligned together as well as OAG24007, OAG23547 and OAG1953 as well asOAG13919, OAG25348 and OAG20922. With the following proteins OAG19173 and OAG23516 not aligning with any of the other *CHS* genes. It can be determined from this that there is potential for sRNA that could silence each of these groups of *CHS* genes together, reducing our sRNA needed to silence all *CHS* genes to 5.

Figure 4.3 Alignment of *A. alternata* open reading DNA frame: alignment of open reading frames for each CHS gene showed similar issies with many CHS genes showing differences and few having overlap making designing a small RNA that targets multiple CHS genes very difficult.

4.3.2 sRNA design

sRNA was designed in Genious using the method shown in 2.5.3 Generation of siRNAS and the sequences are included in [material and methods table 2.2]

4.3.3 Chitin synthase gene comparison

Results about your sRNA design and sRNA production need to be included here.

To determine which of the *CHS* gene products showed the most similarity and therefore could potentially be used as the targets for sRNA that could silence multiple genes at once. The open reading frames of each *CHS* gene were multiple aligned in Genious to compare sequence similarity [Figure 4.3]. Of the *CHS* genes, the one that showed most similarity were OAG23547 and OAG18953, OAG25348 and OAG20922, OAG19178 and OAG19177. However, there was no region commonly shared between these genes meaning designing sRNA that can silence more than two genes at once seems unlikely.

4.3.4 Effect of sRNAs on spore germination

We wanted to see the effect of the sRNA targeting the CHS genes on spore germination. The expectation was that by treating with double stranded sRNA (ds sRNA) for an essential gene that encodes an essential enzyme responsible for the fungal cell wall formation, this would have an inhibitory effect on spore germination. Each of the 5 CHS genes showed a different

degree of inhibition or non-inhibition [Figure 4.4]. With some samples showing higher germination in the treated samples than the control and in many cases the standard error exceeding half of the total inhibition observed, this can be seen in OAG19178 where the total inhibition was near 20 percent but accounting for standard error could be as little as 7 percent. Percentage germination was measured by counting the number of germinated spores in the treated and control samples. This was then used to calculate the relative effect on germination compared to the control using the following calculation; If X is the percentage of germination at a given concentration of sRNA, % inhibition = 100 x [1 - (X - MIN)/(MAX - MIN)]. In this we can see that of the 5 Genes only one showed consistent inhibition which increased with increased concentration of sRNA OAG21670. The others when considering the often-large standard error, would vary between concentrations, sometimes showing higher inhibition at lower concentrations such as in OAG19178. However, in some cases, a negative value for percentage inhibition indicating that the control had less germinated spores than the treated sample as is the case of CHS 5 at 5um [figure 4.4]

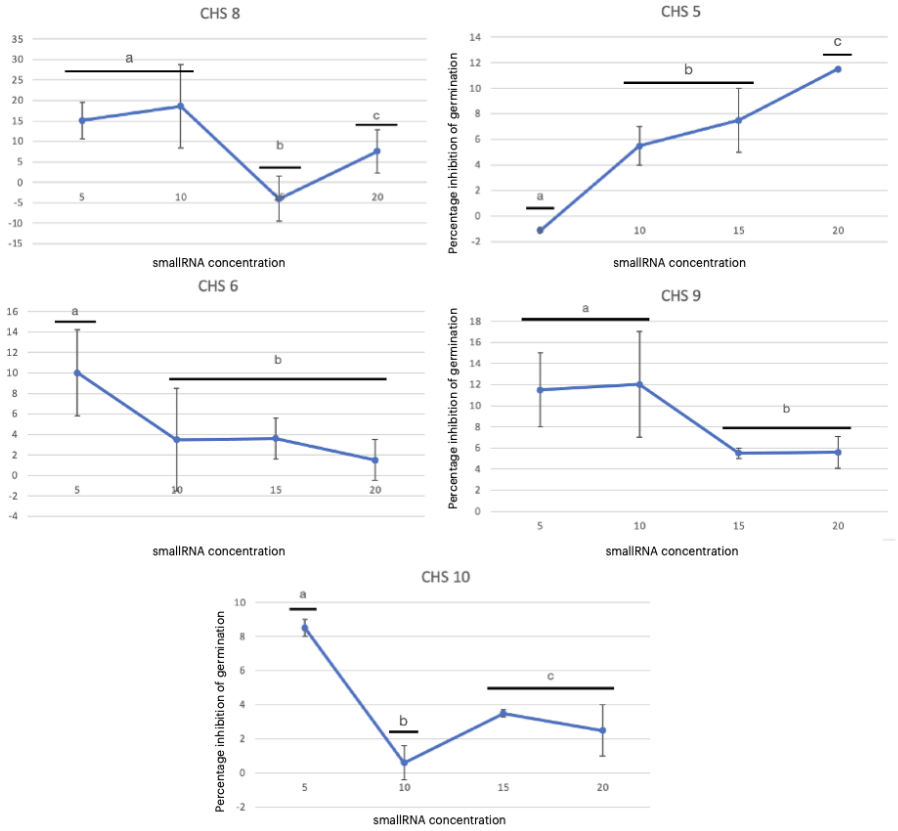


Figure 4.4 sRNA targeting *CHS* genes results in sporadic germination inhibition:

This graph depicts the percentage inhibition of the 5 tested *CHS* sRNAs at 4 different concentrations 5, 10, 15 and 20 µM. each CHS was evaluated with a two-way Anova and the averages for each concentration was grouped into letter categories to denote groups that were statistically similar.

4.3.4 Effect of sRNA targeting *CHS* genes on germ tube length.

It was observed after the first preliminary tests of five *CHS* genes, sRNAs induced gene silencing led to alterations in the germination tube physiology. The germinating tubes often had more branches and were observed to grow faster and longer than the control germinations [Figure 4.5]. This was quantified by measuring the tube lengths from the base of the spore in conidia from each of the sRNA treated samples [Figure 4.6] These measurements were

averaged and used to determine the standard error, and evaluated with a 2-way Anova shown in in figure 4.6. Comparing these average measurements taking into account standard errors, the length of sRNA treated spores was shown to be significantly longer than the non-treated control.

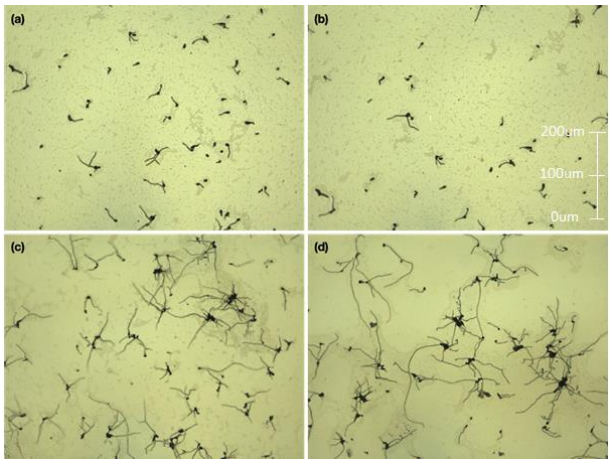


Figure 4.5 *CHS 8* sRNA treatment results in longer germination tubes:

This image depicts 4 pictures taken from different germination assays, in which cellophane strips either treated with a sRNA free control or 5 µM of sRNA were observed 24 hours for any effects on germination length. a) and b) are from control samples that both demonstrate the normal observed length and number of spore germination tubes. c) and d) depict samples that have been treated with sRNA at 5 µM for he *CHS 10* gene which are observed to have significantly longer and more branching germ tubes than the control.

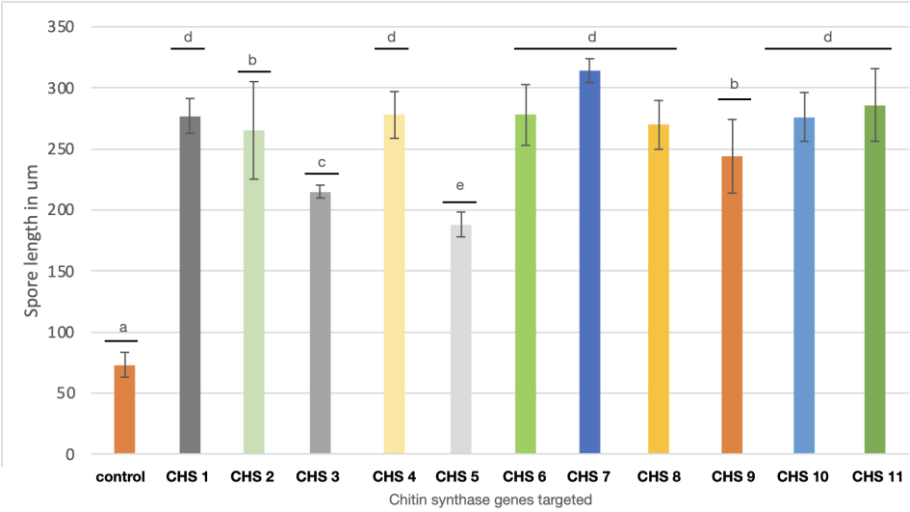


Figure 4.6 *Chitin synthase* targeting sRNAs has an enhancing effect on spore germination tube length:

This graph compares the average spore germination length of *A. alternata* spores treated with one of the sRNA targeting the Chitin synthase genes compared to the control. The sRNA treated samples all show statistically significant increases in germination tube lengths, a two-way Anova was used to determine samples which were statistically similar with each letter on the graph denoting a group of samples which were statistically similar. In the case of OAG23919 which shows the largest increase, of 240 µm compared to the control, with all samples showing a significant difference from the control in group a ($p=0.0001$) with all groups, with group b showing signifigant difference to groups c ($p=0.051$), d ($p=0.015$).and e ($p=0.001$) using replicates of 9 per CHS with each replicate taken as an average of three images from 1 slide. .

4.3.5 sRNAs targeting *CHS* genes influence spore chain production.

After the first experiments with 5 sRNAs targeting the *CHS* genes, we discovered a unique phenotype. Seven days after treatment with the sRNA, the spores were observed to be more numerous and darker, forming long chains additional spores after germination [Figure 4.6]. These spore chains were measured and quantified by counting the number of secondary sporulating structures showing that spore creation after germination increased significantly compared to the control [Figure 4.8] . As the concentration of sRNA increased, the number of secondary structures were also increased. With each increase in sRNA concentration resulting in higher number of spore chain being created, these new spores were counted and

quantified, the standard error calculated to represent the relative increase in sporulation when compared to the control.

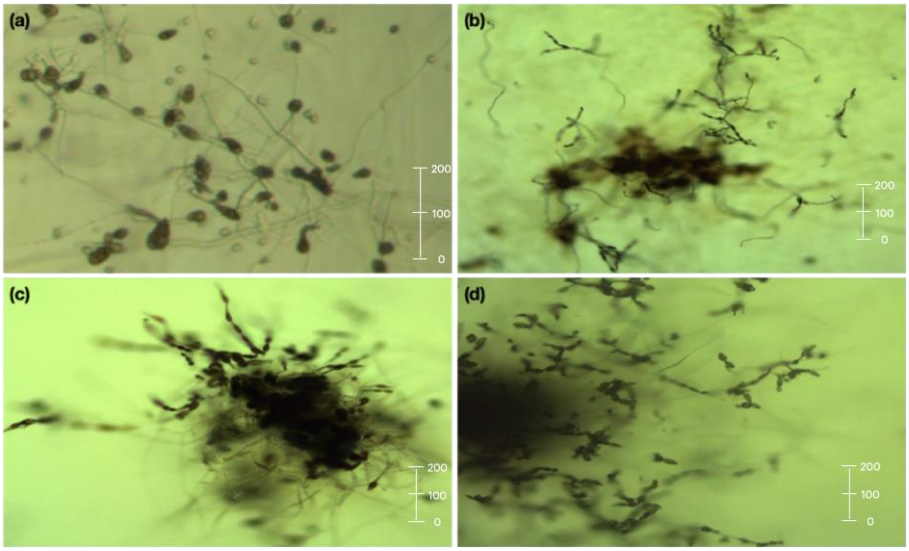


Figure 4.7 *A. alternata* spores produced more secondary sporulating structures after treated with sRNAs targeting *CHS 8*:

In this figure, we can see four germination assays in which spore solutions of *A. alternata* were treated with different concentrations of sRNAs, with a measure of 200 µm for reference. it was observed after a longer period (at least 7 days post treatment) and increased number of spores producing structures were observed. a) shows spores treated with sRNA at 5 µM with little visible change in terms of spore chain creation, b) shows an increase with spore chain visible by the long chains of darker spores attached to the germination tubes. c) shows spores treated with 15 µM of sRNA with a higher number of spore chain and d) showing spores treated with 20 µM with a similarly high number of spore chain. This observation potentially demonstrates an effect of sRNA targeting *Chitin synthase* on the production of spore chains post spore germination.

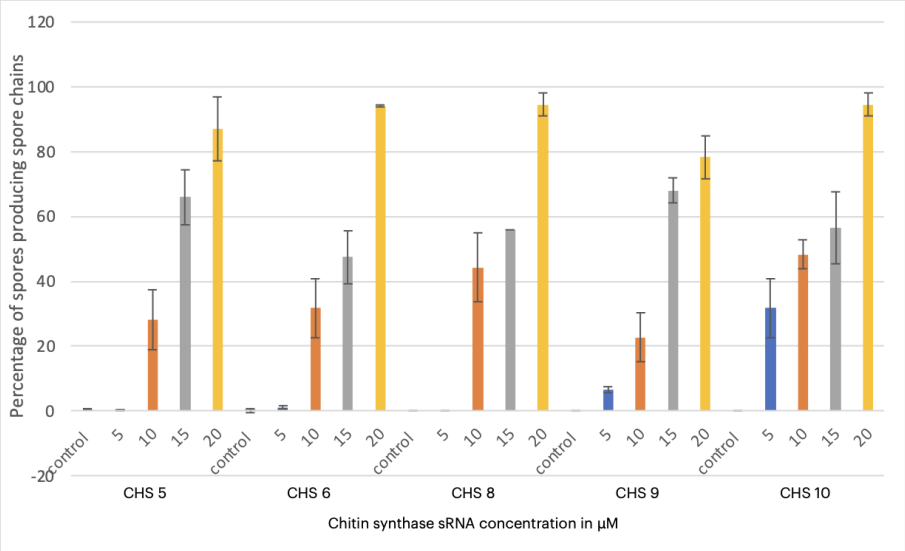


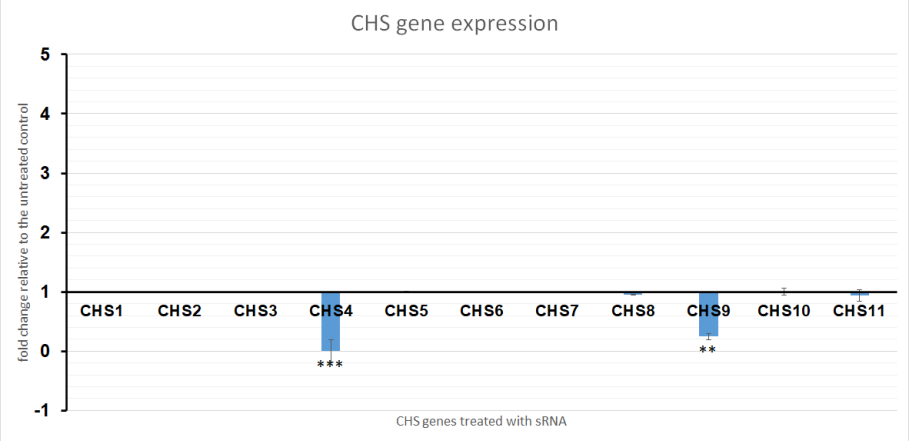
Figure 4.8 Effect of sRNA concentration on the production of secondary structure:

Percentage of germinating bodies producing spore chains increases by up to 90 percent. Spores were drop inoculated onto cellophane strips and were observed over 2 weeks to induce sporulation in the germinated spores. The result was that of the control samples observed even after 2 weeks there was little sign of spore chain production. However, there was significant change in the sRNA treated spores. This change can also be seen to increase with sRNA concentration, the 5uM showing less than 10 percent increase in the first 4 sRNAs with only Oag21670 showing a significant increase in spore chain at that concentration. All sRNA samples show an increase up to at least 20 percent at 10µM, just over 40 percent at 15µM and just under 80 percent at 20 µM.

4.3.6 RT-qPCR data confirms some sRNA had a inhibitory effect on Chitin synthase gene

The observable phenotype of the spores can be explained by a few processes, sRNA could be affecting *CHS* gene expression or the presence of sRNA alone may be enough to cause the observed phenotype. To determine if gene expression is being affected a RT-qPCR was performed to observe the effect of gene expression of the *CHS* targeting ds-sRNA. 11 samples of *A. alternata* spores were taken and treated with one of the 11 sRNA, in addition each sample was split between the control and the treated sample. This process was done using the procedure detailed in the materials and methods [Section 2.4.6] RNA was extracted from the spores using the method detailed in [Section 2.6.1] however the triazol method needed to be modified slightly, using a homogenizer to help break down the hard to lyse spores. These RNA samples

were corrected to the same RNA concentration and tested for the gene expression with primers targeting the *CHS* gene the respective sRNA should target. The effects of this was after calculating for the delta ct values for the controls and treated samples and calculating the fold change. The results show that some of the genes show little effect in expression, however some show decreased expression in the targeted genes. Some by many times fold changes as shown in [Figure 4.9].



Commented [CN(2)]: I redid my analysis and found the original version wasn't done correctly, the new version is done using correct ddct analysis which should show down regulated genes with values between 0 and 1

Figure 4.9 Expression analysis of CHS genes after sRNA treatment:

qPCR results for the 11 chitin synthase genes were analysed using the Comparative CT Method, these values were compared to the untreated control for each sample. this method calculated the fold change difference between the target and the background gene in this case actin. The results show that 8 of the genes showed inhibition, with three of them showing significant inhibition, with values between 0 and 1, compared to the control. Statistical significance was calculated using a student's ***t***-test comparing each sample to its respective control, CHS corresponding to $P \leq 0.00004$ and CHS9 $P \leq 0.01$ with CHS 4 and 9 each showing significant difference to their control.

4.4 Discussion

4.4.1 *Chitin synthase* as a target for sRNA silencing

This study looked at the potential use of sRNA to inhibit *A. alternata* growth and prevent infection in crops. The goal was to assay the effectiveness of sRNAs targeting a potential target gene, *Chitin synthase* and whether sRNA targeting *Chitin synthase* could lead to reduced germination, lesion size, fungal growth, or sporulation. This work was inspired by previous research done on the effect of sRNA targeting *cellulose synthase* in downy mildew [Bilir et al., 2019] which showed a clear inhibitory effect, reducing spore germination in treated spores. We posited that cell wall components could make a perfect target for RNA silencing due to their essential role in both fungal growth, germination, and conservation with the genome.

After treating the spores and observing the effects on germination, spore chain production and spore tube length, we were able to make several conclusions; Under normal untreated conditions, the spore tubules and further mycelia do not produce spore chains. However, it was observed in samples that had been treated with sRNA after 1 week of incubation that there were spore chains forming, distinguished by the darker coloured branches of spores stemming from the meshes of fungi mycelia.

We can hypothesise why this increase in spore chain formation occurs; it may be due to stress induced by the presence of sRNA. It has been shown that increase on stress on the fungi results in higher sporulation [Carvalho et al., 2008] methods such as mycelial abrasion, UV light and lower temperatures all induce higher sporulation through stress. It is possible that this increase in spores is due to redundancy, a process by which copies of a gene increase expression in response to mutation or gene knockdown to compensate for the lower gene expression. It is known that *CHS* plays a role in spore production [Muszkieta et al., 2014] as the chitin is an essential part of the cell wall that forms the spore. So increased *Chitin synthase* production in response to the sRNA may be resulting in more numerous and earlier formation of spores.

The second major observation was the increase in spore tube length observed in response to treatment with sRNA [Figure 3.5]. One explanation for this is that the different genes are capable of compensating for each other, this has also been hypothesised in [Cheng et al., 2015]

who observed that certain RNA constructs used in host-induced gene silencing would result in higher *Chitin synthase* levels and in-turn greater formation of mycelial and progressed infection. This is a process known as redundancy, in which other genes are upregulated in response to knock down of another. This process has been suggested in other *ascomycetes*, [Kong et al., 2012] looked at the role of *CHS* in *Magnaporthe oryzae* including multiple *CHS* mutants, showing that in the *CHS3* mutant expression of *CHS1* was higher, and that double mutants of these genes showed higher sensitivity to oxidative stress, likely indicating a compensatory role in cell responses to stress in *M. oryzae*. This hypothesis would explain all observed phenotypes, the lack of consistent reduction in germination, the increase in spore tube length and increase in spore chain formation.

Chitin synthase is responsible for the synthesis of chitin in fungi. These genes are grouped into seven classes [Fernandes et al., 2014], the chitin machinery is different between dimorphic and filamentous fungi within the ascomycetes; fungi such as *C. albicans* tend to have three to four *CHS* genes the filamentous fungi such as the *Alternaria* genus can contain 10-11 [Choquer et al., 2008]. However, this does not necessarily confirm redundancy, but redundancy has been observed in septa formation in sconce to stress [Louise et al., 2003]. This is because there are 7 types of *Chitin synthase*, with *CHS* 1 and 3 responsible for septum formation in normal conditions, with *CHSs 2* and *8* for the majority of *Chitin synthase* expression but are upregulated in response to echinocandins, which are *β-glucan synthase* inhibitors. This lends more support to theory that the *CHS* possess some degree of redundancy.

CHS genes belonging to classes 3, 5, 6 and 7 are present only in the genomes of filamentous fungi and dimorphic yeast with a higher content of chitin needed for their cell walls. These genes have been shown to perform diverse roles [Kong, 2014]. For fungi, successful reproduction and accurate separation of cellular components is performed by cytokinesis and septation. And two of the main components essential for this process and cell wall integrity are β-glucan and Chitin.

We also know that the effects of *CHS* targeting antifungals have an observable effect on germination. In the presence of nikkomycin Z spores of *A. infectoria* failed to germinate and showed a swollen phenotype [Fernandes et al., 2014]. This is also present in multiple *Chitin synthase* mutants. So, an important question that needs answering, if there is redundancy in *CHS* expression accounting for the increase in tube growth and spore formation, is this only

present when genes are targeted with sRNA or also present when silenced or targeted with antifungals? In contrast we've observed no evidence of consistent inhibition of germination when treated with sRNA. Potentially due to the lack of total prevention of chitin formation that is present when treating the fungi with *Chitin synthase* suppressing antifungals. Another observation made is that fungi shown to have lower cell wall levels of chitin and β-glucan are more susceptible to antifungals that target the cell wall [Fernandes et al., 2014].

It is possible that the phenotype we observe is not due to the silencing effect of the sRNA but of the presence of the RNA itself. To show whether this is the case there will need to be new experiments, testing if the sRNA is causing the effect alone will require a test with scrambled sRNA, sRNA designed with the same length but with no corresponding sequence similarity to a gene in the *A. alternata*. This would show that is the sequence specific nature of the sRNA that is causing the change in phenotype and not some aspect of the RNA alone. This has been done in other sRNA experiments in [Foster et al., 2020] scrambled sRNA was used as a negative control to evaluate the effect of sRNA targeting *Chitin synthase* in *Macrophomina phaseolina* showing that the scrambled sRNA showed no difference to the water only negative control. However, this still does not confirm the same is occurring in our experiment without testing for it.

Additionally, qPCR data of the gene expression of the C*hitin synthase* genes and how those expression levels change because of sRNA treatment is essential in confirming any compensation hypothesis. Until then we cannot claim that either the sRNA resulted in successful gene knockdown and that the phenotype we see is a result of upregulation by other *CHS* genes.

4.4.2 Genomic Data

We took a bioinformatic approach to analysing the whole genome sequence of our *A. alternata* strain. This was needed for several reasons, designing the sRNA, and ensuring that the genes do exist in the genome for our sample and that there were no genetic differences. The results of this are shown in [Figure 4.2 and 4.3] which demonstrates an alignment of the cDNA sequence and potential regions of overlap, which could be used for future sRNA generation. This only demonstrated three pairs of *Chitin synthase* genes with enough genetic similarity to design sRNA that can target both genes. We've found little sequence overlap between *CHS* genes despite many of them sharing highly similar protein domain sequence.

This is likely due to amino acid degeneracy; 20 amino acids can be coded by 64 possible codons. Meaning certain amino acids can be encoded by multiple codons, such as arginine and leucine, this also includes GC content and codon preferences that lead to a much lower level of sequence similarity than protein similarity. You could argue that this could be a sign of an evolutionary pressure against sequence similarity. If sRNAs do form a natural defence used by plants against fungi infection, [Wang et al,. 2016] then it stands to an argument that there would be an evolutionary pressure against multiple essential genes being able to be silenced by a sings sRNA as this would then create a very significant selective pressure against that organism. It also may be impacted by sRNAs use in the fungi genome, specifically in genomic defence, heterochromatin formation and gene regulation. If the *CHS* are regulated by RNAi then there would be an evolutionary pressure towards low genomic overlap between *CHS* genes, as this would allow for a greater degree of control of expression and less likelihood of sRNAs having of target effects. In a paper on *Chitin synthase* regulation [Rogg et al,. 2012] suggested that due to the disconnect between gene expression and Chitin levels in the cells, they are expressed there is evidence to suggest post transcriptional regulation of the *CHS* genes. Further research may shed more light on the situation, confirmation of sRNA playing a role in the *A.alternata*/Tomato pathosystem and sequencing of those RNAs would confirm the potential of *CHS* being a target for Host induced RNA silencing. This would potentially explain the lack of a viable target region for all the *CHS* genes as evolutionary adaption.

4.4.3 Effect of chitin synthase inhibitors in *A. infectoria*

Within the *Alternaria* spp there is very little record of the role of chitin synthase in *A. alternata*, however there have been studied cases in similar *Alternaria* species such as *A. infectoria* which can be used to better understand the role of chitin synthase in the *Alternaria* [Fernandes *et al.*, 2014]. The use of caspofungin and its effect on the Alternaria cells did not lead to a total change in chitin content, at least for the strain IMF006. There was however, an observation of abnormal growth in cells treated with nikkomycin Z. these showed what was described as abnormal balloon like growths. When treating with both caspofungin and nikkomycin there was an observed decrease in the spore germination. During this experiment they found 8 homologues chitin synthase gens in *A.infectoria* it would make interesting future research to look into similarities between *A.infectoria* and *A.alternata* chitin synthase genes and the effect of silencing A.infectoria with sRNA.

The experiments performed in [Fernandes *et al.*, 2014] also quantified the CHS gene expression the tested strains and made the conclusion's that it was the class 5 and class 7 CHS genes that had the most response to the anti-fungals.

4.4.4 Paradoxical effect.

Despite the effectiveness and safety of the Chitin Synthase targeting anti-fungal there are reasons why chitin synthase inhibition has its issues. There has been observation of attenuation at high concentrations. This paradoxical effect which I have shown is an established system in *Candida albicans* means that the exact opposite of what is expected can happen. The fungi grow faster in response to the suppression of chitin synthase. This can even be seen in non-paradoxical effect showing fungi where chitin levels increase in response to Echinocandins which is as result of a salvage process [Walker et al., 2009], these results from the effect on the Fks gene activity being reduced. Due to this there are genes that can result in greater resistance or susceptibility to these anti-fungals [Balashov et al., 2006]. It is not hard to theorise that these processes may play a role in the sRNA silencing of chitin synthase, and that the increased fungal growth observed may be the result of either a paradoxical effect, like *C. albicans*, compensatory expression from other chitin synthase genes, or chitin salvage due to other gene processes.

Chapter 5

sRNA Sequencing and miRNA prediction

5.1 Introduction

Trans-kingdom sRNA silencing is a process that has been observed in fungi/plant pathosystems. This is a biochemical pathway in which s RNAs (including many of the sRNAs such as miRNA and siRNA reviewed in Chapter 1 Introduction) can affect the gene expression. This is done either by blocking the translation of the target mRNA or leading to its degradation by RISC. These sRNAs are usually 21-24 nucleotides (nt) in length and mediate this silencing through sequence complementarity to the target mRNA. These sRNA and are generated through Dicer like protein pathways generating smaller RNA from larger RNA. These mRNAs are usually down regulated so in the case of inter-kingdom sRNA transfer are likely to be involved in the plant immune response [Valinezhad Orang et al., 2014]. This can include the response against viruses, bacteria, insects, parasites, oomycetes, and fungi. Upregulation of an sRNA during infection does not necessarily always mean cross-kingdom silencing, with many genes being downregulated to promote better defence against pathogens, as well as many other processes, response to the environment and developmental processes.

In fungi, specifically the filamentous fungi RNAi is a strongly conserved pathway [lax C et al,. 2020]. These are dependent on AGO and DCL, with many fungi having their own homologs to these genes, and these processes also have roles in protection against foreign RNA or DNA either produced by viruses or transposons with potentially damaging effects.

These pathways are different in fungi compared to many other eucaryotic systems, with some cases of RNAi function being lost in certain processes [Dang et al., 2011] inferring an evolutionary pressure towards losing RNAi in certain circumstances.

Cross-kingdom RNAi is the process by which fungi and the target plant exchange sRNAs for the purpose of silencing genes in theory respective target. This has been studied in the

Botrytis cinerea and *Arabidopsis* pathosystem. These pathosystems demonstrated that the identified sRNA can induce gene silencing and be exchanged between host and pathogen specifically producing sRNA in *B. cinerea* that can silence genes related to plant defence promoting fungal disease progression [Wang M et al,. 2017].

This process has not been confirmed in all fungal pathosystems, for the Ascomycetes such as *Alternaria alternata* there have been some studies. In *Zymoseptoria tritici,* the infectious agent behind Septoria blotch disease in wheat, there has been a study into sRNA transfer [Kettles et al., 2019]. The pathogen causes leaves to be damaged by entering through the stomata and causing lesions and necrotrophy. In this study sRNA was sequenced and studied bioinformatically, sRNA loci were characterised, and targets predicted. These sRNA were also tested experimentally through testing as sRNA can be taken up by the fungi in vitro.

5.1.1 Fungal sRNAs

Fungal siRNAs were some of the first to be discovered and used to characterise the sRNA pathway. *Nueropossa crasa* discovered the sRNA silencing process known as Quelling, which is a form of post transcriptional gene silencing resulting from the presence of transgenes. This phenomenon was first observed in *Neurospora crassa* [Romano, 1992]. In this circumstance, *N. crassa* was transformed with DNA expressing genes that were believed to be silenced endogenously, and their copy number was used to measure the silencing efficiency. This resulted in new mutant strains of *N. crassa* called the quelling deficient mutants. *N. crassa* possesses miRNAs-like sRNAs (milRNAs) through multiple biogenies pathways, and these milRNAs are unique to the fungi in addition to Dicer-independent small non-coding RNAs (disiRNAs) that are generated from endogenous loci and produce both a sense and antisense strand. The process of biogenesis of sRNAs in fungi is well understood; however, the function and role of these RNAs is less well known. One well-documented role that the sRNA play in the fungi is the production of heterochromatin which, is upregulated by the RNAi pathway in *Saccharomyces* [Billmyre et al., 2013).

