Small RNAs in Tomato:

From defence to development

Alex Cantó-Pastor

Department of Plant Sciences
University of Cambridge

This dissertation is submitted for the degree of
Doctor of Philosophy

Darwin College

February 2018

'Podria viure ignorant
la història meva que no vull,
podria viure ignorant
[...]
però encara sento que cada matí,
un sol neix entre els meus dits,
una claror que em fa aixecar: viure! viure!'

Joan Isaac, *Viure* (1977)

A mons pares i ma germana

Declaration of originality

This dissertation is the result of my own work and includes nothing which is the outcome of work done in collaboration except as specified in the text.

It is not substantially the same as any that I have submitted, or, is being concurrently submitted for a degree or diploma or other qualification at the University of Cambridge or any other University or similar institution. I further state that no substantial part of my dissertation has already been submitted, or, is being concurrently submitted for any such degree, diploma or other qualification at the University of Cambridge or any other University or similar institution.

It does not exceed the prescribed word limit for the Biology Degree Committee.

Alex Cantó-Pastor
February 2018

Acknowledgements

All the work accomplished here would have not been possible without the help and assistance of others, directly or indirectly. Firstly, I would like to thank Professor David Baulcombe for the mentorship and support which has guided me throughout this journey, and for the encouragement and freedom to find and follow my own scientific curiosity. I would also like to thank all the members of the Baulcombe and Henderson labs, past and present, for all the great discussions and help. For all the assistance, I would like to thank James, Mel, Antonia, Pawel, and Ombretta. Thanks to my comrades in arms Claire, Bruno, Luke and Catherine. I am also thankful to Sarah and Sasha, who endured my terrible writing and helped me improve it. Finally, I will always be in debt with Sebastian Schornack, who has been both a mentor and a friend, and instrumental to this thesis.

Friends outside the lab turned all stormy weather into smooth sailing. Thanks to DCBC and other Darwinians, for being a great bunch. To Charlie, for putting up with my indecision and quirks. And most importantly, to Poppy and Lucy, who welcomed me into their house and made it a second family.

Nabil Abderrahaman merece una mención especial, imprescindible en los madrugones, las gourmets, y el vinagre. Gracias también a Sara, Adrián y Fran, por las risas y los consejos. A Daniel, Héctor y Joan, por compartir la ropa y las malas decisiones (con los mejores resultados). A Piu, pels pelícans.

Per últim, gràcies a tota la meua família pel seu tarannà i amor incondicional. A la meua super-heroïna particular: ma germana. I finalment, a mons pares per... tot. No hagués pogut aconseguir tot açò sense la vida que ens heu dona't a Andrea i a mi. No vos mereixem. Gràcies. Sempre.

Abstract

RNA silencing is a major regulator of gene expression in plants, controlling from development to transposable element silencing and stress responses. As part of the silencing machinery, micro (mi)RNAs orchestrate silencing of their targets, either directly or through cascades of secondary small interfering (si)RNAs.

To investigate the role of RNA silencing in plant immunity, I chose to focus on the miR482/2118 family, because of its diversity and presence in many plant species since the appearance of seed plants, with most genomes containing several copies, and because its members target sequences conserved in a family of disease resistance genes known Nucleotide biding site leucine-rich repeat (NLR) genes.

In this dissertation, I wanted to address the extent to which the miRNA family and its derived phasiRNAs regulate expression of defence genes as well as contribute to quantitative resistance in crops. I explore the structural differences of miR482/2118 members in *Solanum lycopersicum* and show that they are functionally significant and affect their target preferences. My approach was based on small RNA sequencing and degradome data to characterize targets of these miRNAs, including the recently discovered tomato TAS5 locus.

I also generated transgenic tomatoes constitutively expressing target mimic RNAs that sequester different miR482/2118 members. These tomato mimic RNA lines were less susceptible than their non-transgenic precursors to pathogens *Phytophthora infestans* and *Pseudomonas syringae*.

Additionally, I investigated the role of small RNAs and their effector proteins during vegetative and reproductive development in tomato. I employed transcript and small RNA sequencing and CRISPR-Cas9 techniques of gene editing to investigate the impact of these factors in gamete viability and transposable element silencing in vegetative meristems.

The results presented here provide new evidence about the extent that RNA silencing contributes to the regulation of vital processes in plants. My study primarily explores the extent to which structural differences between the members of the miR482/2118 family affect their range of action, and the use of target mimics against these miRNAs as biotechnological approach for enhancing disease resistance in highly bred cultivars.

Table of contents

List of figures

List of tables

Nomenclature

Acronyms / Abbreviations

AGO Argonaute family protein

ath *Arabidopsis thaliana*

CFU Colony forming units

DAMP Damage-associated molecular pattern

DCL Dicer-like gene

dsRNA Double-stranded RNA

ETI Effector-triggered immunity

GFP *Aequorea victoria* green fluorescent protein

HR Hypersensitive response

HTS High-throughput sequencing

LTR Long-terminal repeat

miRNA micro RNA

MS Murashige-Skoog

NB-ARC NLR domain standing for nucleotide-binding adaptor, APAF-1, Resistance, CED-4

NLR Resistance genes consisting of nucleotide-binding and leucine-rich repeat domains

nt Nucleotide

PAMP Pathogen-associated molecular pattern

phasiRNA phased, secondary siRNA

PHAS phasiRNA-producing loci

Pi 88069 *Phytophthora infestans* strain 88069

pNLR Phasi-producing NLR

PolII RNA polymerase II

PPR Pattern recognition receptor

PTGS Post-transcriptional gene silencing

PTI PAMP-triggered immunity

Pto DC3000 *Pseudomonas syringae* pv. *tomato* strain DC3000

R-gene Resistance gene

RdDM RNA-directed DNA methylation

RdRC RNA-dependent RNA polymerase complex

RDR RNA-Dependent RNA polymerase

R gene Resistance gene

RITSC RNA-induced transcriptional silencing complex

RLK Receptor-like kinase

RLP Receptor-like protein

RPKM Reads per kilobase of transcript per million mapped reads

rRNA Ribosomal RNA

siRNA small interfering RNA

sly *Solanum lycopersicum*

sRNA small RNA

ssRNA single stranded RNA

SS Silencing suppressor

TE Transposable element

VN Vegetative nuclei

Prologue

Life and music are often linked in our culture. Two very distinct processes that originate from a common fundamental concept: dynamics. The nature of life, much like music, resides in constant change and contrast to achieve progress; and the cell, as its conductor, must control these changes to avoid the collapse of the song. The genome and its genes can be imagined as the score and melodies of this symphony. It is through careful orchestration of the melodies that the conductor delivers the music. In the same way, the ability of the cell to regulate gene expression will determine the progress of life. Biologists have spent decades reading notes in the score trying to hear the music, and along the way, several elements that aid the conductor on the interpretation of the piece have been discovered. Among these elements, small RNAs help decide which melodies should be accentuated or silenced.

Chapter 1

Introduction

1.1 The big world of small RNAs

RNA silencing is a major regulator of gene expression in plants, controlling from development to transposable element silencing and stress responses. Among all the components of this machinery, small RNAs (sRNAs) are the cornerstone of the process, acting as the specificity determinants of proteins that trigger RNA degradation, translation inhibition or epigenetic modifications.

1.1.1 Biogenesis of small RNAs

At their simplest level, all RNA silencing pathways follow a basic model. A long single stranded RNA (ssRNA) molecule is converted into a long double stranded RNA (dsRNA), either by folding into itself in a hairpin conformation, by binding another complementary ssRNA, or through synthesis of the complementary strand by an RNA-directed RNA polymerase (RDR) protein. The resulting dsRNA acts as a substrate for Dicer-like (DCL) class RNase-III endonucleases. A DCL protein slices the long dsRNA into short dsRNA molecules, and one of the strands of these short dsRNA molecules is incorporated into an Argonaute (AGO) protein. The loaded AGO will identify its target via base pair

complementarity with its carried sRNA, recruit other effectors, and perform downstream silencing functions (Figure 1.1A). It is the sum of the singularities of all players involved in the process that will ultimately decide the fate of the target.

Dicer RNase-III endonucleases

Most plant sRNAs are 21 to 24 nucleotide (nt) RNA molecules as a result of the particular processing of larger RNA molecules by DCL proteins. DCL proteins are conserved between plants and animals and display a similar domain structure consisting of RNA helicase, PAZ, RNase-III and dsRNA-binding (dsRBD) domains [Margis et al., 2006]. The helicase domain is required for accurate processing of the long dsRNA molecule [Liu et al., 2012], while the PAZ domain and its distance to the catalytic RNase-III domain act as a ruler that determine the size of the small dsRNAs produced [Bologna and Voinnet, 2014] (Figure 1.1B).

Analyses of these proteins in plants are largely based on the model organism *Arabidopsis thaliana*, which coincidentally contains one copy of each of the four main plant DCL clades [Bologna and Voinnet, 2014]. All four DCL proteins have been well characterized and have been shown to participate in distinct RNA silencing pathways [Gasciolli et al., 2005; Henderson et al., 2006] (Figure 1.1C).

DCL1 localizes in the nucleus and is the main processor during plant microRNA (miRNA) biogenesis [Liu et al., 2012]. Plant miRNAs are typically 20 to 22 nt in length, and arise from endogenous genes transcribed by RNA polymerase II (PolII) [Axtell, 2013]. The initial long primary miRNAs (pri-miRNAs) are ssRNA molecules that fold into themselves and form hairpin-like

structures. These hairpin structures have imperfect complementary between the folded regions, which is thought to be one of the determining characteristics for DCL1 processing. Pri-miRNAs are processed by DCL1 into shorter imperfect stem-loop structures called precursor miRNAs (pre-miRNAs), which are then processed again into a mature miRNA duplex consisting of the miRNA and its complementary sequence [1] [Bologna et al., 2009].

The other DCLs (DCL2/3/4) process the vast population of sRNA that are not miRNAs, which are historically referred to as small interfering RNAs (siRNAs) [Axtell, 2013]. Knock out combinations of *DCL2*, *DCL3* and *DCL4* genes have shown that all bind perfect or near perfect dsRNA precursors to produce 22-, 24- and 21-nt siRNAs, respectively. Interestingly, the dsRNA precursors that these DCLs target appear to be partly redundant, and their fate depends largely on the availability of each particularly DCL [Gasciolli et al., 2005; Marí-Ordóñez et al., 2013; Pumplin et al., 2016].

DCL3-derived 24-nt siRNA arise from dsRNAs generated by RDR2, and are subsequently loaded into AGO4 to act in transcriptional gene silencing (TGS). Most of the dsRNA precursors originate from repetitive regions of the genome undergoing RNA-directed DNA-methylation (RdDM), and consequently, DCL3, RDR2 and AGO4 localize predominantly in the nucleus [Pontes et al., 2006].

DCL4 produces phased 21-nt siRNAs (phasiRNAs), including endogenous trans-acting siRNAs (tasiRNAs), and have been shown to trigger post-transcriptional gene silencing (PTGS) in downstream targets by cleavage and/or translational repression [Baumberger and Baulcombe, 2005; Xie et al., 2005]. Interestingly, *DCL4* encodes for two different epigentically regulated isoforms. The main form is localized in the cytoplasm, while the other contains a nuclear localization

[1] the sequence which pairs with the mature miRNA is usually referred to as the mature star or miRNA*

signal and therefore accumulates in the nucleus [Pumplin et al., 2016].

DCL4 and DCL2 are phylogenetically closer to each other than they are to any other DCLs of plants, and both are thought to have partially overlapping or complementary functions (Figure 1.1C). DCL2 produces 22-nt siRNAs from the same set of precursors as DCL4. However, differently to DCL4-derived 21-nt siRNAs, DCL2 22-nt siRNA promote the production of secondary siRNAs at their targets and increase silencing transitivity [Parent et al., 2015].

RNA-dependent RNA polymerases

Long dsRNA that give rise of siRNAs can be produced from folding back into itself, by binding another complementary ssRNA, or through synthesis of the complementary strand. RDRs are the proteins responsible for the latter, as their catalytic RdRP domain is required for copying ssRNA into dsRNA (Figure 1.1B).

Plants contain several RDR proteins that belong to two distinct clades, RDRα and RDRγ, which originated in the eukaryotic ancestor shared with animals and fungi [Zong et al., 2009]. Members of the RDRγ family have not been assigned any role in RNA silencing and remain largely uncharacterised. *A. thaliana* contains three members of the RDRα clade. RDR2 function is associated with TGS through the RdDM pathway. RDR2 proteins localize in the nucleus and convert RNA polymerase IV- (PolIV-) derived transcripts into dsRNA for DCL3 processing [Pontes et al., 2006]. RDR1 is mainly involved in antiviral RNA silencing and interacts exogenous RNAs such as viral genomes [Garcia-Ruiz et al., 2010]. Lastly, RDR6 is involved in the conversion of PolII derived transcripts into dsRNA for the production of 21- and 22-nt siRNA via DCL4/2 processing. Interestingly, RDR6 localizes in discrete cytoplasmic granules, along with other protein effectors of the siRNA pathway [Kumakura

et al., 2009].

Argonautes

AGO proteins were named after the phenotype of the *Arabidopsis thaliana* mutant *ago1*, as its leaves take the shape of little tentacles resembling argonaut octopuses [Bohmert et al., 1998]. AGO proteins are highly specialised RNA-binding molecules that incorporate the sRNAs and coordinate downstream targeting and effects through interacting with other silencing complex effectors [Bologna and Voinnet, 2014]. Different AGO discriminate classes of sRNAs largely based on their 5' nucleotide identity and length [Mi et al., 2008]. This protein family is present in most eukaryotes and some prokaryotes, but with great variation in number and diversity [Zhang et al., 2015].

Most characterized plants contain at least 10 members of the AGO family in their genomes which can be classified into three large clades based on amino-acid homology (Figure 1.1C). Clade I contains the AGO1-like members, involved in 21/22-nt miRNA- and siRNA-directed PTGS [Vaucheret et al., 2004; Zhu et al., 2011]. A distinct sub-clade of AGO5-like members exist within the clade, reported to bind miRNAs and phasiRNAs, and act in germline development [Nonomura et al., 2007; Tucker et al., 2012; Zhai et al., 2015].

Clade II contains AGO4-like members, involved in transcriptional gene silencing of genes and repeats through the RdDM pathway [Gao et al., 2010]. Members of this clade are generally thought of as 'chromatin modifiers', since during their action other effector are recruited to trigger DNA methylation of cytosines.

Lastly, clade III is a very diverse clade containing the AGO2/3/7 type members, with several unrelated and still unclear functions. This clade was initially thought to bind anti-sense siRNAs (both from endogenous and viral loci) and be predominantly involved in immunity, mostly based on the characterization of AtAGO2 [Jaubert et al., 2011]. However, recently it was shown that AtAGO2 also mediates double strand break (DBS) repair [Wei et al., 2012], and AtAGO3 (a very close homolog whose gene is situated 3kb apart from *AtAGO2* in a tandem repeat) has been shown to be capable of participating in epigenetic silencing through RdDM in a similar fashion of members of Clade II [Zhang et al., 2016b]. A highly conserved and specialized sub-clade of AGO7-like exists within in the clade, with an also conserved role in binding miR390 and regulating *TAS3/ARFs* function in leaf patterning [Adenot et al., 2006].

1.1.2 Functions of small RNAs

The different classes of small RNAs present in plants are classified based on the distinct functions they perform, which ultimately depend on their biogenesis pathways [Axtell, 2013]. This classification is generally simplified by dividing actions into two main categories: PTGS or TGS.