Quelling is understood as a means of regulating the expression of exogenous transgenes, by targeting the sequence complimentary to the sRNA for degradation. While siRNAs originate from endogenous sRNA loci and play a role in genome integrity.

milRNA are small RNAs found in *N. crassa,* and ex-siRNAs found in pathogenic fungi mucor circinelleoides are both involved in mRNA degradation of fungal genes. However, the evidence is not as clear as the evidence for conserved miRNA in plants [Lee et al., 2010, Nicolas et al., 2010, Trieu et al., 2015]. There are some examples like *Phytophthora sojae*; in these pathogenic fungi, tRNA-derive sRNAs, tsRNAs, can be used to regulate gene expression by degrading complimentary mRNA of the targeted gene [Wang et al., 2019]. This is obviously similar to miRNAs in plants and animals; these have also been discovered in *Magnaporthe oryzae. However, the role of post transcription gene silencing in fungi is still very poorly understood. [Li et al., 2014]*

Defence against viruses can also be observed using a process like RNAi. Cryphonectria parasitica (causal agent of chestnut blight) has been observed to use sRNA to protect itself from exogenous DNA and RNA coming from a viral infection. The role of RNAi as an antiviral mechanism in fungi and was tested by looking at dicer mutants in C. parasitica. In these mutants, DCL1 and DCL2 were observed to be like same genes from N. crassa and M. orzae and that there were little phenotypic differences observed in mutant. However, infection of the DCL-1 mutant strains with hypovirus CHV1-EP713 resulted in observable differences from the wild type. In these cases, the dcl mutants were more susceptible to mycovirus infection, the dcl mutant strains were severely impacted by the infection compared to the wild type [Kadotani et al., 2014].

5.1.2 Host-induced gene silencing (HIGS)

Barley and wheat have become a growing field in the use of HIGS, this is due to the rise of new fungal strains such as the strain of V6 stem rust in China [Liu T.G] and the failure of plant breeding methods to keep up with new diseases. A series of papers have been done on the viability of HIGS in these systems [Qi et al., 2019]. Wheat is susceptible to multiple different diseases and causes for crop failure, such as grain aphid, greenbug, Barley yellow dwarf virus and Triticum mosaic virus [Figueroa et al., 2014]. Off the fungal diseases, these can be caused by *Fusarium graminearum*, powdery mildew or *Puccinia* species causing rust disease. Nowara et al. [2010] researched the potential for HIGS targeting the *Avra10*, which is a target of the

fungal resistance gene *MLa10*. It has been demonstrated that the knockdown of MLa10 reduced fungal growth, and the silencing of *bgGTf2* and *BgGTF2* reduced fungal development. Rust Disease was studied by Yin et al (2010) using a viral-induced gene silencing method to knock down the fungal *Pst* gene and was used as a method to identify gene targets. Zhu et al. [2014] studied a variety of transgenic wheat that used an sRNA to target MAP kinase gene psFUZ7 resulting in a string highly resistant to *Pst*. Resulting in infections with far lower spore production. Demonstrating the use of HIGS through research is very different from demonstrating their value in commercial use, both accounting for their practical value and their legality.

For HIGS to be commercially viable these methods must be both specific and safe, for both the environment and consumers. Another issue is the adaptation to the resistance mechanisms present in the pathogen. Modifying plants with HIGS constructs is time-consuming and expensive. One solution to this problem has been using adaptations with multiple methods that reduce the possibility of resistance. One example of this is a variety of transgenic corn which has been approved by the Canadian Food Inspection Agency in addition to the United States Environmental Protection Agency both in 2017 [Khajuria et al., 2018]. This transgenic crop is produced by inserting a dsRNA capable of producing sRNA targeting *Dvsnf7* of the western corn worm as well as multiple insecticide proteins. This introduction was based on previous research demonstrating the effect of the presence of this siRNA on the larvae of the western corn worm.

5.1.3 Spray induced gene silencing (SIGS)

HIGS has had a multitude of issues, most predominately being overcoming the regulations related to transgenic plants especially in the European common market [Davison, 2010]. Another method demonstrated to use the RNAi pathway to protect crops from pests has been Spray Induced Gene Silencing (SIGS). However, this method has its issues; sRNA is not naturally stable, and it exists transiently within cells for a short period of time. Although, it is known to travel between organisms using extracellular vesicles [Baldrich et al., 2019]. If there were to be a system designed which could deliver sRNA reliably and efficiently, and those RNAs can be incorporated into fungal or insect and AGO and mediate effecting gene silencing,

this could result in effective fungicides and insecticides using sRNAs. There have been multiple pieces of literature looking into this approach and the problems holding it back. Studies by [Koch et al,. 2013] in *Arabidopsis* showed that expressed dsRNA regulating fungal *CYP51* genes resulted in reduced plant susceptibility to disease, specifically to *F. Graminearum* species. They followed on from this result by using a spray induced method on barley leaves using a long dsRNA construct which included multiple sequences capable of inducing silencing in the CYP51 [Koch et al, 2013]. This was also demonstrated using a GFP contract control in which spraying with the GFP gene silencing RNA resulted in significantly reduced expression. This was able to set a precedent in that theoretically any gene could be targeted by SIGS. However, some were more viable targets than others. Genes with a large degree of redundancy would be poor targets, as well as those unrelated to infection. Other papers, such as that by [Wang et al., 2016], demonstrated the silencing effect could repress grey mould through DICER like proteins in the fungus. However, efficiency and delivery remain the largest drawbacks, with sRNAs likely needing a delivery method such as Bioclay [Worrell et al., 2019] to maintain their silencing effect over long periods and prevent degradation.

With RNAi becoming an increasingly important part of modern research, allowing researchers to study genomics through gene silencing and study gene function, one of its uses is in the development of new resistant plant lines. This is due to the known systems by which the plant host produces sRNAs that can enter pathogens and parasites including nematodes, fungi, and insects. There have been multiple studies that have demonstrated the use of these methods in producing resistant lines.

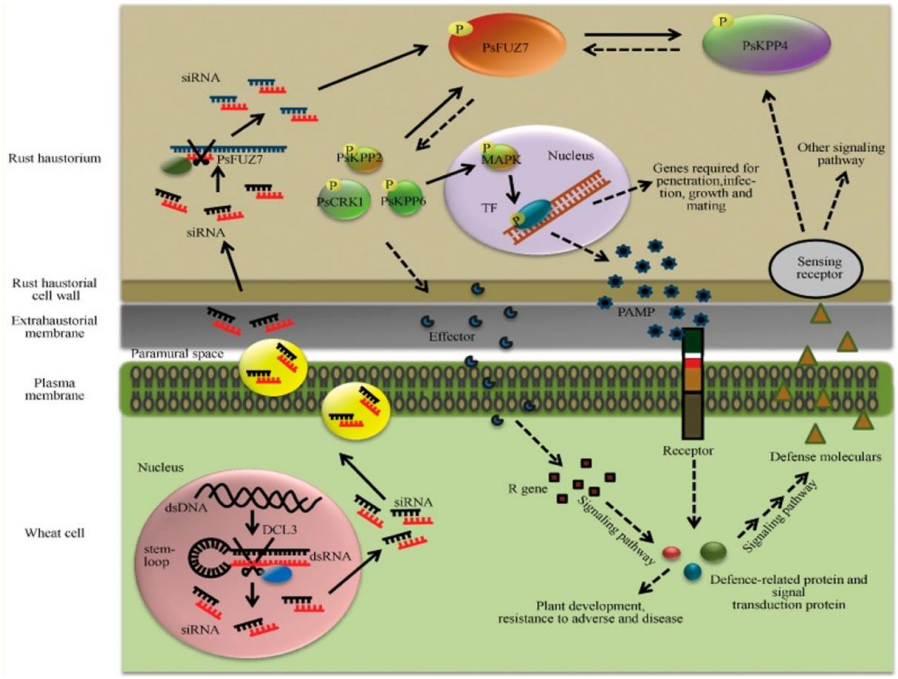


Figure 5.1: Rust Haustorium sRNA diagram from [Zhu et al., 2017]:

This is a diagrammatic representation of a theorised system for the sRNA interactions between transgenic RNAi constructs carrying a strain of wheat, and dsRNA produced by rust haustorium of the fungus Puccinia striiformis f. sp. tritici. The fungal RNA is cleaved by the plant RNA silencing complexes guided by the transgenic added s RNAs targeting these sequences. This double-stranded fungal RNA then produces sRNA, which is transported into the paramural space post the haustorial cell wall. It is also theorised that these silencing complexes induce PAMP triggered immunity, resulting in a stronger immune response to the infection in addition to the gene silencing.

5.2 Aims

The Aim of this chapter is to uncover the sRNA interactions between the fungi and the plant, determining which of the sRNA expressed by both the fungi and the plant are involved in infection. These can be either by cross kingdom silencing or by regulating genes involved in infection.

The objectives for this were as follows.

1. Extract the sRNA either by sRNA extraction alone or by total RNA extraction and sequence the sRNA.
2. Analyse the sRNA by mapping to the Genome of the plant and fungi and determining miRNA loci using miRbase.
3. Categorize both novel miRNA and known miRNA and determine which are up regulated and downregulated during infection.
4. Determine which of the upregulated miRNA have potential targets in fungi using Tapire1.1.

5.3 Results

5.3.1 Optimising the extraction of sRNAs

Before the analysis of the sRNA could begin the RNA needed to be extracted from the *A. alternata* and tomato in three phases, infected tomato tissue, non-infected tomato tissue and *A. alternata* alone. These are referred to in this chapter as the plant, infected and fungi samples respectively. This was first attempted using multiple different kits, such as the RNeasy kit, and several small RNA specific kits. Such as the MiRprimer RNA isolation kit, these work by using filters that can bind to only RNA below a certain size as well as additional steps which filter out genomic DNA and mRNA. However, these kits were found to be ineffective, they would often fail to extract any RNA or if they did, they would produce RNA of small quantities and low quality.

Quality was demined by two methods, observation on electrophoresis gel, where the ratio of the two RNA bands were used to determine the quality of the extraction, this however is only possible in total RNA extraction and measured by a bioanalyiser. For the sRNA extractions quality could only be estimated by nanodrop measurements and looking at the odd measurement at 230/260 of a 2.00 or higher. This method was inaccurate as samples would be evaluated for their quality and viability for sequencing based on RIN quality. RIN quality was measured by a bioanalyiser and was determined by the ratio of the two major RNA bands. This meant that total RNA extraction was used using the Trizol method detailed in [materials and methods 2.4.1] as this extracted the total RNA and from this only the s RNA (<200 bp) would be sequenced. This method still had difficulties with RNA quality, so up to 10 samples of RNA extractions were sent in aliquots of 10ul for bioanalyzer analysis so that the three best samples could be taken for analysis. Small RNAs were sequenced by Novogene and the data was analysed subsequently.

5.3.2 Analysis of sRNAs from infected and non-infected tomato tissues

The first step in the sRNA identification pathway is to align then total sRNA reads to the respective genome. This was done separately for two different analysis methods, miRNA identification performed by novogene, and I carried out the sRNA read distribution analysis using Geneious.

This was performed using the methods shown in [materials and methods 2.7.1] in which the reads were first mapped to the fungi and plant genomes. These results show in both genomes where the sRNA maps to genome showing the likely origins of the sRNAs and how these changes during infection. With peaks in the sRNA expression changing during infection.

The results show a significant difference between sRNA distribution between infected and non-infected samples. However, there is little change in the number of sRNAs with the only individual replicate with a significant difference being INF 1, which disappears when the 3 samples are taken as an average. However, a visible difference shown in [Figure 5.1] is the spread of the sRNAs, the distribution in the plant samples appears to be more spread out in the plant samples while in the infected samples the sRNA total numbers may be the same, but they distribute to a fewer number of genes. The mapped reads can be seen in [Table 5.1] as a raw data showing the number of mapped and un mapped reads, [figure 5.2] shows a graph of the total mapped plant vs infected and [figure 5.3] shows the mapped reads per chromosome.

Table 5.1: Mapped reads per infected and non-infected sample

sRNA sequenced from each of the 6 samples are mapped to the Tomato reference genome. Mapped sRNA is shown both in the number of sRNA reads and percentage of total sRNA mapped to the genome. This was done for both the positive sense strand ("+" mapped sRNA) and the negative sense strand ("-"mapped sRNA)

Sample	Total sRNA	Mapped sRNA	"+" Mapped sRNA	"-" Mapped sRNA
INF1	18749522 (100.00%)	12755550 (68.03%)	10380733 (55.37%)	2374817 (12.67%)
INF2	11883493 (100.00%)	6938412 (58.39%)	5515534 (46.41%)	1422878 (11.97%)
INF4	11763345 (100.00%)	9692035 (82.39%)	8555362 (72.73%)	1136673 (9.66%)
P5	12102622 (100.00%)	10367905 (85.67%)	8049562 (66.51%)	2318343 (19.16%)
P3	11308987 (100.00%)	9969339 (88.15%)	8708178 (77.00%)	1261161 (11.15%)
P4	14173977 (100.00%)	11916149 (84.07%)	10163119 (71.70%)	1753030 (12.37%)

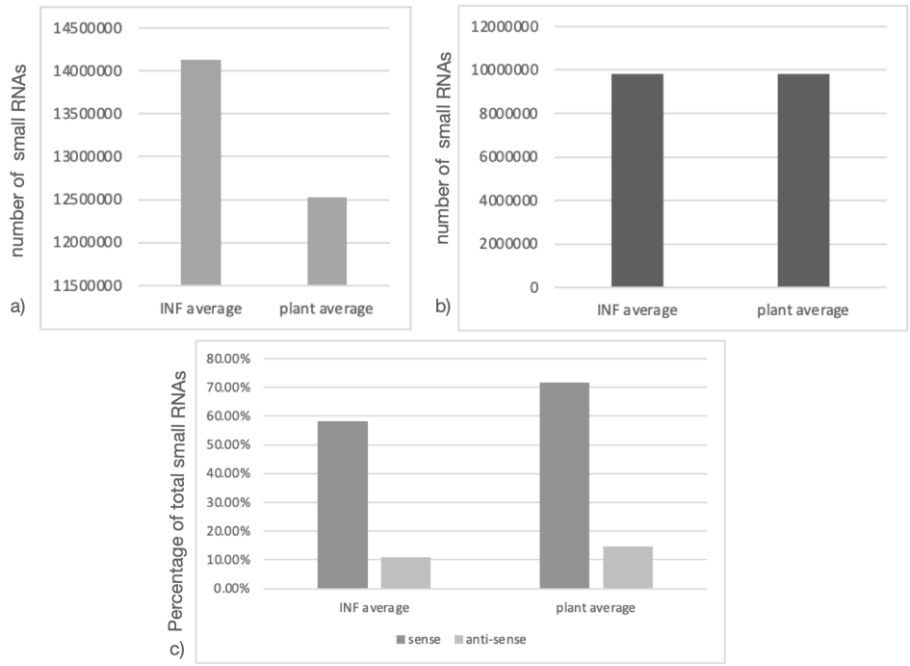


Figure 5.2 Differences between sRNAs mapped to the infected and non-infected tissue:

The sRNA reads from the 6 sRNA libraries were mapped to the plant genome and the average values for mapped sRNAs were plotted into graphs in three states. a) Infected read average and non-infected plant read average totals compared, b) the reads mapped to the genome were compared in both sample averages, and then separated a plotted in c) for both sense and anti-sense strands.

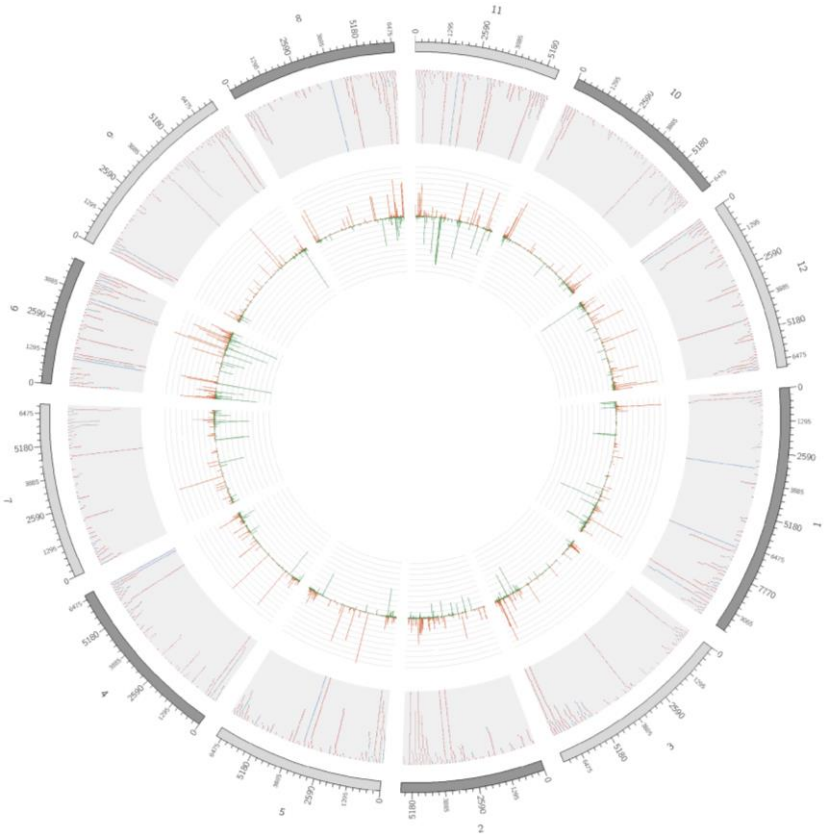


Figure 5.3.1: Reads distribution per chromosome for INF2:

The chromosome as shown is the outer circle. Grey background in the middle area shows the distribution of 10,000 reads on the chromosome. Red represents the sRNA number on the sense strand of the chromosome, and blue represents the number of sRNAs on the antisense strand. All reads are shown in the centre area of the circle. Yellow represents the number of sRNAs on the sense strand of the chromosome, and green represents the number of sRNAs on the antisense strand.

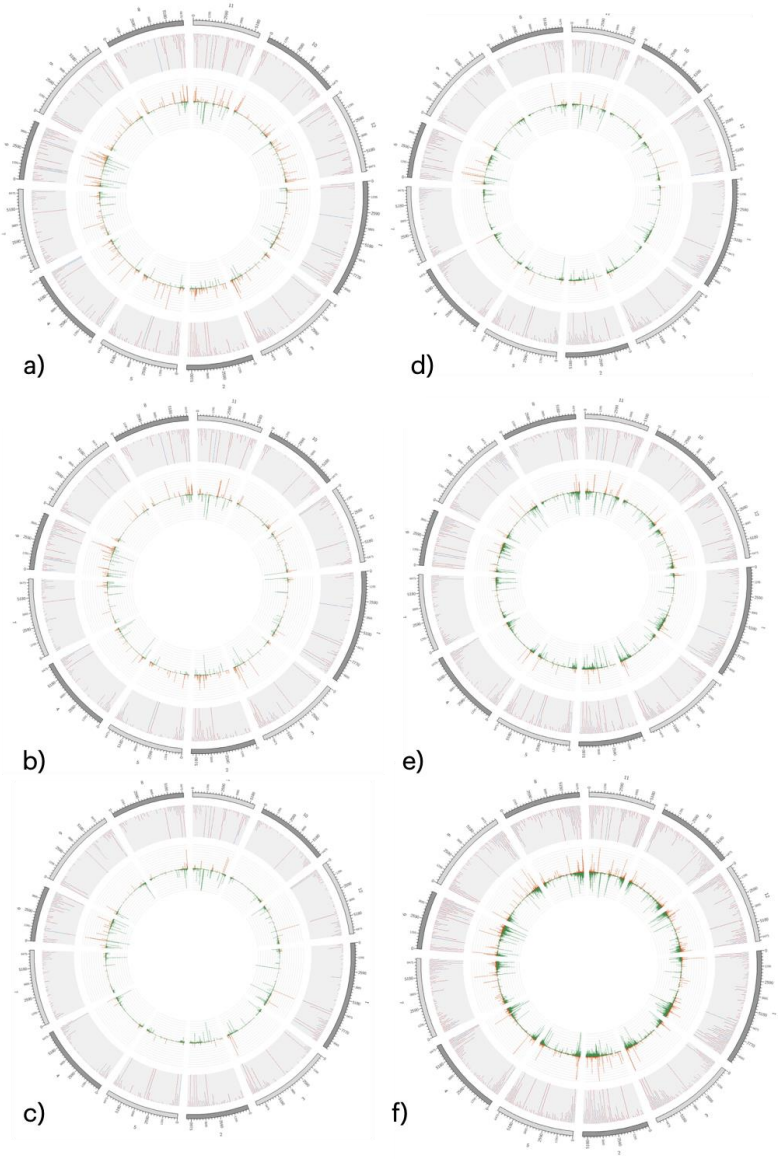


Figure 5.3.1

What can be observed here is that even though the number of sRNA mapped to the genome in both conditions as seen in Figure 5.3.1a,b,c the reads can be observed to distribute differently. With the read mapping in the plant genomes shown in Figure 5.3.1a, b, c seems to be more spread out compared to the infected in Figure 5.3.1

d, e, f, which represent p3,4 and 5 respectively. has the same number of reads but more focused with larger peaks.

5.3.3: Analysis of known miRNA

The reads that were matched to their respective reference genome were compared with specific sequences in the miRbase to obtain detailed information of the mapped reads. This included the secondary structure of the mapped miRNAs, the sequence of miRNAs in each sample, the length, the number of occurrences and other information. Mature miRNAs are developed from precursor through Dicer enzyme digestion. Specificity of enzyme digestion sites makes the first base of the mature miRNA sequence highly biased, the first base distribution was analysed for the miRNAs with different lengths, as well as the distribution of other bases. This analysis could only be performed with the plant and non-infect tissue sRNAs mapped to the plant genome due to the absence of the miRNA database present for *A. alternata* miRNAs. This method was used to narrow down the sRNA mapped to the tomato genome to a list of known miRNAs and novel miRNAs. Known miRNA are shown in [Table 5.2] with a list of mapped mature and hairpin miRNA, unique sRNA, and total sRNA.Showing that of the unique sRNA present only a very small number are miRNA. In addition, it shows that the miRNA numbers are very similar between the infected and the non-infected tissue. This means that any observable difference is not in the numbers of unique miRNA present in the infected. [Table 5.3] lists the read counts for 5 known miRNA and their different expressions in each of the 6 samples.

Table 5.2 Summary of number of known miRNAs in each sample:

The types of miRNAs are distinguished between, mature which are the processed form of the Hairpin. Showing the number of identified miRNA of each type mapping to each genome. This table also includes the unique and total sRNA which does not distinguish the type of sRNA found in the library.

Types	Total	INF1	INF2	INF4	P3	P4	P5

Mapped mature	134	107	103	102	110	116	130
Mapped hairpin	105	100	96	101	101	103	105
Mapped unique sRNAs	13975	2038	1309	1541	1859	2693	4535
Mapped total sRNAs	1912730	606207	170036	273960	128917	186021	547589

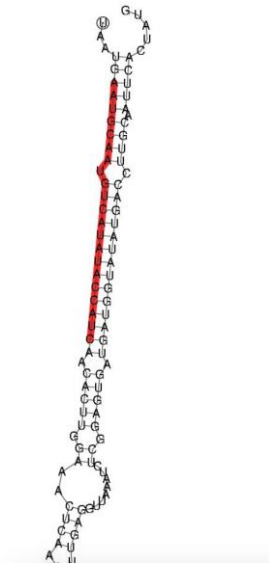


Figure 5.4 Known miRNA secondary structure.

The entire sequence is a miRNA precursor, the red section is the mature sequence generated once the miRNA is processed by Dicer and is the sequence used for targeting the gene. This image shows the position of the known miRNA sly-MIR10528. [mirbase.org]

Table 5.3 Known miRNA expression profile

This table shows an example of the sRNA read counts for 5 novel miRNAs present in each of the 6 tested samples. Showing that some are present in the infected, non-infected or both.
Of the novel miRNA an average read count was taken between infected and non-infected, most of the novel miRNA showed very low average expression so of the novel miRNA the ones with average read counts above 100 were plotted on this graph.

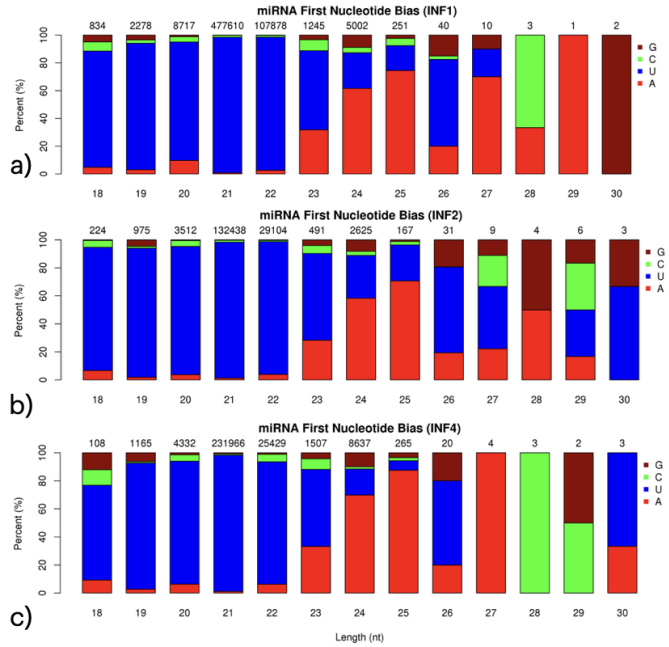
miRNA	INF1	INF2	INF4	P3	P4	P5
sly-miR10528	19.00	0	11.00	162.00	313.00	174.00
sly-miR10529	0	0	0	2.00	9.00	12.00
sly-miR10530	586.00	174.00	107.00	12.00	8.00	48.00
sly-miR10531	13.00	7.00	3.00	0	0	8.00
sly-miR10532	2612.00	1214.00	4299.00	20846.00	27756.00	17929.00

5.3.3.1 The details of mapped known miRNAs

After mapping out the genome there were 134 known mature miRNAs. Five of these miRNA were taken and shown for differential read count in the infected in Table 5.2.2. This demonstrates that in some cases the miRNA was not present in the infected or not present in the plant sRNA library, respectably. Examples include sly-miR10529, which shows no miRNA reads in the infected samples but a few miRNA numbers in the plant tissue without the infection. However, this is an example of where the read counts are so low that is hard to tell if there is any repression or this miRNA during infection has any impact. This is more pronounced for sly-miR10528 where there is a large difference between the expression between infected and non-infected where the read counts in the non-infects are above 100 reads. This is also the case for sly-miR10532 where there is still a significant amount of the miRNA present in the infected but significantly less than the plant tissue.

5.3.3.2 miRNA first nucleotide bias.

The first and last nucleotide of a miRNA has been shown to affect the miRNAs binding to the AGO proteins [M.W. Jones-Rhoades et al,. 2006] so this was also analysed and shown in [Figure 5.4] which shows the first nucleotide biases in each of the 6 samples, the three replicates of the infected and the three replicates of the plant. Observations present here include that there is a change in the 22 bp miRNA from infected to non-infected, with the infected showing far higher levels of uracil in the first base than in the non-infected which had higher levels of cytosine and adenine. This effect can also be seen at 24 nt and higher, which showed consistently higher distribution of adenine in the first nucleotide. There was also a far higher preference for cytosine in the infected, especially at 26nt long.



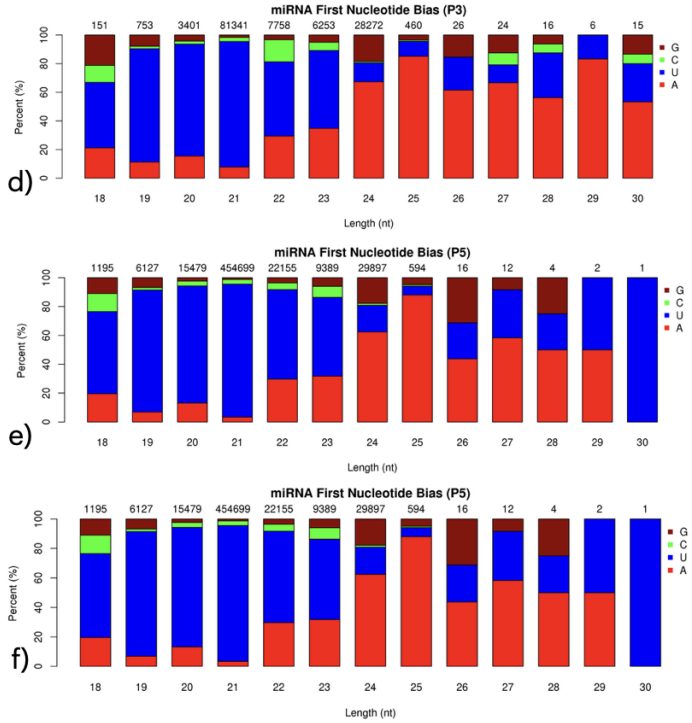


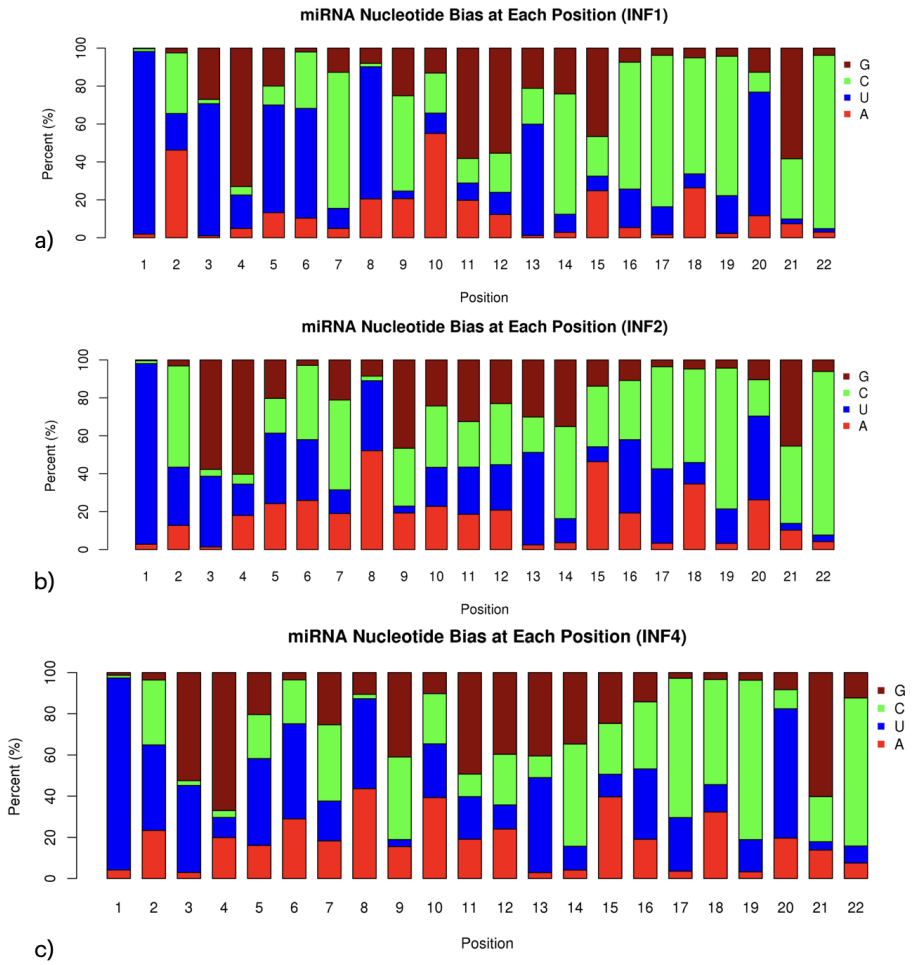
Figure 5.4 miRNA first nucleotide bias

The first nucleotide bias is the percentage of sRNA that one nucleotide in the first base pair, this allows immediately to see which samples showed a bias to a particular base pair over another. In this case all three plant samples show visibly more adenine at the first base in sRNA between 24 and 30. The plant samples also show more guanine percentages in sRNA between 18 and 22 bp in length. (a) (b)(c) show results for the infected samples INF1 INF2 and INF4 respectively with (d)(e) and (f) showing the results for P5,P3 and P4 respectively.

5.3.3.3 miRNA all nucleotide biases

In an addition to the first nucleotide there was also analysis performed to analyse the biased at least individual base of the miRNAs. This allows you to clearly see the increase in adenine in the non-infected miRNA over the infected miRNA. It also shows a significant change in the last base with a significant increase in cytosis content in the infected over the plant miRNA. Overall, for the rest of the positions you can observe an

overall noticeable increase in cytosine in the infected miRNA and a decrease in the percentage of uracil. These are represented in graphical form in [figure 5.5].



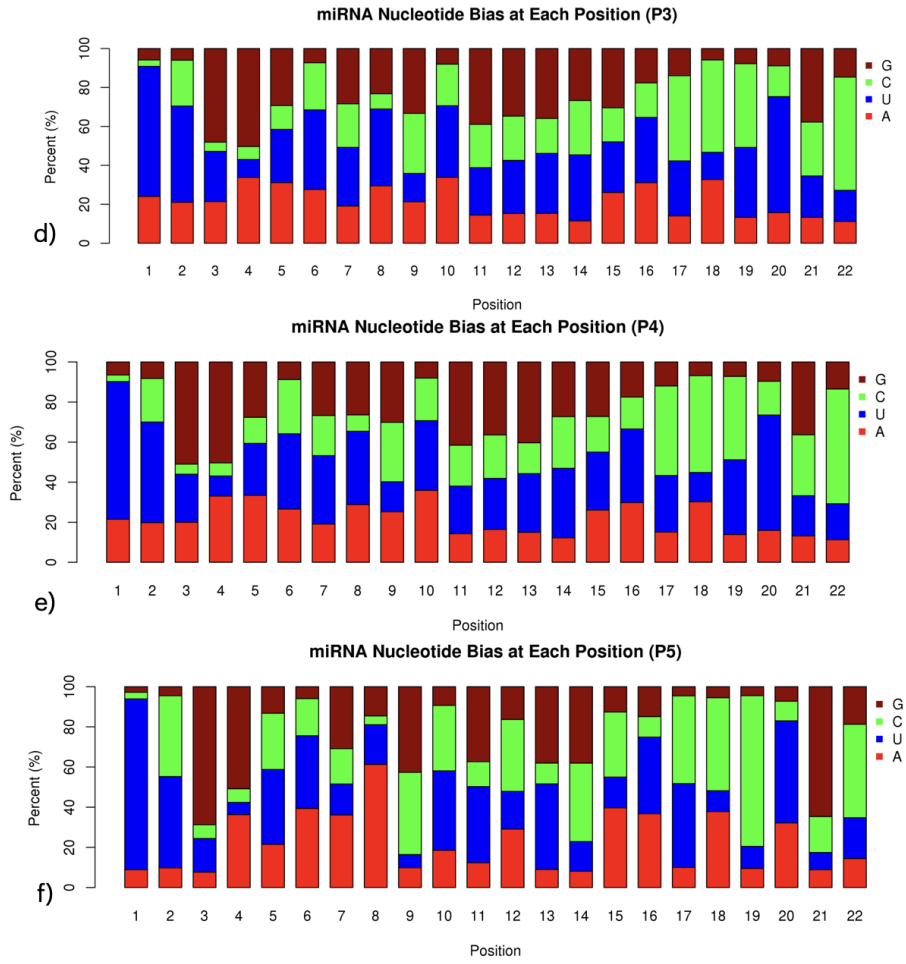


Figure 5.5 miRNA nucleotide bias at each position:

The X axis shows each position of miRNA nucleotide, The Y axis shows the percentage. (a) (b)(c) show results for the infected samples INF1 INF2 and INF4 respectively with (d)(e) and (f) showing the results for P5,P3 and P4 respectively.

5.3.4 non-coding (ncRNA) Analysis

Small RNA reads are annotated with sequences from the non-coding transcript sequences of the species from its genome reference. The matched reads from rRNA, tRNA, snRNA, and snoRNA are removed as these are non-silencing small RNA. These ncRNAs are shown in [Table 5.4]

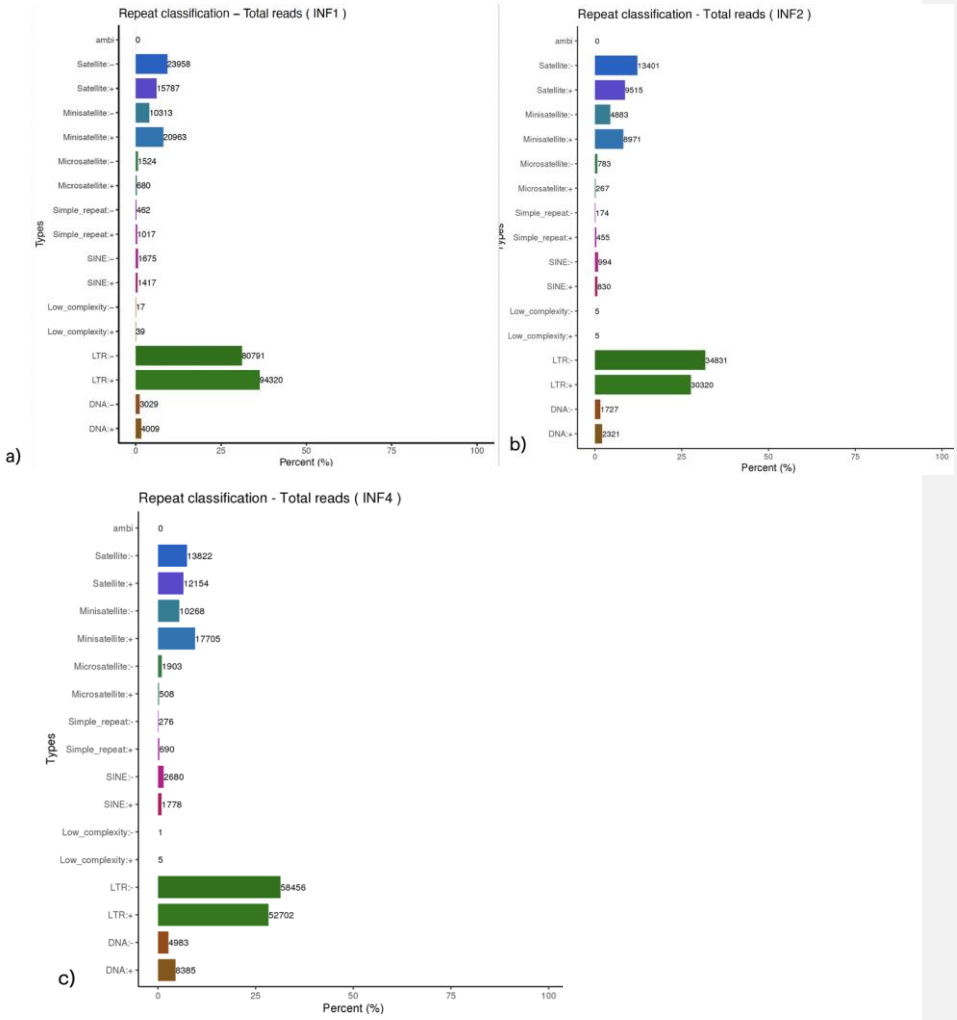
Table 5.4 Statistics of annotated ncRNA:

Annotated sRNA are listed by the read count for each of the 6 samples, three for the infected designated INF1,INF2 and INF4, and three for the plant sRNA designated P3,P4 and P5. Some of the annotated sRNA are separated by positive and negative sense strand by "+" and "-" for possive and negative sense strand respectively.

Types	INF1	INF2	INF4	P3	P4	P5
rRNA	1328629	495836	243211	231288	571860	561038
rRNA:+	1296139	484326	234129	229746	565007	522473
rRNA:-	32490	11510	9082	1542	6853	38565
tRNA	23612	4766	4598	3518	15851	47408
tRNA:+	4046	662	1693	334	1655	7525
tRNA:-	19566	4104	2905	3184	14196	39883
snRNA	69000	20637	14754	4425	21950	38529
snRNA:+	56593	18114	13076	3522	15316	27498
snRNA:-	12407	2523	1678	903	6634	11031
snoRNA	348979	74953	140222	26259	128076	257666
snoRNA:+	319282	69553	138332	25398	123342	240309
snoRNA:-	29697	5400	1890	861	4734	17357

5.3.5 Repeat Sequences Alignment

If a species has repeats transposon information, this repeated sequence is used to annotate the sRNA. Otherwise, we will predict repeated sequences from the beginning based on reference genome. The sRNA is then aligned with the predicted repeating sequences and potential repetitive reads of sRNA are removed. Statistics on the various repeats of uniquely expressed sRNAs and total number of sRNAs were then calculated.



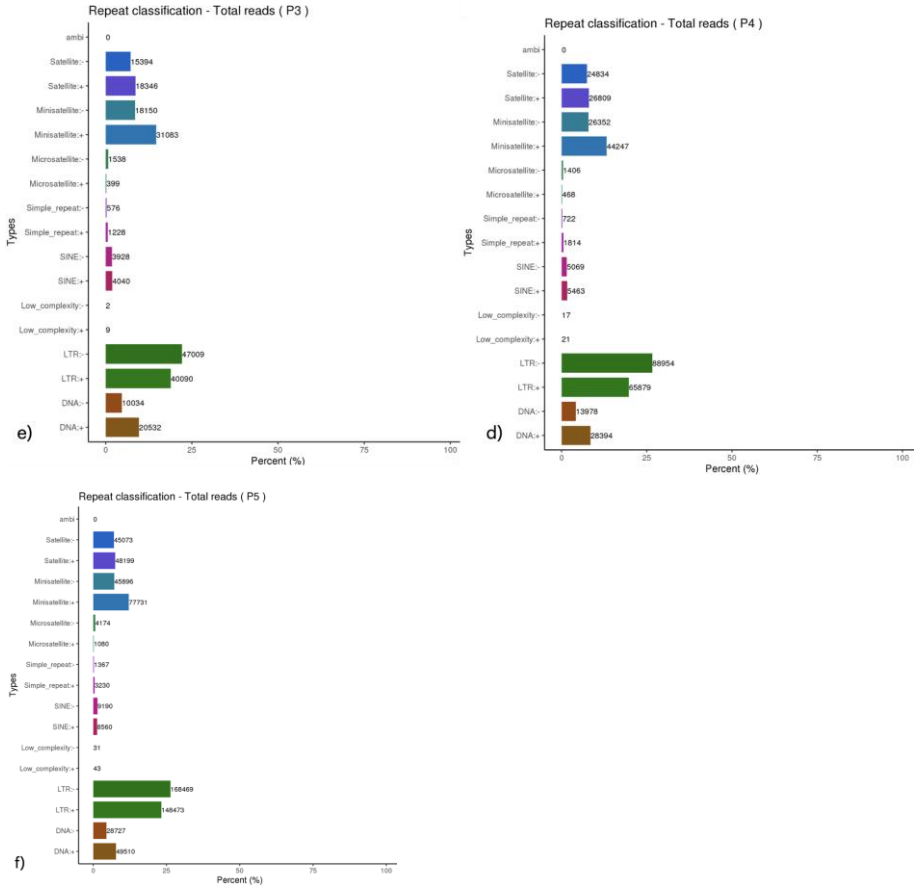


Figure 5.6 Total repeat reads:

This figure shows the total numbers of reads which map to repeat regions for the three infected samples. Figure 5.6 a) shows the distribution for sample INF2 showing the majority distributing to the Long terminal repeat (LTR) regions and this can be seen also in 5.6 b and 5.6 c for the INF 1 and INF 4 samples respectively. Distribution can also be seen for P3,4 and 5 in images 5.6 d,e and f.

5.3.6 Plant NAT-siRNA

Natural antisense transcripts (NATs) are a type of RNA used for regulation. They can be found predominately in plant genomes. They have a significant role in pathology making them a very interesting RNA for plant pathologists . NAT-siRNA can be found in two forms,

cis-NATs and trans-NATs. The RAW read data for sRNAs mapped to NATs are shown in Table 5.5 with the distribution of NAT-siRNA shown in [figure 5.7] where the distribution is graphically represented for both infected and plant trans-NAT siRNA.

Cis-NAT pairs are transcribed from the opposing DNA strands but are at the same genomic locus. They have a range of different orientations and different lengths and degrees of overlap between exact sequence complementary to its target. This contrasts with trans-NAT pairs which are observed to be transcribed from different loci and are only capable of forming partial complementarily with its target. NATS were first discovered using the Platinated Database [Chen et al., 2011] which allowed researchers to identify known NATs and novel NATs.

Table 5.5 sRNAs mapped to NATs

sRNAs from each sample were mapped to available sRNA databases to identify NATs, with the types of NATs shown on the left column and the number of reads mapping to them from each sample shown under the sample names.

Types	INF1	INF2	INF4	P3	P4	P5
trans-NAT	2802458	1556467	442446	593931	1633106	1001376
trans-NAT:+	2791307	1550053	437248	588769	1622800	982986
trans-NAT:-	11151	6414	5198	5162	10306	18390

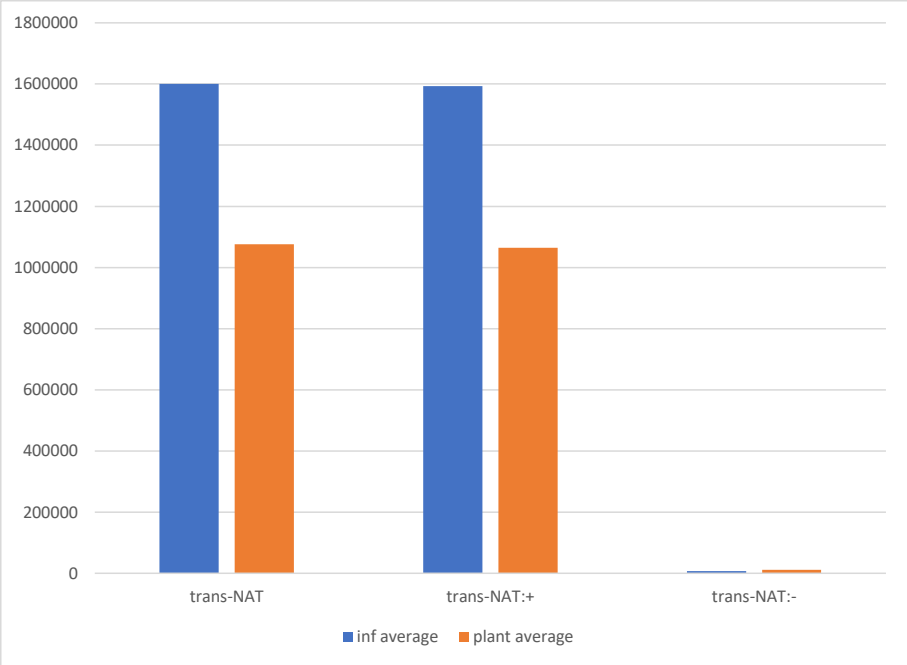


Figure 5.7. Trans-NAT sRNAs from infected or non-infected plant tissues.
Trans-Nat's were identified by sRNA reads from each library mapping to known trans-Nat louses. These were then split between total trans Nat in both plant and infected and these were separated and plotted between positive sense strand aligned and negative sense strand aligned. This graph clearly shows the large difference between negative sense and positive sense trans-Nat's and some difference between infected and non-infected with the infected showing a larger distribution of trans-NATs.

5.3.7 Exon and Intron Alignment

The sequences of sRNA sequencing may have degraded fragments of mRNA. This part of the analysis is to annotate the sequencing reads as much as possible. This allows us to remove the sequences from these genes before the new miRNA predicted. As these regions might be the product of degraded fragments. These regions can predict relatively few miRNAs compared

to the gene intergenic region. This data is shown in [table 5.6] in which the sRNA aligning to exons and introns are recorded for plant and infected tissue.

Table 5.6 Number of sRNAs mapped to exon and intron
sRNAs were mapped to the Tomaot genome and those that mapped perfectly to exon or intron regions were quantified. The first column shows types of regions including positive and negative sense exon or intron and all exons in total. The numbers represent the numb er of reads mapped to those regions.

Types	INF1	INF2	INF4	P3	P4	P5
exon	4839046	2782951	6558286	6917644	6101787	3210279
exon:+	3222314	2652013	6279105	6814093	5883489	2934775
exon:-	1616732	130938	279181	103551	218298	275504
intron	447401	237049	344156	506824	867739	1216120
intron:+	275811	149326	199742	294018	462556	660593
intron:-	171590	87723	144414	212806	405183	555527

5.3.8 Novel miRNA Prediction

miRNA is separated into three categories, hairpin, which includes the entire hairpin sequence before processing by DICER, the mapped star miRNA, which is a miRNA that maps to another more expressed miRNA and likely complementary to it [Feng X et al,. 2014] and the mapped mature miRNA, the miRNA that is like predominant to that mature miRNA precursor. This method allows for prediction of novel miRNA no present in the sRNA database by prediction based owns tractate, having already eliminated known miRNA and potential misleading reads from those mapping to introns, NAT-sRNA and repeat sequences.

Table 2.7 Number of novel miRNAs

sRNA could be mapped to miRNA regions that are either mature, star or hairpin regions of the miRNA, here the sRNA read numbers are shown which map to any of those three regions including unit and total sRNA.

Types	Total	INF1	INF2	INF4	P3	P4	P5
Mapped mature	230	131	107	173	200	206	220
Mapped star	104	19	19	23	37	45	70
Mapped hairpin	240	144	133	189	214	220	228
Mapped unit sRNA	5619	688	443	631	919	1260	1678
Mapped total sRNA	64451	16277	3666	4640	7653	12566	19649

Figure 5.8 The secondary structure of the novel miRNAs on partial schematic matches.

The entire sequence is a miRNA precursor, the red section is the mature sequence [mirbase.org]

After identifying the novel milRNA the next step is to evaluate the difference in expression and the properties and targets of those miRNA. Expression can be initially looked at in terms of read count between infected and non-infected tissue. This is shown in [Table 5.8] in which the number of reads for 5 novel miRNA are shown for each of the 6 samples.

First nucleotide bias was observed to have several differences in the novel to known miRNA. Mostly a far higher percentage of guanine than cytosine, this difference was also pronounced compared to the plant tissue novel miRNA ethic showed a higher percentage of adenine and uracil but still a larger percentage of cytosine and guanine compared to the known miRNA. The large number of miRNAs also showed a far higher bias towards adenine compared to the infected novel miRNAs. The first base pair is an indicator of target as the first 5 bp of a miRNA is the seed region and determines target. Meaning that significant difference in expressed miRNA first bas pair bias is a strong indicator in a change of miRNA targets in the two conditions. The nwuclotide distribution at the first base is indicated in [figure 5.9] indicating with different colours which base has the higher number of sRNAs possessing it as the first base as a percentage of the total sRNAs, this is also shown for every base in the sRNA in [figure 5.10] with a similar graphical model.

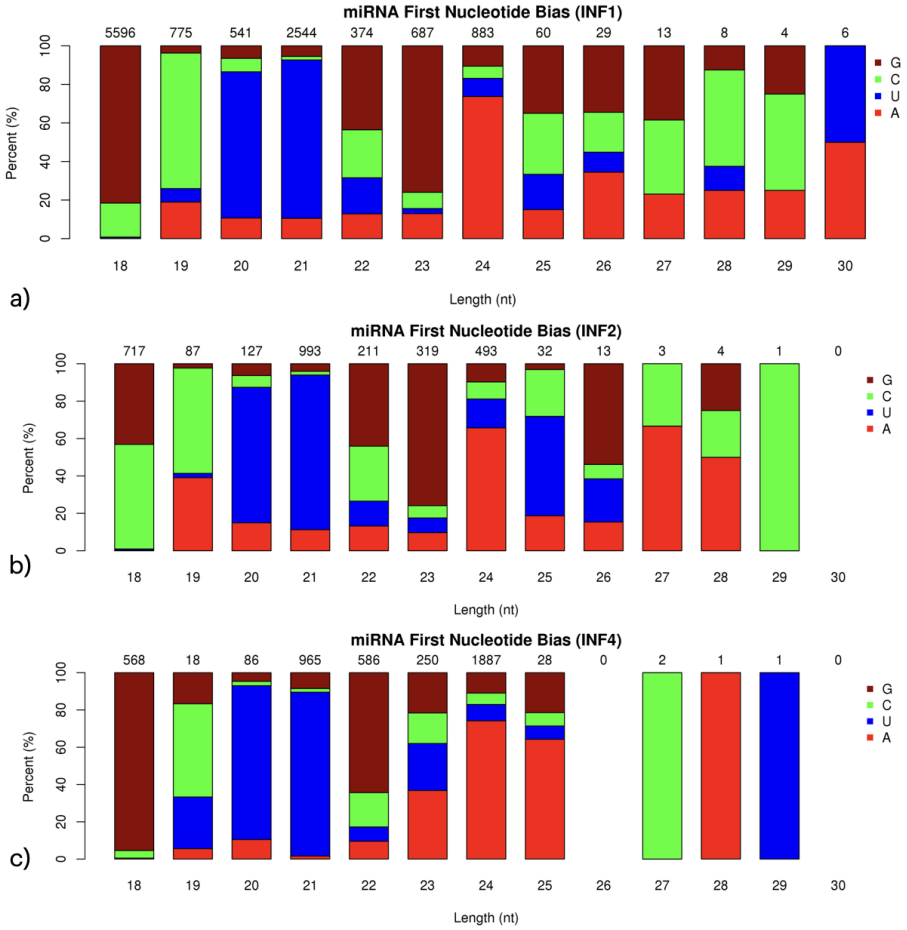
Table 5.8 Novel miRNA expression profile of 5 novel miRNAs:

Example of 5 novel miRNA and the number of reads mapping to these miRNA found in each sample.

miRNA	INF1	INF2	INF4	P3	P4	P5
novel_1	2315.00	877.00	875.00	37.00	45.00	305.00
novel_10	9.00	6.00	16.00	19.00	1.2.00	75.00
novel_100	0	5.00	17.00	9.00	8.00	28.00
novel_101	514.00	165.00	0	0	0	0

novel_102	0	0	0	6.00	5.00	4.00

This table shows the read count differences in the 6 evaluated samples for 5 novel miRNA some clearly showing zero or reduced expression in the fungi tissues. However, some miRNA like novel_10 has so few miRNAs in both samples that determining if there is a significant difference in both conditions.



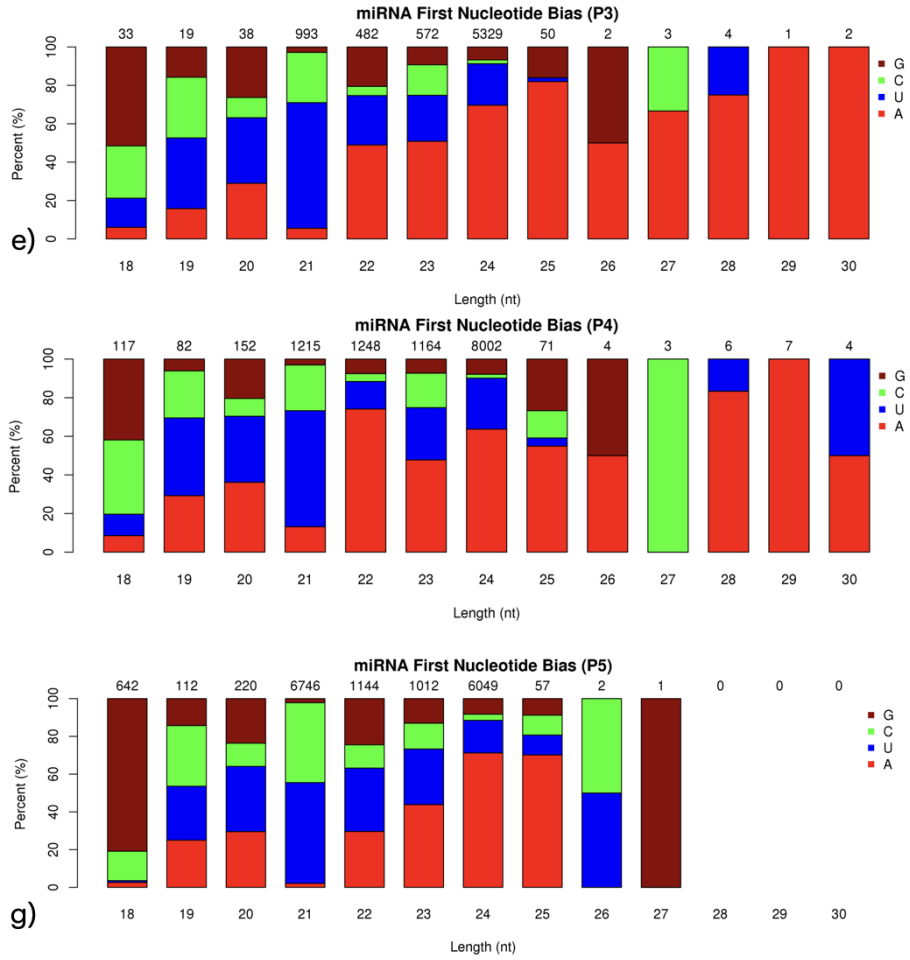
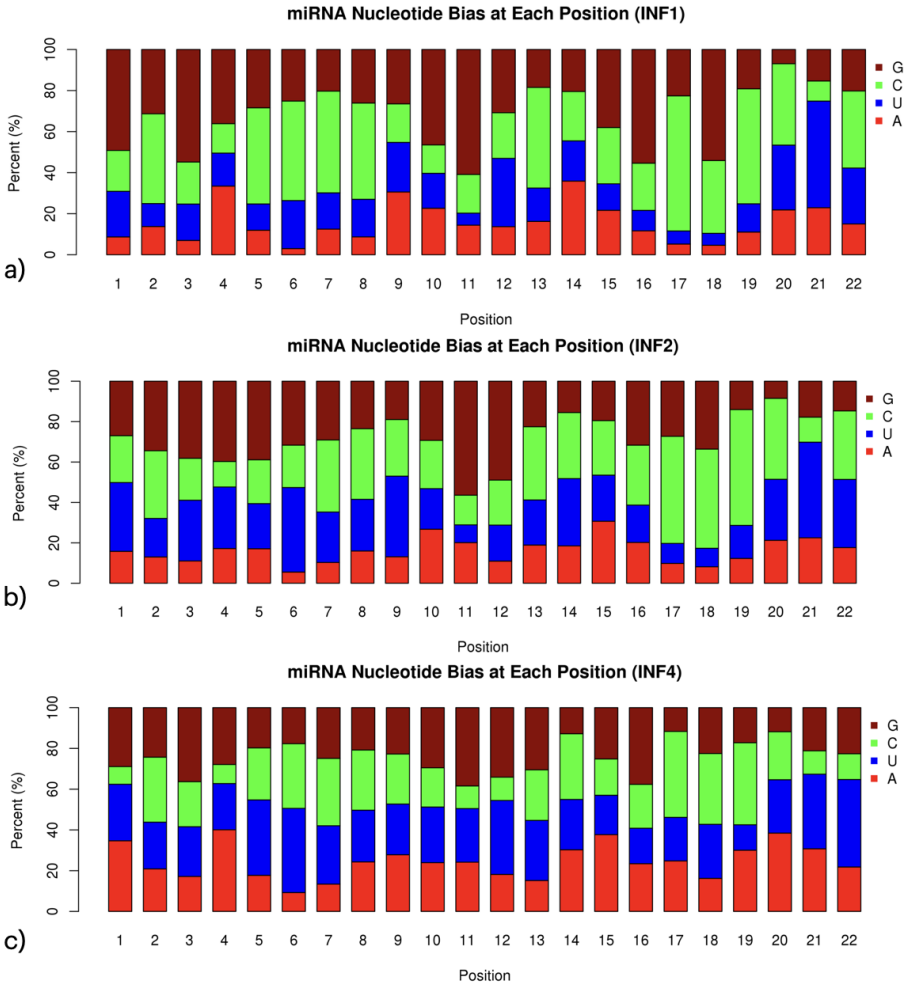


Figure 5.9 Novel miRNA first nucleotide bias:

The length of miRNAs is shown in the X axis, the Y axis is the percentage. (a) (b)(c) show results for the infected samples INF1 INF2 and INF4 respectively with (d)(e) and (f) showing the results for P5,P3 and P4 respectively.


miRNA Nucleotide Bias at Each Position (INF1)
Percent (%)
0
20
40
60
80
100
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22
Position
G
C
U
A
a)
miRNA Nucleotide Bias at Each Position (INF2)
Percent (%)
0
20
40
60
80
100
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22
Position
G
C
U
A
b)
miRNA Nucleotide Bias at Each Position (INF4)
Percent (%)
0
20
40
60
80
100
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22
Position
G
C
U
A
c)

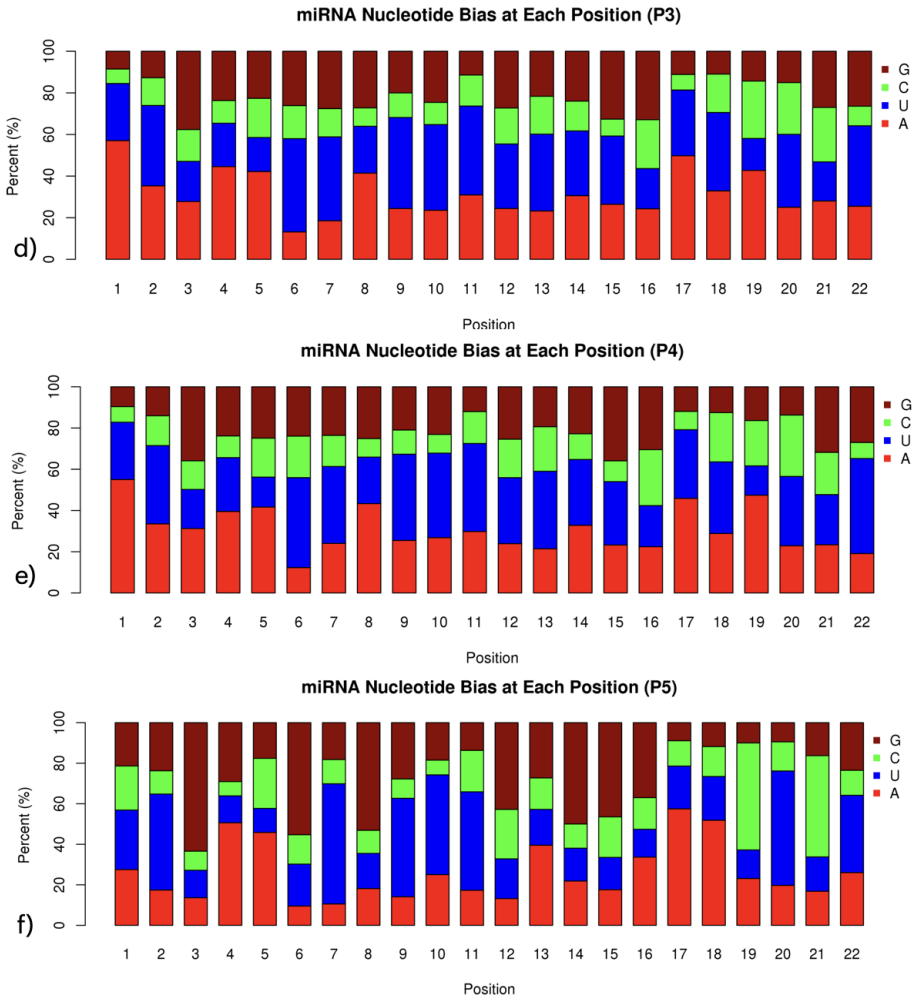


Figure 5.10: novel miRNA nucleotide bias at each position:

The X axis shows each position of miRNA and the Y axis shows the percentage of miRNA showing this bias at this position. (a) (b)(c) show the samples for the INF 1,INF2 and INF 4 respectively and (d) (e) and (f) show the results for the P2,P4 and P5 respectively.