Post-transcriptional gene silencing and miRNAs

Plant 21- and 22-nt miRNAs participate in RNA silencing through PTGS, which results in inhibiting translation or degrading their target mRNA. Upon processing of the pre-miRNA and loading into AGO1, the passenger strand of a mature miRNA/miRNA* duplex is removed. miRNA-loaded AGO1 finds its target through Watson-Crick base pair complementarity, bringing along other members of the RNA-induced silencing complex (RISC) for subsequent silencing. Plant miRNA-target interactions have been shown to require the whole miRNA mature sequence to be near perfectly aligned with its mRNA

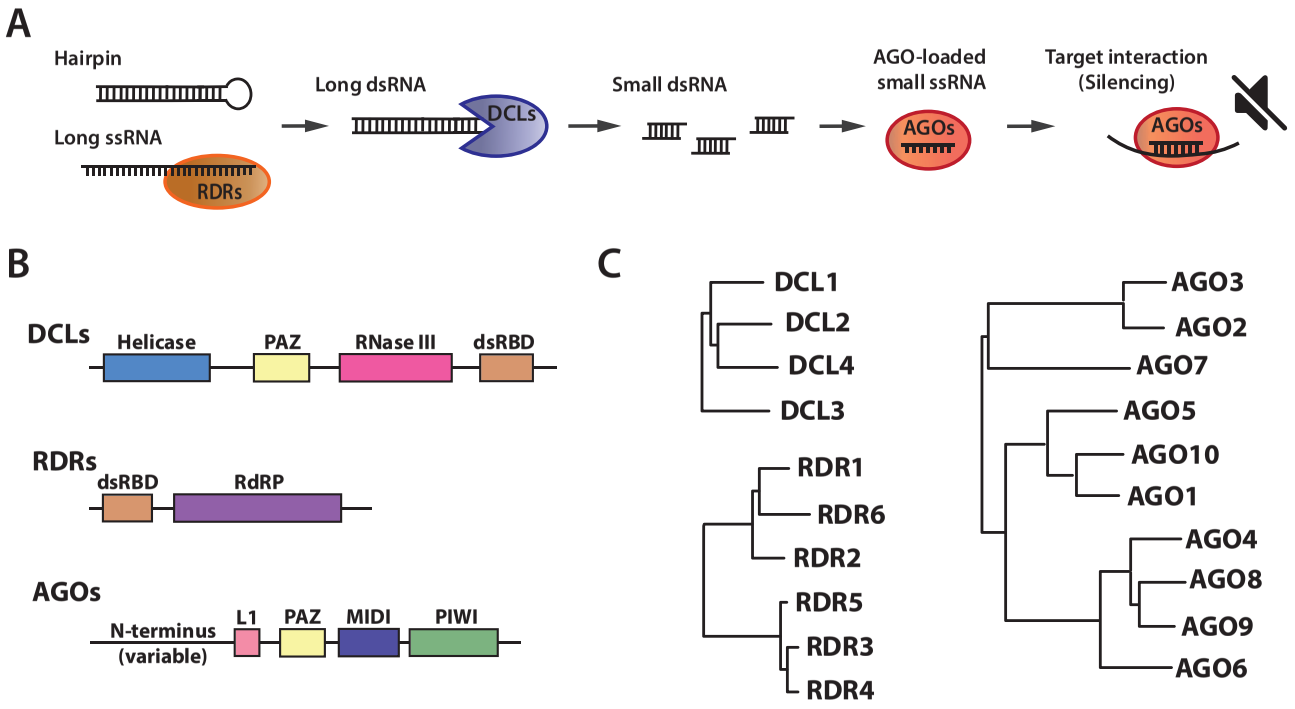

Fig. 1.1 **Simplified model and core components of RNA silencing.** (A) Simplified model of the general production and function of small RNAs. A long dsRNA molecule is originated through complementary binding (hairpin or IR) or RDR-derived synthesis. The product acts as substrate for a DCL protein to produce small dsRNA molecules. One of the strands of the RNA molecule is incorporated into an AGO protein and is used to identify the target through base pair complementarity and perform silencing functions. (B) Conserved domain structure and (C) phylogenetic classification of each class of core proteins in RNA silencing. Figure based on Bologna and Voinnet [2014].

target, differently to their animal counterparts where only loose complementarity in about first eight nucleotides of the miRNA is required [Dai and Zhao, 2011].

Most of the PTGS occurs through the slicing activity of AGO1 at positions 10-11 of the paired RNA molecules. The major relevance of the slicing activity of AGO1 for PTGS is indicated by severe developmental defects of a non-catalytic *ago1* mutant and the negative dominant effects of ectopically expressing that same allele in a wild type plant [Carbonell et al., 2012]. However, in particular situations with imperfect pairing around the centre of miRNA/mRNA complementary region, slicing is not produced and transcripts undergo translational repression instead [Brodersen et al., 2008]. Translational repression in plants remains largely unexplored due to its technical challenges, and there have been reports of AGO1-mediated repression of translation without any mRNA decay [Iwakawa and Tomari, 2013].

A particular class of 22-nt miRNAs have also been found to additionally trigger the formation of phasiRNAs [Chen et al., 2010a]. The production of this 21- and 22-nt siRNAs requires the recruitment of RDR6, which occurs through the reprogramming of AGO1-RISC upon loading a 22-nt miRNA. Not only a 22-nt size, but a particular asymmetric structure of the miRNA/miRNA* duplex is required to trigger secondary siRNA biogenesis [Manavella et al., 2012]. Interestingly, in vitro analysis have sown that a 1-nt deletion at the 3' end of a phasiRNA-trigger miRNA or a mismatch at the 3' end of the target site abolishes the formation of phasiRNAs. These changes lead to a failure of the AGO1-RISC to recruit the factors necessary for the formation of the complex, such as SUPPRESSOR OF GENE SILENCING3 (SGS3) which in normal conditions stabilizes the 3' cleavage transcript and avoids its degradation (Figure 1.2) [Yoshikawa et al., 2013].

Transcriptional gene silencing and 24-nt siRNAs

DCL3-dependent 24-nt siRNAs[2] regulate transcription by inducing methylation of genomic cytosines through the RdDM pathway. Most of these siRNAs originate from a specialized transcriptional machinery that requires plant-specific RNA polymerases PolIV and PolV [Pikaard et al., 2012]. The most current model for the RdDM pathway suggests that PolIV transcripts are converted into dsRNA by RDR2 and subsequently processed into 24-nt by DCL3 and exported outside the nucleus. In the cytoplasm, these 24-nt siRNA are loaded into AGO4 (or other members of the AGO4-like clade) and then reintroduced in the nucleus. Back in the nucleus, loaded AGO4s target nascent transcript of PolV and recruit methyltransferase DRM2 to establish *de novo* DNA methylation at cytosines in any sequence context (reviewed in [Matzke and Mosher, 2014]).

Where PTGS and TGS meet: phasiRNAs and IR-siRNAs

PhasiRNAs are originated from both coding and non-coding RNA transcripts via the action of a miRNA. The miRNA recognises a single stranded ssRNA, cleaving the target and triggering the conversion of this ssRNA into dsRNA. DCL4 protein is then recruited and slices the dsRNA into 21-nt siRNAs. Since the miRNA always cleaves the ssRNA at the same position and DCL4 generally cuts in regular 21-nt intervals, most sRNAs produced from this process appear in particular register (Figure 1.2). The resulting sRNA signature of this process is what gives rise to the term "phased" [Fei et al., 2013]. In a small number of cases, these siRNAs can arise from endogenous inverted repeats, in which folded-back dsRNA do not require miRNA targeting and RDR6 processing and are incorporated only in the later stages of this pathway (and therefore do not

[2]Many authors refer to these as heterochromatic- (hc-)siRNAs, as the mostly originate from repetitive regions of the genome. However, I do not favour the term since transcribed regions (generally regarded as euchromatic) can also rise to this class of siRNAs by processing the PolII/RDR6-dependent dsRNAs through DCL3 upon saturation of DCL4/2 [Marí-Ordóñez et al., 2013].

present a phased signature) [Kasschau et al., 2007].

TAS genes were the first phasiRNA loci described due to their highly evolutionary conservation, the non-coding nature of the loci, and their important roles in development which ultimately easy to identify [Allen et al., 2005]. Due to the ability to target complementary mRNAs *in trans* and induce silencing, these TAS-derived siRNA were originally named trans-acting small interfering RNAs (tasiRNAs) [Allen et al., 2005]. However, many other phasiRNA-generating loci (named *PHAS*) have then been described since, some of which with no obvious targets or function [Zhai et al., 2015]. Some of these *PHAS* loci include protein-coding gene families with important roles in development, such as MYB transcription factors, PPRs (involved with organelle RNA processes), or NLR genes (involved in plant defence) [Howell et al., 2007; Shivaprasad et al., 2012; Wang et al., 2011; Xia et al., 2012; Zhai et al., 2011]. For those reasons, the term '*TAS*' is nowadays relegated to describe only non-coding *PHAS* transcripts with no reported function other than to give rise to phasiRNAs involved in PTGS [Fei et al., 2013].

A long suggested link between PTGS and TGS was recently uncovered. An epigenetic study on a large population of *A. thaliana* mutants showed that the level of methylation in *dcl3* mutants is only slightly reduced, while in *dcl2 dcl4 dcl3* triple mutants is more strongly decreased [Stroud et al., 2014]. The overall results suggest that 21- and 22-nt siRNAs derived from DCL4/2 participate in RdDM and compensate for the loss of DCL3. These results also are in concordance with previous observations of hyper-methylation at TAS genes [Wu et al., 2012]. Further analyses discovered that these 21- and 22-nt siRNAs are incorporated into AGO6, which interacts with PolV to initiate methylation [McCue et al., 2014].

Based on current evidence, it has been suggested that RDR6-DCL2/4-dependent phasiRNAs could initiate the establishment of *de novo* DNA methylation that would then be maintained and enhanced by the canonical RdDM-dependent 24-nt siRNAs [Matzke and Mosher, 2014]. Therefore, while miRNAs regulate expression through PTGS and 24-nt siRNAs act through TGS, the role of phasiRNAs is more loosely defined and provide a link between these two processes in RNA silencing. Given their capacity to trigger silence of *cis* and *trans* targets through PTGS and/or TGS, and the difficulty to confidently identify targets, their exact functional relevance remains a matter of great speculation. This is particularly true for plant immunity, as many PHAS loci from genes involved in defence response have been identified in various plant species [Li et al., 2012b; Liu et al., 2014; Shivaprasad et al., 2012; Xia et al., 2015; Zhai et al., 2011].

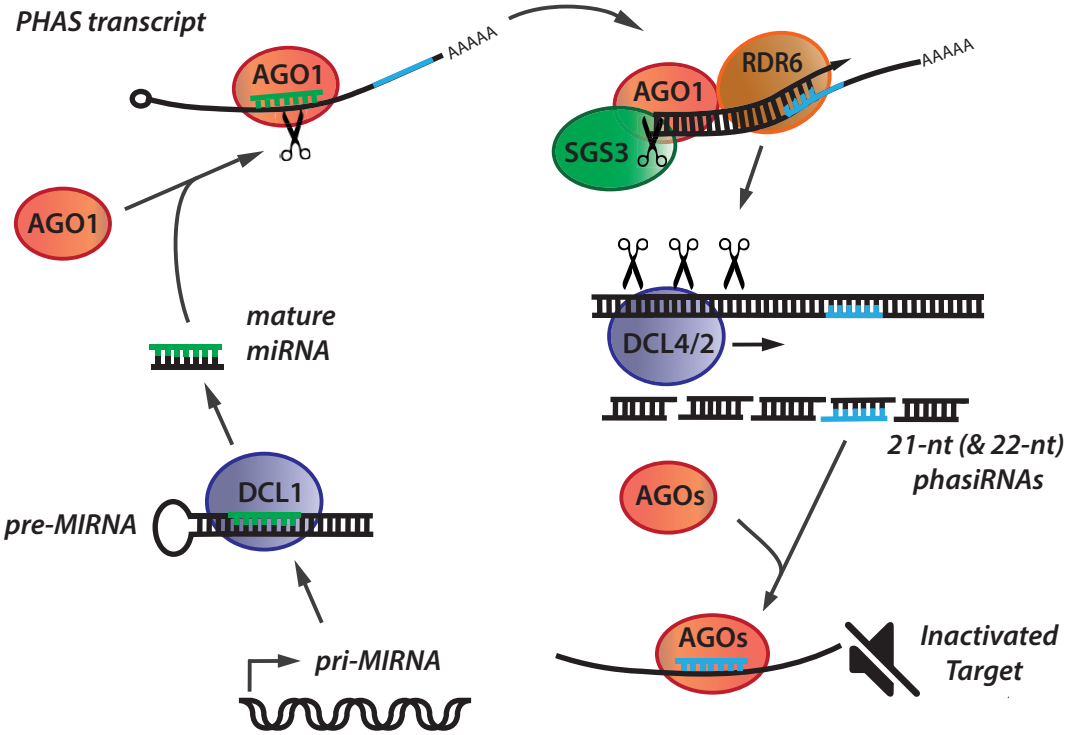

Fig. 1.2 **Phased small RNAs biogenesis and function.** Schematic model of the production and action of 21-nt phasiRNAs. In brief, a microRNA gene (*MIRNA*) gets transcribed and processed by DCL1 into its mature form. This microRNA will be loaded into AGO1 and target a long poly-adenylated RNA (with or without coding potential) that has been transcribed from the genome by PolII. The long ssRNA undergoes an initial AGO1-RISC mediated cleavage guided by the miRNA. SGS3 is recruited to stabilize the cleaved transcript and RDR6 then converts the ssRNA into a long dsRNA fragment. Lastly, this dsRNA gets processed by DCL4 (or DCL2) to generate a phased array of 21-nt (or 22-nt siRNA) starting at the miRNA cleavage site that will target complementary sequences downstream.

1.2 The plant immune system

1.2.1 An overview

Unlike animals, plants lack a circulating immune system. In contrast, the immunity of a plant relies mostly on a cell-autonomous response to fight pathogens. Each cell must be genetically equipped to respond to pathogen attacks, and in order to orchestrate such responses, plants have evolved a highly sophisticated immune system. This system is based on a combination of external and intracellular receptors that recognise infection-associated molecules and provide resistance to attacks by pathogens such as viruses, bacteria, fungi, oomycete, nematodes and insects [Jones and Dangl, 2006].

On the external surface of the cells, host-encoded pattern recognition receptors (PRRs) detect in the apoplastic space various ligands such as pathogen-associated molecular patterns (PAMPs)[3] or damage-associated molecular patterns (DAMPs) derived from the host and released during cell damage [Couto and Zipfel, 2016]. Membrane-bound PPRs are mostly receptor-like kineses (RLKs) and receptor-like proteins (RLPs). RLKs are generally formed by a ligand-binding ectodomain, a transmembrane domain and an intracelular kinase domain. RLPs share similar features with RLKs, but lack a intracellular signal domain such as the kinase and are thought to act in conjunction with other receptors [Couto and Zipfel, 2016]. Upon recognition of a ligand, activated PPRs trigger downstream an initial immune response known as PAMP-triggered immunity (PTI) (Figure 1.3b). PTI effectively prevents infection from most non-adapted pathogens and constitutes the basal layer of immunity during infection.

[3]Many authors use the terms PAMP and microbe-associated molecular pattern (MAMP) interchangeably. This is based on the fact that most microbes, not only pathogens, express the molecules detected by these receptors.

Adapted pathogens that have co-evolved with a particular host, secrete molecules known as effectors that can interfere with host immunity or manipulate the host metabolism for virulence [Jones and Dangl, 2006]. These effectors are generally delivered to the intracellular space of infected cells to interfere with the PTI and enable infection. For that reason, host plants have develop a second layer of defence mediated intracellular receptor proteins. These receptors are classically encoded by resistance (R) genes, which recognise delivered effectors, (also referred as avirulence or Avr proteins). Recognition of these effector can occur through direct interaction of the R gene and the effector or by sensing perturbations caused by the effector on host components [Cui et al., 2015] (Figure 1.3c-d). R gene activation trigger an immune signal that often leads to localized cell death, referred to as hypersensitive response (HR). This highly amplified version of defence response is known as effector-triggered immunity (ETI). In turn, pathogens and host have evolved multiple layers of effector action and recognition by the host, involving iterating cycles of defence and counter-defence (Figure 1.3e-f). Ultimately, this process has turned into an evolutionary arms race between the pathogen and the host, which explains why the protein families involved in this process are among the most rapidly evolving in plants [Chen et al., 2010b; Jacob et al., 2013].

Plants have also evolved an additional defence mechanism mediated by RNA silencing, initially discovered for its important role in antiviral immunity [Baulcombe, 2004]. Parallel to protein-protein interactions of the PTI and ETI layers, the plant RNA silencing machinery detects the presence of foreign RNA and directs cleavage or inhibition of genome replication (in the case of RNA viruses) or other pathogen derived transcripts [Deleris et al., 2006]. As a counter-defence mechanism, many pathogens encode effector proteins that tamper with RNA silencing, known a silencing suppressors (SS) [Pumplin and Voinnet, 2013].

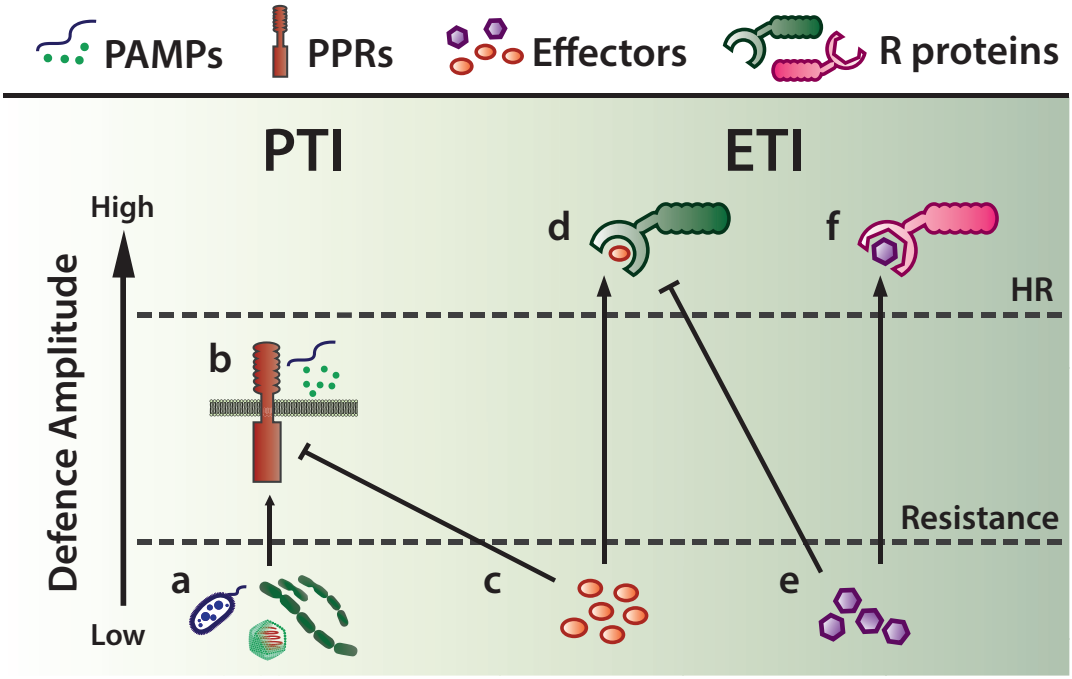

Fig. 1.3 **Simplified model of classic plant immunity.** Overview of the two levels of microbial recognition. (a) Upon infection of a pathogen, (b) the presence of extracellular pathogen-associated molecular patterns (PAMPs) is perceived by pattern-recognition receptors (PPRs) with an recognition ectodomain. Activation of PPRs triggers a downstream signalling cascade known as PAMP-triggered immunity (PAMP), which includes oxidative bursts, callose depositions and the induction of defence genes. (c) Specialised pathogens deliver effectors to interfere with the PTI and, more generally, manipulate the host to promote microbial growth and disease. (d) At the same time, plants encode for resistance (R) proteins that recognize pathogen effectors directly or by surveilling the integrity of the effector's targets. Activation of R proteins triggers a highly amplified defence response that generally leads to localized cell death, known as hypersensitive response (HR). This second layer is known as effector-triggered immunity (ETI). (e & f) Highly specialized pathogen and host interactions have evolved multiple levels of counter-defence and counter-counter-defence to confer optimal virulence or recognition, respectively. Image based on Pumplin and Voinnet [2013].

Although fundamentally distinct, RNA silencing and PTI/ETI pathways have been shown to act interactively and/or with a great deal of crosstalk [Bhattacharjee et al., 2009; Navarro et al., 2008; Shivaprasad et al., 2012].

1.2.2 R gene-mediated immunity

R gene-mediated resistance has been historically thought as the recognition of an avirulence factor of a pathogen by a host R protein, subsequently triggering active defence response. This was initially explained on the basis of a 'gene-for-gene' model based on the seminal findings of H. H. Flor when studying the rust of flax. Flor showed that the inheretance of both the flax resistance to the fungus and the ability of the fungus to cause disease were controlled by a matched pair of genes [Flor, 1956]. For that reason, the model originally postulated that every R gene in a plant corresponds to an avirulence gene in a pathogen [Flor, 1971].

To date, many R genes have been identified, cloned and classified based on the characteristics of their sequences. Most known R genes encode cytoplasmic nucleotide-binding, leucine-rich repeat (NLR)[4] proteins. NLRs are one of the largest family proteins in both plants and animals, and in both instances they are involved in immune responses and apoptosis and present a modular domain architecture [Jones et al., 2016]. NLRs are in turn part of a larger protein family of signal transduction AAA+ ATPases with numerous domains (STAND). A typical plant NLR contains a central nucleotide-binding site (NBS) domain which is usually involved in self-regulation and oligomerization, a C-terminal leucine-rich repeats (LRR) domain, which plays a role in ligand sensing and regulation, and a variable N-terminal domain usually containing a coiled-coil (CC) or Toll-interleukin-1 receptor (TIR) domain which often acts as a signalling

[4]Also commonly referred to as NB-LRRs or NBS-LRRs.

domain [Zhang et al., 2017c].

In the absence of a pathogenic cue, NLRs are tightly regulated and remain in an auto-inhibited state [Ade et al., 2007; Rairdan and Moffett, 2006]. Their activation involves major structural conformation changes in the protein upon detection of a pathogen signal, that lead from auto-inhibition to activation and downstream signalling [Cui et al., 2015]. The following sections explore the role of the different NLR domains in the different stages of this process.

The Nucleotide-Binding Domain and its role in NLR activation

The central plant NBS is also referred to as NB-ARC, standing for nucleotide-binding adaptor, APAF-1 (apoptotic protease-activating factor 1), Resistance, CED-4 (cell-death protein 4). This is because the domain is shared with animal proteins involved in apoptosis such as human APAF-1, *Drosophila* DARK (drosophila apaf-1 related killer) and *C. elegans* CED4 , which probably evolved from a class of prokaryotic ATPases [Yue et al., 2012].

The NB-ARC generally consists of three subunits: The nucleotide binding (NB) subdomain, and two helical subdomains called ARC1 and ARC2. Although no structure is available for an NB-ARC, biochemical evidence and information from crystal structures of homologous domains in animals seem to indicate that the NB-ARC controls the activation state of NLRs based on the nature of the nucleotide bound to the domain [Zhang et al., 2017a]. In the presence of the protein elicitor (the avirulence molecule), the NB subdomain is thought to exchange adenosine diphosphate (ADP) for adenosine triphosphate (ATP), driving downstream signalling and/or possibly oligomerization. Several plant NLRs have been shown to hydrolyse ATP to ADP, which has been suggested to promote the changes that bring an active NLR to its resting state, like in animals [Tameling et al., 2002, 2006; Zhang et al., 2017c]. A highly conserved

motif in the NB subdomain in animal and plants known as the P-loop contains a lysine that interacts with the phosphate of ADP or ATP and involved in nucleotide binding and hydrolysis [Leipe et al., 2004]. Mutations of this domain results a signalling-dead NLR incapable of triggering an HR, such as potato Rx [Bendahmane et al., 2002], flax M [Williams et al., 2011], or tomato I2 and NRC1 [Sueldo et al., 2015; Tameling et al., 2002]. Another highly conserved motif present in the ARC2 subdomain known as MHD also plays an important role in NLR activation through interacting with the ADP/ATP. Mutations in this subdomain often lead to a constitutively active (auto-activated) state of the NLR, in which the presence of the elicitor is no longer required [Sueldo et al., 2015; Williams et al., 2011].

The C-terminal LRR domain and its role in recognition, regulation and auto-inhibition

The LRR domain is usually comprised of repeats consisting of a β-strand and α-helix connected by loops, although there is significant variation between NLRs [Kobe and Kajava, 2001]. The most recent models reveal a compact 'horseshoe'-like structure that interacts with the NB-ARC and dissociates upon activation [Sela et al., 2012; Slootweg et al., 2013]. These intra-molecular interactions are thought to stabilize an inactive signalling conformation [Zhang et al., 2017c]. In many cases, deletions or swapping of the LRR domain can result in auto-activation in the absence of the elicitor [Ade et al., 2007; Bendahmane et al., 2002; Howles et al., 2005]. Additionally, individual domains of *A. thaliana* RPS5 co-immunoprecipitate with each other and with the full-length protein, which suggests that all domains contribute to intra-molecular interactions [Ade et al., 2007]. Ultimately, it is likely that extensive (inter- and) intra-molecular interactions are required to stabilize the signal and conformation of NLRs.

Other studies have suggested that the LRR domain was also responsible for recognition specificity. Domain swaps between closely related NLRs in barley (RAR1/SGT1), potato (Rx/Gpa2), and flax (L5/L6) showed that their recognition specificity was localized in the LRR domain [Rairdan and Moffett, 2006; Ravensdale et al., 2012; Shen et al., 2003]. Additionally, random mutagenesis of the LRR domain of Rx created novel variants that recognised a wider variety of strains of its pathogen target, the Potato Virus X (PVX), than the wild type Rx [Farnham and Baulcombe, 2006].

In some cases, the LRR domain has been shown to directly interact with the elicitor using yeast two hybrid assays, as in the case of rice Pita with the avirulence protein AvrPita, and flax L with AvrL [Ellis et al., 2007; Jia et al., 2000]. However in other NLRs, different domains (and not the LRR) have also been shown to interact with their elicitors. The N-terminal domains of rice Pik, tobacco N and potato RB physically interact with fungus *Magnaporthe oryzae* Avr-Pik, tobacco mosaic virus (TMV) p50, and oomycete *Phytophthora infestans* IPI-O, respectively [Burch-Smith et al., 2007; Chen et al., 2012; Kanzaki et al., 2012]. Similar findings were reported for the non-LRR C-terminal domain of rice RGA5, as it was sufficient to bind *M. oryzae* Avr-Pia and Avr1-CO39 [Cesari et al., 2013]. All the evidence suggests that the LRR greatly contributes to proper recognition and limit inappropriate activation. However, since these functions essentially depend on major structural changes, it is unsurprising to observe that other domains also contribute to these. Ultimately, it is through the intra-molecular interactions between all domains that function is probably achieved in most NLRs.

The N-terminal domains and their role in signalling

Plant NLRs are broadly classified based on the type of domain contained in their N-terminal domain regions. Canonical NLRs commonly fall into two different classes: CC- and TIR-domain-containing NLRs and referred to as CNLs and TNLs, respectively. Some authors classify NLRs in TNLs and non-TNLs (nTNLs), as they prefer to make a clear distinction between 'canonical' CNLs and members of a sister family they refer to as RNL. RNLs contain a CC-like domain similar to the *A. thaliana* RESISTANT TO POWDERY MILDEW 8 (RPW8) protein [Shao et al., 2016]. However, since it is thought that RNLs arose from an ancestral CNL, other authors simply refer to this group as the CC_R sub-class of CNLs [Collier et al., 2011].

Plant NLRs are similar to the mammalian NOD class proteins[5] [Jones et al., 2016]. Upon activation, NOD proteins oligomerise to trigger inflamatory responses [Jones et al., 2016]. As well as in animals, oligomerisation of some plant NLRs represents one of the earliest steps following activation [Zhang et al., 2017c]. Several studies have shown how CC and TIR N-terminal domains are necessary and sufficient for this process. N-terminal domain self-association is essential in many NLRs, including TNLs such as flax L6, or *A. thaliana* RPP1, RPS4 and SNC1, and CNLs such as barley MLA. Isolated expression of these domains leads to self-association and elicitor-independent HR signalling, and mutations that abolish their self-association suppress their activity [Bernoux et al., 2011; Maekawa et al., 2011; Schreiber et al., 2016; Williams et al., 2014; Zhang et al., 2017b]. There are exceptions like the potato Rx and *A. thaliana* RPS5, where their CC domains do not trigger HR but the NB-ARC does, suggesting additional and alternative functions for different N-terminal domains [Ade et al., 2007; Rairdan et al., 2008].

[5]Some plant pathologists refer to the NLR acronym as "NOD-like receptor proteins".

Aside from self-association, some N-terminal domains have also been shown for heterogeneous interactions, which can influence the action either positively or negatively [Zhang et al., 2017c]. *A. thaliana* RPS4 TIR domain can heterodimerise with RRS1, which suppresses auto-activity [Williams et al., 2014]. A similar effect is found in rice paired CNLs RGA4 and RGA5, although RGA4 CC domain by itself self-associates but is insufficient for signalling [Cesari et al., 2013]. Whether oligomerisation, homo- and hetero-, is a common feature of plant NLRs is a topic for further study.

1.2.3 NLR function defined by its sensing nature

Pathogen recognition of NLRs through functional analysis have yielded multiple models of action. As sensors, it was initially suggested that NLRs acted through direct recognition of effectors. Although this model holds true for some NLRs, others have been shown to act by surveilling the effector targets, other NLRs or downstream signalling pathways [Zhang et al., 2017c].

Direct recognition

In support to H. H. Flor's first hypothesis, the flax TNLs L5/6/7 and M directly recognise the fungal flax rust effectors AvrL567 and AvrM by physical interactions [Catanzariti et al., 2010; Dodds et al., 2006]. Such is the case for the *A. thaliana* RPP1 and the oomycete downy mildew effector ATR1 [Krasileva et al., 2010]. However, the evidence for direct interactions of NLRs with effector is limited. This is possibly due to the evolutionary inefficiency of such system, since pathogens (with their shorter life cycles) would be able to diversify their effectors faster than plants could their NLRs. It would also be highly costful from a genomic perspective, as it would require maintaining an NLR repertoire

as broad as potential pathogens exist for the plant.

Indirect recognition

Independently evolved effectors from diverse pathogens have been shown to target the same plant proteins to enable virulence and facilitate their infection strategies [Mukhtar et al., 2011; Weßling et al., 2014]. Some NLRs recognize infection by monitoring these disease-associated changes in the host, effectively allowing a broader range of recognition with the minimal amount of genes.

When the host proteins monitored by NLRs are involved in cellular processes of the host immunity, these are called 'guardees'. One example includes the *A. thaliana* RPM1-interacting protein 4 (RIN4), a negative regulator of basal immunity, that is targeted by various bacterial *P. syringae* effectors. RIN4 was identified via its interaction in yeast with AvrB [Mackey et al., 2002]. In normal conditions, RIN4 functions as a negative regulator as it is required for accumulation of NLRs such as RPM1 [Mackey et al., 2002]. Cleavage or phosphorylation of RIN4 by three unrelated *P. syringae* type III effectors (AvrRpm1, AvrRpt2, and AvrB) is sensed by CNLs RPS2 and RPM1, which ultimately triggers immunity [Axtell and Staskawicz, 2003; Mackey et al., 2002].

In other cases, the monitored protein is only mimicking an effector target, acting as a bait, and are generally referred to as 'decoys'. A classic example of a 'decoy' is the *A. thaliana* protein kinase PBS1, which is cleaved by the bacterial *P. syringae* effector AvrPphB. The effect is monitored by CNL RPS5 to trigger HR [Ade et al., 2007]. However, unlike other PB1-like kinases also cleaved by AvrPphB (such as BIK1, a central components of PTI signaling), mutations in PBS1 kinase domains do not affect virulence, suggesting that other kinases

(and not PBS1) are the intended targets of the effector [Zhang et al., 2010].

Indirect recognition through integrated domains and paired NLRs

In certain cases, NLRs have integrated non-canonical domains such as BED zinc fingers, hydrolases, protein kinases, WRKYs, or HMAs. In rice, CNLs RGA5 and Pik-1 contain HMA domains that directly bind different rice blast fungus effectors [Cesari et al., 2013; Kanzaki et al., 2012]. Both domains resemble rice blast susceptibility protein Pi21, and it is likely that they are acting as a target decoy for the fungal effectors [Fukuoka et al., 2009]. Similarly in *A. thaliana*, TNL RRS1 contains a WRKY domain that binds bacterial effectors AvrRPS4 from *P. syringae* and Pop2 from *Ralstonia solanacearum* (and in some *A. thaliana* ecotypes, RRS1 also recognises fungus *Colletotrichum higginsianum*) [Narusaka et al., 2009]. Both AvrRPS4 and Pop2 have been shown to also target several WRKY transcription factors involved in the regulation of defence genes, so it is also like that the integrated domain in RRS1 is acting as a decoy for these effectors [Le Roux et al., 2015; Sarris et al., 2015]. Whether these domains act exclusively a 'decoy' for pathogen effectors that target TFs or could also function as transcriptional regulators is currently unknown.