5.3.9 Plant Trans-acting siRNAs (TAS)

The identification of known TAS gene based on *Arabidopsis* and *Oryza sativa* database. We used UEA sRNA tools (Mixon et al., 2008) to predict new TAS gene. This Table shows the small number of TAS genes observed in the two small RNA library indicating that if these TAS genes are expressed in tomato they are likely not expressed in large quantities during infection and therefore not likely to be of interest for this project. The mapped sRNAs to TAS genes can be seen in [Table 5.9] indicating the number of TAS gene mapped sRNA in each infected and plant sample.

Table 5.9 Statistics of sRNA mapped to TAS gene

sRNA mapped to TAS regions are shown with the numbers representing the number of reads per sample.

Types	INF2	INF4	P3	P4	P5
TAS	1	2	1	4	6
TAS:+	1	2	1	4	6
TAS:-	0	0	0	0	0

5.3.10 Annotation of small RNA reads

Some s RNA reads may be mapped to multiple regions and annotated with multiple features. To denote each sRNA with a unique annotation, we follow below priority for annotation known miRNA, rRNA, tRNA, snRNA, snoRNA, repeat gene, novel miRNA. The quantity of sRNA reads mapped to genome is shown in the total column, known miRNA, which is the number and percentage of sRNAs reads mapped to known miRNA, rRNA/tRNA/snRNA/snoRNA, which is the number and percentage of sRNAs reads mapped to rRNA/tRNA/snRNA/snoRNA, the repeat column which is the number and percentage of sRNAs reads mapped to repeat region. The total expression of the different annotated reads can be seen in raw data form in [table 5.10] and this data is shown in graphical form in [figure 5.11] where the expression is compared for each annotated type of sRNA between the average plant and infected samples.

Table 5.10.1 Number of all sRNA annotation from the plant and infected samples mapped to the plant genome.

All possible regions for the sRNA reads to map to and the number of reads from each sample that mapped to them.

Types	INF1	INF2	INF4	P3	P4	P5
total	12755550	6938412	9692035	9969339	11916149	10367905
known miRNA	606207	170036	273960	128917	186021	547589
rRNA	1328629	495836	243211	231288	571860	561038
tRNA	23612	4766	4598	3518	15851	47408
snRNA	69000	20637	14754	4425	21950	38529
snoRNA	348979	74953	140222	26259	128076	257666
repeat	260001	109482	186316	212358	334427	639753
NAT	2802458	1556467	442446	593931	1633106	1001376
novel miRNA	16277	3666	4640	7653	12566	19649
TAS	1	2	1	4	6	0
exon: +	3222314	2652013	6279105	6814093	5883489	2934775
exon: -	1616732	130938	279181	103551	218298	275504
intron: +	275811	149326	199742	294018	462556	660593
intron: -	171590	87723	144414	212806	405183	555527

other	2013939	1482567	1479445	1336518	2042760	2828498

Table 5.10.2 Number of all sRNA annotation from the plant and infected samples mapped to the plant genome.

Continuation of table 5.10.1

Types	INF1(percent)	INF2(percent)	INF4(percent)	P3(percent)	P4(percent)	P5(percent)
total	100.00%	100.00%	100.00%	100.00%	100.00%	100.00%
known miRNA	4.75%	2.45%	2.83%	1.29%	1.56%	5.28%
rRNA	10.42%	7.15%	2.51%	2.32%	4.80%	5.41%
tRNA	0.19%	0.07%	0.05%	0.04%	0.13%	0.46%
snRNA	0.54%	0.30%	0.15%	0.04%	0.18%	0.37%
snoRNA	2.74%	1.08%	1.45%	0.26%	1.07%	2.49%
repeat	2.04%	1.58%	1.92%	2.13%	2.81%	6.17%
NAT	21.97%	22.43%	4.57%	5.96%	13.70%	9.66%
novel miRNA	0.13%	0.05%	0.05%	0.08%	0.11%	0.19%
TAS	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
exon: +	25.26%	38.22%	64.79%	68.35%	49.37%	28.31%
exon: -	12.67%	1.89%	2.88%	1.04%	1.83%	2.66%
intron: +	2.16%	2.15%	2.06%	2.95%	3.88%	6.37%
intron: -	1.35%	1.26%	1.49%	2.13%	3.40%	5.36%
other	15.79%	21.37%	15.26%	13.41%	17.14%	27.28%

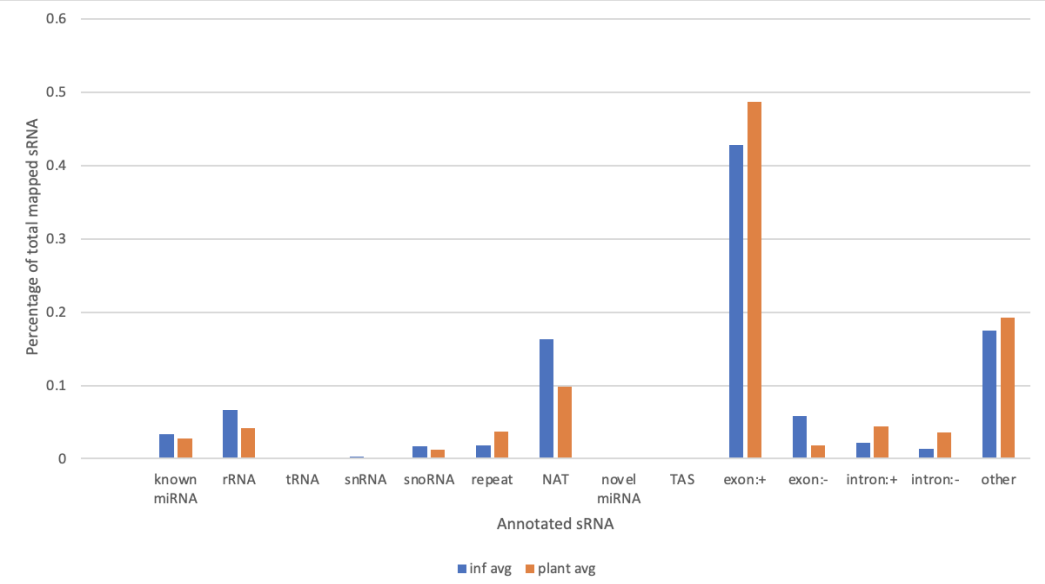


Figure 5.11. Average annotated sRNA percentages shows little difference between infected and non-infected in plants:

This figure shows the total annotated sRNA was calculated as a percentage of the total mapped sRNA to show which of the known sRNAs were seen to be upregulated during infection, this was seen to have little effect on the sRNA relative percentages, with only NATs and exon mapped sRNA showing any visible difference.

5.3.11 miRNA Base Edit

The mature miRNA has a Position at 2-8 known as the seed region, this region is highly conserved and many studies have shown play a significant role in target identification. Targets of a miRNA can be different depending on change of nucleotides in seed region. However, RNA can be the target of a process of RNA editing where bases can be changed, usually A to I [Park et al., 2021] these can be verified in the miRNA bioinformatically. In this analysis, miRNAs with base edits can be found by aligning unannotated sRNA reads with mature miRNAs found using mirbase. The results showed a surprising high percentage of miRNA showing base editing. With up to 80% of miRNA in some cases being found with base edits and in some positions showing significantly higher rates of base edits. As an example, in Novel_1 its precursor miRNA showed a rate of 49% of RNAs containing a base

edit, while the mature showed 80% with site 4 being a common target with 47% possessing a g to c base edit.

5.3.12 miRNA Expression and Differential Expression analysis

5.3.12.1 miRNA Expression analysis

The expression of known and unique miRNAs in each sample was calculated based on read count and used to determine the normalised expression, these are statistically analysed and normalised by TPM using a method shown in [Zhou et al., 201]. The normalised expression = (read count*1,000,000)/libsize. Libsize is the sample miRNA read count. The results for these are shown in [table 5.11]

Table 5.11 Result table showing a sample of 10 novel miRNAs:

This sample of 10 novel miRNA shows the differences in read counts for each of the infected and plant samples, showing the TPM for each infected and plant sample. Indicating some miRNA with no expression in the infected compared to the plant such as novel_102 and some with higher expression in the infected such as novel_101

sRNA.readcount	INF1.tpm	INF2.tpm	INF4.tpm	P3.tpm	P4.tpm	P5.tpm
novel_1	3852.94563102595	5206.1072327492	3394.34095476022	296.495740878749	252.962471611989	584.203892904892
novel_10	14.9790542890858	35.6176093460607	62.0679488870441	152.254569640439	67.4566590965305	143.656694976613
novel_100	0	29.6813411217172	65.9471956924843	72.1205856191552	44.9711060643537	53.6318327912687
novel_101	855.470433843342	979.484257016669	0	0	0	0
novel_102	0	0	0	48.0803904127702	28.106941290221	7.66169039875268
novel_103	6.65735746181589	17.8088046730303	50.4302084707233	112.187577629797	33.7283295482652	30.6467615950107
novel_105	0	5.93626822434345	0	8.01339840212836	16.8641647741326	22.985071196258
novel_106	0	0	0	0	0	17.2388033971935

novel_107	19.9720723854477	29.6813411217172	100.860416941447	3397.68092250242	4592.67420682212	430.970084929838
novel_108	9.98603619272384	0	0	16.0267968042567	11.2427765160884	24.9004937959462

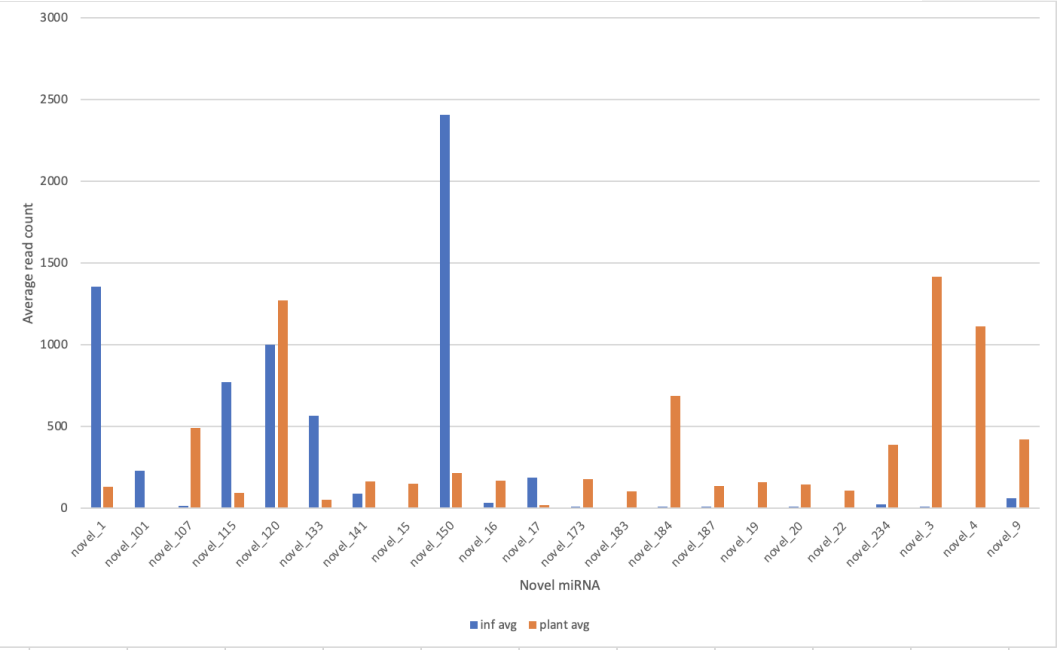


Figure 5.12. Novel miRNAs showing notable differences in expression:

Of the novel miRNAs, an average read count was taken between infected and non-infected, most of the novel miRNA showed very low average expression so of the novel miRNA the ones with average read counts above 100 were plotted on this graph. The graph shows that multiple novel miRNAs show upregulation during infection which many only show expressions during infection.

5.3.12.2 miRNA TPM distribution

TPM density distribution can represent the gene expression mode of samples. Both FPKM and TPM are methods for normalizing data for use in sRNA and mRNA analysis [Zhao et al., 2020]. The statistical meanings of these two methods however are different. FPKM (the known number of Fragments Per Kilobase transcript sequence per Millions base pairs

sequenced) taking the effect of sequencing depth and gene length on fragment counts into account. TPM (transcripts per million reads) = (read-Count*1,000,000)/total read-Count.

TPM does not need to consider the length of s RNA, because the sRNA fragment obtained by sequencing is a complete sRNA for the purpose of calculation. In the transcriptome, a read may be a part of a gene, so the expression level of the genes needs to be normalized by the length of the genes. Therefore, TPM in this case is the method used for calculating expression level in miRNA/sRNA and normalising the data.

TPM density can be observed in [figure 5.13] with the density curve shifting from infected and non-infecting, with the non-infected showing a peak in TPM density at 1.5 and 2 log 10TPM+1 while the infected peaks at 0.1-05 indicating a significant change in miRNA density distribution during infection.

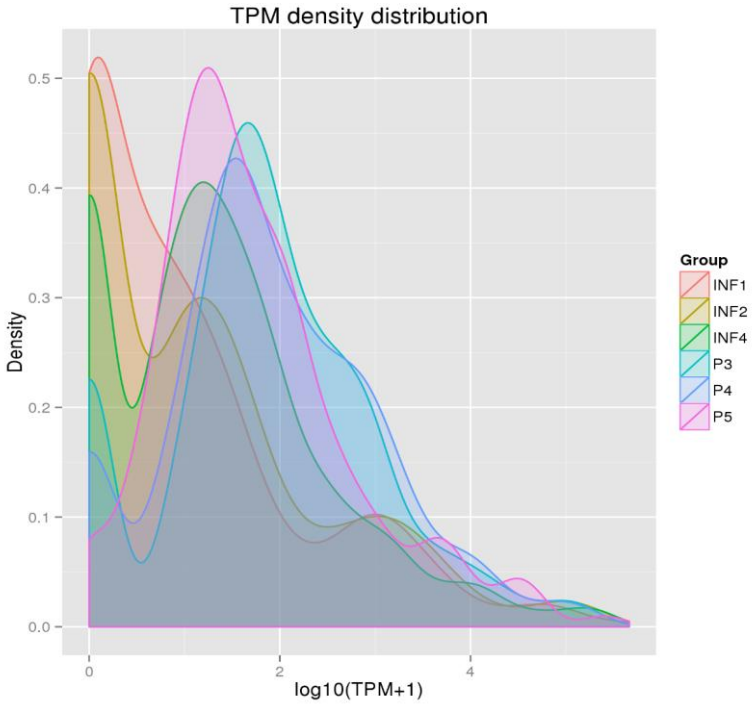


Figure 5.13: TPM distribution:

The X axis shows sample density with colours depicting the sample names and y axis is the value of miRNA log_{10}(TPM+1).

5.3.13 RNA-Seq Correlation

Performing correlation analysis requires biological replicates. These can be used to determine gene expression levels between samples and if these changes play an important role, and whether this change is statistically significant. This method is used to determine correlation coefficient, the closer correlation coefficient is to 1, the more similar the samples are to each other. The square of the Pearson correlation coefficient should be larger than 0.92 and the R2 should be at least larger 0.8. this correlation is showed in [figure 5.14]

First Classical Pearson's correlation tests were done to verify the relationships between expression profiles of miRNA and mRNA.

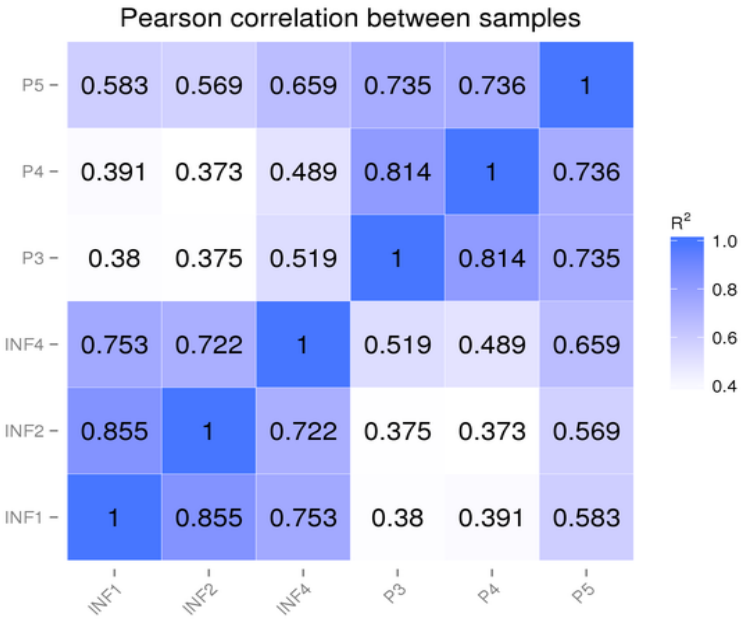


Figure 5.14 RNA-Seq Correlation:

The x and y axis represent the names of the samples and the values and colours represent the $\log_{10}$(TPM+1) (R^2):Pearson RSQ; Rho: spearman coefficient of association. Where each sample compared to itself gives a value of 1 for R^2.

5.3.14.1 Differential expression

Differential expression is calculated using the miRNA read counts, samples and their replicates are analysed using DESeq2 based on methods in [Michael et al., 2014] this was first done to generate the correlation cluster. These results are shown in [table 5.12] where the average of the infected and plant alone differential miRNA expression is shown for 5 miRNA.

Table 5.12 Different miRNA expressions result

miRNA expression fold change was calculated, and this table represents a sample of 5 miRNA and their respective p values comparing the infected to the plant samples. Here showing 3 sly-miR168a-5p, Novel _1 and sly-miR482b with positive values indicating up regulation.

sRNA	infected_readcount	p_alone_readcount	log2FoldChange	pval	padj
sly-miR6024	27.1792644766253	868.536467759299	-4.732	4.2224e-17	1.2625e-14
sly-miR482b	45116.3265222699	655.74310579307	5.6119	1.7304e-16	2.587e-14
novel_3	14.2975013854648	666.185673420298	-5.0865	1.4056e-13	1.4009e-11
novel_1	2094.0773558291	57.2448190893237	4.8001	2.7074e-13	2.0238e-11
sly-miR168a-5p	10362.695131155	272.811339053146	4.8415	4.5173e-13	2.7013e-11

sRNA column corresponds to the miRNA mature id. read count values of sample1 after normalised read count values of sample2 after normalised. $\log_2$.Fold_change. which is calculated by $\log_2$(Sample1/Sample2). the p.value in the hypergeometric test. q.value which is the p.value after being normalised.

5.3.14.2 Filtering the Different Expression miRNA

A volcano plot infers the overall distribution of different miRNA expressions. For an experiment with no biological replicate, the threshold is normally set as follows: $|\log_2(\text{FoldChange})| > 1$ and qvalue < 0.01. this differential expression is shown in [figure 5.15] as a volcano plot.

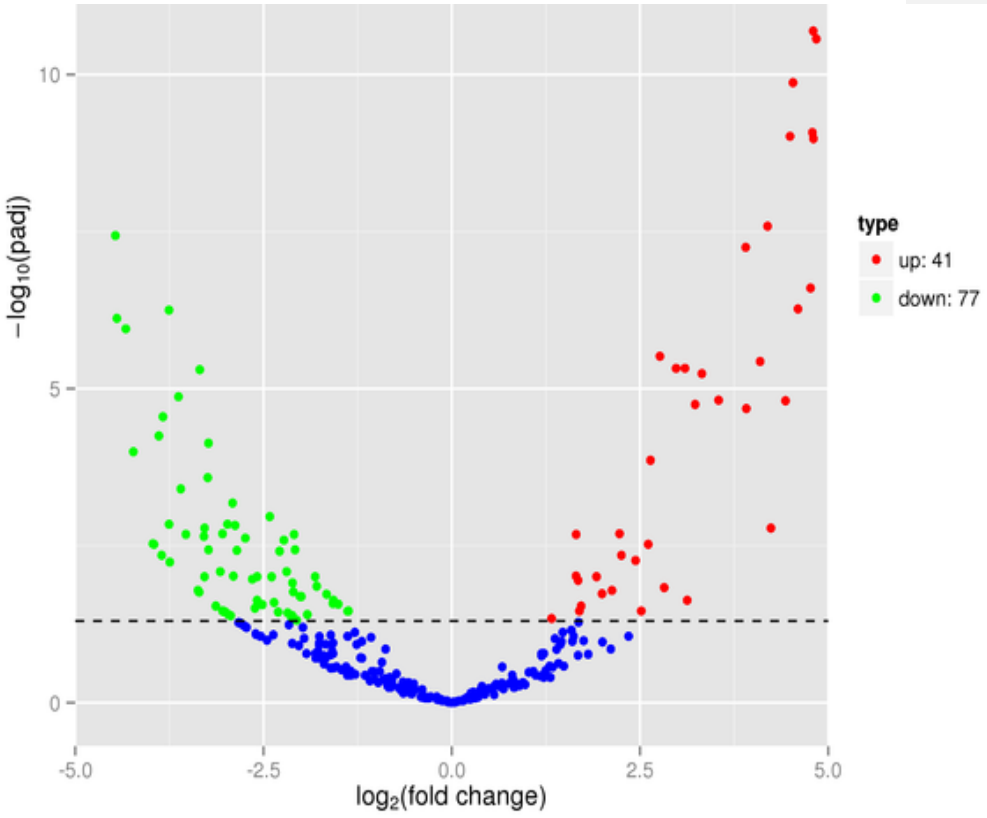


Figure 5.15 Volcano Plot.

For this diagram the x-axis denotes the fold change observed in miRNA expression between various samples, and the y-axis is used to show the statistical significance of the differences. The red dots show samples which are significantly different.

5.3.14.3 Cluster Analysis of the Differences between miRNAs Expressions

Cluster analysis is used to find miRNA expression patterns under a variety of experiment conditions. By clustering miRNAs with similar expression patterns, it is possible to recognize unknown functions of miRNAs and/or the function of unknown miRNAs.
In hierarchical clustering, differently coloured areas represent different groups of the cluster. miRNAs within each group may have similar functions or take part in the same biological process. In addition to the TPM cluster, K-means and SOM were also used to cluster the log2(ratios). miRNAs within the same cluster have the same changing trend in expression levels under different conditions.

Clustering based on the expression level of s RNA. We calculate log10 (TPM+1) for expression level to determine the clustering pattern of differential miRNA expression under different experimental conditions. Clustering uses the pheatmap package in R. The relative expression level in the union of differential miRNAs is clustered with log2 using the distance algorithm. The distance between each miRNA is calculated to obtain the relative distance by iterative iteration. H-cluster, K-means, and SOM are all clustering methods.

Cluster analysis is used to evaluate the expression patterns of differential miRNAs under different experimental conditions. The cluster can aggregate miRNAs with the same or similar expression patterns to identify unknown miRNA functions or unknown functions of known miRNAs. These miRNAs may share similar features.

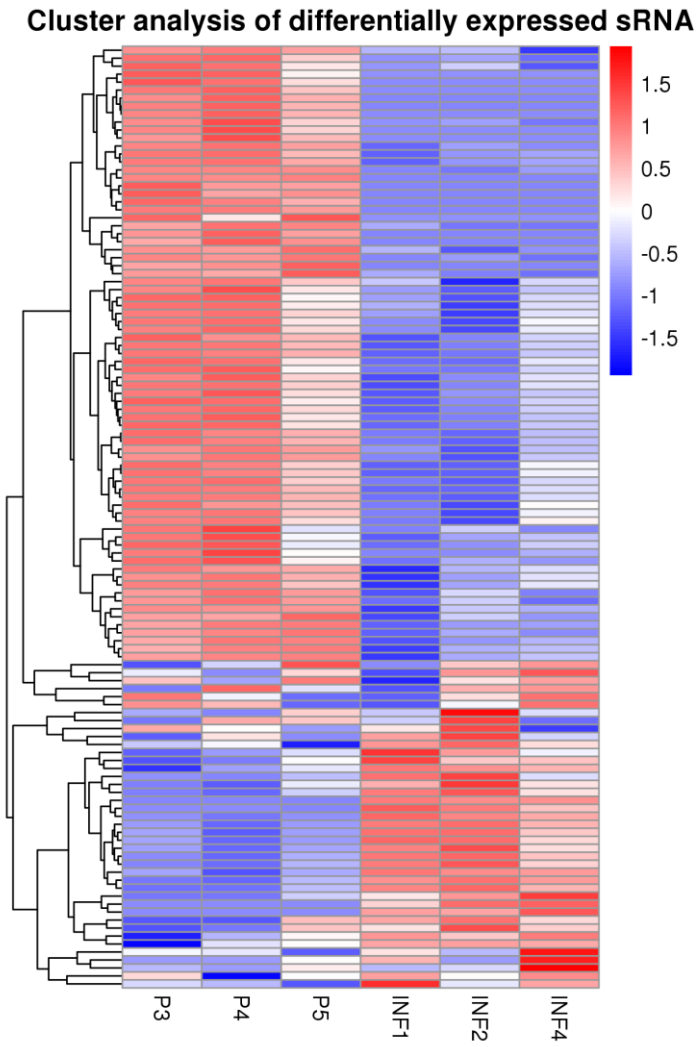


Figure 5.16 Cluster Analysis.

TPM cluster analysis is done by clustering the samples by their log10(TPM+1) value, the red samples represent the miRNAs with a high expression level, while the blue represents miRNAs with low expression level.

5.3.15 Target Gene Prediction for Known and Novel miRNAs

The target genes of known and novel miRNAs are predicted, and the relationships between miRNAs and the corresponding target genes are found. The results shown in Table 5.13

Table 5.13 Predicted target of 10 novel miRNAs

miRNA name	predicted target
novel_101	Solyc06g050760.1
novel_103	Solyc11g066330.1
novel_105	Solyc02g078210.3
novel_106	Solyc05g013930.2
novel_106	Solyc06g009490.3
novel_106	Solyc09g008010.3
novel_108	Solyc01g006380.2
novel_108	Solyc01g087240.3
novel_108	Solyc01g098520.3
novel_108	Solyc01g099200.3

5.3.16.1 Gene Ontology Enrichment analysis

Gene Ontology is a form of bioinformatic classification system, this allows the samples to unify gene properties across a variety of species. GO includes three main branches, the cellular compares, the molecular function, and biological process. It has been established that GO terms with padj < 0.05 represent significant enrichment. The results of the GO enrichment are shown in [figure 5.17] and [table 5.14]

GO enrichment analysis provides target gene candidates of potential known and novel miRNAs, this can be used to reference background and biological functions. The results can be used to reveal the functions related to the predicted target gene candidates of known and novel miRNAs. This method of calculating gene numbers [Young et al., 2010] starts with mapping all target gene candidates to GO terms in the database.

Table 5.14 GO Enrichment Result

GO accession	Description	Term type	Over_represented_pValue
GO:0043531	ADP binding	molecular_function	1.3581e-102
GO:0032559	adenyl ribonucleotide binding	molecular_function	7.8964e-37
GO:0030554	adenyl nucleotide binding	molecular_function	2.4057e-36
GO:0001883	purine nucleoside binding	molecular_function	8.2891e-36
GO:0032549	ribonucleoside binding	molecular_function	8.2891e-36

The GO accession number corresponds to the Gene Ontology entry, this is followed by its description, the Detailed description of Gene Ontology, its Term type which constitutes GO types, over represented pValue, its Corrected pValue for which GO with Corrected P-value < 0.05 are significantly enriched in DEGs, CAD item this corresponds to the number of target gene candidates related to this term and the CAD list which is the number of target gene candidates with GO Annotation, the Bg item which is the number of reference genes related to this test, and the Bg list which is the number of all genes in GO.

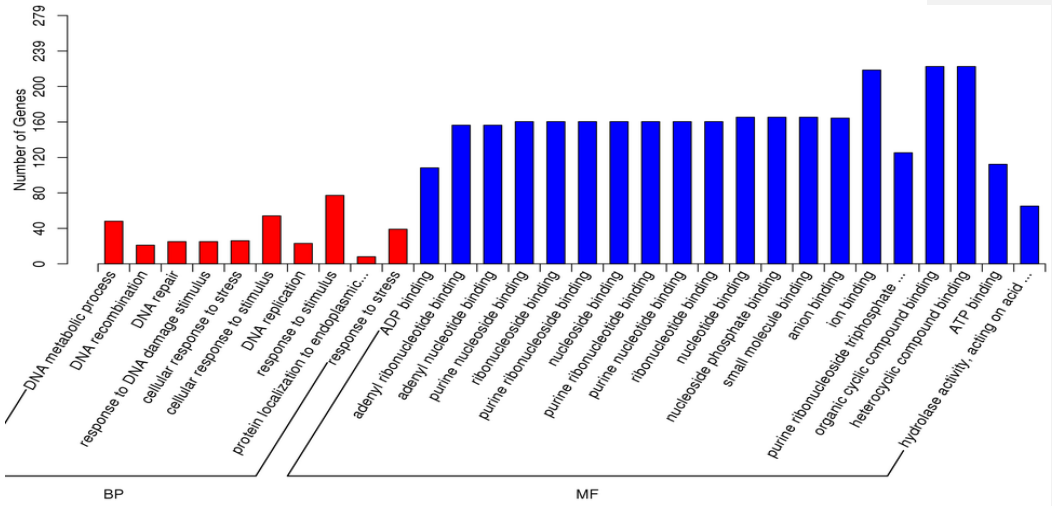


Figure 5.17 The histogram of target candidate genes:

The x axis shows the 3 GO ontologies' next GO term, the y axis shows the number and percentage of target gene candidates annotated in this GO term.