Interestingly, the NLRs presented in the previous paragraph share another characteristic other that containing an integrated domain: they require another NLR to initiate the signalling cascade. The NLR containing the integrated domain acts a sensor of the pathogen, and the other acts as a 'sensor' of the first NLR and as signal transducer. Rice CNLs RGA5 and Pik-2 recognise RGA4 and Pik-1, respectively, while *A. thaliana* TNL RPS4 recognises RRS1 [Cesari et al., 2013; Kanzaki et al., 2012; Narusaka et al., 2009]. Strikingly, pairs that interact physically are linked genomically in a head-to-head arrangement,

sharing a common promoter, which also suggest co-regulation of expression.

Helpers

Recently, certain NLRs have been identified acting downstream in the signalling cascade. These cooperative NLRs have been termed as 'helpers' and only three distinct sets have been identified so far. The first two sets comprise the members of the CC_R sub-class of CNLs. As mentioned earlier, this corresponds to NLRs containing an RPW8-like CC-domain similar to the *A. thaliana* non-NLR R protein RPW8 which confers broad spectrum resistance to powdery mildew [Collier et al., 2011]. These are found as a distinct basal clade of the CNLs and appear in low numbers in all angiosperms [Shao et al., 2016]. The few CC_R members that have been characterized are *N. benthamiana* NRG1 and *A. thaliana* ADR1 (and other ADR1-like members), which in both cases are suggested to function as positive regulators of several NLR-mediated immune responses [Bonardi et al., 2011; Peart et al., 2005].

The third set of 'helpers' correspond to the members Solanaceae-specific CNLs of the NRC family (NLR Required for Cell death). In *N. benthamiana*, members of the NRC family NRC2/3/4 redundantly contribute to immunity mediated by various canonical CNL sensors [Wu et al., 2017]. The redundancy of their functions in the signalling network is suggested to enhance the robustness and 'evolvability' of the system. An important difference between 'helpers' and the signal transducers of paired NLRs is that the former have not been found to physically interact with the NLRs they assist.

1.2.4 Regulation of NLR expression

The negative implications of plant immunity in overall plant fitness due to pleiotropic effects of an activated defence response have been the focus of numerous studies over the recent years (reviewed in Karasov et al. [2017]). Several mechanisms keep NLR genes under tight regulation by limiting their expression and/or activation in unchallenged conditions while still being able to induce a rapid immune response in the presence of pathogens. These mechanism range from alternative splicing (AS) to non-sense mediated decay (NMD), epigenetic regulation, and RNA silencing (described in detail in the next section).

One of the first evidence of mechanism of NLR regulation was the expression of different functional variants of the tobacco TNL N gene achieved through AS mechanisms [Dinesh-Kumar and Baker, 2000]. In plants, AS can occur either constitutively or under particular developmental or stress conditions. Two AS variants of N dynamically were found in tobacco, a short N_S mRNA encoding for the functional N protein and a long N_L mRNA that lead to a frame shift and an early stop codon that excluded the LRR from the protein. N_S was found prevalently before infection, whereas N_L took over after TMV infection. Interestingly, the ability to produce both variants was required for full resistance against TMV [Dinesh-Kumar and Baker, 2000].

Some NLR transcripts generated as a consequence of AS, mutations, or transcription errors have been shown to recruit the NMD machinery, an RNA surveillance mechanism that exists in all eukaryotes. The NMD machinery detects mRNAs with premature termination codons, upstream open reading frames (uORF), long 3' untranslated regions (3' UTRs), or other aberrant transcripts, and trigger mRNA decay. Several AS TNLs have been shown to be regulated by the NMD pathway, and accumulate in mutants of the pathway or during infection, leading to an (auto)immune response [Gloggnitzer et al., 2014;

Riehs-Kearnan et al., 2012]. Both AS and NMD-mediated regulation of NLR expression have only been found in TNLs. CNLs are probably preferentially regulated through other mechanisms, suggested by the general absence of conserved introns in this class type [Shao et al., 2016].

Recently, epigenetic regulation has also been shown to contribute to NLR regulation. In rice, the *Pigm* locus confers resistance to rice blast fungus. The locus contains two CNLs, PigmR and PigmS that have opposing functions. PigmR is responsible for the resistance via homodimerization, while PigmS antagonizes PigmR-mediated resistance by forming a heterodimer that prevents PigmR activation. In leaves, expression of PigmS is suppressed by methylation of transposon sequences in its promoter through the RdDM machinery. Conversely, PigmS expression is released in pollen, where it contributes to seed set and yield [Deng et al., 2017].

1.2.5 RNA silencing as part of plant immunity

Direct action of RNA silencing

It is generally suggested that RNA silencing evolved as a defence mechanism against foreign nucleic acids of RNA viruses, and it was only later co-opted into an endogenous regulatory mechanism [Pumplin and Voinnet, 2013]. The numerous reports that ultimately led to the discovery of RNA silencing were mostly based on observations of responses to foreign RNA. One of the first accounts was the necrotic symptoms in tobacco plants caused by the tobacco ring spot virus, and how these symptoms were relieved in newly emerging leaves [Wingard, 1928]. Given our current knowledge, it is more than likely that the recovery was due to the activation of the antiviral RNA silencing that inhibited the spread of infection by targeting the viral RNA [Baulcombe, 2004]. Viral dsRNA produced by the action of viral or host RDR1/6 from ssRNA

molecules, are perceived and targeted by DCL4/2. This targeting leads to the production of viral siRNAs that are subsequently loaded into AGO1/2 to reinforce defence through the antiviral RNA silencing [Pumplin and Voinnet, 2013].

Role of miRNAs in direct regulation of PTI

Additionally, roles for RNA silencing in PTI have also been uncovered. Upon bacterial infection, accumulation of the 22-nt miR393 is induced in *A. thaliana*. The miRNA targets F-box auxin receptors through direct interaction and phasiRNA production [Navarro et al., 2006]. Auxin-responsive elements in promoters have been shown to activate transcription of genes involved in PTI, so transcriptional repression of these auxin receptors ultimately leads to repression of PTI [Navarro et al., 2006]. The miRNA and its targets are conserved in many plants and have been shown to play a similar role upon oomycete infection in soybean [Wong et al., 2014]. Several other miRNAs have been shown to enhance resistance by regulating diverse classes of transcription factors, during infection of bacterial and fungal pathogens [Campo et al., 2013; Li et al., 2014; Niu et al., 2016; Soto-Suárez et al., 2017]. These results reinforce the importance of RNA silencing in biotic stresses.

Indirect regulation of ETI: Counter-counter-defence

Pathogen secrete an array of molecules to interfere and disarm the host defence mechanisms, and RNA silencing is no exception. Examples of suppressors of silencing are present for many pathogens from virus to bacteria, oomycete, and fungus [Anandalakshmi et al., 1998; Navarro et al., 2008; Qiao et al., 2013; Weiberg et al., 2013]. As part of the arms race between the pathogen and the host, plants have evolved mechanism to neutralize the negative effects of the suppressors. A common strategy has been uncovered in several plant species

based on a negative regulation of NLRs directed by various miRNA families, which is relieved by the pathogen suppressors upon infection [Fei et al., 2015].

The first evidence for phasiRNA-derive transcriptional regulation of NLRs came from three independent studies in legumes and nightshades. Numerous 22-nt miRNAs have been shown to directly target NLR transcripts and trigger phasiRNA production [Li et al., 2012b; Shivaprasad et al., 2012; Zhai et al., 2011]. Since then, examples of miRNAs that target NLRs have been found in both gymnosperms and angiosperms. However, the number of those miRNAs present in the genome vary greatly between species: from tens in spruce to just two in *A. thaliana* or none reported so far in melon [Chen et al., 2010a; González et al., 2015; Xia et al., 2015].

An unusual level of redundancy has been observed in most cases. In Solanaceae species, several miRNAs including miR482, miR6019, miR6027, miR5300, miR6024, and miR6026 have been reported to target NLRs and trigger large numbers of phasiRNAs [Li et al., 2012b; Shivaprasad et al., 2012]. Same has been reported for hundreds of NLRs in Fabaceae species, induced only by miR1050, miR2109 and miR2118 [Arikit et al., 2014; Zhai et al., 2011]. In monocots, however, these same families of miRNAs have been reported a different role, seemly unrelated to NLR regulation, during the formation of reproductive tissues [Johnson et al., 2009; Zhai et al., 2015]. Nevertheless, a 22-nt grass-specific miRNA, miR9863, has been described regulating NLR genes in barley and wheat through direct targeting and phasiRNAs [Liu et al., 2014].

Indirect evidence of the importance of this regulation mechanism can also be found in pathogen studies. Upon *Fusarium oxysporum* fungal infection of a resistant tomato cultivar, endogenous miR482 and miR5300 levels have been shown to decrease and their NLR targets are up-regulated. If targets were re-silenced via virus-induced gene silencing (VIGS), resistance to *F. oxysporum*

was compromised [Ouyang et al., 2014]. Another study showed that overall levels of expression defence genes is relatively reduced in potato cultivars that are sensitive fungal pathogen *Verticillium*, compared to resistant cultivars [Tai et al., 2013]. In *A. thaliana*, *RDR6* expression decreased rapidly upon treatment with bacterial PAMP fgl22. Additionally, *rdr6* and *miR472*^{m} (which targets CNL RPS5) mutants were reported to have enhanced resistance to bacterial pathogen *P. syringae* [Boccara et al., 2014]. Interestingly, the importance of the 22-nt nature of these miRNAs was uncovered in Liu et al. [2014]. The 22-nt miR9863 in barley was shown to more efficiently suppress CNL MLA1 compared to its 21-nt counterpart. The authors suggested that this effect was due to the *cis* action of the phasiRNA produced by the 22-nt miRNA.

Despite the abundant reports of NLR regulation through RNA silencing in plants, we still lack a clear understanding of the mechanistic process by which these miRNAs provide robustness to plant immunity. By studying one of the most extended miRNA families involved in this regulation, the miR482/2118, the work presented here aims to provide new clues towards its elucidation.

1.3 Aims and objectives

In my PhD project I chose to identify the role of RNA silencing throughout tomato development. In this dissertation I present data based on a combination of molecular and bioinformatic approaches. I use high throughput sequencing (HTS) data to identify miRNA targets and changes in small RNA accumulation and RNA expression. I use short tandem target mimics to specifically sequester miRNAs and prevent their action. I study the effect of these target mimics on the regulation of the immune response by challenging transgenic lines carrying these with bacterial and oomycete pathogens. I also study a reproductive AGO protein and the effect of its mutants in male gamete development. Lastly, I used

HTS data to study TE activation in the vegetative meristem and its influence on the small RNA population.

Chapter 2

Materials and Methods

2.1 Bacterial strains

Chemically competent Escherichia coli strains Subcloning Efficiency™ DH5α™ and One Shot ® TOP10 were used for plasmid amplification. Both strains were supplied by Invitrogen™ (Life technologies™).

Electro-competent Agrobacterium tumefaciens strain AGL1 carrying the different insertion vectors developed in this project was used for agro-transformation of tomato. Electro-competent Agrobacterium tumefaciens strain C58C1 carrying insertion vectors developed in this project was used for agro-infiltration of *N. benthamiana* leaves.

2.2 Plant strains and growth conditions

Tomato (*Solanum lycopersicum*) cultivars M82 were raised from seeds in compost (Levington ™ M3) and maintained in a growth room with 16/8h light/dark periods at 22°C (day) and 18°C (night), with 60% relative humidity, at a light intensity of 300 µmol photons $m^{-2} \cdot s^{-1}$. Unless indicated otherwise, leaves where collected from 1-month-old plants.

2.3 Seed extraction

Tomato seeds were treated for 5 minutes in 18% HCl and 10 minutes in 10% Na_3PO_4, followed by overnight drying on absorbent paper. For in-vitro germination, a prior sterilisation treatment of 10 min in 4% bleach followed by four washes in sterile water was performed.

2.4 Tomato stable transformation

Agrobacterium tumefaciens-mediated transformation of the standard processing tomato cultivar M82 were performed using a modified version of Van Eck et al. [2006]. In brief, sterilized seeds were germinated on 1/2 strength Murashige-Skoog medium, 1X Nitsch&Nitsch vitamins, 0.8% agar, 1.5% sucrose, pH 6. Cotyledon segments from one-week-old seedlings were inoculated with *A. tumefaciens* strain AGL1 containing constructs of interest in a solution of MS, 3% sucrose at $OD_{600} = 1$. The explants were then dried quickly on Whatman paper and placed on a plate without selection under low light (1X MS medium, 1X Nitsch&Nitsch vitamins, 0.6% agar, 3% sucrose, 100 $mg{\cdot}l^{-1}$ myo-inositol, 0.5 $mg{\cdot}l^{-1}$ 2,4-D, 0.1 $mg{\cdot}l^{-1}$ kinetin, pH 5.7). Following a 2-day co-cultivation, the cotyledon segments were transferred to a selective regeneration medium (1X MS medium, 1X Nitsch&Nitsch vitamins, 0.4% agargel, 2% sucrose, 100 $mg{\cdot}l^{-1}$ myo-inositol, 2 $mg{\cdot}l^{-1}$ zeatin, 320 $mg{\cdot}l^{-1}$ timentin, pH 6) containing Kanamycin (100 $mg{\cdot}l^{-1}$) for selection. When shoots pushed through the lid, they were transferred to a selective media containing cefotaxime (250 $mg{\cdot}l^{-1}$). After five weeks the shoots were transferred to rooting media (1/2 strength MS medium, 1X Nitsch&Nitsch vitamins, 0.25% gerlite, 0.5% sucrose, 50 $mg{\cdot}l^{-1}$ kanamycin, 320 $mg{\cdot}l^{-1}$ timentin, pH 6). Regenerants with well-developed roots were transferred to peat bags and grown under high humidity until they could be transferred to M3 compost and grown under normal conditions.

2.5 Meristem collection

Tomato M82 seeds were sown and grown for 10 days under normal conditions in the glasshouse. Whole apices where collected and fixated in 100% Acetone (30 min at 400 mmHg). To ensure that only vegetative meristems where collected, only plants presenting the 5 th leaf primordial where used. Meristems were dissected using an Ultra Fine Micro Knive (FST ®) following the procedure detailed in Park et al. [2012] and stored at -80 °C. A pool of 30 meristems typically yielded between 600ng and 1µg of total RNA.

2.6 *Nicotiana benthamiana* leaf infiltration

Nicotiana benthamiana seeds were sown in a plate with compost under high humidity. A week later, germinated seedlings were transferred to pots with M3 compost and grown for two weeks. Once plants were ready for infiltration, *Agrobaterium* cells carrying the vectors of interest were cultured in LB with the appropriate set of antibiotics. Once cultures reached approximately OD_{600} = 0.4, they were spinned down, re-suspended in Infiltration buffer (10 mM $MgSO_4$, 10 mM MES [pH 5.6], 150 µM acetosyringone), and adjusted to a final OD_{600} = 0.5. Two leaves per plant, and three plants per construct combination were infiltrated with 1 ml syringes.

Three days after infiltration, leaves were detached from each plant and examined under a UV light microscope (Leica DFC 310FX). Images were taken of each leaf at the site of infiltration, allowing for comparison of relative fluorescence between treatments. Leaf discs around the infiltrated area were collected at this stage for protein detection.

2.7 Molecular biology methods

2.7.1 DNA extraction

DNA was extracted from leaves, or pools of collected meristems. Samples were collected in 1.5 ml eppendorf tubes with 3 x 3 mm silica balls, and freeze in liquid nitrogen. Tissue was disrupted with the use of tissue-lyser. DNA was subsequently extracted with the with the Puregene™ kit, following manufacturer's instructions (QIAGEN™).

2.7.2 RNA extraction

For the extraction of RNA from tissue of the various tomato tissues used in this work, Trizol™ RNA protocol (Ambion™) or QuickRNA™ Micro/MiniPrep (Zymo Research Corp.) were used following the instructions of the manufacturer.

2.7.3 Transcript quantification

In order to detect different RNA expression levels, RNA was copied into single copy and high-quality cDNA. For that purpose, SuperScript ® IV (Invitrogen™) was used following the exact instructions of the manufacturer. The reverse transcription was performed using 5 µg of total RNA per sample and a 1:1 mix of random hexamers and oligo dT primers.

Quantitative PCR was performed in BioRad CFX96 machines using Eva-Green detection chemistry. Relative quantities of each transcript were calculated using standard curves quantification method. Three biological samples and three technical replicates per sample were used for each experiment.

2.7.4 Small RNA northern blot

Small RNA detection was performed using the northern blot technique. In brief, 5 µg of total RNA per sample were prepared in 10 µl, added equal volume of 2X loading buffer (95% deionized formamide, 18 mM EDTA, 0.025% SDS, xylene cyanol FF, bromophenol blue), and boiled at 65°C for 5 min. Then placed in ice for 1 min and loaded and run in a 15 % polyacrylamide 7 M urea gel, using 0.5X TBE running buffer. RNA was then transferred to a positively charged nylon membrane (Amershan Hybond-N+, GE Healthcare^{TM}) using overnight capillary system: gels were soaked for 10 min in 20X SSC, then placed on a clean glass plate, membrane on top, 2 pieces of 3MM paper soaked in 20X SSC, and finally 3-5cm of thick paper on top. Another glass plate on top and 1kg of weight on top of the upper plate. RNA was cross-linked two times per side with 0.12 J of UV light in a Stratalinker® (Agilent^{TM}). Oligonucleotides and Locked Nucleic Acid (LNA^{TM}; by Exiqon) probes radiolabelled with γ-^{32}P-ATP were hybridized in ULTRAhyb-Oligo buffer (Thermo Fisher Scientific^{TM}) for 12 hours at 40°C or 2 hours at 57°C, respectively. Then washed three times with 2X SSC, 0.2% SDS. Phosphoimager plates (Fujifilm) were exposed and then imaged with a Typhoon 8610 (Molecular Dynamics).

2.7.5 Protein western blot

To detect eGFP accumulation in *N. benthamiana* agro-infiltrated leaves, western blotting was used. In brief, protein extracts were made by grinding $4 \cdot 5$ mm^2 leaf discs in protein extraction buffer (125 mM Tris-HCl, 2% SDS, 6 M urea, 0.5% β-meracptoethanol, 0.02% bromophenol blue). Protein extracts were run in 10% NuPAGE® Bis-Tris precast polyacrylamide gels with 1X NuPAGE® MES running buffer (Life technologies^{TM}). Gels were subsequently blotted to an Immun-Blot PVDF (BioRad®) using wet transfer method (Transfer buffer: 48mM Tris, 39mM Glycine, 20% MetOH). Membranes were blocked overnight at 4 °C in 10 ml of blocking solution (1X TBS, 0.1% Tween-20, 5% milk). An

anti-GFP rabbit polyclonal antibody (sc-8334; Santa Cruz Biotechnology, INC) was then used for GFP detection. Membranes were incubated for 1 h at RT with a 1:1,000 dilution of the primary antibody in blocking solution. After washes, a 1:5,000 dilution goat anti-rabbit IgG fused to HRP was used to later proceed with enhanced chemiluminescence (ECL) detection.

2.8 Pollen viability quantification

A modified version of Alexander staining was performed to asses pollen viability based on the protocol developed by Peterson et al. [2010]. Tomato inflorescences (stage 12 flower buds, based on Brukhin et al. [2003]) were fixed in Carnoy's fixative (ethanol:chloroform:acetic acid [6:3:1]) for at least 2 hours at room temperature. Under a stereomicroscope, mature anthers were dissected and placed in a drop of water on a microscopic slide. Pollen was released from anthers with the edge of a forceps and 2-3 drops of stain solution (10% alcohol, 25% glycerol, 4% glacial acetic acid, 0.05% acid fuchsin, 0.01% malachite green, 0.005% orange G) were added. The slides were then covered with a cover slip which was sealed with a rubber solution (Fixogum). Slides were incubated on an oven at 37°C overnight and then inspected under an Olympus BX-41 microscope.

2.9 Pathogen assays

2.9.1 *Phytophthora infestans* infection

The *Phytophthora infestans* strain in this study is 88069, originated from the Netherlands in 1988. Cultures were stored in liquid nitrogen and grown on rye sucrose medium. Infection assays were performed on detached tomato leaves, measuring lesion sizes, based on a modified version of Du et al. [2015]. In brief, four well developed leaves per plant and four plants per condition were detached

from four-week-old plants and placed in water-saturated paper in a tray. Spore suspensions of *P. infestans* were prepared by rinsing two-week-old plates covered with mycelium with cold water and incubating the sporangiophore at 4°C for 1-2 hours. After release of zoospores, the concentration was adjusted to $5 \cdot 10^4$ spores·ml^{-1}. *P. infestans* were spot-inoculated on the abaxial side of the leaf, by placing six 10 µl droplets on various locations right and left of the midvein. The trays were covered and incubated at room temperature at constant light a photoperiod. Disease assessments were performed at 3 days post inoculation (dpi) under blue ligth using a DarkReader ® Transilluminator (Clare Chemical Research) and a Nikon COOLPIX P520. Lesion diameters were measured using ImageJ software, followed by statistical analysis and plotting performed in R.

2.9.2 *Pseudomonas syringae* infection

The *Pseudomonas syringae* pv. *tomato* strain in this study is DC3000, which is a pathogen of tomato developed in 1986 as rifampicin-resistant derivative of Pst DC52. Cultures were stored in liquid nitrogen and grown on King's B medium. Infection assays were performed *in planta*, inoculating mature tomato leaves and measuring bacterial growth similarly to the procedure described in Liu et al. [2015b] (based on the original work of Katagiri et al. [2002]). In brief, bacteria were grown on Kings B solid medium containing 100 µg·ml^{-1} rifampicin at 28°C. Cultures were harvested when their OD_{600} was between 0.2 and 0.4, spinned down, and re-suspended in Infiltration buffer (10 mM $MgCl_2$) to a final OD_{600} = 0.2 (corresponding to $5{\cdot}10^8$ CFU·ml^{-1}). Serial dilutions in infiltration buffer were made to generate bacterial suspensions of $5{\cdot}10^4$ CFU·ml^{-1} and proceeded with syringe inoculation of 4-week-old tomato plants (three leaflets per leaf, two leaves per plant, three plants per condition). We monitored bacterial growth within leaf tissue at the day of infection and day 3 days post-infection by grinding six leaf disks (6 mm of diameter) per sample and plating 10 µl of 1:10 serial dilutions of the ground material on King's B media. The levels of

bacterial proliferation were calculated using the following formula:

$$CFU/mm^2 = \frac{N\,(colonies/drop) \cdot 10^{\,ser.\,dilut.}}{10\frac{\mu l}{drop} \cdot 8\ discs \cdot \frac{\pi \cdot (6\ mm/2)^2}{disc}}$$

Statistical analysis and plotting performed in R.

2.10 Bioinformatic analyses

2.10.1 Transcriptome analyses

Chapter 5

RNAseq libraries were prepared using the Truseq ® RNA Sample Prep Kit v2 (Illumina). In brief, four biological replicates each of 1 month old tomato leaf RNA were prepared using 1 µg of total RNA per sample. PolyA bead selection and non-strand-specific RNA-seq libraries were made and indexed according to manufacturer instructions. Finalized libraries were sequenced as a pool on one lane of a NextSeq 500/550 High Output Kit v2 (150 cycles). Sequences were demultiplex, and trimmed and filtered using Trim Galore! (Babraham Bioinformatics) with default parameters and reads were concordantly aligned to the Heinz genome SL3.00 version and the ITAG3.2 transcriptome using Kallisto [Bray et al., 2016], with the parameter `-b 100`. Differential expression analysis was performed on kallisto counts using the Bioconductor package Bayseq [Hardcastle et al., 2012].

Chapter 7

Total RNA samples were prepared using QuickRNA™ microprep. Four biological replicates each of 1 month old tomato leaf RNA and of pools of 30 tomato vegetative meristems (as described in 2.5) were prepared using 300 ng of total RNA. RNA was depleted of ribosomal RNA using the Ribo-Zero

rRNA removal kit (Epicenter). Strand-specific RNA-seq libraries were made and indexed with the Script-Seq v2 kit (Epicentre) according to manufacturer instructions. Finalized libraries were sequenced as a pool on one lane of a HiSeq 2000 100PE. Sequences were demultiplex, and trimmed and filtered using Trim Galore! (Babraham Bioinformatics) with default parameters and reads were concordantly aligned to the Heinz genome SL2.50 version using Bowtie2 v2.2.4 [Langmead and Salzberg, 2012], with default parameters. Differential expression analysis was performed on raw counts on ITAG2.4 annotation and LTRs extracted from Xu and Du [2014] using the Bioconductor package Bayseq [Hardcastle et al., 2012].

2.10.2 sRNA-seq analyses

Chapter 4

Small RNAs libraries were prepared using the NEBNext ® Small RNA Library Prep (New England Biolabs). In brief, three biological replicates each of 1 month old tomato leaf RNA were prepared using 1 µg of total RNA per sample. After preparation, size selection of libraries was performed using BluePippin 3% agarose cassettes (Sage Science). Each library was barcoded, pooled, and sequenced using a single NextSeq 500/550 High Output Kit v2 (75 cycles). Sequences were demultiplex, and trimmed and filtered using Trim Galore! (Babraham Bioinformatics) with default parameters and reads were concordantly aligned to the Heinz genome SL3.00 version using Bowtie v1.2.0 [Langmead et al., 2009], with the bowtie modifiers `-v 1 -m 50 --best --strata`. Identification of sRNA loci and differential expression was performed using the segementSeq package [Hardcastle and Kelly, 2013].

Chapter 7

Small RNAs were enriched using the novel method of fully denaturing total RNA prior to selection [Harris et al., 2015]. In brief, sRNAs where extracted

from a Formaldehyde/Formamide polyacrylamide gel using 300ng of total RNA as starting material. Four biological replicates each of 1 month old tomato leaf RNA and of pools of 30 tomato vegetative meristems were prepared. Libraries were produced using the TruSeq ® Small RNA Sample Preparation Kit (Illumina). Each library was barcoded and 50SE sequenced on one lane of a HiSeq 2000. Sequences were demultiplex, and trimmed and filtered using Trim Galore! (Babraham Bioinformatics) with default parameters and reads were concordantly aligned to the Heinz genome SL2.50 version using Bowtie v1.1.1 [Langmead et al., 2009], and sequences that did not perfectly align were discarded. To best handle multi-mapping sequences generated from repetitive regions of the genome, the bowtie modifiers `-v 0 -k 1 --best` were employed. Identification of sRNA loci and differential expression was performed using the segementSeq package [Hardcastle and Kelly, 2013].

2.10.3 Phylogenetic analysis

BLASTN analyses were performed using genomic sequences of tomato genes and miRNA precursor sequences against the genomes of all plant model organisms and all available genome assemblies of major *Solanaceae* species. The threshold expectation value was defined at 10^{-3} to filter out any spurious hits. Any hits were then curated manually.

2.10.4 Target prediction

All tomato miRNA mature sequences were downloaded from miRBase (v21) [Kozomara and Griffiths-Jones, 2013]. MicroRNAs targeting genes were predicted by psRNATarget v1 [Dai and Zhao, 2011] and TargetFinder v1.7[Allen et al., 2005]. Cut-off values were established using modifiers `Max UPE 50.9` & `Expectation 4` for psRNATarget v1, and `-c 6` for Targetfinder, based on the optimal scores reported in Akhtar et al. [2016]. Highly confident interactions were defined as those with a psRNATarget expectation score ≤ 3 and UPE ≤ 25.

Tomato NLR sequences were retrieved from [Andolfo et al., 2014] and curated using the ITAG3.2 annotation [The Tomato Genome Consortium, 2012].

2.10.5 Degradome (PARE) analysis

Parallel analysis of RNA ends (PARE) was performed using the software sPARTA [Kakrana et al., 2014]. The analysis was done on publicly available datasets from the only tomato degradome data from leaf tissue, which corresponds to Lopez-Gomollon et al. [2012]. For the analysis, sPARTA modifiers `-tarPred E -tarScore S -minTagLen 17` were used.

2.10.6 Phasing analysis

Identification of phased loci overlapping with genes was performed by calculating the log P value from the hypergeometric distribution developed in Santos [2014], based on the previous method of Chen et al. [2010b].

2.11 Oligonucleotide sequences

Experiment	Name	Sequence (5'-3')
Northern blot	*sly-miR168*	GTCCCGACCTGCACCAAGCGA
	sly-U6	GGCCATGCTAATCTTCTCTGTATCGTT
	sly-miR482b	GGTATGGGAGGAGTAGGAAAGA
	sly-miR482a (LNA)	GGCATGGGCGGTGTAGGCAAGA
	sly-miR2118a	TAGGAATGGGTGGAATTGGAAA
	sly-miR2118b (LNA)	TTGGCATGGGTGGAATAGGAAA
qRT-PCR	*TAS5_qPCR_1_F*	GGTTTGGTTCGGGTTGTTTA
	TAS5_qPCR_1_R	TCAACATTGCTTTCCCACTTT
	LRR2_qRT_B_F	CTAGCGAAGCGTGGTCTTGA
	LRR2_qRT_B_R	TGAGCACAAAAGAGTTGTAGCTT
	NRC1_qRT_A_F	GCAGCCTCCAAGTGATGGAT
	NRC1_qRT_A_R	AACCGTGGATCAGTCGTTCC
	Bs4_qPCR_F3	TCTGTGCTTGAGACCACCAA
	Bs4_qPCR_R3	TCCGCCTCTGCTCCCTATATT
	EXP_qPCR_F	GCTAAGAACGCTGGACCTAATG
	EXP_qPCR_R	TGGGTGTGCCTTTCTGAATG
RT-PCR	*mimicry_482_F*	AAGGCATGGGCGGCACTGT
	mimicry_482_R	GAATTCTCTTTCCTACTATGC
	mimicry_2118b_F	ACCAATTGGCATGGGTGATA
	mimicry_2118b_R	GAATTCTTTCCTATTCAGTCA
Mimicry cloning	*miR2118b_mimic_Fw*	CACCAATTGGCATGGGTGATAGAATAGGAAAGTTGTTGTTGTTATGGTCTAATTT
	miR2118b_mimic_Rv	AAATATGGTCTAAAGAAGAAGAATTTGGCATGGGTGACTGAATAGGAAAGAATTC
	miR2118b_target_Fw	GAATTCTTTCCTATTCAGTCACCCATGCCAAATTCTTCTTCTTTAGACCATATTT
	miR2118b_target_Rv	AAATTAGACCATAACAACAACAACTTTCCTATTCTATCACCCATGCCAATTGGTG
	miR482_mimic_Fw	CACCTAAAATGATTTGGCATGGGTGGAATAGGAAAAGAAAAAATTTGTTGGCAT
	miR482_mimic_Rv	GGGTGGAATAGGAAAAAAGAAAAGCGTTTGGCATGGGTGGAATAGGAAAGAATTT
	miR482_target_Fw	AAATTCTTTCCTATTCCACCCATGCCAAACGCTTTTCTTTTTTCCTATTCCACCCATG
	miR482_target_Rv	CCAACAAATTTTTTCTTTTCCTATTCCACCCATGCCAAATCATTTTAGGTG
	MIR2118b_cloning_F	CACCAAGGCATGGGCGGCACTGTAGGCAAGAGTTGTTGTTGTTATGGTCTAATTT
	MIR2118b_cloning_R	AAATATGGTCTAAAGAAGAAGAATGGTATGGGAGGCATAGTAGGAAAGAGAATTC
	MIR482f_cloning_F	GAATTCTCTTTCCTACTATGCCTCCCATACCATTCTTCTTCTTTAGACCATATTTAAA
	MIR482f_cloning_R	TTAGACCATAACAACAACAACTCTTGCCTACAGTGCCGCCCATGCCTTGGTG
		CACCAATGATGGCATGGGCGGTGTAGGCAAGAAGAAAAAATTTGGGTATGGGAG
		GAGTAGGAAAGAAAAGAAAAGCGTGGAATGGGCGGTATTGGCAAGAGAATTTATG
		CATAAATTCTCTTGCCAATACCGCCCATTCCACGCTTTTCTTTTCTTTCCTACTCCTC
		CCATACCCAAATTTTTTCTTCTTGCCTACACCGCCCATGCCATCATTGGTG
		CACCTGATGAGATTTAGAGATTGTCATGC
		GTTGCATCTTTCATCCCAC
		CACCCATTACAGCAGCAACACACATTTC
		ATCCAACAAGTTCAATTAAGGG
CRISPR	*sgRNA_AGO5_1_1*	tgtggtctcaATTGGTCGAGGAACTTACTCCAGgttttagagctagaaatagcaag
	sgRNA_AGO5_1_2	tgtggtctcaATTGGTTTCATCGATGACTCGGGgttttagagctagaaatagcaag
	sgRNA_AGO5_2_1	tgtggtctcaATTGCAGATATGAAGTAGTGGGAgttttagagctagaaatagcaag
	sgRNA_AGO5_2_2	tgtggtctcaATTGGTGGGTTGGAATACTGGAGgttttagagctagaaatagcaag
slyAGO5	*AGO5_prom_R*	TGTTAACGAGAGAGAGAAAAAGAGT
	AGO5_prom_F	GACGTGTCACTGCCAACTG
	AGO5_Cloning_R_NoStop	GCAGTAAAACATGACCTCCG
	AGO5_Cloning_F	ATGTCGGAACGTGGACGAG

Table 2.1 Oligonucleotides used in this work

Chapter 3

The miR482/2118 family in tomato

3.1 Introduction

PhasiRNAs were originally described as trans-acting small interfering RNAs (tasiRNAs) due to their ability to target complementary mRNAs *in trans* and silence them [Allen et al., 2005]. Since then, numerous other examples have been identified (reviewed in Axtell [2013]). The importance of these negative regulators is reflected in their abundance and diversity. There are numerous families of miRNAs that negatively regulate defence genes of the Nucleotide-Binding Leucine-Rich Repeat (NLR) family [Halter and Navarro, 2015]. Among these, the miR482/2118 family is the most extended as it first appeared during the emergence of seed plants [Xia et al., 2015]. This family is present in most seed plants, although in some instances the apparent conservation may reflect repeated rounds of convergent evolution [Xia et al., 2015]. As mentioned in the first section of this dissertation, the miR482/2118 family and several other NLR specific miRNAs are unusual amongst miRNAs because they are 22-nt rather than 21-nt in length [Li et al., 2012b; Shivaprasad et al., 2012; Zhai et al., 2011]. This size difference is what triggers their targets to be converted

into a dsRNA and processed through the phasiRNA pathway [Chen et al., 2010a].