The Directed Acyclic Graph (DAG) is a graph used to visualise the GO enrichment; on this graph the different branches represent the inclusion of two GO terms. In normal cases, the top 10 results from GO enrichment are selected as main nodes in directed acyclic graphs, where the associated terms are also represented, and the depth of colours indicates enrichment level. DAGs for a biological process, molecular function and cellular component are shown respectively.

5.3.16.2 KEGG Pathway Analysis

The KEG (Kyoto Encyclopaedia of Genes and Genomes) is a collection of curated databases related to information including genomes, pathways, diseases, drugs, and chemical substances. KEGG enrichment collection is mainly utilised for bioinformatic analysis.

This analysis resulted in corrected p-values for each of the miRNA and their respective pathways determining if they are statistically significantly enriched in the infected samples denote 5 miRNA and their respective pathways which are shown to not be enriched in the

infected samples. Table 5.15 shows an example of 5 miRNA which are enticed in the infected samples with a corrected p-value less than 0.05

Table 5.15 Pathway annotation result

Term	ID	Sample number	Background number	P-Value	Corrected P-Value
Monobactam biosynthesis	sly00261	2	10	0.0144364361771	0.548405166671
Other glycan degradation	sly00511	2	14	0.0251938776484	0.548405166671
Phosphatidylinositol signalling system	sly04070	4	73	0.0323889440416	0.548405166671
Selenocompound metabolism	sly00450	2	17	0.0348193756616	0.548405166671
ABC transporters	sly02010	2	24	0.0616470302962	0.677387138297

The term column denotes the Description of this KEGG pathway followed by the Unique ID of this pathway in the KEGG database, the Sample number which is the Number of target genes related to this pathway, the Background number which is the Number of reference genes related to this pathway, the P-value: P-value generated by a hypergeometric test and finally the Corrected P-value which determines statistical significance, Corrected P-value smaller than 0.05 are considered as significantly enriched in target gene candidates.

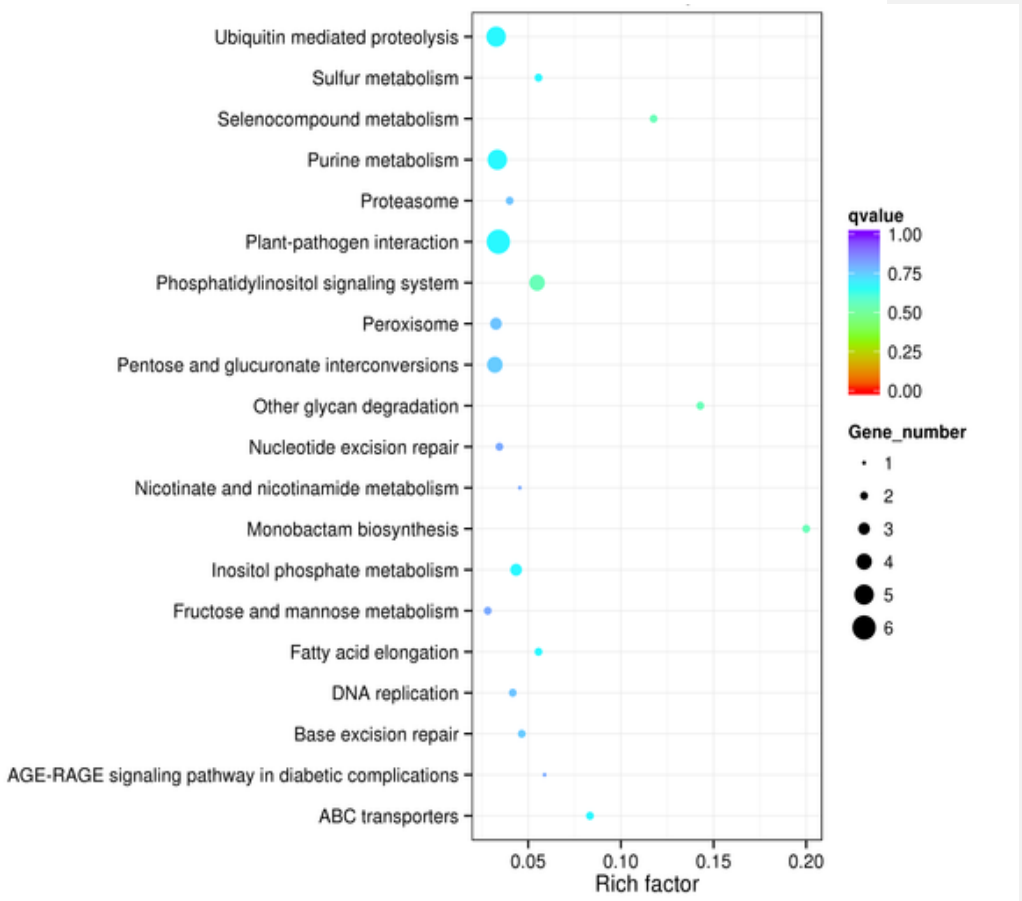


Figure 5.18 KEGG enrichment scatter plot of DEGs.

In this plotted graph the y-axis denotes the name of the pathway, and here the x-axis shows the Rich factor. Dot size is represented by the number of target genes, and the colour indicates the q-value. This scatter plot shows the 20 pathways for which enrichment of miRNA can be observed from the KEGG enrichment calculations.

5.3.16.3 Glycan degradation pathway enriched with miRNA

One of the pathways which showed the most significantly significant enrichment of miRNA is the glycan degradation pathway. The diagram of this pathway is shown in [Figure 5.19] and the map of enrichment shown in [5.18] with glycan degradation showing one of the highest rich factors, including other pathways such as monobactam synthesis and Selenocompound metabolism.

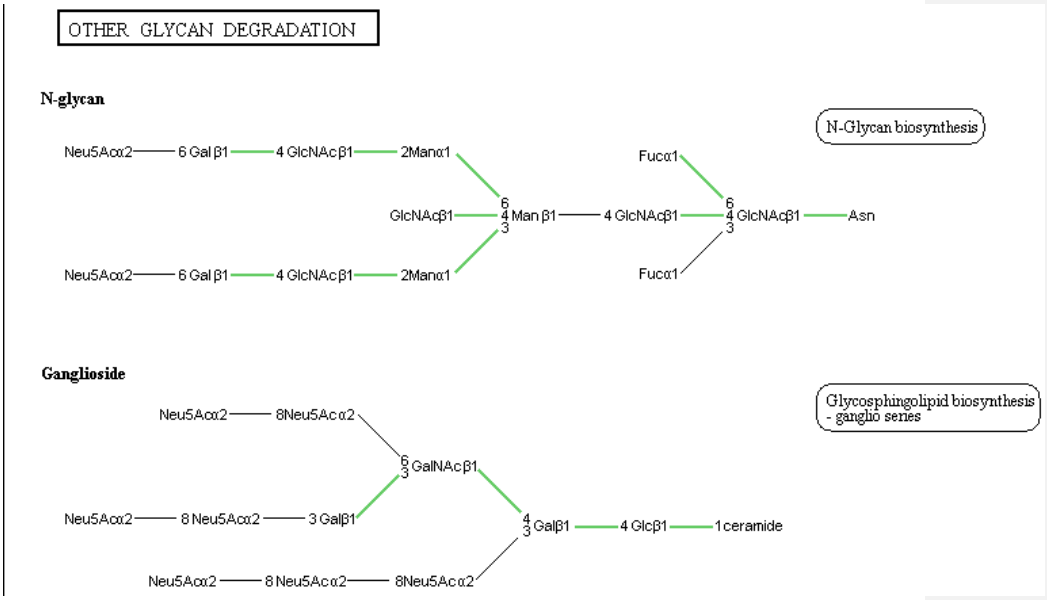


Figure 5.19 Metabolic map of target genes in the Glycogen degradation pathway:

This diagram of Glycan degradation pathways enriched with miRNAs during infection, protein glycosylation is an essential post translational modification in membrane proteins present in eukaryotic organism, a process which is highly conserved in plants, mammals, and dimorphic fungi [Chaliha. C. et al,. 2018] N-glycan functions in plants are not well understood, so the presence of miRNA degradation of genes expressed in the n-glycan degradation pathway may indicate a role of N-Glycan in plant disease

defence.

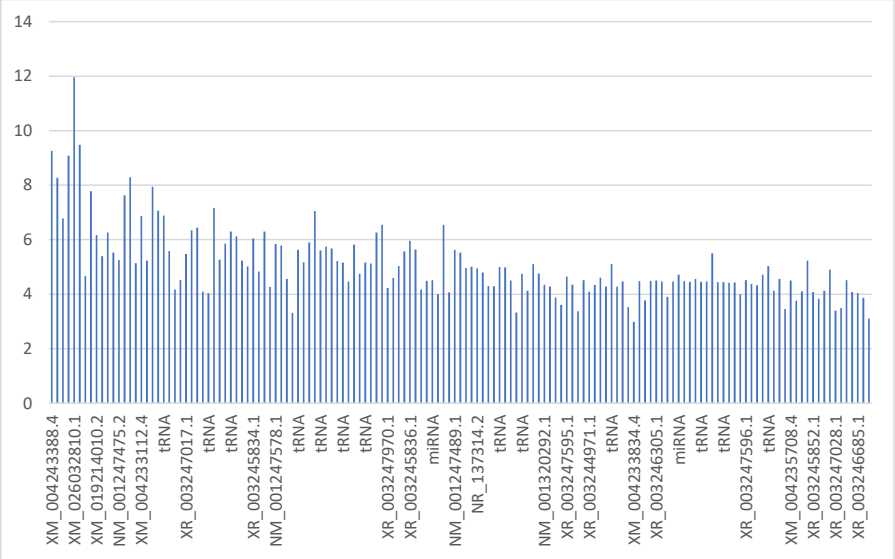


Figure 5.20 : sRNAs from the plant and infected samples mapped to the plant genome and filtered for upregulation during infection:

In addition to miRNA analysis sRNAs were mapped to the plant genome using Geneious. These were then filtered for absolute certainty at above 6, mean base read count above 100 and log2 ratio greater than 1 to filter sRNA read count greater in the infected, and a high enough read count to be considered significant. These sRNAs were aligned to the fungal genome and analysed using TAPIR.1 for potential targets.

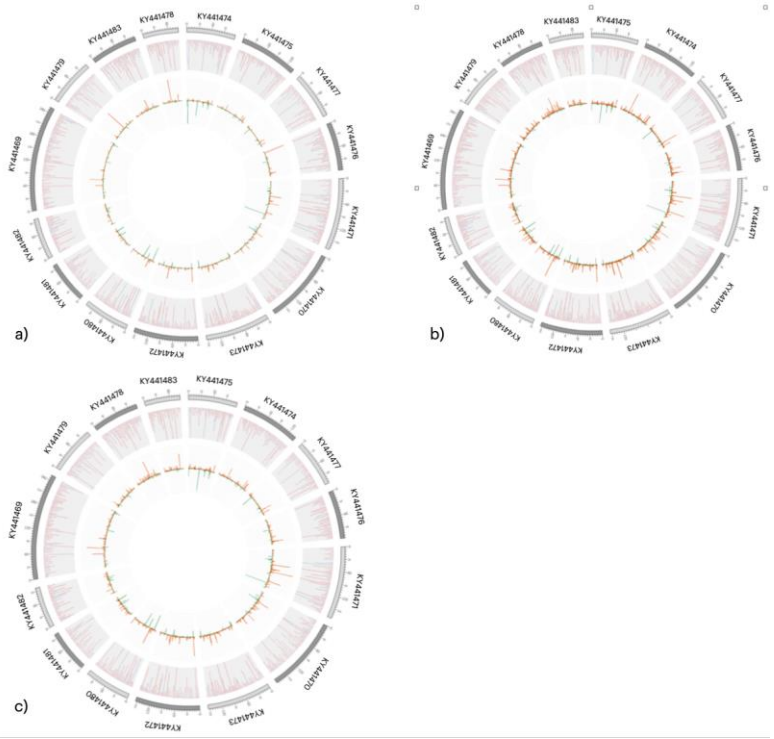
5.3.17 *A. alternata* and infected tomato tissue sRNA analysis

sRNA reads are mapped to the genome using Bowtie total mapped reads are listed on [table 5.16] demonstrated in graphical form in [figure 5.21], to analyse their expression level and distribution on genome. Density of small RNA reads on each chromosome is determined by each sample. Circus is used to graphically demonstrate distribution of reads on each chromosome. The longest 10 contigs or scaffolds were chosen for analysis

Table 5.16 Statistics of mapping results:

Sample Iditol sRNA: Number of total sRNA after length filtering, Mapped sRNA: Number and percentage of sRNA mapped to genome, + Mapped sRNA: Number and percentage of mapped sRNA in the same direction as the genome, – Mapped sRNA: Number and percentage of mapped sRNA in the opposite direction to the genome

Sample	Total sRNA	Mapped sRNA	"+" Mapped sRNA	"-" Mapped sRNA
INF2	11883493 (100.00%)	207401 (1.75%)	122495 (1.03%)	84906 (0.71%)
INF4	11763345 (100.00%)	259533 (2.21%)	213571 (1.82%)	45962 (0.39%)
INF1	18749522 (100.00%)	737776 (3.93%)	494842 (2.64%)	242934 (1.30%)
F9	25082088 (100.00%)	1476808 (5.89%)	1067832 (4.26%)	408976 (1.63%)
F11	28567126 (100.00%)	2409137 (8.43%)	1782893 (6.24%)	626244 (2.19%)
F10	27137999 (100.00%)	3804695 (14.02%)	3020945 (11.13%)	783750 (2.89%)


a)
b)
c)

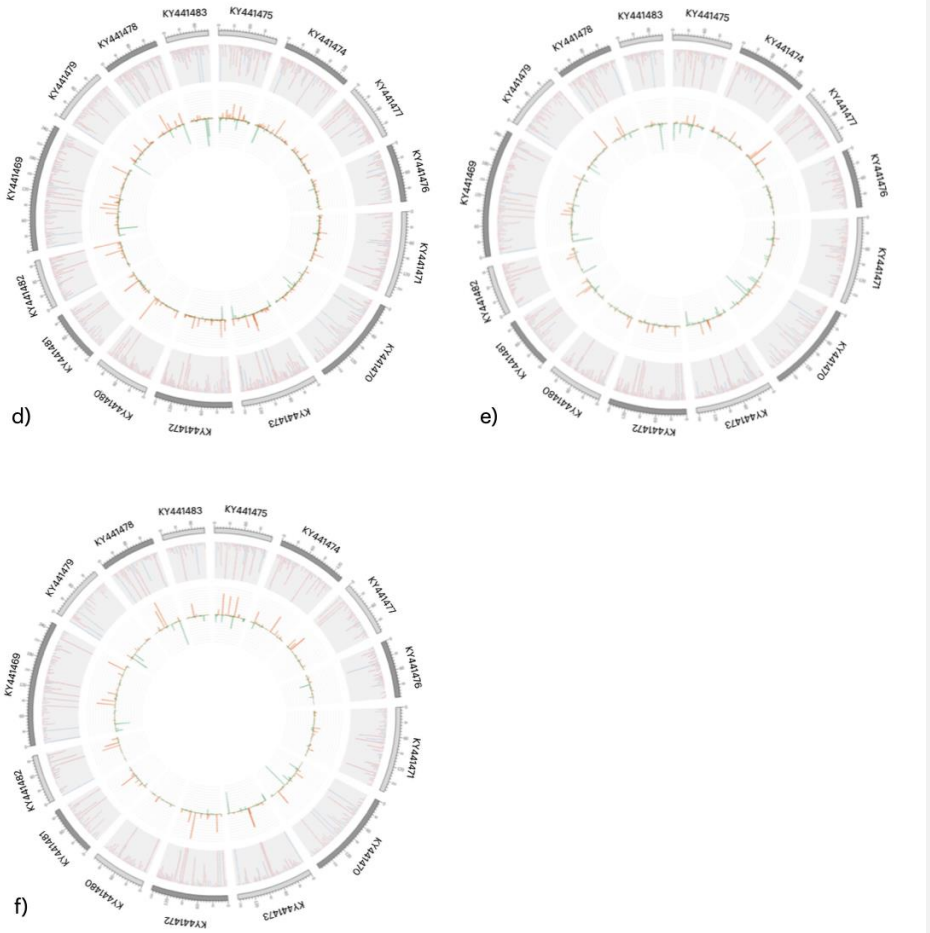


Figure 5.21 Read distribution per chromosome. The chromosome is shown as the outer circle. Grey background in the middle area shows the distribution of 10,000 reads on the chromosome. Red represents the number of sRNAs on the sense strand of the chromosome, and blue represents the number of sRNAs on the antisense strand. All reads are shown in the centre area of the circle. Yellow represents the number of sRNAs on the sense strand of the chromosome, and green represents the number of sRNAs on the antisense strand. Figures a) b) and c) correspond to F9, 11 and 10 respectively and d) e) and f) correspond to INF 1, 2 and 4 respectively.

5.3.18 Mapping infected and fungal sRNAs to the *A. alternata* genome

To analysis the sRNA library of the fungi without any information present in the plant genome a simpler method had to be applied. For this the sRNAs were mapped to the reference genome using Geneious and this provided contigs with the mapped sRNA showing as peaks at specific locations. This was then calculated in using Geneious using deseq to calculate the read quantity to annotated genes in the genome. These were then compared for expression to identify genes with sRNAs mapped to it at higher rates in the infected than the *A. alternata* tissue alone. This resulted in the following table of genes where sRNA read counts were mapped to them, had a larger than 100 mean read count, higher than 6 absolute confidence and greater than 1 log 2 ratios. This filtered for genes with an RNA accumulation which may indicate sRNA of interest. These sRNA that maps out the plant genome may indicate fungal sRNA that are being expressed to silence genes in the plant genome.

Table 5.17 sRNA mapped to the fungi and infected genome calculated for differential expression

Sequence Name	Name	Length	Differential Expression Absolute Confidence	Differential Expression Log2 Ratio	Differential Expression p-value
NW_017306215.1	XM_018526915.1	1021	33.29	13.12827162	9.71E-37
NW_017306226.1	XM_018530071.1	1341	30.52	16.28536416	1.34E-33
NW_017306227.1	XM_018532387.1	1064	13.29	8.928651139	6.85E-16
NW_017306200.1	XM_018530146.1	1657	13.22	12.92193984	8.52E-16
NW_017306211.1	XM_018531507.1	3317	12.72	10.64836217	2.91E-15
NW_017306211.1	XM_018529948.1	1277	11.96	10.89931556	1.79E-14
NW_017306203.1	tRNA	71	11.05	10.61755425	1.59E-13
NW_017306195.1	XM_018526525.1	1831	10.84	9.072117612	2.77E-13
NW_017306212.1	XM_018527758.1	1575	10.7	9.34837949	3.91E-13
NW_017306198.1	XM_018533397.1	4270	9.38	11.18369469	9.59E-12

NW_017306198.1	XM_018533444.1	2111	8.57	7.745270302	6.87E-11
NW_017306201.1	tRNA	72	8.07	8.697955304	2.46E-10
NW_017306192.1	XM_018532116.1	4999	7.01	7.149438306	2.98E-09
NW_017306225.1	XM_018532357.1	1725	6.79	9.361562772	5.10E-09
NW_017306190.1	XM_018530756.1	2109	6.65	10.50521347	7.10E-09
NW_017306191.1	XM_018524221.1	1960	6.62	9.087730845	7.75E-09
NW_017306214.1	XM_018531715.1	1421	6.55	8.973415025	9.31E-09
NW_017306192.1	XM_018532073.1	1242	6.4	8.717295927	1.34E-08
NW_017306230.1	XM_018532674.1	941	6.3	10.34896267	1.74E-08

Differential expression was calculated using log2 ration comparing the normalised read counts of the srRNA mapped to that region. This is demonstrated in the graph in [figure 5.17] showing the gene regions with sRNA mapped to them and statistically significant increase to the *A. alternata* alone. These were also used to generate a volcano plot which demonstrates the overall number of mapped gens with sRNA and how these distribute compared between the fungi and the infected. Showing the relatively few numbers in the fungi which are statistically significantly over expressed during infection.

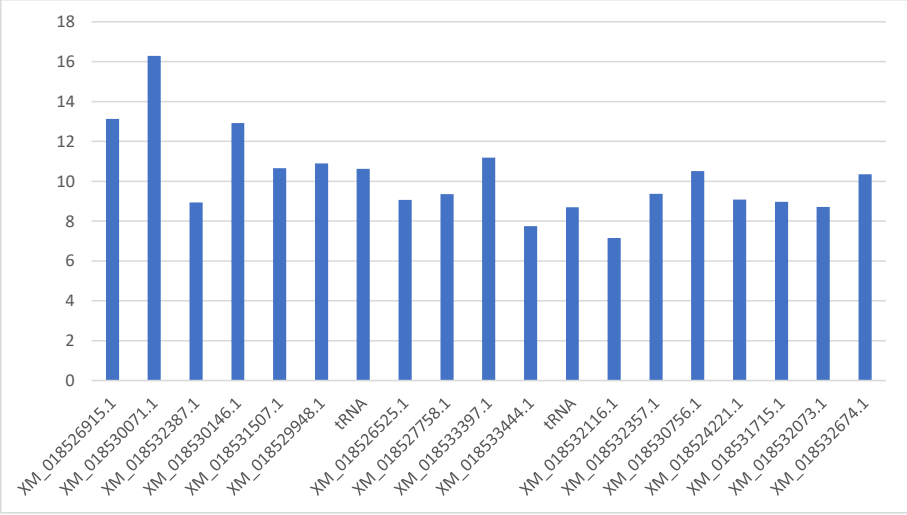


Figure 5.22: Genes for which mapped sRNAs showed upregulation during infection.

Figure plotting all genes with sRNA with absolute confidence above 0, listing the genes which showed significantly higher sRNA accumulation and so are likely to be down regulated.

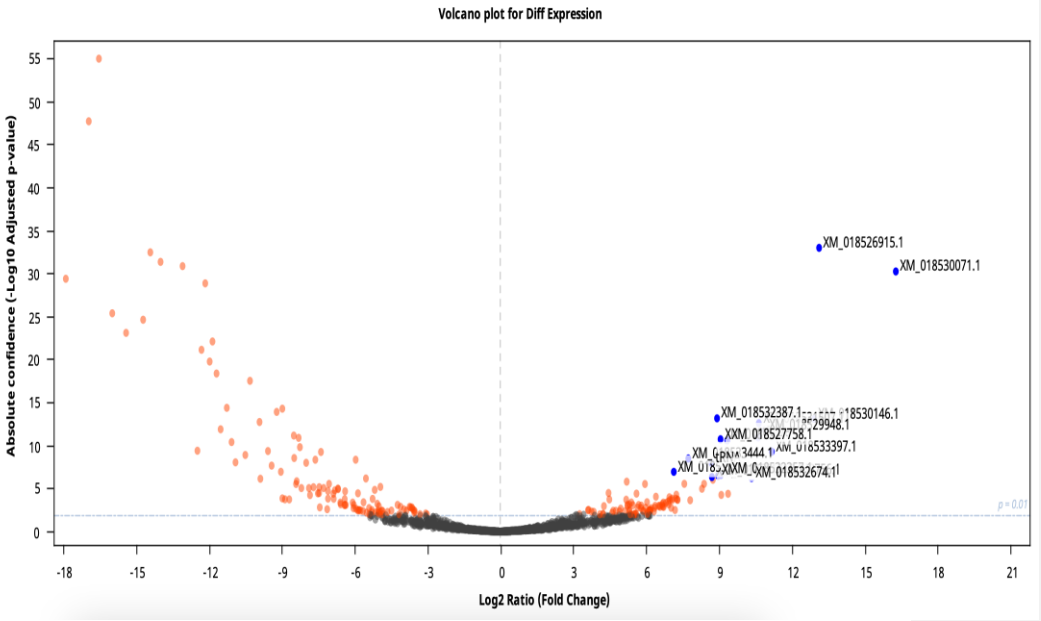


Figure 5.23 Volcano plot of the sRNAs mapped to the *A. alternata* genome shows upregulated sRNA mapped to exons.

The volcano plot compares sRNAs by their absolute confidence and Log2 Ratio, the fold change determines the degree of up or down regulation with possive numbers being upregulated and negative down regulated and the Absolute confidence denoting p value confidence value above 2 equates to a p value of less than 0.01.

5.3.19 Tapir1.1 analysis

Tapir is a software publicly available for the prediction of miRNA targets that can also be used to predict targets in any genome, rather than just in plants. However, the system is designed for plants, and as there is no known parameters for miRNAs targeting fungi such as *A.alternata.* I chose to use set parameters based on previous work in [weiburg et al,. 2012]. The target prediction parameters allowed no gap within the alignment, the 10th nucleotide

must match, at most 2 mismatches were permitted, and a score of 4.5 within the Tapir.1 algorithm.

The novel miRNAs shown in Table Table 5.11.1 were put into the Tapir system to detect potential targets in the fungi, in this case finding no results. However, the miRNAs were aligned to the genome. In this case 130 miRNA were found to align to the fungal genome, however in some cases with too many mismatches to be viable.

Table 5.18: Plant miRNAs showing to have potential fungal targets.
miRNAs which matched the s=desired qualities, the read count was above 100, the read count showed significant increase in the infected over the plant alone and the mapped miRNA to the fungal genome showed a high enough degree of match suggest potential silencing ability. This narrowed down the miRNAs to 6 candidates. These targets were evaluated for their function using the Ensembl fungal database, showing that of the six, two were found to have known conserved gene function, one with Catherin heavy chain domains and the other with a domain for vascular transport chaperone 4.

Table 5.18: miRNA with predicted targets in the fungi genome.

plant miRNA name	miRNA sequence	predicted fungal target	gene domains
sly-miR159b	UUGGAAAGAAGGGA-GCUCUAC	CC77DRAFT_1057609 hypothetical protein	Clathrin, heavy chain/VPS, 7-fold repeat
sly-miR9473-5p	UGGCUGUAAAUCUAA-ACUCGU	CC77DRAFT_1097266	unknown
novel_166	UAGCAUGCCAUGUAGAACACGUGU	none	none
sly-miR399b	CAUGCCAAUAGAGAGUAGCCC	CC77DRAFT_1030436	unknown
sly-miR159	UUUGGAUUGAAGGGAGCUCUA	AA0111_g11549 gene	Vasculor trasnport chaparone 4
novel_112	ACAUAGUGCUACAUAGACCGAAAA	none	none

5.4. Discussion

Total RNA extraction, sRNA sequencing and genomic analysis were performed to better understand the sRNA interactions taking place within the *A.alternata*-tomato pathosystem. With a goal towards identifying novel miRNAs, which might play a role in infection, whether that is by cross kingdom silencing of gene regulation. This was done using the approach outlined in this chapter beginning with total RNA extraction and sequencing, followed by mapping to the genome and using the miRNA database to identify novel and known miRNAs from the samples and analyse their expression. For the fungi, sRNAs were analysed by mapping to the genome reference for *A. alternata* and the gene expression calculated based on read count and upregulated expression during infection.

5.4.1 sRNA reads mapped to the plant genome from the infected and non-infected plant tissue showed varied changes between conditions.

The mapping of sRNAs to the plant genome allowed for both identification of the miRNA as well as other forms of sRNA such as trans-acting sRNA and repeat elements. This also gave useful information about the overall change of the sRNAomes during infection. However, it is important to note that total sRNA read counts are unreliable as a direct measure of sRNA activity, as many reads that align to exons may in fact be the result of degraded mRNA. The sRNA libraries for the non-infected tomato tissue, infected and fungi alone were mapped to both the tomato and *A. alternata* genome. The result showed how little of the sRNA sequenced mapped to the genome. For the Tomato sRNA 80% of the reads mapped to the tomato genome while the infected sRNA showed 70% mapping to the tomato genome. while the fungi sRNA reads showed less than 14% of the sRNA mapped to the fungal genome, and for the infected sRNA reads less than 4%. Showing significant difference, meaning that the vast majority of the sRNAs likely originate from the tomato.

5.4.2 Differential expression of multiple miRNAs

miRNA analysis for the plant samples was very successful, identifying many novel miRNAs in addition to known miRNAs and learning much about the structure and function of the miRNA and their role in infection. The first analysis looked at the first bases of the sRNA in each library. Including both the novel miRNA and the known miRNA, this showed that the first base changes significantly between infection and non-infection. As the first base often plays a role in mIRNA target this means there is likely a change in which miRNA are beign expressed and which genes targeted during infection.

In addition, we identified the presence of extensive post-transcriptional RNA base edits allowing for the plant to tailor miRNA by altering the bases at essential locations to change its potential targets. This is a potential future study to look in detail at the effects of these base edits on the target genes and whether the increase in base edits in infected tissue is an active response to infection.

For the known miRNA of the total 14000 unique sRNA, 105 miRNA hairpins were identified and over 134 mature miRNAs produced from them. In addition, a total of 230 novel miRNA were identified by predicting miRNA structures using miRNA prediction software.

5.4.3 Differentially expressed miRNAs target and regulate multiple gene pathways during *A.alternata* infection of tomato

Using gene target predictions software and multiple methods of target expression and enrichment calculation methods. These methods include DAG and KEGG distribution analysis. We were able to identify the target genes and those which were upregulated in the infected tissue.

The results from this give a picture of multiple genes which are targeted by miRNA during infection and are like down regulated as part of the plant defence response. The known and novel miRNA once identified were analysed for which nucleotide is most distributed at the

first base pair and at every base pair respectively. These metabolic pathways include glycan degradation [figure 5.19], monobactam biosynthesis, and selenocompound metabolism. N-glycans are known to play a role in defence against filamentous fungi [Chaliha et al., 2018]. N-glycans can attach to the plasma membrane PAMP recognizing receptors, these make up the first early recognition system for plants during a fungal infection. The role of these N-Glycans was looked at in [Beihammer.G et al,. 2021] in Arabidopsis where N-glycosylation mutants were evaluated for their ability to detect flagellin and bacterial elongation factor. This means it makes sense that N-glycan degradation pathways are downregulated during infection. This is likely to facilitate more availably of N-glycans for PAMP detection begins as results can show that a single N-glycan can be essential for PAMP recognition at the cell surface and that this down regulation of glycan degradation is done by miRNA.

Monobactam biosynthesis relates to defence against bacterial infection, monobactams are a naturally produced beta-lactam antibiotic. Synthesis of nocardicin A, a monobactam produced in tomato is done by processing L-4-hydroxyphenylglycine (L-pHPG), L-arginine and L-serine. In this pathway ATP sulfurylase 1 is down regulated by miRNA during infection. In the case of filamentous fungi, it means down regulating production of a natural anti-biotic .Seleno Compounds are metallic compounds of the element selenium, the role of selenocompounds in plants is poorly understood as most of the research focussing on them relates to their potential as a cancer treatment [Bartolini et al., 2017], however one observable role they may play is as an antioxidant [Tapiero et al., 2003] and as reactive oxygen species play a significant role in plant immunity [Muhammad et al., 2018].

5.4.4 Identification of *A.alternata*-targeting plant miRNAs

From the miRNA identified in the plant there were 6 miRNAs with all the required characterises, of these two were found to have potential gene targets with known functions. The other 4 had predicted genes which encode hypothetical proteins but very little was understood and available on their function and role in infection. As for the two genes with targets one of them has domains for Catherin a protein with multiple research papers looking into its role in infection. Catherin is a protein which requires 3 heavy chain associated with 3 light chains. The role of this protein is intake of extracellular and membrane compounds.

Mutants have been generated to study the role of Cathrin, such as in T*rypanosoma brucei* where depletion of Catherin using antisense RNA led to the parasite being non-viable [Allen et al., 2003]. *Toxoplasma gondii* has also been studied showing that mutants were able to invade cells effectively but upon invasion showed unusual development.

Within fungi, cryptococcus neoformans that had been mutated in CHC was unable to uptake haemoglobin and showed deficiency in the production of virulence factors [Bairwa et al., 2019] however the role it plays in filamentous fungi is poorly understood, RNAi of CHC in *S. sclerotiorum* showed that the protein plays a role in endocytosis of dsRNA [Wytinck et al., 2020] which I have discussed have been shown to play a role in infection but not confirmed with *Alternaria* and tomato.

Another proposed fungal target is a vascular transporter chaperone. these proteins are not well understood in filamentous fungi, with almost no papers discussing their role in the *Alternaria* Spp or similar fungi. in *Saccharomyces cerevisiae* it has been studied, they are high molecular weight membrane proteins, which are present in the vacuole membrane. These have several roles related to ATPases, endocytosis, ER-Golgi trafficking, and endocytosis [Tomashevsky et al., 2020]. The VTC4 gene is highly conserved in fungi, performing the role of polyP synthesis in yeast VTC4 is part of the vacuole transporter Chaperone complex which consists of 5 subunits, with VTC4 having homologs in many other species of fungi [Gomes-Vieira et al., 2018].

Chapter 6

Genomic analysis of *A.alternata*

6.1 Introduction

Over the last 10 years there has been an incredible advance in the field of bioinformatics and genomics, with the advances in genome sequencing and bioinformatics [M Younus Wani et al,. 2018]. The power of these systems to generate vast quantities of data is incredible, and it has developed new fields in handling and making sense of this data. With biological data existing in multiple levels, Genomic, RNA and epigenetic, more and more tools and methodologies have been developed to understand the underlying biological systems to mine these datasets. This allows researchers in their field to make connections between data and apply these findings in practical biology.

One of the certainties of biology is that biological systems are extremely complex, with next generation sequencing leading to a revolution in the study of genomics the depth of this complexity began to be truly understood. As next generation sequencing (NGS) overtook sanger sequencing as the preferred method of genome sequencing, projects like the human genome project highlighted the need for increased knowledge and methodology in bioinformatics to understand and translate bioinformatic data.

6.1.1 De novo assembly

The first stage in the analysis of genetic data is to receive the raw genomic data, through next generation sequencing, which produces large quantities of short DNA information. De novo assembly is the process by which this is assembled into contigs without the need of a reference genome [Sohn JI et al,. 2018]. This process is essential because in many cases a reference cannot be used, such as comparing genomes of different strains looking for genetic differences between it and the reference. For new genome sequencing, there needs to be a method to remove redundancies in the data and collate regions that can closely correlate to chromosomes.

De novo assemblies require no prior knowledge of the source DNA length, composition, or layout. Reads can vary from tens to thousands of base pairs in length depending on the method used [Modi A et al,.2021]. Illumina sequencing as an example, can result in reads from 36-150 [Modi A et al,. 2021] these can be either single or paired end reads. Paired ends proceed from larger fragments that are read, resulting in two paired lengths of DNA read, one from the sense strand and another from the antisense strand and a separation distance that can be calculated. The aim of assembly is to turn these reads into contigs, which can then be ordered with each other to form scaffolds.

These assemblies are performed by software, such as Velvet and Spades [Abbas MM et al,. 2014] done by de Bruijn graph. This is a very difficult to perform computational program, complicated even further by repeats, which can occur sporadically within a genome. Differences between the platforms include spades finding higher N50 values than Velvet and often resulting in longer contig length [Bankevich A et al,. 2012] both systems were analysed for assembling sequences for *S. sonnies* resulting larger number of contigs greater than 500bp in Spades which Velvet showed lower n50 values.

6.1.2 Quast

The current technology of sequence and assembly suffers from many problems, making the reconstruction for whole chromosomes more difficult. Ther can be variations between assembly programs, resulting in radically different results, due to the different Heuristic algorithms used by each program. Because of this assembled genome much be assessed and evaluated on quality. This can be done by many different programs, such as GAGE, Assemblathon,Plantagora and QUAST [Gurevich A et al,. 2013]. However, there are limitations, GAGE and Plantagora for example require a reference genome. Quast has several advantages over the other methods and so has become the standard for assembly evaluation. It evaluates multiple areas and is easy to use.

The issue of needing a reference is still a problem in areas of bioinformatic analysis[Lischer HEL et al,. 2017]. Sequencing produces a series of fragments with no context, mapping the reads to an available reference is a valuable tool. However, this also has issues, as reads can

be short, have no indication of location other than their sequence, and the reference genomes can be very large.

6.1.3 Fungal genomics

Fungal genomes are unique of the eukaryotic genomes, they are smaller than the average eukaryotic genome, with the largest being in the oomycetes with an average genome of 75 megabase pairs (MBP) [Lee JH et al,. 2021] with small genomes in some families that are similar size to bacterial genomes. Because of this, fungal genomes were one of the first genomes to be fully sequenced [Galagan JE et al,. 2005]. Genomics in fungi has been used to compare different fungal genomes and study their evolution and the emergence of different fungal groups. Qualities such as chromosomal rearrangements were discovered in fungi more often than other eukaryotes [Galagan JE et al,. 2005] with the exception being filamentous ascomycetes like *Alternaria alternata*, which have conserved genes within chromosomes. Whole genome sequencing (WGS) is a valuable tool in the identification of fungal strains, finding genetic variations that may impact the phenotype of the strain. This is especially important in the field of host-pathogen interaction to find genes that are different in specific strains and to use RNA sequencing to identify expression differences during infection.

6.1.4 Comparing genome sequences.

One important aspect of genomics is comparing genomes and visualising the differences. There is multiple software that can compare genomes, including Mummer/DNAdiff and Mauve. Mauve is a method of genome comparison that identifies conserved regions in genomes, inversions, and conserved regions. The comparison between genomes can also be used to determine nucleotides substitutions, small insertions and indels. This system integrates separate analysis methods into one system. DNAdiff is a part of the Nicer and Mummer alignment program [Khelik K et al,. 2017], which measures and quantifies the changes between aligned genomes, as well as other differential qualities [Khelik k et al,. 2017]. This can be used to assess the quality of genome assemblies, but cane also be used for comparison of genomes. This does not provide a visualisation of the comparison like Mauve but does provide a table of the detected differences.

6.2 Aims

In chapter 3, it was demonstrated that there were new sRNAs and miRNAs identified by mapping the sRNA sequence to the reference genome. However, there are likely differences between the reference and the strain genome used for the sRNA sequencing. So further bioinformatic analysis was carried out to compare the reference genome to the sequence of the Cabi strain of *A. alternata* .

The objectives for this were as follows.

5. Isolate the whole genome of the Cabi *A.alternata* isolate, sequence with Illumina and perform *de novo* assembly with the raw reads.
6. Evaluate the quality of the assembly
7. Compare the Cabi *A.alternata* genome to the reference by mapping reads to the reference and by mapping the de novo assembled genome to the reference.
8. Verify the presence of identified sRNAs and miRNAs in the assembled genome.

6.3 Results

6.3.1 Isolation of the gDNA of *A.alternata*

The first step in this process is to isolate and sequence the genomic DNA of the *A. alternata* strain that was used for the analysis. The Cabi strain would likely have many differences to the reference genome which would be important to learn to explain various findings in the sRNA chapter. Such as why so few of the sRNA mapped to the reference genome. To do this the total genomic DNA must be extracted and the genome sequenced and bioinformatically analysed.

The agarose gel of the whole genome (gDNA) of the Cabi *A.alternata* strain showed good quality and high molecular weight bands at the top of the gel (ANY PICTURE). There was no observed contamination with low molecular weight degradation. the *A.alternata* gDNA achieved a concentration of 22.07 ng/µl and a quality of 2.08 (260/280) this sample was sent to BGI genomics for whole genome sequencing.

The sequenced raw data from the NGS service was sent in sets of two fasta quality files fastq (R1 and R2), which corresponded to the sequence reads for the forward and reverse strands of the genome. This included adaptor trimming. Both files contained 15 million 100 base pair small reads. On inspecting the multiQC and Fasta QC reports, raw data improved in quality post trimming and went from low Phre score less than 20 to PHre score greater than 20.

6.3.2 Analysis of the NGS raw reads for the *A.alternata* genome

To be used for bioinformatic analysis the genome reads needed to be assembled into contigs by *de Novo* assembly first using SPAdes, using the server provided by Galaxy. This is the process in which the 100 bp reads are assembles into contiguous contigs. This was done using K-Mer values (127, 99, 77, 55, 33 and 21) which produced contigs and scaffolds from the reads. The K-Mer is the number of sequences of which the length values is equal to K, and which can produce a contig. The best of the K-Mer value assemblies is provided for the final assembly. In addition, the same process was taken with Velvet to produce a *De Novo*

assembly to compare to the SPAdes a shown in [table 6.1] .The best of which was to be used as the final assembly.

Quality assessment of the assembly is done based on three qualities: number of contigs with the fewest and closest number to the number of chromosomes in the reference genome as possible, total genome size with the best being the closest to the reference genome, and N50 value. N50 is the minimum contig length needed to cover 50% of the genome and it is in most cases a good indicator of a good assembly to have a larger N50 and a very small N50 compared to genome size is a likely indication of high fragmentation.

For the two assemblies, they were assessed for quality using QUAST a tool for measuring the metrics of comparing genome assemblies. QUAST generated in [table 6.2], in this case the N50 was large for both indicating a successful genome assembly, however, are not comparable between assemblies due to not having the same combined length values [Miller JR et al,. 2010]. The Velvet assembly produces fewer contigs, however the closest to the reference in terms of genome length is the SPAdes assembly. Due its total length being closer to reference genome size of 33.6 Mb [Bihon W et al.,2016], however, still over 1Mb lower, this makes it a better choice for the final assembly. Most of this project will use the genome to verify the presence of the designed *CHS* sRNAs and confirm the presence of sRNA interest in our unique strain. Therefore, it is undesirable to have a smaller than expected genome as there is a higher likelihood that an area, which corresponds to our sRNAs of interest, have been removed from the assembly.

This table shows the results of analysis of the two genome assemblies in QUAST, this includes the N50, the sequence length of the shortest contig at 50% of the total genome size. This data also includes total genome size. The N75 which is calculated the same as N50 but for 75% of the genome. The L50 is the smallest number of contigs whose length can sum up to half the genome size and L75 which is the same as the L50 but for 75% of the genome. This also includes the N per 100 kbp, which is the measure of the number of uncalled base pairs per 100000 base pairs. 1) analysis of the genome assembly generated using Velvet, this assembly gave 91 total contigs, a total length of 30.2 million bp and an N50 value of 792991 bp; 2) QUAST analysis of the SPAdes assembly of the *A. alternata*

genome sequences, this resulted in more contigs at 544, N50 at 2.2 million bp and a total length of 33.03 million [table 6.1]. The conclusion was that both could be considered successful assemblies from the large N50 values, but the chosen assembly would be the one closest to the reference genome size of 33.6 Mb [Bihon W et al.,2016] and so the Spades was chosen as the best assembly due it being the most likely to include the sequence for all our identified sRNAs.

Table 6.1: Analysis of *A. alternata* genome assemblies using QUAST.
This table depicts the different values determined by QUAST in both the reference and the Spades assembled sequence. With the spades having significantly more contigs however showed a significantly larger N50.

Assembly name	*A.alternata* in Velvet	*A.alternata* in SPAdes
no of contigs	91	1921
total length	30209590	33036416
GC (%)	51.02	51.15
N50	792991	2286370
N75	441351	1791136
L50	14	5
L75	27	9
Ns per 100 kbp	9.92	25.74

6.3.3 Comparison of Cabi strain and reference using mauve

The resulting genome assembly showed many differences from the reference genome, a significantly larger number of contigs compared to the chromosomes. This can be seen in [Table 6.2] where the reference and the *de novo* assembly are compared. To better understand the degree of difference the two genomes were compared using mauve [Darling et al., 2010]. Mauve is a tool for comparing two genomes in a way that the differences can be visualised. The conclusions that can be drawn from this analysis is that the reference end the

genome sequence of the Cabi fungi are almost completely different genomes, With large regions of non-homology and many areas of homology with translocations. This makes many finding in the sRNA analysis make sense such as low percentages of sRNA mapping to the reverence, in some cases less than 10 percent. To further very this I used the Mauve alignment to compare the genome to a second know *Alternaria alternata* reference. this one being the AlatI reference genome. The alignment in mauve is shown in [Figure 6.1] for this the genomes are still very different however there are less regions allowing large areas of the spades genome not aligning to the reference.

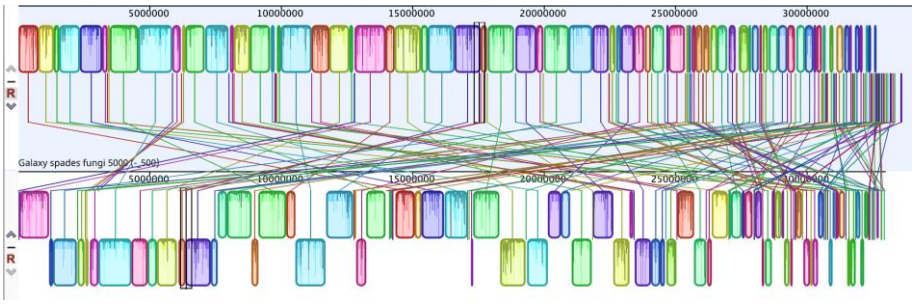


Figure 6.1 Mapping the reads to the reference

Mauve alignment shows that the genome is significantly different between the *de novo* assembly and the reference sequence. With large areas of the de novo assembly not being present in the reference and regions which match being rearranged.

Mapping reads to reference using Bowtie2. The results found that there was a significant difference between the first reference and the spades assembled genome. These differences can be seen in [Table 6.1] and the mauve alignment in [Figure 6.1] with reg denoting regions that do not align to either genome. The data is shown for all chromosomes and unassigned scaffolds. Each line assigned to the reference and spades respectively in order that they appear. Each colour represents an LCB. Each distinct LCB represents a localised Co-linear Blacks. The concept for these is that genomes are divergent from their ancestral species into multiple modern species. Chromosomes can be rearranged and include changes such as inversion, deletions, and translocations. Much of the genome may stay the same and remain conserved in different arrangements. LCBs are regions of the chromanones that appear to be conserved across the regions analysed. Analysis between

two identical genomes would result in a single LCB across the whole genome. In this case the genomes are a mosaic of multiple different LCBs, many regions being found on difference t chromosomes indicating translocations.

6.3.4 DNAdiff analysis of the references and spades assembly.

The next stage of analysis was to look for a method of comparing the two genomes which gave specific quantifiable differences. This should include factors such as the total sequences, those that align and the number that do not, these values as a percentage of the total sequences. As well as values such as total base pair alignment, SNPs, insertions, translocations and inversions. The tool I chose for this was DNAdiff, this program is a wrapper software which is open source and used for building alignments, based on a default series of parameters. This includes the nucmer scripts to process the information outputs and for the report of the alignment information. This included all the information I wanted to learn and was easy to perform being accessible by Galaxy. DNAdiff is a common tool for analysing two sequences with a degree of similarity [Khelik K et al,. 2017].

The results from the DNAdiff analysis highlights the similarities and the difference between the two genomes. This is in comparison with the total aligned bases and the unaligned bases. These sections show that 95% of the reference bases align to the Cabi genome while only 75% of the Cabi genome aligns with the reference. Indicating that there is a significant region of the Cabi genome different to the reference which those regions that do match seem to match very well. This may explain the large discrepancy in the sRNA alignment. With only a small percentage of both the infected and the fungi sRNA sequence aligning to the reference and may be an indicator that sRNAs overall may not be very well conserved between *Alternaria* genomes.

The report produced by DNAdiff [table 6.2]is separated into two columns, with rows split by the different metrics. TotalSeqs refers to the total number of sequences, aligned denotes the number of sequences with at least one alignment, and unaligned shows the number with no alignment. This also includes the total aligned bases and unaligned bases showing that 95% of the bases in the reference did align to the spade's assembly but 75% of the spades assembly did not align. Indicating that there were large regions for the Cabi *A. alternata* genome does not present in the reference.

Table 6.2 : DNAdiff statistics comparing the reference to the spade's assembly.

	Reference	Spades assembly
TotalSeqs	79	10851
AlignedSeqs	73(92.4051%)	87(0.8018%)
UnalignedSeqs	6(7.5949%)	10764(99.1982%)
TotalBases	32990834	41568504
AlignedBases	31375220(95.1028%)	31404430(75.5486%)
UnalignedBases	1615614(4.8972%)	10164074(24.4514%)
1-to-1	1180	1180
TotalLength	31280100	31280050
AvgLength	26508.5593	26508.5169
AvgIdentity	98.1964	98.1964

This shows counts of structural elements such as rearrangements, breakpoints being the number of non-maximal alignment endpoints, relocations, breaks in an alignment block in the same sequence, translocations, Inversions which are breaks in alignment where the adjacent 1-1 alignment is inverted with respect to the other. Insertions, and small Nucleotide polymorphisms (SNPs) (Table 6.3).

Table 6.3 : Table of alignment features.

	Reference	Spades assembly
Breakpoints	3066	3111
Relocations	22	16
Translocations	155	148
Inversions	83	89
Insertions	1162	1324
InsertionSum	1715430	2221889
InsertionAvg	1476.2737	1678.1639
TandemIns	5	5
TandemInsSum	537	461
TandemInsAvg	107.4	92.2
TotalSNPs	501963	501963

6.3.5 sRNA mapped to the Spades genome showed far greater coverage

Upon discovering that the mauve alignment that there were large regions of the genome not mapping to the reference this posed a question to be answered from the sRNA alignment. The sRNA alignment needed to use the reference genome as the available RNA software needed an annotated genome to calculate expression levels. This resulted in a very low level of sRNA mapping to the reference. 7690640 sRNA reads mapped to the reference genome from the sRNA libraries in total, compared to the spades in which 45,230,432 reads from the three sRNA libraries mappers to the spades genome. Meaning that there is a difference of 5 times, 56% to 10% in the spade's assembly of the Cabi genome to the reference genome.

This clearly demonstrates that compared to the reference there is significant differences as shown in the mauve alignment, which is clearly affecting the identification of fungal s RNAs. What is also interesting is the change in the infected samples. The infected samples show between 1.75 and 3.93% alignment of sRNAs to the reference. While when this was done for the cabi genome assembly there was a marked increase, going to 22.24% to 49.13% [Table 6.4] . This could lead to several conclusions, that the sRNAs expressed during infection are not conserved between markedly different genomes within the *Alternaria* genus, indicating that they are likely not conserved.

Table 6.4 sRNA showed significant greater number of mapped reads in the de novo assembly

The sRNA from the three replicates of the fungi and infected samples were mapped to both the reference and the de novo assembly of the genome showing the difference in sRNA conserved between the two genomes. The F6 sample shows 5 times increase in the number of sRNA mapped to the de novo assembled Cabi genome to reference. Showing that the lack of identification of fungal sRNA is due predominantly to the lack of sequence and annotation of the correct fungal genome.

Table 6.4: Total sRNA mapped difference between reference and assembled genome.

Sample	Total sRNA	Mapped sRNA to Reference	Mapped sRNA to assembly
INF2	11883493 (100.00%)	207401 (1.75%)	5838195(49.13%)
INF4	11763345 (100.00%)	259533 (2.21%)	2616722(22.24%)
INF1	18749522 (100.00%)	737776 (3.93%)	8985176(47.92%)
F9	25082088 (100.00%)	1476808 (5.89%)	6390391(25.48%)
F11	28567126 (100.00%)	2409137 (8.43%)	15372400(53.81%)
F10	27137999 (100.00%)	3804695 (14.02%)	12693627(46.78%)

6.4 Discussion

6.4.1 De novo assembly showed most success in SPAdes

The aim of this genomics chapter was to correct some of the issues that arrived from using genomic methods to sequence the genome of the sample, in this case *A.alternata* and how to use bioinformatic tools to assemble the reads from this and to compare this the reference genome. This was to look for genomic differences to explain the lack of sRNA mapping from the *A.alternata* sRNA library to the reference genome. This I looked at in more detail 6.3.5 which showed that more sRNA mapped to the Cabi genome assemblies, my conclusion from this is that this could indicate that many of the sRNAs are not well conserved between *Alternaria alternata* strains. If this si the cae you could assume that the sRNAs are not essential but may be a form of adaptation to new pathogens that are required by that role to not be well conserved. There has been evidence of conserved miRNA in fungi [Ma X et al,. 2019] so one explanation for this is that the majority of these sRNAs mapping to the cabi genome are not relevant for infection. There are many explanations for sRNAs that are unique to that genome, this can include regulatory sRNA and exon/intron fragments, and repeat sequences [Silvestri A et al,. 2019].

The results of the *de Novo* assembly were successful, this was tested with two assembly tools SPAdes and Velvet to determine which of these provided the better result. The quality of this was assessed using Quast. The results of this showed that despite the Spades

showing much larger number of contigs, the N50 values showed significantly higher value in the spades assembly and the L50 and L75 values were much lower. The results from this were that the assemblies could both be considered successful assemblies based on parameters shown in other papers [Simpson JT et al,. 2012] but the spades were chosen as the preferred assembly.

6.4.2 Comparison to the reference showed many differences and similarities.

The results of this genomic analysis showed several details. The Cabi genome when assembles showed many differences from the reference genome, the DNAdiff analysis showed many differences. Despite 92% of the reference sequence aligning to the Cabi sequence, the vast majority of the Cabi assembly did not align to the genome. This could likely be to a few reasons, genomic diversity between strains has been studied [Peris D et al,. 2022] where genetic differences between fungal strains have been studied in addition to papers looking at strains in the *Alternaria Spp* such as [Gannibal PB,. Et al 2022]. Demonstrating genetic diversity between strains has been done but more focussed on individual locus that can be used as markers of individual strains and not in large scale genomic differences.

This explains the lack of sRNA identifiable from its mapping to the reference howler the Cabi assembly could not be used to identify miRNA in the sRNA library. This was for multiple reasons, one is the lack of miRbase database for *alternaria alternata*, but also that the alignment to quantify and compare the sRNA requites an annotated genome.

6.4.3 Mauve alignment and sRNA mapping shows that sRNA expression is significantly different between the strain and the reference.

The mauve alignment clarifies much of what is observed in the DNAdiff alignment. That there are many regions that match between the two genomes however there are many rearrangements as well as regions present in the Cabi genome that are not present in the reference genome. As can be seen in the mauve alignment in [figure 6.1] the *A.alternata* reference and the Cabi genome still contain many of the same LCBs but often with small nucleotides changes and gene inversions. This was what first indicated to me that this difference may be one of the underlying causes for the lack of alignment of sRNA to the

reference. Prompting me to realign the sRNA to the de Novo assembly and compare this the alignment to reference. This is seen in [table 6.4] showing that the increase in alignment is significant in some cases over 5 times increase in percentage of sRNA aligned to the Cabi genome compared to the reference. Demonstrating that the issue was not in the sRNA sequencing but in the genome used to align them too. An assembled genome si not enough to identify novel miRNA, however. To do so would need the use of an annotated genome and miRNA prediction software.

6.4.4 How annotation of the genome assembly can be used to identify sRNA in *A.alternata*

I have established so far that without an annotated version of the cabi strain genome and it is not possible to fully identify miRNAs to the same extent as to what was found for the plant genome. Annotation is the process of identifying the genes of a novel sequenced genome. With the correct software and methodology, you can determine gene locations and identify novel miRNA loci. The details vary but many tools will align the novel sequence to the databases of protein coding genes and transcripts to the genome and that gaps are allowed in the place of introns. This method can also make use of spliced alignment, which is used to generate evidence for gene structure complete exons, gene locations.

This sequencing can take the form of expressed sequence tags, full length cDNA sequences or RNA-seq which is cDNA from next generation transcriptome sequencing. Full length cDNA has the advantage of the transcriptional start, exons of fully spliced transcripts and polyadenylation cleaving sites. This allows all the needed genome structures to be revealed, ORF, UTRS and exons. [Cook DE et al,.2019] The ideal situation for identifying fungal sRNAs then would be a full-length cDNA sequencing annotation of a chosen strain. In this case the Cabi *A.alternata* strain, in addition to genome sequencing, sRNA and transcriptome sequencing. This would lead to the creation of an annotated reference accurate to your genome, which has been shown to be needed in this chapter. Software tools such as EST_GENOME to align the FL_cDNA to the genome sequence post de novo assembly.

Accurate annotation can be done with homology-based methods using prediction algorithms to reveals exons and introns [Harrison PM,. 2021] however repeat sequences can

make this difficult and as was seen through the mauve and DNAdiff comparison there are many large differences between this genome and the reference. Making homology-based prediction unlikely. There has been new progress made ink repeat element identification. There has been much progress in gene prediction suing not just the genome sequence but also using the external data. Such as transcript databases.

An annotated genome will allow a similar process as performed for the plant sRNA, the sRNA can be mapped to the genome and elements such as exons, nc-RNA, etc can be removed by the same methods shown in chapter 5. Resulting in sRNA that can be predicted for novel miRNA transcripts in the same method as the novel miRNA using novel miRNA prediction software. Potentially with its parameters altered to match the miRNA structure of fungal miRNA. This approach would then provide a catalogue of fungal miRNA, with predicted targets, expression differential between the infected and non-infected, predicted targets in the plant if present in addition RNA-seq data would be able to apply differences between target gene expression between infection and non-infection, to determine if the proposed targets do become de-regulated in correlation with increased miRNA expression.

An annotated genome for the exact strain sequenced allows for the identification of novel sRNA genes and sRNA mapping with the ability to annotate the locations and target genes of these sRNA. This combined with miRNA prediction could result in creating a miRNA identification method for *A.alternata* and allowing the prediction of miRNA targets in the plant. Additional data can support this such as mRNA seq, which can be used to correlate the effect on the expression of the potential targeting genes and whether this correlates with increased expression of candidate miRNA in vivo. This can also be used with regulate genes within the fungi, which can be of just as much importance as identifying cross-kingdom sRNA.

Chapter 7

General discussion

7.1 Aims of the thesis

This study aimed to look in closer detail at the *A. alternata/* tomato pathosystem with a focus on the role of sRNA and the potential for sRNAs as a method for targeting and silencing essential fungal genes for pathogenesis. In addition, I looked in detail at the role of *Chitin synthase* genes and the potential they may have as a target for sRNA silencing, and whether the targeting of this gene with sRNA would have a negative effect on spore viability and fungal pathogenicity.

The hypothesis of this thesis is that sRNA interacts have been confirmed to exist in fungi and plant pathosystems, and that by studying these interactions we can better understand and utilise this system. However not all fungal plant pathosystems have been studied for these interactions, many plants and fungi have not been sequenced for sRNA in this manner. Nor have they been evaluated for the sRNA upregulated during infection or had their miRNA identified regarding cross kingdom silencing. The aim of this project was to do this analysis for tomato and Alternaria alternata and to evaluate potential targets for sRNA silencing. With sRNA becoming an increasingly more viable option for fungal control new gene targets will need to be identified. One of the best options for this is to identify gene targets already used by the plant, as these have been naturally selected for their effectiveness in defence. These can then be further up regulated and bred for in resistance strains or enhanced through either host induced gene silencing or spray induced gene silencing.

To this there were several steps needed to be taken which I have outline in this thesis. Pathology analysis to determine the optimum Alternaria strain, and its optimum conditions for producing spores for analysis, chitin synthase was chosen to be the best candidate for sRNA and novel sRNAs were designed and used to target these genes and determine their effects. sRAN sequencing was used to determine miRNA present in the plant and determine which if any were upregulated during infection and if any of these had predicted targets in the fungi genome. This discussion will evaluate these results and their overall conclusions as well as look at future work that may be done in this field to improve and further elaborate on these results.

7.2 the Cabi isolate determined the most infective in tomato and the best methodology found for spore production.

Of the three available Alternaria alternata strains available it was the strain CABI LSHB SM-O430 that showed the largest lesion size in tomato leaves compared to the other strain. this was likely due host specificity expression changes because of multiple subculturing. Alternaria host specificity is determined by host specificity toxins, whose expression levels can be severely altered by multiple culturing [Raghunath, T et al,. 1968]. This is especially the case as was seen in the A001 and FERA which had re-cultures on potato and parsnip respectively. Resulting in done regulation of tomato HSTs and reduced infectivity. However, there are still larger lesions that the non-host specific *A.brassicola* and the control.

This *A.alternata* strain was then further cultured and evaluated for its spore production, having found during the culturing process that spore production in alternaria is inconsistent. With some sample plates producing abundant spores and others producing none. A methodology needed to be applied to determine the most effective method of culturing spores' production. Studying the literature, conidia production in the *Alternaria spp* is heavily influenced by factors including media, temperature, humidity, and stress. With examples of procedures including mycelial stress, cold shock and nutrient deprivation all having been shown in other papers to have results in inducing sporulation. The only method found in the literature unable to be tested was Black UV light which has been reported as very effective in inducing sporulation [Pruss S et al,. 2014]. In the absence of this method, I needed to test all the known methods to see which was the most effective in my sample. This was essential a regular and abundant source of spores was needed to analyse sRNA silencing and its effect of germination. Mycelial tissue alone was not sufficient for sRNA analysis and there was no way of determining its effectiveness and for sRNA to be a viable antifungal it must be functional on plant leaves. Requiring its function to be seen either in spore germination or appressoria formation.