Although there have not been any reports of the direct action of these NLR derived phasiRNAs, indirect evidence from various authors suggests these phasiRNAs may play a role in what is known as "counter-counter defence" (first described in Shivaprasad et al. [2012]; reviewed in Pumplin and Voinnet [2013]). Upon infection, pathogens secrete effectors that suppress gene silencing in plants [Anandalakshmi et al., 1998; Navarro et al., 2008; Qiao et al., 2013]. These effectors down-regulate the action of miRNAs, including the miR482/2118 members [Shivaprasad et al., 2012]. The current hypothesis proposes that these effectors also release the phasiRNA-mediated silencing of NLR, and the accumulation of NLR transcripts is boosted. Ultimately, this leads to an up-regulation of defence genes and therefore enhances plant immunity. (Figure 3.1).

In many plant genomes, large numbers of phasiRNAs are produced from NLR genes [Li et al., 2012b; Shivaprasad et al., 2012; Zhai et al., 2011]. However, the extent to which sRNA, and in particular the miR482/2118 family, modulate NLR expression is not very well understood. Comprehensive analysis of the action of this sRNA is necessary in order to understand how NLR regulation originates and is maintained.

The focus of this chapter is to explore the differences in the ancient plant miRNA family miR482/2118, its derived phasiRNAs and their targets. The results presented here address previously ignored features in the members of this miRNA family in tomato (*Solanum lycopersicum*) that appear to influence their action.

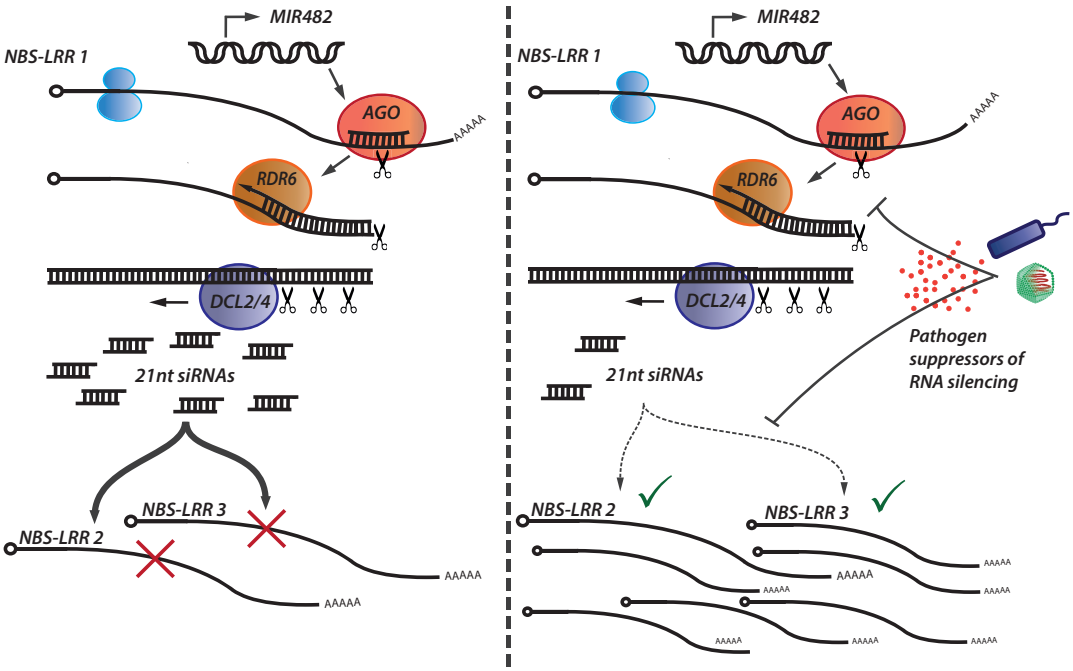

Fig. 3.1 **Defense and counter-defence mechanisms via RNA silencing in plants.** (Left panel) Several NLR resistance genes are constitutively down-regulated by the basal expression of miR482. This targeting triggers the production of phasiRNAs that further silence these and other resistance gene family members. (Right panel) Upon infection, these mechanisms are impaired by the pathogen suppressors of RNA silencing, which releases the expression of the resistance genes and defence response.

3.2 Results

3.2.1 Nucleotide divergence of tomato miR482/2118 members

To investigate the miRNA mediated silencing of defence genes, I chose to focus on the miR482/2118 family due to its ancestral origin, appearing first in an ancestor of seed plants and still persisting in most lineages due to either conservation or convergent evolution (Figure 3.2) [Xia et al., 2015]. Previous studies identified several different tomato miRNAs and classified all of them as miR482 homologs [de Vries et al., 2015; Shivaprasad et al., 2012; Zhang et al., 2016a] (Table 3.1). Upon closer inspection, it is clear that two members of this family align in different registers to the classic miR482 sequence, starting at position 3 of the consensus (Figure 3.3). This characteristic sequence is reminiscent of the other structural form of this miRNA family, commonly referred as miR2118. This other form has been reported primarily in gymnosperms, monocots and legumes, where no miR482-like members have been described [Johnson et al., 2009; Zhai et al., 2011]. The presence of both structural forms in tomato provides a valuable scenario in which to study this family. In the inherently short sequence of a miRNA, such structural variation will fundamentally have a major impact on functionality and target preference. I decided to further explore these differences and, to simplify the narrative, I renamed miRNAs as miR2118 if the 5' position aligns to position 3 in the consensus (Table 3.1 and Figure 3.3). I referred to the remaining members using their miR482 nomenclature in the miRBase database.

The present analysis focused on the five sequences that could be validated in our sRNA datasets from leaves of 1 month old tomato plants. The alignment shows that most deviation from the consensus sequence (5 variable sites) corre-

*In Karlova et al. [2013], all miR482 members were pulled for the analysis without any specifications.

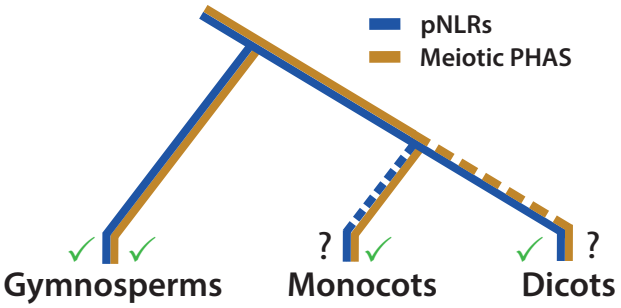

Fig. 3.2 **Evolutionary history of the miR482/2118 family across plant lineages.** Up to this point, nembers of the miR2118/482 family have been sequenced in 34 plant species, including 4 gymnosperms. PhasiRNA production in NLRs (pNLRs) derived from miR482/2118 targeting has only been observed in gymnosperms and dicots, while pre-meiotic miR2118-dependent PHAS loci have only been found in gymnosperms and monocots (reviewed in Fei et al. [2016]).

Sequence	This work	miRBase	Zhang et al. 2016	DeVries et al. 2015	Shivaprasaed et al. 2012	Karlova et al. 2013	Notes
UUUCCAAUUCCACCCAUUCCUA	mi2118a	sly miR482a	miR482a	miR482	miR482	miR-Y	-
UUUCCUAUUCCACCCAUGCCAA	mi2118b	sly-miR482d	miR482d	miR482g	-	-	-
UCUUGCCUACACCGCCCAUGCC	miR482b	sly-miR482b	miR482b	miR482a	miR482a		-
UCUUUCCUACUCCUCCCAUACC	miR482e	sly-miR482e	miR482e	miR482f	miR482f	miR482*	-
UCUUGCCAAUACCGCCCAUUCC	miR482c	sly-miR482c	miR482c	miR482b	miR482b		-
UUACCAAUUCCACCCAUUCCUA	-	-	-	miR482h	-	-	No evidence of expression
UCUUUCCUACUCCUCCCCUACC	-	-	-	-	miR482c	-	Misidentified as new
UCUUUCCUACUCCUCCCAUGCC	-	-	-	-	miR482e	-	members, probably due
UCUUUCCUACUCCUCCCAUCCC	-	-	-	-	miR482d	-	to sequencing errors

Table 3.1 **Summary of identified members of the miR482/2118 family in tomato.** All available nomenclatures of miR482/2118 members in tomato in the different studies present in the current literature.

sponds to the wobble position of codons in the conserved P-loop amino-acid motif (GMGGVGKT) of NLR proteins (Figure 3.3). This feature is consistent with previous reports and with the proposal that this family of miRNAs could target many mRNAs encoding proteins with the NLR consensus for this motif irrespective of synonymous coding sequence changes in the mRNA [Shivaprasad et al., 2012].

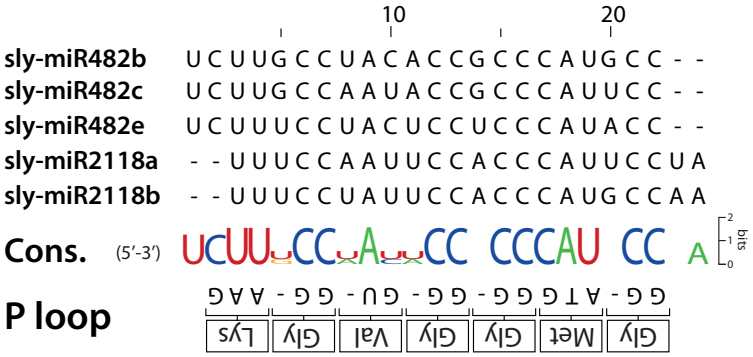

Fig. 3.3 **Distinction between miR482 and miR2118 members in tomato.** Nucleotide sequence alignment of mature miR482/2118 members in tomato. The consensus sequence of each position in the alignment versus the P-loop motif is shown at the bottom.

3.2.2 Targeting potential of tomato miR482/2118 members

Target prediction of plant miRNAs is complicated. Several limitations and caveats still exist when addressing 'real candidates'. What defines a miRNA target? Historically, the most important defining factor has been the sequence similarity necessary for pairing of the miRNA and the transcript. All target prediction tools correctly address this issue and rank predicted targets based on the result of this evaluation. Despite that, many other factors such as spatio-temporal expression of both miRNA and target transcript are rarely taken into account. A nevertheless valid *in silico* analysis does not possess the necessary information to determine the true regulatory function of the miRNA. Integrating other types of data in the analysis, such as degradome sequencing and sRNA sequencing are key to refine the findings and extract valid conclusions.

In the case of the tomato miR482/2118, previous analyses have failed to provide a clear picture of the extent of targeting of these miRNAs. de Vries et al. [2015] only reported the very top NLR hits of each miRNA in order to elaborate a shared regulatory network of the miRNAs. Additionally, the parameters used in their analysis are far more stringent than the optimal cut-off for retrieving most true targets [Srivastava et al., 2014]. A more general analysis using the *Arabidopsis thaliana* genome sequence indicated that the miR482 family in tomato targeted mRNAs for a subset of the NLR mRNAs encoding proteins with coiled-coil domains at the N terminus (CNLs) [Shivaprasad et al., 2012]. However, there are several potential drawbacks in this analysis as a consequence of the large evolutionary distance between Solanaceae and Brassicae. These two families diverged over 100 million years ago (mya) during the split between asterids and rosids in the eudicots [Moore et al., 2010]. Since then, their genomic composition of NLR-encoding genes has differed significantly. First, the *A. thaliana* genome contains a third fewer genes that encode NLRs compared

to the tomato (207 vs 326) [Andolfo et al., 2014; Meyers et al., 2003]. Among these NLRs, the predominant classes also differs. In *A. thaliana*, 71 % (147) of NLRs contain TIR domains (TNLs), while in tomato these merely represent 12 % (40) of the total composition. CNLs are the major class of this family in tomato with 64 % (209) of the total [Andolfo et al., 2014; Meyers et al., 2003].

One additional difference between the *A. thaliana* and the tomato genome lies in the number of miR482/2118 members and their derived phasiRNAs. In Solanaceae, several miRNA members have been identified alongside with dozens of phasiRNA loci potentially deriving from these [Li et al., 2012b; Shivaprasad et al., 2012]. In *A. thaliana*, only one member of this microRNA family exists (miR472) and NLR-derived phasiRNA loci have only been detected in transgenic miR472 over-expression assays, and not in normal conditions [Boccara et al., 2014; Howell et al., 2007]. In summary, any conclusions drawn from using a very limited dataset of targets or the *A. thaliana* genome to characterize the tomato miR482/2118 should be treated with utmost caution. Predictions using components from different species could be misleading due to the interlink between the diversification of NLR genes and the evolution of microRNAs that target them [Zhang et al., 2016a].

As a first step, I set to re-evaluate these results using the the current assembly of the tomato genome (SL3.0; The Tomato Genome Consortium [2012]). To distinguish defence resistance genes from the total repertoire of the tomato genes, I used a curated annotation based on the findings of Andolfo et al. [2014]. Their work also provided information about evidence for expression of these resistance genes in RNAseq datasets in both normal conditions and under pathogen *Phytophthora infestans* infection. I further created high-throughput sequencing (HTS) data of small RNAs in 1-month old tomato leaves under normal conditions, and used it to define phasiRNA-producing NLRs (pNLRs). I interpreted these as being the NLRs yielding sufficient small RNAs to make a

significant contribution to the small RNA population, and defined it by the total number 21-nt siRNAs mapping to their genomic region being greater or equal to 10 normalized counts per million (nRPM) (Figure 3.4). Out of 301 NLRs present in the curated Andolfo et al. [2014] dataset †, 161 presented evidence of expression and only 32 had sufficient counts to be considered as pNLRs (Table 3.3). All of my 32 pNLRs overlapped with the 162 NLRs present in their RNAseq data. Since the phasiRNA production cycle requires an initial step of transcription, this overlap indicates good correlation between datasets. Only 2 of the 32 pNLRs belonged to the TNL protein class, whereas most belonged to the CNL class. This observation is in agreement to the findings of the less suitable *A. thaliana* dataset, where the vast majority (71 %) of NLRs are TNLs. My results show a clear predilection for CNLs within the species, and further reinforces previous hypothesis that miR482 members preferentially target CNL-type NLRs.

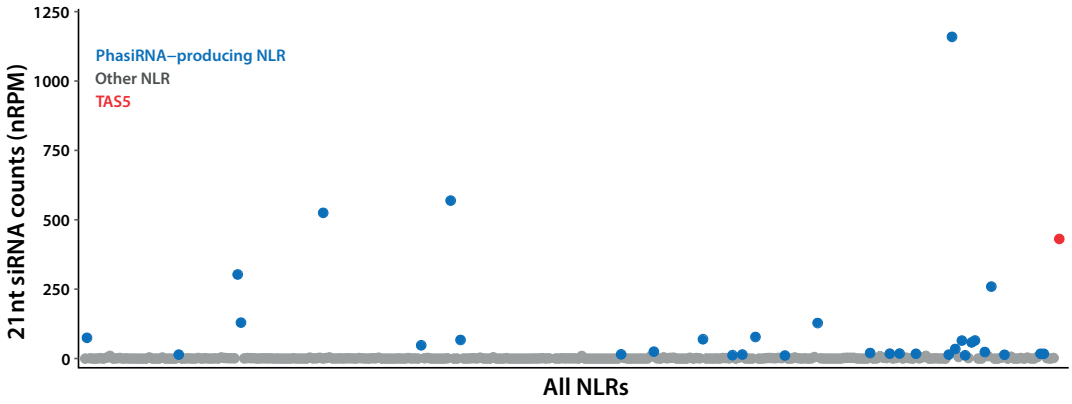

Fig. 3.4 **Defining the phasiRNA-producing NLRs in tomato.** Dot-plot representing the sum of 21-nt sRNA normalized counts per million (nRPM) aligning to an individual NLR. A total of 32 NLRs presented $\geq$ 10 nRPM counts and were defined as pNLRs (blue). *TAS5* (red) is added as a reference.

To evaluate the targeting preferences of the miR482/2118 members, I used bioinformatic target prediction tools. Recent work by Srivastava et al. [2014]

†25 genes initially annotated in the study have been (1) rejected due to correct reassembly of genomic sequences, or (2) fused with other genes, and dropped in the current version of the tomato genome.

suggested psRNATarget [Dai and Zhao, 2011] and Targetfinder [Fahlgren and Carrington, 2010] as the most suitable programs for microRNA target prediction in plants due to high 'efficiency', explained as optimal balance between 'precision' (accuracy of prediction) and 'recall' (sensitivity) in finding 'true-positive' targets in miRNA-mRNA interactions in non-model plants. I employed both programs using the suggested parameters by Srivastava et al. [2014] and the mature sequences of the miR482/2118 members (Table 3.1) against a dataset containing the latest cDNA sequences of all tomato genes (ITAG3.2; The Tomato Genome Consortium [2012]). Both tools performed similarly, presenting almost perfect overlap between the two outputs (>90 % in all cases), so I continued the analysis with psRNATarget (being the only one of the two still currently maintained).

All microRNAs yielded relatively similar total number of targets using loose parameters (Table 3.2). Out of all predicted results, miR482b and miR482e targets were greatly enriched for NLRs, and more so under stringer parameters (scores ≤ 3, where precision is predicted to be >75 % but with recall of <40 %). The enrichment of NLRs in the target list was not as evident for miR482c and miR2118a. Unlike predicted before, not all the members of the family seem to contribute significantly to the production of NLR-derived phasiRNAs.

The ability of a 22-nt miRNA to trigger phasiRNAs is intrinsic to its sequence [Chen et al., 2010a]. The structure of the miRNA:miRNA* duplex when loading into AGO has been shown to be the defining factor for recruitment of the machinery that causes the biogenesis of phasiRNAs [Manavella et al., 2012]. Therefore I used the HTS sRNA data as another layer to narrow down true targets.

The miR482 members (particularly miR482b and e) were predicted to target (and trigger) almost three quarters of all pNLRs in my analysis. Included in those members were all the NLRs that produced the greatest quantity of sRNAs

(>100 nRPM) (Tables 3.2 & 3.3). The overlap between predicted targets of the different microRNAs revealed that only miR482b and miR482e (and miR2118b to a lesser extent) share a significant amount of hits (Figure 3.5) while miR482c and miR2118a have more unique hits. However, given that in miR482c the number of high confidence targets compared to total targets was very low (7 vs 45), it is safe to assume that overlap between miR482c and the other miR482 could be biased and shadowed by the presence of false positives within the miR482c hits.

Although these results validate to a certain extent the concept that the miR482/2118 family in tomato preferentially targets NLRs, noticeable differences appeared in my analysis. Different members varied in their abilities to target (and therefore initiate) the pNLRs. Members of the miR482 clade accounted for the majority of these pNLRs. Of the miR2118 clade, miR2118b preferentially target a TAS gene (*TAS5*) and contribute to the pNLRs in a less significant manner (Table 3.3). The remaining member, miR2118a, preferentially target very few pNLRs. All together suggesting that the structural differences intrinsic in the sequences of these two clades have functional significance.

	All genes (n=35768)				**NLRs (n=302)**				**Exp. NLRs (n=162)**				**Exp. & phasiR NLRs (n=32)**			
MIRNA	*Total pred.*		*High Conf.*		*Total pred.*		*High Conf.*		*Total pred.*		*High Conf.*		*Total pred.*		*High Conf.*	
miR482b	64	(0.18 %)	25	(0.07 %)	54	(17.9 %)	25	(8.3 %)	40	(24.7 %)	15	(9.3 %)	18	(56.3 %)	6	(18.8 %)
miR482e	71	(0.20 %)	20	(0.06 %)	55	(18.2 %)	19	(6.3 %)	40	(24.7 %)	12	(7.4 %)	20	(62.5 %)	7	(21.9 %)
miR482c	45	(0.13 %)	7	(0.02 %)	15	(5.0 %)	3	(1.0 %)	12	(7.4 %)	3	(1.9 %)	5	(15.6 %)	1	(3.1 %)
miR2118a	65	(0.18 %)	24	(0.07 %)	17	(5.6 %)	9	(3.0 %)	12	(7.4 %)	7	(4.3 %)	2	(6.3 %)	1	(3.1 %)
miR2118b	55	(0.15 %)	17	(0.05 %)	31	(10.3 %)	16	(5.3 %)	25	(15.4 %)	14	(8.6 %)	7	(21.9 %)	2	(6.3 %)

Table 3.2 **Summary of target prediction analysis.** The analysis was divided between all genes in the tomato genome, all annotated NLRs from Andolfo et al. 2014, all NLRs with any evidence of expression, and all NLRs that we have identified as producing siRNAs in our dataset. High confidence targets were defined as target score ≤ 3 in our psRNATarget results.

General information			Target Prediction				
Gene_ID	NLR class	sRNA prod.	miR482b	miR482e	miR482c	miR2118a	miR2118b
Solyc11g065780	CNL	1159.4	2	3	4	-	-
Solyc05g008070	CNL	569.4	1.5	2.5	-	-	3.5
Solyc02g036270	CNL	514.6	1.5	3.5	3.5	-	3.5
Solyc04g005540	CNL	303.2	3.5	3.5	3.5	-	4
Solyc11g071995	CNL	259.4	4	3.5	-	-	-
Solyc04g005550	CNL	129.6	3.5	3.5	3.5	-	4
Solyc10g051050	CNL	128.0	3	3.5	-	-	-
Solyc09g064610	CNL	77.7	-	-	-	-	-
Solyc01g008800	TNL	74.7	-	3.5	-	4	-
Solyc08g007630	CNL	70.0	-	-	-	-	-
Solyc05g009630	CNL	66.9	-	-	-	-	-
Solyc11g069990	CNL	65.3	4	3.5	-	-	-
Solyc11g069620	CNL	64.9	3.5	3.5	-	-	-
Solyc11g069925	CNL	58.4	3	3	-	-	-
Solyc05g005330	CNL	48.2	-	-	-	-	-
Solyc11g068360	CNL	34.4	3.5	3.5	-	-	-
Solyc07g049700	CNL	25.0	3.5	2	-	-	-
Solyc11g071410	CNL	23.3	-	-	-	-	-
Solyc10g085460	CNL	19.9	-	-	-	-	-
Solyc11g011350	TNL	18.2	-	3.5	-	3	-
Solyc11g006640	CNL	17.8	3.5	3	-	-	-
Solyc12g044190	CNL	17.3	3.5	3.5	-	-	-
Solyc11g020100	CNL	17.3	3	2.5	-	-	-
Solyc12g044200	CNL	17.2	3.5	3.5	-	-	-
Solyc07g005770	CNL	15.5	3	3.5	-	-	4
Solyc09g018220	CNL	14.7	-	-	-	-	-
Solyc11g064770	CNL	14.3	-	-	-	-	-
Solyc02g032650	TNL	14.3	3.5	4	3.5	-	2
Solyc12g006040	CNL	13.7	4	3	-	-	3.5
Solyc08g076000	CNL	12.5	-	-	1.5	-	2.5
Solyc11g069660	CNL	12.0	-	-	-	-	-
Solyc09g098130	CNL	11.1	-	-	-	-	-
Solyc06g005410	TAS5	431.2	-	-	-	-	1 / 2

Table 3.3 **Target prediction for all phasiRNA-producing NLRs.** Summary of pNLRs, with their gene id, class of NLR protein based on the phylogenetic analysis of Andolfo et al. [2014] and not in the presence of representative domains, total counts for 21-nt sRNAs (nRPM), and targeting scores for each individual microRNA (red indicates stronger targeting prediction). *TAS5* (bottom) is added for reference.

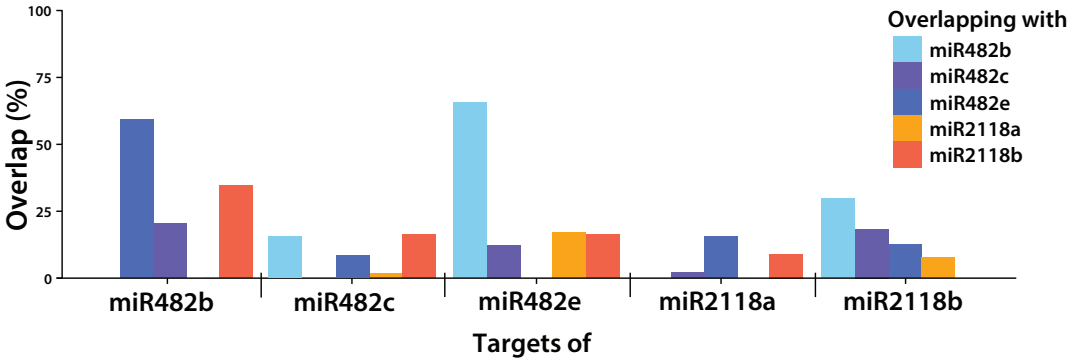

Fig. 3.5 **Overlapping predicted targets between microRNA members.** Bar-plot representing percentage of shared findings between total predicted targets for each microRNA compared with the other members of the family.

3.2.3 Phased small RNA signatures in tomato

In the previous section of this chapter I have defined phasiRNA producing NLRs (pNLRs) purely based on the abundance of sRNAs without addressing the existence (or not) of a phasing signature in their sRNAs. This responds to a rationalized personal call that is exposed below.

The first problem one encounters when looking for phased siRNA loci is the definition of a sRNA producing loci. Historically, sRNA loci have been defined as genomic regions with a discrete number of distinct matching sRNA reads, and with a maximum distance (in nt) between those matches [Kasschau et al., 2007; Lelandais-Brière et al., 2009]. It is important to keep in mind the limitation of this approach. Sequencing depth and quality of mapping introduces biases rarely accounted for. However, ultimately one can make a binary call on whether a loci produces sRNAs or not (based on the depth of our observations). This is not the case when identifying loci producing phasiRNAs. All methods compute a phasing score somewhat based on the abundance of sRNA in phased positions compared to the non-phased positions [Axtell et al., 2006; Shivaprasad et al., 2012; Zhai et al., 2011]. These two dimensions of the data makes the analysis more suited for a quality call of 'how phased is this loci?' and less so for a defining call. While all the current methods have been fairly successful in addressing the quality of phasing, I suggest that secondary siRNA loci should be considered in terms of the degree of phasing without a fundamental point of cut-off. I propose the use of phasing as a informative quality and not a defining term of a sRNA locus.

To identify and evaluate the register of phasiRNAs in tomato leaves, I ran my HTS data of small RNAs through PhaseR [Santos, 2014], a bioinformatic tool that addresses the quality of phased sRNA loci based on the sRNA abundance and register matching to the genomic location. In total, 13 of the top 20 phased

signatures (including the highest one) corresponded to pNLRs (Table 3.4). And within those, the great majority appeared in my data prediction for members of the miR482-clade. *TAS5*, the TAS gene targeted by miR2118b appeared as the second strongest signature among all genes. As expected, most candidates have already been reported to be targeted by different 22-nt miRNA in tomato. Such is the case for TIR1 and miR393 [Lin et al., 2013], or DLC2a and miR6026 [Kravchik et al., 2013]. Interestingly, all three miRNAs (miR482/2118, miR393, and miR6026) have been involved at different stages of plant immunity in tomato or other plant species [Kravchik et al., 2013; Navarro et al., 2006; Shivaprasad et al., 2012].

Gene ID	Start	End	phaseR score	Annotation
Solyc02g036270	278	3514	-26.9	Disease resistance protein (NLR class) family
Solyc06g005410	361	1513	-24.6	TAS5
Solyc10g051050	938	2624	-22.9	Disease resistance protein (NLR class) family
Solyc09g074520	409	2264	-21.0	Transport inhibitor response 1 (TIR1)
Solyc05g008070	287	2479	-18.8	Disease resistance protein (NLR class) family
Solyc04g005540	431	3965	-18.3	Disease resistance protein (NLR class) family
Solyc05g009630	701	2948	-17.4	Disease resistance protein (NLR class) family
Solyc11g069990	763	3397	-16.5	Disease resistance protein (NLR class) family
Solyc11g065820	2293	6253	-16.0	Disease resistance protein (NLR class) family
Solyc12g099870	1691	2072	-15.5	LRR RLK
Solyc11g011350	780	3629	-15.5	Disease resistance protein (NLR class) family
Solyc06g048960	518	3675	-15.3	Dicer-like 2a (DCL2a)
Solyc05g051230	2225	2876	-14.2	MOCS3-like
Solyc04g051190	154	1809	-14.2	P450 carotenoid β-hydrolase (CYP97A29)
Solyc11g071995	506	3586	-14.1	Disease resistance protein (NLR class) family
Solyc11g020100	957	2602	-14.0	Disease resistance protein (NLR class) family
Solyc01g058100	0	126	-13.9	TAS3-1
Solyc09g018220	562	2549	-13.5	Disease resistance protein (NLR class) family
Solyc02g032650	376	3091	-13.3	Disease resistance protein (NLR class) family
Solyc11g065790	12	409	-13.2	Disease resistance protein (NLR class) family

Table 3.4 **Summary of phasing signatures.** Top 20 phasing signatures in my sRNA dataset mapping to tomato genes. Colour code indicates when the gene is a predicted to be a preferential target of (blue) miR482 or (red) miR2118.

These results suggest that miR482-pNLR regulatory networks are the greatest contributors to the phasiRNA population in tomato leaves. Other miRNAs might be involved the phasiRNA production from other NLRs. But if these were

to target the same NLRs, the phasing signature would otherwise be compromised since the siRNAs would be distributed in several registers (assuming that any other miRNAs did not initiate the exact same register as miR482/2118b). Further experiments described in the next chapter aim to provide more conclusive evidence the major role of miR482 members in this process. Notably, *TAS5* appeared as the most clear phasiRNA signature outside of the pNLR family. In a following section I explore the relationship of this TAS gene with the sub-functionalization of the miR482/2118 members.

3.2.4 Degradome data analysis of miR482/2118 targets

PTGS in plants mainly occurs via the cleavage of a transcript, although translational regulation has not been thoroughly assessed [Axtell, 2013]. The PTGS process leaves a degradome signature of uncapped, mRNA cleavage or decay products [Baulcombe, 2004]. Parallel analysis of RNA ends (PARE) is a technique that utilizes high-throughput sequencing to profile these products on a genome-wide basis. One of the computational tools that predicts miRNA targets and validates those targets using PARE data is sPARTA [Kakrana et al., 2014]. I decided to make use of these resources by analysing publicly available datasets from previously published work in tomato. The only limitation of this technique is, as with most HTS approaches, the depth of the datasets elaborated. I ran sPARTA on the only available tomato PAREseq data from leaf tissue, which corresponds to Lopez-Gomollon et al. [2012].