7.3 silencing CHS induced spore tube growth and spore formation

The first steps of analysis were to synthesise sRNA targeting 5 CHS genes, as a preliminary test to see if there is any viability in targeting the genes. The results from this found only one of the genes having consistent inhibition with increased concentration. The rest when treated showed some inhibition in some cases, but this did not increase with concentration. However, there were other observations, including increased germ tube lengths and increased spire chain formation. The results for all 11 CHS genes showed that in all cases the spore germ tubes increased in length by up to 3 times after treatment with the sRNA. I discussed the possible reasons for this in addition to increases in spore chain formation. With the Qpcr analysis showing that the sRNA was successfully degrading the mRNA for the CHS genes means that we can reasonably assume that this phenotype is due to the sRNA silencing. There are multiple examples of inhibiting chitin synthase having a paradoxical effect. Examples of antifungals that inhibit chitin synthase have led to fungi growing faster and chitin levels increasing in response to the Echinocandins. That chitin salvation process and potential signs of compensation expression by other genes makes CHS an unlikely candidate for sRNA silencing to control infection. Increases germ tube length and especially increase in spore formation will likely exasperate infection. Especially as spores produced by the affected fungi will likely not be affected by the sRNA even if it does reduce germination. However, the results give a methodology for identification of new gene targets which could lead to future work in mass screening of new potential targets.

7.4 sRNA sequencing revealed novel miRNA and potential fungal targets in plant miRNA.

sRNA sequencing makes up most of the results for this thesis, with most of the information of the sRNA interactions present in the pathosystem making up the annotation of the miRNA present in the plant genome and produced in response to the infection. There were issues with identifying fungal miRNA. With the fungi lacking miRNA database for Alternaria alternata or for a similar organism making it not possible to annotate the miRNA present in the Alternaria sRNA library the same way as the tomato. For the tomato the sRNA was sequenced, and the miRNA base library was used to annotate the mapped sRNA, determining silencing from non-silencing sRNA, and annotating from them the miRNA. This includes the known miRNA annotated by previous sequencing and novel miRNA newly discovered in this thesis.

sRNA annotation of the known and novel miRNA predominantly looked to discover the miRNA role in regulation of genes involved in Alternaria infection. The miRNA with predicted targets in the plant genome that are upregulated during infection likely knockdown the target through miRNA mediated gene silencing. This case is seen in the monobactam biosynthesis and the

glycan degradation pathway. Both processes from their role would make sense to be down regulated during infection, but it was not known before now that miRNA performed the role of down regulating these genes during fungal infection. This expands on the role miRNA plays during infection, potentially being able to perform more specific gene regulation in response to infection.

miRNA was also identified to have predicted targets in the fungal genome, this was done by identifying the miRNA with a read count above 100 bp, mapping to the plant genome and upregulating it in the infected sRNA libraries. The results were multiple miRNA that I could map to the fungal genome and strict criteria outlined in chapter 5, to determine if any of the miRNA had potential targets. Then the potential target genes were identified on ensembl fungi to determine if there was a protein coding gene the miRNA aligned to and in two of the cases there were, and the genes were viable targets. VTC4 had research showing it plays a role in infection and trafficking of virulence factors outside of fungal cells however there is no research on the role for these proteins in Alternaria. The same is for Catherin, there are multiple studies in other fungi showing the essential role of Catherin for fungi but very little related to Alternaria. This leads on to potential further studies to look at gene mutants of these gens or by designing sRNA to target these genes and look at their effect on infection.

7.5 The Cabi strain showed significant differences to the reference genome

This data was further looked at in the genomics chapter. One limitation posed by the methods used to analyse the sRNA was that both miRNA identification and sRNA mapping required an annotated genome. This wasn't ideal as the sRNA may show significant differences between different strains. This isn't an issue for the plant genome due to the lack of large genome differences between plants within a species [Sahu, K.K. et al,. 2017]. To do this first the whole genome of the Cabi strain was sequenced by extracting the genomic DNA and having the genome sequenced, this resulted in two strands in millions of small reads. This needed to be assembled and that assembly needed to be evaluated for quality. This could have been done by two methods, assembly to the reference or by de novo assembly. I chose to do a de novo assembly as there was a large chance of there being major differences to the reference. De novo assembly was done by both spades and velvet with both assemblies evaluated by Quast to determine the effectiveness of the assembly, with the spade's assembly being chosen as the preferred assembly. This was to be used to verify the sRNA results in the Cabi genome. This was successful however the mapping to the genome revealed some flaws in the data. With large

areas of the genome not mapping to the reference, this may explain the sRNA that did not map to the reference but was still present in the sRNA library. As the sRNA must originate from somewhere, whether degraded mRNA or sRNA themselves. Meaning that there is likely more sRNA to be revealed in the interaction. However, to find these other miRNAs, there would need to be an annotated version of the assembled Cabi genome. To test this hypothesis, I mapped the sRNA libraries to both the reference genome and the spades assembly.

Overall significance of the research

The end goal of all PHD thesis is to contribute meaningfully to the field, so in this section I wanted to summarise the major findings this project has made for the field of plant pathology

• It was demonstrated that although cell structural protein synthases like cellulose synthase and chitin synthase may work for some fungi this is not the case for all. It is important that for a target gene to be confirmed as a potential target for sRNA silencing it must be observed to inhibit germination with no major non-beneficial side effects like germ tube length increase or increased sporulation.

• It was shown that many of the methods of sporulation induction in Alternaria alternata are not effective in all strains. With the methodology needing to be tailored for the unique fungi you intend to work. This can preferably lead to less issues in future research for others struggling with the same issue in Alternaria alternata or similar fungi.

• It was demonstrated in a step-by-step method how novel sRNAs can be generated targeting a gene in Alternaria and how these sRNA can be evaluated for their effectiveness as an sRNA for use in HIGS or SIGS. Potentially paving the way for more papers screening multiple gene targets using this simple and easily replicated methodology.

• It was demonstrated that despite the different functions and roles of different chitin synthase genes that sRNA targeting them did result in their expression decreasing through QPCR and that this reduction in expression resulted in a similar phenotype. Potentially concluding that the paradoxical effect and its mechanisms occur at the level of post transcriptional regulation and not at the protein level.

• This study contributed to the annotation of novel miRNA in the Alternaria species, with over 100 novel miRNAs discovered and the functions and targets of many of the, found in addition to this. Many of them play a role in infection by down regulating specific genes.

• This study is the first to look at the potential presence of trans kingdom miRNA present in the Alternaria tomato pathosystem identifying two potential miRNAs with viable fungal targets that are also upregulated during infection. Potentially leading to more viable research in the pathosystem and the Alternaria spp and into the miRNA's potential anti-fungal activity.

• Finally, I believe this project has given an overview in to pitfalls and problems that arise when studying sRNA interaction in fungal plant pathosystems not explored in other papers, as well as details review of the literature and field that may prove very useful for future studies in the area.

Future Directions

This study has both generated answers to questions on the interactions between fungal and plant sRNA but has equally posed more questions. This question can be answered in future which I'm sure would inevitably result in a continuing rabbit hole of future research. In this section I would like to outline my views on where future work in this topic could lead.

Firstly, on Chitin synthase there are many aspects of the findings that could be investigated further. For example, the cause of the observed paradoxical effect. Future work could verify the findings on the germination, looking at new sRNAs and observing their effect on germination for all 11 CHS genes, as well as for use in combination. As silencing multiple CHS may be more effective. Similarly work may investigate the genetic processes underlying the changes observed, such as a mRNA sequence to see which genes are altered in response to the sRNA excluding the target. What knock on effects may be silencing one CHS gene.

This work can also go on to look at other potential gene targets, a few were looked at in chapter 5, such as Catherin and VTC4. Both of which had potential miRNA targets form the plant miRNA. Other viable targets could be tested, in theory any gene found to be upregulated during infection for the fungi could be a viable target. I can foresee a viable project in which mRNA sequencing can identify upregulated genes during infection, either in this a pathosystem or another, and these screened for sRNA viable targets by designing sRNA for each gene and testing on spores for effects on germination. Similar tests can be done to see the effect on lesion

size rather than germination. As lesion size covers a larger array of essential processes that may be inhibited such as appressoria formation.

Many areas of the sRNA sequencing are also left to be studied. The miRNA which was shown to be upregulated during infection can be looked at in more detail. Such as mutant strains created which do not contain these miRNAs or where these miRNAs are upregulated could be studied. Including mutant strains in which the target processes themselves are mutated or artificially down regulated and that's effect on infection. A notable absence from this study is the look at fungal miRNAs as the same methods used to identify the plant miRNA could not be used. A novel strategy could be developed. This would require first annotating the spades assembled Alternaria alternata genome to remove any potential errors from using the reference. miRNA would have to be predicted purely based on ion structure rather than any reference form the miRNA base. In addition, more work would need to be done to verify the presence of transkingdoms miRNAs. Such as labelling the miRNAs that are throes to play this role and attempting to observe if they are taken up by fungal cells. Looking at their effect on the fungi and the proposed target in vitro with synthesised miRNA or looking at the effect of upregulating these miRNAs on infection.

Finally, future work could look at the viability of the identified miRNAs as a prey induced fungicide against alternaria, if signs in lab work indicate they do have a fungicidal value or improve plant defence against the onset of infection. Future would or could be field test their use as a spray induces sRNA fungicide.

References

Abbas MM, Malluhi QM, Balakrishnan P. (2014) Assessment of de novo assemblers for draft genomes: a case study with fungal genomes. BMC Genomics. 2014;15 Suppl 9(Suppl 9):S10. doi: 10.1186/1471-2164-15-S9-S10. Epub Dec 8. PMID: 25521762; PMCID: PMC4290589.

Avenot H. F.; Michailides T. J. (2007) Resistance to boscalid fungicide in *Alternaria alternata* isolates from pistachio in California. *Plant Dis.* 2007, 91 (10), 1345–1350. 10.1094/PDIS-91-10-1345

Abuley I. K., B. J. Nielsen,R. (2017) Resistance status of cultivated potatoes to early blight (Alternaria solani) in Denmark Labouriau.

Aravin AA, Naumova NM, Tulin AV, Vagin VV, Rozovsky YM, Gvozdev VA (2001) Double-stranded RNA-mediated silencing of genomic tandem repeats and transposable elements in the D. melanogaster germline". Current Biology. 11 (13): 1017–1027 .

Adachi H, Yu YT (2014) Insight into the mechanisms and functions of spliceosomal snRNA pseudouridylation . World Journal of Biological Chemistry. 5(4): 398–408 .

Baldrich, P., San Segundo, B (2016) MicroRNAs in Rice Innate Immunity. *Rice* **9,** 6, .

Bankevich A, Nurk S, Antipov D, Gurevich AA, Dvorkin M, Kulikov AS, Lesin VM, Nikolenko SI, Pham S, Prjibelski AD, Pyshkin AV, Sirotkin AV, Vyahhi N, Tesler G, Alekseyev MA, Pevzner PA. (2012) SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. J Comput Biol. May;19(5):455-77.

Bartolini D, Sancineto L, Fabro de Bem A, Tew KD, Santi C, Radi R, Toquato P, Galli F. (2017) Selenocompounds in Cancer Therapy: An Overview. Adv Cancer Res. 2017;136:259-302. doi: 10.1016/bs.acr.2017.07.007. Epub : Sep 28. PMID: 29054421.

Bennett JW, Klich M. (2003) Mycotoxins. Clin Microbiol Rev. Jul;16(3):497-516. doi: 10.1128/CMR.1 97-516.2003. PMID: 12857779; PMCID: PMC164220.

Belkhadir Y, Subramaniam R, Dangl JL.(2004) Plant disease resistance protein signaling: NBS-LRR proteins and their partners. Curr Opin Plant Biol. Aug;7(4):391-9. doi: 10.1016/j.pbi.2004.05.009. PMID: 15231261.

Bartel DP, (2009) MicroRNAs: target recognition and regulatory functions". Cell. 136 (2): 215–33, .

Brian D. Rutter Roger W. Innes (2020) Growing pains: addressing the pitfalls of plant extracellular vesicle research.

Borges, F., and Martienssen, R. A. (2015) The expanding world of small RNAs in plants. *Nat. Rev. Mol. Cell Biol.* 16, 727–741 .

Baldrich P, Brian D. Rutter, Hana Zand Karimi, Ram Podicheti, Blake C. Meyers, Roger W. Innes. (2019) Plant Extracellular Vesicles Contain Diverse Small RNA Species and Are Enriched in 10- to 17-Nucleotide "Tiny" RNAs .

Buhtz A, Springer F, Chappell L, Baulcombe DC, Kehr J. (2008) Identification and characterization of small RNAs from the phloem of Brassica napus. Plant J. Mar;53(5):739-49. doi: 10.1111/j.1365-313X.2007.03368.x. Epub 2007 Nov 14. PMID: 18005229.

Barkai-Golan, R.. (2001) Postharvest Diseases of Fruits and Vegetables: Development and Control. Netherlands, Elsevier Scienc .

Bills. G.F, Chamuris G. P, Farr.D.F and Rossman.A.A.Y.(1987) Fungi on plants and plant-products in the United States - a sourcebook for plant pathogenic fungi. Phytopathology, 77, 1701-1701 .

Beihammer G, Maresch D, Altmann F, Van Damme EJM, Strasser R. Lewis (2021)A Glycans Are Present on Proteins Involved in Cell Wall Biosynthesis and Appear Evolutionarily Conserved Among Natural *Arabidopsis thaliana* Accessions. Front Plant Sci. 2021 Mar 11;12:630891. doi: 10.3389/fpls. .630891. PMID: 33777069; PMCID: PMC7991798.

Bellemain, E., Carlsen, T., Brochmann, C (2010) . ITS as an environmental DNA barcode for fungi: an *in silico* approach reveals potential PCR biases. *BMC Microbiol* **10,** 189

Burdett H, Bentham AR, Williams SJ, Dodds PN, Anderson PA, Banfield MJ, Kobe B. The Plant "Resistosome": Structural Insights into Immune Signaling. Cell Host Microbe. 2019 Aug 14;26(2):193-201. doi: 10.1016/j.chom.2019.07.020. PMID: 31415752.

Borges, F., Martienssen, R. (2015)The expanding world of small RNAs in plants. *Nat Rev Mol Cell Biol* **16**, 727–741.

Bonmatin J.-M.; Giorio C.; Girolami V.; Goulson D.; Kreutzweiser D.; Krupke C.; Liess M.; Long E.; Marzaro M.; Mitchell E. (2015) A. Environmental fate and exposure; neonicotinoids and fipronil. *Environ. Sci. Pollut. Res.* 22 (1), 35–67. 10.1007/s11356-014-3332-7.

Bowen AR, Chen-Wu JL, Momany M, Young R, Szaniszlo PJ, Robbins PW. (1992) Classification of fungal chitin synthases. Proc Natl Acad Sci U S A. Jan 15;89(2):519-23. doi: 10.1073/pnas.89.2.519. PMID: 1731323; PMCID: PMC48270.

Brakhage AA. (1998) Molecular regulation of beta-lactam biosynthesis in filamentous fungi. Microbiol Mol Biol Rev. 1998 Sep;62(3):547-85. doi: 10.1128/MMBR.62.3.547-585 . PMID: 9729600; PMCID: PMC98925.

Chaliha C, Rugen MD, Field RA, Kalita E. (2018) Glycans as Modulators of Plant Defense Against Filamentous Pathogens. Front Plant Sci. 2018 Jul 4;9:928. doi: 10.3389/fpls. 00928. PMID: 30022987; PMCID: PMC6039678.

Cook DE, Valle-Inclan JE, Pajoro A, Rovenich H, Thomma BPHJ, Faino L (2018) Long-Read Annotation: Automated Eukaryotic Genome Annotation Based on Long-Read cDNA Sequencing. Plant Physiol. 2019 Jan;179(1):38-54. doi: 10.1104/pp.18.00848. Epub Nov 6. PMID: 30401722; PMCID: PMC6324239.

C.J. Combrink, J.C., Kotze, J.M., Wehner, P.C. and Grobbelaar. (1985) Fungi associated with core rot of Straking apples in South Africa.

Chung KR. Stress Response and Pathogenicity of the Necrotrophic Fungal Pathogen (2012) Alternaria alternata. Scientifica (Cairo). 2012;2012:635431. doi: 10.6064/2012/635431. Epub Dec 10. PMID: 24278721; PMCID: PMC3820455.

Catalanotto C, Cogoni C, Zardo G. (2016) MicroRNA in Control of Gene Expression: An Overview of Nuclear Functions. Int J Mol Sci. Oct 13;17(10):1712. doi: 10.3390/ijms17101712. PMID: 27754357; PMCID: PMC5085744.

Chen X. (2010) Small RNAs - secrets and surprises of the genome. Plant J. Mar;61(6):941-58. doi: 10.1111/j.1365-313X.2009.04089.x. PMID: 20409269; PMCID: PMC3062250.

Carthew RW, Sontheimer EJ. (2009) Origins and Mechanisms of miRNAs and siRNAs. Cell: Feb 20;136(4):642-55. doi: 10.1016/j.cell.2009.01.035. PMID: 19239886; PMCID: PMC2675692.

Chen D., Yuan C., Zhang J., Zhang Z., Bai L., Meng Y (2011) PlantNATsDB: a comprehensive database of plant natural antisense transcripts. Nucleic Acids Research 40:D1:D1187–D1193. (PlantNATsDB)

Cock, P.J.A., Fields, C.J., Goto, N., Heuer, M.L., and Rice, P.M (2010). The Sanger FASTQ file format for sequences with quality scores, and the Solexa/Illumina FASTQ variants. Nucleic acids research 38, 1767-1771.

"Crops". FAOSTAT. Retrieved 9 August 2019. Countries – Select All; Regions – World + (Total); Elements – Production Quantity; Items – Tomatoes; Years – 2017 + 2016

Chitwood DH, Timmermans MC. (2010) Small RNAs are on the move. Nature. Sep 23;467(7314):415-9. doi: 10.1038/nature09351. PMID: 20864994.

Callum N. Watson, Antonio Belli, and Valentina Di Pietro, (2019) Small Non-coding RNAs: New Class of Biomarkers and Potential Therapeutic Targets in Neurodegenerative Disease.

Chen, C., Zeng, Z., Liu, Z (2018) Small RNAs, emerging regulators critical for the development of horticultural traits. *Hortic Res* **5,** 63 .

Cuperus, J. T., Fahlgren, N. & Carrington, J. C. (2011) Evolution and functional diversification of miRNA genes. *Plant Cell* **23**, 431–442 .

Chen, C., Zeng, Z., Liu, Z. (2018) Small RNAs, emerging regulators critical for the development of horticultural traits. *Hortic Res* **5,** 63 .

Davison, J.(2010)"GM plants: Science, politics and EC regulations". *Plant Science*. **178** (2): 94–98 .

Deng P, Muhammad S, Cao M, Wu L. (2018) Biogenesis and regulatory hierarchy of phased small interfering RNAs in plants. *Plant Biotechnol J* .

Dang Y, Yang Q, Xue Z, Liu Y.(2011) RNA interference in fungi: pathways, functions, and applications. Eukaryot Cell. 2011 Sep;10(9):1148-55. doi: 10.1128/EC.05109-11. Epub Jul 1. PMID: 21724934; PMCID: PMC3187057.

Doehlemann G, Ökmen B, Zhu W, Sharon A. (2017) Plant Pathogenic Fungi. Microbiol Spectr. Jan;5(1). doi: 10.1128/microbiolspec.FUNK-0023-2016. PMID: 28155813.

Dukowic-Schulze S, Anitha Sundararajan, Thiruvarangan Ramaraj, Shahryar Kianian, Wojciech P. Pawlowski, Joann Mudge and Changbin Chen1.(2016) miRNAs and Indications for a Role of PhasiRNAs in Meiosis Front. Plant Sci.

Dang Y, Yang Q, Xue Z, Liu Y. (2011)RNA interference in fungi: pathways, functions, and applications. Eukaryot Cell. Sep;10(9):1148-55. doi: 10.1128/EC.05109-11. Epub 2011 Jul 1. PMID: 21724934; PMCID: PMC3187057.

Dodds PN, Rathjen JP. (2010) Plant immunity: towards an integrated view of plant-pathogen interactions. Nat Rev Genet. 2010 Aug;11(8):539-48. doi: 10.1038/nrg2812. Epub Jun 29. PMID: 20585331.

Erlich, Y., and Mitra, P.P. (2008) Alta-Cyclic: a self-optimizing base caller for next-generation sequencing. Nature methods 5, 679-682.

Feng X, Wang Z, Fillmore R, Xi Y. MiR-200 (2013) a new star miRNA in human cancer. Cancer Lett. 2014 Mar 28;344(2):166-73. doi: 10.1016/j.canlet.2013.11.004. Epub Nov 19. PMID: 24262661; PMCID: PMC3946634.

Figueroa M, Hammond-Kosack KE, Solomon PS (2017) A review of wheat diseases-a field perspective. Mol Plant Pathol. 2018 Jun;19(6):1523-1536. doi: 10.1111/mpp.12618. Epub Dec 26. PMID: 29045052; PMCID: PMC6638159.

Friesen TL, Faris JD, Solomon PS, Oliver RP (2008) Host-specific toxins: effectors of necrotrophic pathogenicity. Cell Microbiol. Jul;10(7):1421-8. doi: 10.1111/j.1462-5822.2008.01153.x. Epub 2008 Apr 1. PMID: 18384660.

Fonseca JP, Mysore KS (2018) Genes involved in nonhost disease resistance as a key to engineer durable resistance in crops. Plant Sci. 2019 Feb;279:108-116. doi: 10.1016/j.plantsci. PMID: 30709487.

Fisher MC, Henk DA, Briggs CJ, Brownstein JS, Madoff LC, McCraw SL, Gurr SJ. (2012) Emerging fungal threats to animal, plant, and ecosystem health. Nature. 11;484(7393):186-94. doi: 10.1038/nature10947. PMID: 22498624; PMCID: PMC3821985.

Falcone M, Ferrara S, Rossi E, Johansen HK, Molin S, Bertoni G. (2018) The Small RNA ErsA of *Pseudomonas aeruginosa* Contributes to Biofilm Development and Motility through Post-transcriptional Modulation of AmrZ. Front Microbiol. Feb 15;9:238. doi: 10.3389/fmicb.2018.00238. PMID: 29497413; PMCID: PMC5819304.

Friedlander M.R., Mackowiak S.D., Li N., Chen W., Rajewsky N (2011) miRDeep2 accurately identifies known and hundreds of novel microRNA genes in seven animal clades. Nucleic Acids Res 40:37-52. (miRDeep2)
Fei Q, Xia R, Meyers BC (2013) Phased, secondary, small interfering RNAs in posttranscriptional regulatory networks. Plant Cell: Jul;25(7):2400-15.

Gannibal PB, Orina AS, Kononenko GP, Burkin AA (2022) Distinction of *Alternaria* Sect. *Pseudoalternaria* Strains among Other *Alternaria* Fungi from Cereals. J Fungi (Basel). Apr 20;8(5):423. doi: 10.3390/jof8050423. PMID: 35628679; PMCID: PMC9142887.

Glawe DA (2008) The powdery mildews: a review of the world's most familiar (yet poorly known) plant pathogens. Annu Rev Phytopathol.;46:27-51. doi: 10.1146/annurev.phyto.46.081407.104740. PMID: 18680422.

Galagan JE, Henn MR, Ma LJ, Cuomo CA, Birren B (2005) Genomics of the fungal kingdom: insights into eukaryotic biology. Genome Res. Dec;15(12):1620-31. doi: 10.1101/gr.3767105. Erratum in: Genome Res. 2006 Feb;16(2):304. PMID: 16339359.

Gandía M, Garrigues S, Bolós B, Manzanares P, Marcos JF (2019) The Myosin Motor Domain-Containing Chitin Synthases Are Involved in Cell Wall Integrity and Sensitivity to Antifungal Proteins in *Penicillium digitatum*. Front Microbiol. Oct 18;10:2400. doi: 10.3389/fmicb.2019.02400. PMID: 31681248; PMCID: PMC6813208.

Gomes-Vieira AL, Wideman JG, Paes-Vieira L, Gomes SL, Richards TA, Meyer-Fernandes JR. (2018) Evolutionary conservation of a core fungal phosphate homeostasis pathway coupled to development in Blastocladiella emersonii. Fungal Genet Biol. 2018 Jun;115:20-32. doi: 10.1016/j.fgb.2018.04.004. Epub Apr 5. PMID: 29627365.

Gong Z, Qi J, Hu M, Bi G, Zhou JM, Han GZ. (2022) the origin and evolution of a plant resistosome. Plant Cell. Apr 26;34(5):1600-1620. doi: 10.1093/plcell/koac053. PMID: 35166827; PMCID: PMC9048963.

Grover ND (2010) Echinocandins: A ray of hope in antifungal drug therapy. Indian J Pharmacol. Feb;42(1):9-11. doi: 10.4103/0253-7613.62396. PMID: 20606829; PMCID: PMC2885632.

Gurevich A, Saveliev V, Vyahhi N, Tesler G (2013) QUAST: quality assessment tool for genome assemblies. Bioinformatics. 2013 Apr 15;29(8):1072-5. doi: 10.1093/bioinformatics/btt086. Epub Feb 19. PMID: 23422339; PMCID: PMC3624806.

Grassini P, Eskridge KM, Cassman KG (2013) Distinguishing between yield advances and yield plateaus in historical crop production trends. Nat Commun. ;4:2918. doi: 10.1038/ncomms3918. PMID: 24346131; PMCID: PMC3905725.

Godfrey P. Apangu, Carl A. Frisk, Beverley Adams-Groom, Jack Satchwell, Catherine H. Pashley, Carsten A. Skjøth, (2020) Air mass trajectories and land cover map reveal cereals and oilseed rape as major local sources of Alternaria spores in the Midlands, UK, Atmospheric Pollution Research, Volume 11, Issue 9.

Goodman, Novacky, A. J., & Novacky, A. J. (1994) The hypersensitive reaction in plants to pathogens : a resistance phenomenon /R.N. Goodman and A.J. Novacky. .

Gullino, M.L., Garibaldi, A. (2018). Environment Modification for Disease Management. In: McGovern, R., Elmer, W. (eds) Handbook of Florists' Crops Diseases. Handbook of Plant Disease Management. Springer, Cham. https://doi.org/10.1007/978-3-319-39670-5_5

Harrison PM. (2023) Computational Methods for Pseudogene Annotation Based on Sequence Homology. Methods Mol Biol: 24:35-48. doi: 10.1007/978-1-0716-1503-4_3. PMID: 34165707.

Haas BJ, Zeng Q, Pearson MD, Cuomo CA, Wortman JR. (2011) Approaches to Fungal Genome Annotation. Mycology. Oct 3;2(3):118-141. doi: 10.1080/21501203.2011.606851. PMID: 22059117; PMCID: PMC3207268.

Hanley BA, Schuler MA. (1991) Developmental expression of plant snRNAs. Nucleic Acids Res.Nov 25;19(22):6319-25. doi: 10.1093/nar/19.22.6319. PMID: 1956791; PMCID: PMC329146.

Hahn M. (2014) The rising threat of fungicide resistance in plant pathogenic fungi: Botrytis as a case study. J Chem Biol. May 28;7(4):133-41. doi: 10.1007/s12154-014-0113-1. PMID: 25320647; PMCID: PMC4182335.

Häweker H, Rips S, Koiwa H, Salomon S, Saijo Y, Chinchilla D, Robatzek S, von Schaewen A (2010) Pattern recognition receptors require N-glycosylation to mediate plant immunity. J Biol Chem. Feb 12;285(7):4629-36. doi: 10.1074/jbc.M109.063073. Epub 2009 Dec 11. PMID: 20007973; PMCID: PMC2836068.

Hyun MW, Yun YH, Kim JY, Kim SH (2011) Fungal and Plant Phenylalanine Ammonia-lyase. Mycobiology. Dec;39(4):257-65. doi: 10.5941/MYCO.2011.39.4.257. Epub 2011 Dec 7. PMID: 22783113; PMCID: PMC3385129.

Hyde, K.D., Jones, E.B.G., Liu, JK. *et al.* Families of Dothideomycetes. *Fungal Diversity* **63,** 1–313, 2013.

Huen, A., Bally, J. & Smith, P (2018) Identification and characterisation of microRNAs and their target genes in phosphate-starved *Nicotiana benthamiana* by small RNA deep sequencing and 5'RACE analysis. *BMC Genomics* **19,** 940.

Hutvagner, Gyorgy. (2005) Small RNA asymmetry in RNAi: Function in RISC assembly and gene regulation. FEBS Letters. 579 (26): 5850–5857 .

Houwing S, Kamminga LM, Berezikov E, Cronembold D, Girard A, van den Elst H, Filippov DV, Blaser H, Raz E, Moens CB, Plasterk RH, Hannon GJ, Draper BW, Ketting RF (2007) A role for Piwi and piRNAs in germ cell maintenance and transposon silencing in Zebrafish. Cell. 129 (1): 69–82. .

Jones-Rhoades MW, Bartel DP, Bartel B. (2006) MicroRNAS and their regulatory roles in plants. Annu Rev Plant Biol. ;57:19-53. doi: 10.1146/annurev.arplant.57.032905.105218. PMID: 16669754.

Jindo K, Evenhuis A, Kempenaar C, Pombo Sudré C, Zhan X, Goitom Teklu M, Kessel G. (2021) Review: Holistic pest management against early blight disease towards sustainable agriculture. Pest Manag Sci. Sep;77(9):3871–3880. doi: 10.1002/ps.6320. Epub 2021 Feb 27. PMID: 33538396; PMCID: PMC8451811.

Jeseničnik T, Štajner N, Radišek S, Mishra AK, Košmelj K, Kunej U, Jakše J. (2022) Discovery of microRNA-like Small RNAs in Pathogenic Plant Fungus Verticillium nonalfalfae Using High-Throughput Sequencing and qPCR and RLM-RACE Validation. Int J Mol Sci. Jan 14;23(2):900. doi: 10.3390/ijms23020900. PMID: 35055083; PMCID: PMC8778906.

Jain A, Sarsaiya S, Wu Q, Lu Y, Shi J. (2019) A review of plant leaf fungal diseases and its environment speciation. Bioengineered. Dec;10(1):409-424. doi: 10.1080/21655979.2019.1649520. PMID: 31502497; PMCID: PMC6779379.

Jindo K, Evenhuis A, Kempenaar C, Pombo Sudré C, Zhan X, Goitom Teklu M, Kessel G. (2021) Review: Holistic pest management against early blight disease towards sustainable agriculture. Pest Manag Sci. 2021 Sep;77(9):3871-3880. doi: 10.1002/ps.6320. Epub Feb 27. PMID: 33538396; PMCID: PMC8451811.

Jaiswal, S., Iquebal, M.A., Arora, V (2019) Development of species specific putative miRNA and its target prediction tool in wheat (*Triticum aestivum* L.). *Sci Rep* **9,** 3790, .

Jiang, L., Schlesinger, F., Davis, C.A., Zhang, Y., Li, R., Salit, M., Gingeras, T.R., and Oliver, B (2011) Synthetic spike-in standards for RNA-seq experiments. Genome research 21, 1543-1551.

Kanehisa M, Araki M, Goto S, Hattori M, Hirakawa M (2008) KEGG for linking genomes to life and the environment. Nucleic Acids research36:D480–484. (KEGG) Moxon S., Schwach F., MacLean D., Dalmay T., J Studholme D., and Moulton V. . A toolkit for analysing large-scale plant small RNA datasets. Bioinformatics 24 (19): 2252-2253. (UEA sRNA tools)

Kadotani N, Nakayashiki H, Tosa Y, Mayama S (2004)One of the two Dicer-like proteins in the filamentous fungi Magnaporthe oryzae genome is responsible for hairpin RNA-triggered RNA silencing and related small interfering RNA accumulation. J Biol Chem. Oct 22;279(43):44467-74. doi: 10.1074/jbc.M408259200. Epub 2004 Aug 10. PMID: 15304480.

Khelik K, Lagesen K, Sandve GK, Rognes T, Nederbragt AJ. (2017) NucDiff: in-depth characterization and annotation of differences between two sets of DNA sequences. BMC Bioinformatics. Jul 12;18(1):338. doi: 10.1186/s12859-017-1748-z. PMID: 28701187; PMCID: PMC5508607.

Kettles GJ, Hofinger BJ, Hu P, Bayon C, Rudd JJ, Balmer D, Courbot M, Hammond-Kosack KE, Scalliet G, Kanyuka K. (2019) sRNA Profiling Combined With Gene Function Analysis Reveals a Lack of Evidence for Cross-Kingdom RNAi in the Wheat - *Zymoseptoria tritici* Pathosystem. Front Plant Sci. Jul 4;10:892. doi: 10.3389/fpls.2019.00892. PMID: 31333714; PMCID: PMC6620828.

Kufel J, Grzechnik P. (2019) Small Nucleolar RNAs Tell a Different Tale. Trends Genet.Feb;35(2):104–117. doi: 10.1016/j.tig.2018.11.005. Epub Dec 16. PMID: 30563726.

Kettles GJ, Hofinger BJ, Hu P, Bayon C, Rudd JJ, Balmer D, Courbot M, Hammond-Kosack KE, Scalliet G, Kanyuka K.(2019) sRNA Profiling Combined With Gene Function Analysis Reveals a Lack of Evidence for Cross-Kingdom RNAi in the Wheat - *Zymoseptoria tritici* Pathosystem. Front Plant Sci. Jul 4;10:892. doi: 10.3389/fpls.2019.00892. PMID: 31333714; PMCID: PMC6620828.

Kwon-Chung KJ, Sugui JA. (2009) What do we know about the role of gliotoxin in the pathobiology of Aspergillus fumigatus. Med Mycol. ;47 Suppl 1(Suppl 1):S97-103. doi: 10.1080/13693780802056012

Khajuria C, Ivashuta S, Wiggins E (2018) Development and characterization of the first dsRNA-resistant insect population from western corn rootworm, Diabrotica virgifera virgifera LeConte. *PLoS One*. 2018;13(5):e0197059

Klattenhoff C, Bratu DP, McGinnis-Schultz N, Koppetsch BS, Cook HA, Theurkauf WE. (2007) Drosophila rasiRNA pathway mutations disrupt embryonic axis specification through activation of an ATR/Chk2 DNA damage response. Developmental Cell. 12 (1): 45–55.

Xiuzhen Kong, Meng Yang, Brandon H. Le, Wenrong He, Yingnan Hou (2022)The master role of siRNAs in plant immunity.

Koch A., Kumar N., Weber L., Keller H., Imani J., Kogel K.H. (2013) Host-induced gene silencing of cytochrome P450 lanosterol C14 -demethylase-encoding genes confers strong resistance to *Fusarium* species. *Proc. Natl. Acad. Sci. USA.* ;110:19324–19329. doi: 10.1073/pnas.1306373110.

Komiya R. (2017) Biogenesis of diverse plant phasiRNAs involves an miRNA-trigger and Dicer-processing. J Plant Res. Jan;130(1):17-23. doi: 10.1007/s10265-016-0878-0. Epub 2016 Nov 29. PMID: 27900550; PMCID: PMC5219027.

Khadempour L, Burnum-Johnson KE, Baker ES, Nicora CD, Webb-Robertson BM, White RA 3rd, Monroe ME, Huang EL, Smith RD, Currie CR. (2016) The fungal cultivar of leaf-cutter ants produces specific enzymes in response to different plant substrates. Mol Ecol.

Nov;25(22):5795-5805. doi: 10.1111/mec.13872. Epub 2016 Oct 26. PMID: 27696597; PMCID: PMC5118115.

Keller H, Boyer L, Abad P. Disease susceptibility in the Zig-Zag model of host-microbe interactions: only a consequence of immune suppression? Mol Plant Pathol. 2016 May;17(4):475-9. doi: 10.1111/mpp.12371. PMID: 26788791; PMCID: PMC6638450.

Lax C, Tahiri G, Patiño-Medina JA, Cánovas-Márquez JT, Pérez-Ruiz JA, Osorio-Concepción M, Navarro E, Calo S (2020) The Evolutionary Significance of RNAi in the Fungal Kingdom. Int J Mol Sci.Dec 8;21(24):9348. doi: 10.3390/ijms21249348. PMID: 33302447; PMCID: PMC7763443.

Lenardon MD, Munro CA, Gow NA. (2010) Chitin synthesis and fungal pathogenesis. Curr Opin Microbiol. 2010 Aug;13(4):416-23. doi: 10.1016/j.mib.2010.05.002. Epub Jun 8. PMID: 20561815; PMCID: PMC2923753.

Lee JH, Siddique MI, Kwon JK, Kang BC (2021) Comparative Genomic Analysis Reveals Genetic Variation and Adaptive Evolution in the Pathogenicity-Related Genes of Phytophthora capsici. Front Microbiol. Aug 13;12:694136. doi: 10.3389/fmicb.2021.694136. PMID: 34484141; PMCID: PMC8415033.

Lischer HEL, Shimizu KK. (2017) Reference-guided de novo assembly approach improves genome reconstruction for related species. BMC Bioinformatics. Nov 10;18(1):474. doi: 10.1186/s12859-017-1911-6. PMID: 29126390; PMCID:PMC5681816.

Lefrancq M.; Imfeld G.; Payraudeau S.; Millet M (2012) Kresoxim methyl deposition, drift and runoff in a vineyard catchment. *Sci. Total Environ.* 2013, 442, 503–508. 10.1016/j.scitotenv.

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

Liu, N., Tu, L., Wang, L (2017) MicroRNA 157-targeted *SPL* genes regulate floral organ size and ovule production in cotton. *BMC Plant Biol* **17,** 7 .

Lenardon MD, Munro CA, Gow NA (2010) Chitin synthesis and fungal pathogenesis. Curr Opin Microbiol. 2010 Aug;13(4):416-23. doi: 10.1016/j.mib.2010.05.002. Epub Jun 8. PMID: 20561815; PMCID: PMC2923753.

Li L, Chang SS, Liu Y (2010) RNA interference pathways in filamentous fungi. Cell Mol Life Sci. Nov;67(22):3849-63. doi: 10.1007/s00018-010-0471-y. Epub 2010 Aug 1. PMID: 20680389; PMCID: PMC4605205.

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

Liang D, White RG, Waterhouse PM. Gene silencing in Arabidopsis spreads from the root to the shoot, through a gating barrier, by template-dependent, nonvascular, cell-to-cell movement. Plant Physiol. 2012 Jul;159(3):984-1000. doi: 10.1104/pp.112.197129. Epub 2012 May 11. PMID: 22582134; PMCID: PMC3387722.

Lee, Hyang Burm (2015) Alternaria in Food: Ecophysiology, Mycotoxin Production and Toxicology." *Mycobiology* vol. 43,2 93-106. doi:10.5941/MYCO.2015.43.2.93
Li R, Oliver RA, Townsend CA. Identification and Characterization of the Sulfazecin Monobactam Biosynthetic Gene Cluster. Cell Chem Biol. 2017 Jan 19;24(1):24-34. doi: 10.1016/j.chembiol.2016.11.010.

Li Y, Lu YG, Shi Y, Wu L, Xu YJ, Huang F, Guo XY, Zhang Y, Fan J, Zhao JQ, Zhang HY, Xu PZ, Zhou JM, Wu XJ, Wang PR, Wang WM. (2014) Multiple rice microRNAs are involved in immunity against the blast fungus Magnaporthe oryzae. Plant Physiol. Feb;164(2):1077-92. doi: 10.1104/pp.113.230052. Epub 2013 Dec 13. PMID: 24335508; PMCID: PMC3912081.

Liu R, Xu C, Zhang Q, Wang S, Fang W. (2017) Evolution of the chitin synthase gene family correlates with fungal morphogenesis and adaption to ecological niches. Sci Rep. Mar 16;7:44527. doi: 10.1038/srep44527. PMID: 28300148; PMCID: PMC5353729.

Li M, Jiang C, Wang Q, Zhao Z, Jin Q, Xu JR, Liu H. (2016) Evolution and Functional Insights of Different Ancestral Orthologous Clades of Chitin Synthase Genes in the Fungal Tree of Life. Front Plant Sci. Feb 1;7:37. doi: 10.3389/fpls.2016.00037. PMID: 26870058; PMCID: PMC4734345.

M Younus Wani, NA Ganie, S Rani, S Mehraj, MR Mir, MF Baqual, KA Sahaf, FA Malik and KA Dar (2018) Advances and applications of Bioinformatics in various fields of life. Int. J. Fauna Biol. Stud. ;5(2):03-10.

Mao, X., Cai, T., Olyarchuk, J.G., and Wei, L (2005). Automated genome annotation and pathway identification using the KEGG orthology (KO) as a controlled vocabulary. Bioinformatics 21, 3787–3793. (KOBAS)

Michael I Love,Wolfgang Huber,Simon Anders (2014) Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2.Genome Biology,DOI 10.1186/s13059-014-0550-8. (DESeq2)

Ma X, Wiedmer J, Palma-Guerrero J (2020) Small RNA Bidirectional Crosstalk During the Interaction Between Wheat and Zymoseptoria tritici. Front Plant Sci. Jan 8;10:1669. doi: 10.3389/fpls.2019.01669. PMID: 31969895; PMCID: PMC6960233.

Modi A, Vai S, Caramelli D, Lari M (2021)The Illumina Sequencing Protocol and the NovaSeq 6000 System. Methods Mol Biol. ;2242:15-42. doi: 10.1007/978-1-0716-1099-2_2. PMID: 33961215.

Marchese S, Polo A, Ariano A, Velotto S, Costantini S, Severino L (2018) Aflatoxin B1 and M1: Biological Properties and Their Involvement in Cancer Development. Toxins

(Basel). May 24;10(6):214. doi: 10.3390/toxins10060214. PMID: 29794965; PMCID: PMC6024316.

Meier UT (2017) RNA modification in Cajal bodies. RNA Biol. Jun 3;14(6):693-700. doi: 10.1080/15476286.2016.1249091. Epub 2016 Oct 24. PMID: 27775477; PMCID: PMC5519239.

Merzendorfer H (2011) The cellular basis of chitin synthesis in fungi and insects: common principles and differences. Eur J Cell Biol. 2011 Sep;90(9):759-69. doi: 10.1016/j.ejcb.2011.04.014. Epub Jun 22. PMID: 21700357.

Magan, A. Medina Aldred. (2011) Possible climate-change effects on mycotoxin contamination of food crops pre- and postharvest N.

Moghaddam GA, Rezayatmand Z, Nasr Esfahani M, Khozaei M (2019) Genetic defense analysis of tomatoes in response to early blight disease, Alternaria alternata. Plant Physiol Biochem. 2019 Sep;142:500-509. doi: 10.1016/j.plaphy.2019.08.011. Epub Aug 16. PMID: 31445475.

Meister G, Tuschl T (2004) Mechanisms of gene silencing by double-stranded RNA. *Nature*. ;431:343–349.

Montgomery TA, Yoo SJ, Fahlgren N, Gilbert SD, Howell MD, Sullivan CM, Alexander A, Nguyen G, Allen E, Ahn JH (2008) AGO1-miR173 complex initiates phased siRNA formation in plants. Proc Natl Acad Sci USA 105: 20055–20062, .

Muhammad T, Zhang F, Zhang Y, Liang Y.(2019) RNA Interference: A Natural Immune System of Plants to Counteract Biotic Stressors. Cells. Jan 10;8(1):38. doi: 10.3390/cells8010038. PMID: 30634662; PMCID: PMC6356646.

Pruss S, Fetzner R, Seither K, Herr A, Pfeiffer E, Metzler M, Lawrence CB, Fischer R. (2014) Role of the Alternaria alternata blue-light receptor LreA (white-collar 1) in spore formation and secondary metabolism. Appl Environ Microbiol. Apr;80(8):2582-91. doi: 10.1128/AEM.00327-14. Epub 2014 Feb 14. PMID: 24532063; PMCID: PMC3993185.

Park, E., Jiang, Y., Hao, L (2021) Genetic variation and microRNA targeting of A-to-I RNA editing fine tune human tissue transcriptomes. *Genome Biol* **22,** 77

Pinto VE, Patriarca (2017) A. Alternaria Species and Their Associated Mycotoxins. Methods Mol Biol. ;1542:13-32. doi: 10.1007/978-1-4939-6707-0_2. PMID: 27924529.

Piqué N, Miñana-Galbis D, Merino S, Tomás JM. (2015) Virulence Factors of Erwinia amylovora: A Review. Int J Mol Sci. Jun 5;16(6):12836-54. doi: 10.3390/ijms160612836. PMID: 26057748; PMCID: PMC4490474.

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

Pérez-Reytor D, Plaza N, Espejo RT, Navarrete P, Bastías R, Garcia K. (2017) Role of Non-coding Regulatory RNA in the Virulence of Human Pathogenic *Vibrios*. Front Microbiol. Jan 11;7:2160. doi: 10.3389/fmicb.2016.02160. PMID: 28123382; PMCID: PMC5225090.

Nanda AK, Andrio E, Marino D, Pauly N, Dunand C. (2010) Reactive oxygen species during plant-microorganism early interactions. J Integr Plant Biol. Feb;52(2):195-204. doi: 10.1111/j.1744-7909.2010.00933.x. PMID: 20377681.

Nicolás FE, Ruiz-Vázquez RM.(2013) Functional diversity of RNAi-associated sRNAs in fungi. Int J Mol Sci. Jul 24;14(8):15348-60. doi: 10.3390/ijms140815348. PMID: 23887655; PMCID: PMC3759863.

O'Brien J, Hayder H, Zayed Y, Peng C. (2018) Overview of MicroRNA Biogenesis, Mechanisms of Actions, and Circulation. Front Endocrinol (Lausanne). Aug 3;9:402. doi: 10.3389/fendo.2018.00402. PMID: 30123182; PMCID: PMC6085463.

Ozata, D.M., Gainetdinov, I., Zoch, A. (2019) PIWI-interacting RNAs: small RNAs with big functions. *Nat Rev Genet* **20,** 89–108, .

Qi T, Guo J, Peng H, Liu P, Kang Z, Guo J (2019) Host-Induced Gene Silencing: A Powerful Strategy to Control Diseases of Wheat and Barley. Int J Mol Sci. Jan 8;20(1):206. doi: 10.3390/ijms20010206. PMID: 30626050; PMCID: PMC6337638.
Qiao Y, Shi J, Zhai Y, Hou Y, Ma W. Phytophthora effector targets a novel component of small RNA pathway in plants to promote infection. Proc Natl Acad Sci U S A. 2015 May 5;112(18):5850-5. doi: 10.1073/pnas.1421475112. Epub 2015 Apr 20. PMID: 25902521; PMCID: PMC4426467.

Raghunath, T. (1968) Behaviour of Alternaria poonensis inciting blight of coriander in artificial culture. *Mycopathologia et Mycologia Applicata* **36**, 347–352 https://doi.org/10.1007/BF02050381

Rogg LE, Fortwendel JR, Juvvadi PR, Steinbach WJ. (2011) Regulation of expression, activity and localization of fungal chitin synthases. Med Mycol. 2012 Jan;50(1):2-17. doi: 10.3109/13693786. .577104. Epub 2011 Apr 28. PMID: 21526913; PMCID: PMC3660733.

Robiglio AL, López SE. Mycotoxin production by Alternaria alternata strains isolated from red delicious apples in Argentina. Int J Food Microbiol. 1995 Jan;24(3):413-7. doi: 10.1016/0168-1605(94)00035-5. PMID: 7710918.

Rutter BD, Rutter KL, Innes RW. Isolation and Quantification of Plant Extracellular Vesicles. Bio Protoc. 2017 Sep 5;7(17):e2533. doi: 10.21769/BioProtoc.2533. PMID: 34541189; PMCID: PMC8413617.

Rotem, J. (1990) Overwintering of alternaria macrospora in Cotton Debris. *Phytoparasitica* **18**, 143–152 https://doi.org/10.1007/BF02981231

Sahu, K.K., Chattopadhyay, D (2017) Genome-wide sequence variations between wild and cultivated tomato species revisited by whole genome sequence mapping. *BMC Genomics* **18**, 430 https://doi.org/10.1186/s12864-017-3822-3

Simpson JT, Durbin R. (2011) Efficient de novo assembly of large genomes using compressed data structures. Genome Res. 2012 Mar;22(3):549-56. doi: 10.1101/gr.126953.111. Epub Dec 7. PMID: 22156294; PMCID: PMC3290790.

Silvestri A, Fiorilli V, Miozzi L, Accotto GP, Turina M, Lanfranco L. (2019) In silico analysis of fungal small RNA accumulation reveals putative plant mRNA targets in the symbiosis between an arbuscular mycorrhizal fungus and its host plant. BMC Genomics. Mar 4;20(1):169. doi: 10.1186/s12864-019-5561-0. PMID: 30832582; PMCID: PMC6399891.

Storz, G (2002) An expanding universe of noncoding RNAs.

Stevens DA. (2000) Drug interaction studies of a glucan synthase inhibitor (LY 303366) and a chitin synthase inhibitor (Nikkomycin Z) for inhibition and killing of fungal pathogens. Antimicrob Agents Chemother. 2000 Sep;44(9):2547-8. doi: 10.1128/AAC.44.9.2547-2548. . PMID: 10952614; PMCID: PMC90104.

Si-Ammour A, Windels D, Arn-Bouldoires E, Kutter C, Ailhas J, Meins F Jr, Vazquez F. (2018) miR393 and secondary siRNAs regulate expression of the TIR1/AFB2 auxin receptor clade and auxin-related development of Arabidopsis leaves. Plant Physiol. 2011 Oct;157(2):683-91. doi: 10.1104/pp.111.180083. Epub 2011 Aug 9. PMID: 21828251; PMCID: PMC3192580.

Sohn JI, Nam JW(2018) . The present and future of de novo whole-genome assembly. Brief Bioinform. Jan 1;19(1):23-40. doi: 10.1093/bib/bbw096. PMID: 27742661.

Tapiero H, Townsend DM, Tew KD. (2003) The antioxidant role of selenium and seleno-compounds. Biomed Pharmacother. May-Jun;57(3-4):134-44. doi: 10.1016/s0753-3322(03)00035-0. PMID: 12818475; PMCID: PMC6361120.

Takashi Tsuge, Yoshiaki Harimoto, Kazuya Akimitsu, Kouhei Ohtani, Motoichiro Kodama, Yasunori Akagi, Mayumi Egusa, Mikihiro Yamamoto, Hiroshi Otani (2013) Host-selective toxins produced by the plant pathogenic fungus *Alternaria alternata*, *FEMS Microbiology Reviews*, Volume 37, Issue 1, January, Pages 44–66,

Tsuda K, Somssich IE.(2015) Transcriptional networks in plant immunity. New Phytol. May;206(3):932-947. doi: 10.1111/nph.13286. Epub 2015 Jan 26. PMID: 25623163.

Treitschke S, Doehlemann G, Schuster M, Steinberg G. (2010) The myosin motor domain of fungal chitin synthase V is dispensable for vesicle motility but required for virulence of the maize pathogen Ustilago maydis. Plant Cell. Jul;22(7):2476-94. doi: 10.1105/tpc.110.075028. Epub 2010 Jul 27. PMID: 20663961; PMCID: PMC2929105.

Thomma BP. Alternaria spp.: from general saprophyte to specific parasite.(2003) Mol Plant Pathol. Jul 1;4(4):225-36. doi: 10.1046/j.1364-3703.2003.00173.x. PMID: 20569383.

Treitschke S, Doehlemann G, Schuster M, Steinberg G. (2010) The myosin motor domain of fungal chitin synthase V is dispensable for vesicle motility but required for virulence of the maize pathogen Ustilago maydis. Plant Cell. Jul;22(7):2476-94. doi: 10.1105/tpc.110.075028. Epub 2010 Jul 27. PMID: 20663961; PMCID: PMC2929105.

Valinezhad Orang A, Safaralizadeh R, Kazemzadeh-Bavili M.(2014) Mechanisms of miRNA-Mediated Gene Regulation from Common Downregulation to mRNA-Specific Upregulation. Int J Genomics. 2014;2014:970607. doi: 10.1155/2014/970607. Epub Aug 10. PMID: 25180174; PMCID: PMC4142390.

Van Den Bosch F.; Paveley N.; van den Berg F.; Hobbelen P.; Oliver R. (2014) Mixtures as a fungicide resistance management tactic. *Phytopathology* , 104 (12), 1264–1273. 10.1094/PHYTO-04-14-0121-RVW.

Wang, Q., Li, T., Zhong, C. (2019) . Small RNAs generated by bidirectional transcription mediate silencing of RXLR effector genes in the oomycete *Phytophthora sojae*. *Phytopathol Res* **1**, 18, .

Wang L., Feng Z., Wang X., Wang X., Zhang X. (2010). DEGseq: an R package for identifying differentially expressed genes from RNA-seq data. Bioinformatics 26, 136-8. (DEGseq) Storey, J. D. 2003. The positive false discovery rate: A Bayesian interpretation and the q-value, Annals of Statistics. 31: 2013-2035

Wen M., Shen Y., Shi S., and Tang T (2012) miREvo: An Integrative microRNA Evolutionary Analysis Platform for Next-generation Sequencing Experiments. BMC Bioinformatics 13:140.

Wu HJ, Ma YK, Chen T, Wang M, Wang XJ (2012) PsRobot: a web-based plant small RNA meta-analysis toolbox. Nucleic Acids Res 40:W22–W28. (psRobot)

Wang M, Weiberg A, Dellota E Jr, Yamane D, Jin H (2017) Botrytis small RNA Bc-siR37 suppresses plant defense genes by cross-kingdom RNAi. RNA Biol. Apr 3;14(4):421-428. doi: 10.1080/15476286.2017.1291112. Epub 2017 Mar 7. PMID: 28267415; PMCID: PMC5411126.

Watson CN, Belli A, Di Pietro V.(2019) Small Non-coding RNAs: New Class of Biomarkers and Potential Therapeutic Targets in Neurodegenerative Disease. Front Genet. Apr 26;10:364. doi: 10.3389/fgene.2019.00364. PMID: 31080456; PMCID: PMC6497742.

Wild CP, Gong YY. (2010) Mycotoxins and human disease: a largely ignored global health issue. *Carcinogenesis* ;31(1):71-82. doi:10.1093/carcin/bgp264

Wytinck N, Sullivan DS, Biggar KT, Crisostomo L, Pelka P, Belmonte MF, Whyard S. (2020) Clathrin mediated endocytosis is involved in the uptake of exogenous double-stranded RNA in the white mold phytopathogen Sclerotinia sclerotiorum. Sci Rep. Jul 29;10(1):12773. doi: 10.1038/s41598-020-69771-9. PMID: 32728195; PMCID: PMC7391711.

Whanger PD. (2002) Selenocompounds in plants and animals and their biological significance. J Am Coll Nutr. 2002 Jun;21(3):223-32. doi: 10.1080/07315724. .10719214. PMID: 12074249.

Wiest PM, Wiese K, Jacobs MR, Morrissey AB, Abelson TI, Witt W, Lederman MM. (1987) Alternaria infection in a patient with acquired immunodeficiency syndrome: case report and review of invasive alternaria infections. Rev Infect Dis. Jul-Aug;9(4):799-803. doi: 10.1093/clinids/9.4.799. PMID: 3326127.

Wang, Hailing Jin (2016) Spray-Induced Gene Silencing: a Powerful Innovative Strategy for Crop Protection Ming Jan; 25(1):4-6. .

Wang, Y., Liu, W., Wang, X. (2020) MiR156 regulates anthocyanin biosynthesis through *SPL* targets and other microRNAs in poplar. *Hortic Res* **7,** 118.

Wang M, Weiberg A, Lin FM, Thomma BP, Huang HD, Jin H. (2016) Bidirectional cross-kingdom RNAi and fungal uptake of external RNAs confer plant protection. Nat Plants. 19;2:16151. PMID: 27643635; PMCID: PMC5040644 .

Wei H, Zhou B, Zhang F (2013) Profiling and identification of small rDNA-derived RNAs and their potential biological functions. *PLoS One*. ; 8(2):e56842.

Werner S, Sugui JA, Steinberg G, Deising HB (2007) A chitin synthase with a myosin-like motor domain is essential for hyphal growth, appressorium differentiation, and pathogenicity of the maize anthracnose fungus Colletotrichum graminicola. Mol. Plant Microbe Interact. **20**:1555–1567. 10.1094/MPMI-20-12-1555

Worrall Elizabeth A., Bravo-Cazar Ana, Nilon Alexander T., Fletcher Stephen J., Robinson Karl E., Carr John P., Mitter Neena, (2019) Exogenous Application of RNAi-Inducing Double-Stranded RNA Inhibits Aphid-Mediated Transmission of a Plant Virus. Frontiers in Plant Science.

Xia, R., Chen, C., Pokhrel, S (2019) 24-nt reproductive phasiRNAs are broadly present in angiosperms. *Nat Commun* **10,** 627,

Xia, R., Zhu, H., An, Yq. (2012) Apple miRNAs and tasiRNAs with novel regulatory networks. *Genome Biol* **13,** R47 .

Xiao, J., Tang, X., Li, Y (2011) Identification of microRNA precursors based on random forest with network-level representation method of stem-loop structure. *BMC Bioinformatics* **12,** 165, .

Yuan M, Ngou BPM, Ding P, Xin XF. (2021) PTI-ETI crosstalk: an integrative view of plant immunity. Curr Opin Plant Biol. Aug;62:102030. doi: 10.1016/j.pbi.2021.102030. Epub 2021 Mar 5. PMID: 33684883.

Annex

Annexes 1: list of *CHS* genes with exon and intron locations

CHS gene name	Genome location	Exon / Intron	Start	End	Length in Nt
CC77DRAFT_1092832	KV441473:1469399-1469882	*OAG*23516-1	1,469,449	1,469,832	384
		Intron 1-2	1,469,833	1,469,886	54
		OAG23516-2	1,469,887	1,470,458	572
		Intron 2-3	1,470,459	1,470,510	52
		OAG23516-3	1,470,511	1,470,959	449
CC77DRAFT_1002156	KV441504: 30,804-32,636	OAG13919-1	30,804	30,916	113
		Intron 1-2	30,917	30,965	49
		OAG13919-2	30,966	31,492	527
		Intron 2-3	31,493	31,545	53
		OAG13919-3	31,546	32,636	1,091
CC77DRAFT_987140	KV441476: 702,395-706,717	OAG21670-1	702,395	705,438	3,044
		Intron 1-2	705,439	705,485	47
		OAG21670-2	705,486	705,976	491

		Intron 2-3	705,977	706,027	51
		OAG21670-3	706,028	706,357	330
		Intron 3-4	706,358	706,411	54
		OAG21670-4	706,412	706,717	306
CC77DRAFT_1027903	KV441470:1,078,709-1,081,788	OAG25348-1	1,078,709	1,078,727	19
		Intron 1-2	1,078,728	1,078,869	142
		OAG25348-2	1,078,870	1,079,084	215
		Intron 2-3	1,079,085	1,079,149	65
		OAG25348-3	1,079,150	1,081,702	2,553
		Intron 3-4	1,081,703	1,081,773	71
		OAG25348-4	1,081,774	1,081,788	15
CC77DRAFT_929028	KV441472:1,112,837-1,120,382	OAG24007-1	1,120,382	1,120,209	174
		Intron 1-2	1,120,208	1,120,054	155
		OAG24007-2	1,120,053	1,119,646	408
		Intron 2-3	1,119,645	1,119,583	63

		OAG24007-3	1,119,582	1,118,964	619
		Intron 3-4	1,118,963	1,118,911	53
		OAG24007-4	1,118,910	1,118,357	554
		Intron 4-5	1,118,356	1,117,738	619
		OAG24007-5	1,117,737	1,117,457	281
		Intron 5-6	1,117,456	1,116,429	1,028
		OAG24007-6	1,116,428	1,116,180	249
		Intron 6-7	1,116,179	1,116,106	74
		OAG24007-7	1,116,105	1,114,160	1,946
		Intron 7-8	1,114,159	1,114,111	49
		OAG24007-8	1,114,110	1,112,837	1,274
CC77DRAFT_1031975	KV441481: 41,704-44,373	*OAG19*173-1	41,704	41,808	105
		Intron 1-2	41,809	41,858	50
		OAG19173-2	41,859	42,025	167
		Intron 2-3	42,026	42,071	46

		OAG19173-3	42,072	44,373	2,302
CC77DRAFT_1021806	KV441482: 466,038-469,661	OAG18953-1	466,038	466,478	441
		Intron 1-2	466,479	466,534	56
		OAG18953-2	466,535	467,012	478
		Intron 2-3	467,013	467,062	50
		OAG18953-3	467,063	467,424	362
		Intron 3-4	467,425	467,473	49
		OAG18953-4	467,474	469,144	1,671
		Intron 4-5	469,145	469,195	51
		OAG18953-5	469,196	469,661	466
CC77DRAFT_1021349	KV441481: 49,098-55,161	OAG19177-1	55,161	54,369	793
		Intron 1-2	54,368	54,320	49
		OAG19177-2	54,319	50,681	3,639
		Intron 2-3	50,680	50,634	47
		OAG19177-3	50,633	49,098	1,536
CC77DRAFT_1062346	SuperContig KV441481: 56,446-62,079	OAG19178-1	56,446	56,726	281
		Intron 1-2	56,727	56,776	50

		OAG19178-2	56,777	61,948	5,172
		Intron 2-3	61,949	61,997	49
		OAG19178-3	61,998	62,079	82
CC77DRAFT_1061482	KV441478: 1,105,915-1,107,721	OAG20922-1	1,105,915	1,106,111	197
		Intron 1-2	1,106,112	1,106,167	56
		OAG20922-2	1,106,168	1,107,420	1,253
		Intron 2-3	1,107,421	1,107,473	53
		OAG20922-3	1,107,474	1,107,721	248
CC77DRAFT_982264	KV441472: 18,545-21,977	OAG23547-1	18,545	18,906	362
		Intron 1-2	18,907	18,972	66
		OAG23547-2	18,973	19,189	217
		Intron 2-3	19,190	19,253	64
		OAG23547-3	19,254	20,454	1,201
		Intron 3-4	20,455	20,509	55
		OAG23547-4	20,510	21,581	1,072
		Intron 4-5	21,582	21,636	55
		OAG23547-5	21,637	21,977	341