Due to the small size of the dataset, only highly abundant cleavage products (limited by the expression of the transcript and the action of the miRNA) could be detected. Due to the two nucleotide deviation between miR2118 and miR482 members, degradation products could be unequivocally assigned to one clade or the other. The most abundant degradome signatures included both *TAS5* predicted sites by miR2118b (Table 3.5), indicating a clear directed action of miR2118b towards regulating this TAS gene. Only miR482-clade members were confirmed to cleave pNLRs. All of the confirmed targets where NLRs that produced sRNAs, further supporting the idea that the ability to produce phasiRNAs is intrinsic to the miRNA. Therefore, if the target is cleaved by such miRNA, it will always produce a detectable sRNA signature.

Similar to previous works, I was unable to identify strong validated NLR targets for miR2118a [Karlova et al., 2013; Lopez-Gomollon et al., 2012]. The two predicted targets for miR2118a, an asp dehydrogenase and a glycine de-

carboxylase, had acceptable signatures for the miRNA but the score was very poor (both with 7). Therefore it is unlikely that AGO would be able to cleave and they are probably not targets of this miRNA.

miRname	Target	Score	Cleavage Position	Reads	Perc. in 10nt	Peak category	Corrected p-value	Annotation
miR2118b/a	Solyc06g005410	1 / 4	513	62	1	0	0.0001/0.0058	TAS5
miR482b/c	Solyc02g036270	1.5 / 3	1017	12	1	2	0.0058/ 0.0155	Disease resistance protein (NLR class) family
miR2118b	Solyc06g005410	2	2536	18	0.3	2	0.0061	TAS5
miR482c	Solyc04g005540	3	864	7	1	2	0.0155	Disease resistance protein (NLR class) family
miR482e	Solyc11g013750	4	1778	9	1	2	0.0401	Disease resistance protein (NLR class) family
miR2118a	Solyc06g064550	7	6499	8	1	2	0.042	Aspartokinase-homoserine dehydrogenase
miR482c	Solyc06g062440	3.5	607	7	1	2	0.0475	Disease resistance protein
miR2118a	Solyc08g065220	7	4795	4	1	3	0.0487	Glycine decarboxylase p-protein

Table 3.5 **Summary of degradome (PARE) signatures.** Summary of validated degradome products for miR482/2118 members on the whole tomato transcriptome (p-value < 0.05). Peak category refers to PARE read abundances of that position correspond to (0) $>$ 90th percentile, (1) $>$ 75th, (2) $>$ 50th percentile, (3) $<$ 50th percentile of total PARE read signatures in the genome.

3.2.5 A novel TAS gene is targeted by miR2118b

For all miR482 members the top predicted targets were NLRs, while the top two targets sites of miR2118b were contained in a poorly characterized TAS gene in tomato named *TAS5*. In 2012, Li and colleagues developed a webserver[‡] containing bioinformatic tools for miRNA/tasiRNA analysis and identification. As part of their proof of concept analysis, they validated in tomato several TAS genes that are conserved across many plant species [Li et al., 2012a]. Additionally, they predicted the existence of a novel TAS gene triggered by sly-miR482d (miR2118b in my proposed nomenclature) and named it *TAS5*. Until now, this is the only confidently described target of the miR482/2118 family in Solanaceae that is not an NLR gene. The results from Li et al. [2012a] are the only previous characterization of *TAS5*. In my datasets, *TAS5* was one of the most prolific 21-nt siRNA producers (Figure 3.4) and the only validated target of miR2118b through degradome analysis (Table 3.5), making it a major contributor to the phasiRNA population and an interesting candidate to investigate.

Historically, the model by which 21-nucleotide phasiRNAs are triggered is divided in two mechanisms. These are known as the 'one-hit' and 'two-hit' pathways. In 'the one-hit' mechanism, a single miRNA directs cleavage of the mRNA target triggering the production of phasiRNAs in the fragment 3' to (or downstream of) the target site [Allen et al., 2005]. This 'one-hit' miRNA trigger is 22-nt in length, such as miR173 for the *TAS1 / TAS2* [Chen et al., 2010a] or miR482 for the pNLRs. For the 'two-hit' model, a pair of 21-nt miRNA target sites is used, of which cleavage occurs at only the 3' target site, triggering the production of phasiRNAs 5' to the target site [Axtell et al., 2006]. Interestingly, a previously unreported feature of *TAS5* is that it presents two miRNA target sites, a feature unlike any of the pNLRs triggered by the miR482 family. Until now, *TAS3* was the only well-described 'two-hit' locus in eudicots. Intriguingly,

[‡]The webserver was discontinued and is no longer available.

both cleavage signatures for target sites of *TAS5* appeared as the significant findings of sPARTA (Table 3.5), whereas in *TAS3* only the 3' target site is highly complementary to the miR390 and sliced. This characteristic makes *TAS5* unique in eudicots. Given the 22-nt nature of miR2118b but the presence of two cleavage sites in *TAS5*, I investigated which of the two models applied to the phasiRNA production of this TAS gene.

TAS5 is long transcript (1.5kb) produced from a locus in chromosome 6 of the tomato genome. The gene spans a region of 2.7kb with 4 exons (Figure 3.6A). After evaluating the register of phasiRNAs through PhaseR, I observed *TAS5* was one of the most significant and clear phasing signatures in the genome (Table 3.4 & Figure 3.6B). I calculated the percentages of sequenced sRNAs from each register. The phasing registers of sRNAs from the antisense strand were corrected to account for the 2-nt, 3' overhangs characteristic of DCL4 cleavage. Most sRNA were in phase with in the 5' cleavage site of miR2118b or in the immediately adjacent register (Figure 3.6C). This shift of the register to adjacent sites has already been reported previously in other TAS genes and is thought to be due to DCL4 mis-cleavage possibly caused by RNA secondary structures [Axtell et al., 2006; Li et al., 2012a] (It is worth noting that if this shift was due to the interference of any miR482 members, it would have occurred in the opposite direction). From this pattern, I conclude that *TAS5* is the most relevant target of miR2118b. This targeting triggers the production of highly abundant phasiRNAs 3' to the upstream target site. Therefore, although cleavage occurs at both target sites, it is unclear whether the second cleavage site contributes in any way to phasiRNA production and remains a topic for further investigation (Table 3.5).

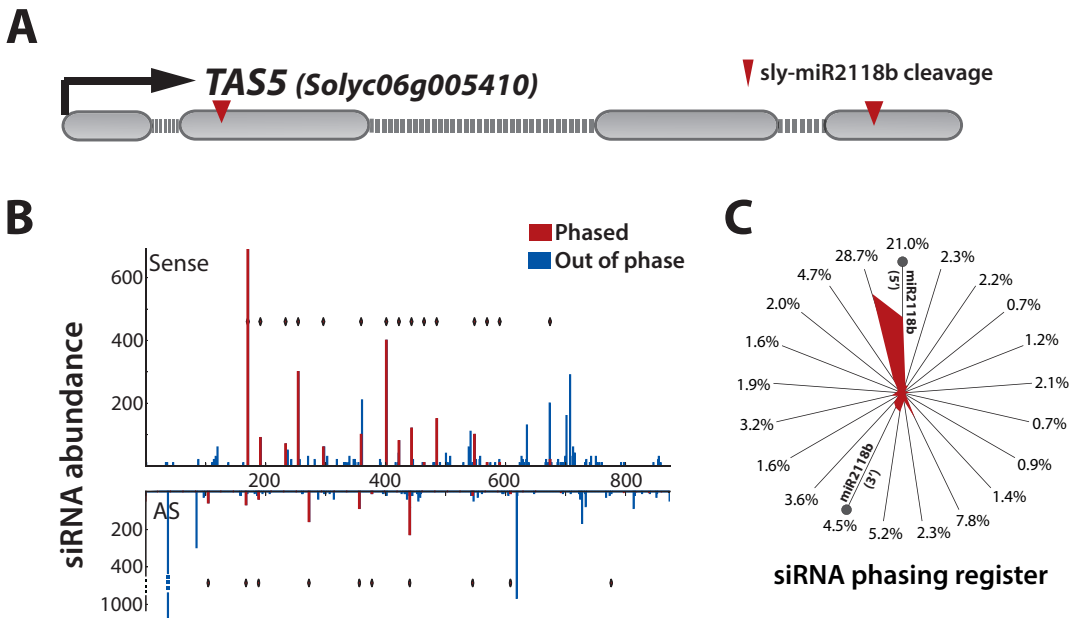

Fig. 3.6 ***TAS5* genomic features.** (A) Gene diagram of *TAS5* locus. A 2.7kb region containing 4 exons (grey boxes) and 3 introns (grey dotted lines). The arrow on the left indicates transcription start site and direction. (B) Number of sequenced small RNAs with 5' residues at each position between the cleavage sites of the *TAS5* transcript. Red bars indicate phased sequences while blue indicate out of phased. Red diamonds indicate that expected phased siRNA is present in the sample. (C) Distribution of the phasing of small RNAs at the *TAS5* locus. Each spoke of the wheel represents 1 of the 21 possible registers, with the percentage of small RNAs mapping plotted as distance from the centre (correction of 2-nt 3' overhangs of DCL cleavage was applied when assigning register in the anti-sense strand). The specific registers predicted from 5' and 3' cleavage sites of miR2118b are indicated with grey circles.

3.2.6 *TAS5* is a long non-coding RNA composed of degenerated NLR sequences

So far, I have presented evidence for sub-functionalization of miR482/2118 members in tomato, in particular for miR2118b and *TAS5*. However, an evolutionary explanation for the co-option of miR2118b to the targeting of *TAS5* and away from the targeting of pNLRs has not been explored. Li et al. [2012a] reported a stretch of *TAS5* sequence was similar to resistance protein coding sequences. By performing a BLAST search against the tomato transcriptome, I discovered that the whole transcribed region of *TAS5* is comprised of scrambled NLR-like sequences (Figure 3.7A & B). *TAS5* had similarity to both CNL and TNL types of NLRs on both the sense and antisense orientations. Overall, the whole transcript appears to be formed by three distinct regions. There are two tandem regions similar to the TIR motifs of TNL-type NLRs, and a region similar to a reversed NB-ARC motif reminiscent of the ones encoded in CNL-type NLRs (but in antisense orientation) inserted between them. Using bioinformatic tools [Kong et al., 2007; Yi et al., 2014], I predicted the transcript is unlikely to be translated into a functional protein (coding potential score of -1.146). Therefore, I concluded that this is a long non-coding RNA comprised of rearranged and degenerated sequences of multiple NLR genes.

I further explored the degree of similarity of these pseudo domains with all TNL and CNL genes of the tomato genome. Several NLR genes presented high degree of sequence similarity to *TAS5*, including well characterized resistance genes (3.7C). This similarity means that TAS5-derived phasiRNA could potentially target a great number of tomato NLRs. Notably there were only two members of the TNL family among the pNLRs (Table 3.3), and *TAS5* could therefore be the main source of phasiRNAs capable of regulating this NLR group.

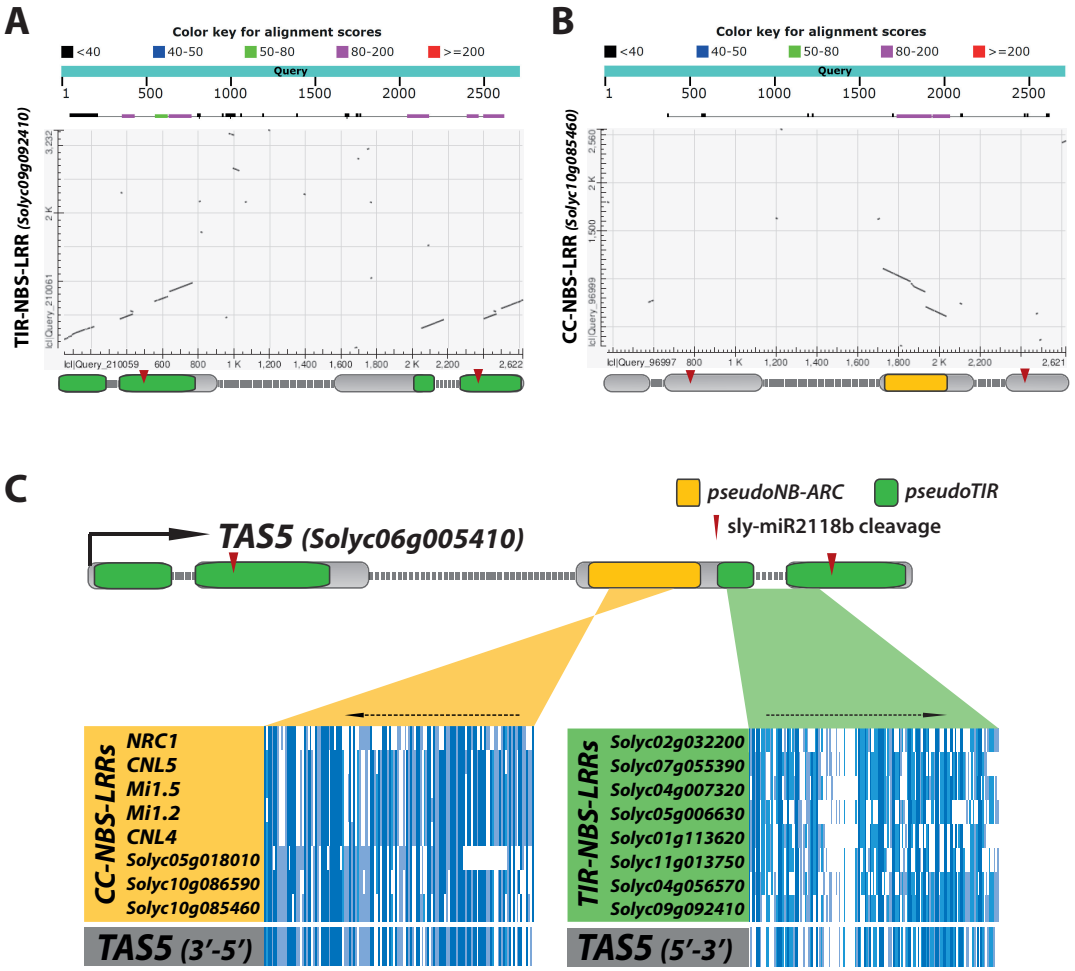

Fig. 3.7 **Sequence similarity between NLR domains and *TAS5*.** BLAST (Basic Local Alignment Search Tool) summary and dot plot matrix of *TAS5* versus (A) the closest TNL and (B) the closest CNL in the tomato genome. Gene diagram of *TAS5* locus is placed underneath the matrix indicates in (A) similarity to a TIR domain in green and (B) similarity to a NB-ARC in yellow. (C) Regions within exons with most significant sequence similarity with known NLR domains are highlighted in yellow and green. Nucleotide sequence alignment of these regions and known tomato NLRs are shown below. The degree of conservation for each nucleotide along the region is represented by the colour, with a dark blue denoting a high level of conservation and a light blue denoting a low level. Dotted arrows indicate direction of the sequence similarity.

3.2.7 *TAS5* and miR2118b are only conserved in *Solanum* species

To investigate the evolutionary history of *TAS5* and miR2118b, I decided to search for homologous sequences in other plant species. I performed BLASTN analyses using genomic sequences of both tomato genes against the genomes of all plant model organisms and all available genome assemblies of major Solanaceae species. The threshold expectation value was defined at 10^{-3}, a value set to filter out any spurious hits. Any hits were then curated manually for a match with the TNL and CNL regions in the same transcript to exclude NLR genes that were partially matching different regions of *TAS5*. For miR2118b I investigated the precursor sequence as well as the mature miRNA, in order to exclude other members of the miR482/2118 family. Similarly to the results published in de Vries et al. [2015], I was unable to find homologues of miR2118b outside of the *Solanum* clade. I was also unable to find homologues of *TAS5* outside of the *Solanum* clade, which suggests a coevolution scenario. It supports the hypothesis of an interlink between the evolution of microRNAs and the diversification their target genes [De Felippes et al., 2008; Xia et al., 2015], as with the co-evolution of transcription factors and their binding sites [Dermitzakis and Clark, 2002].

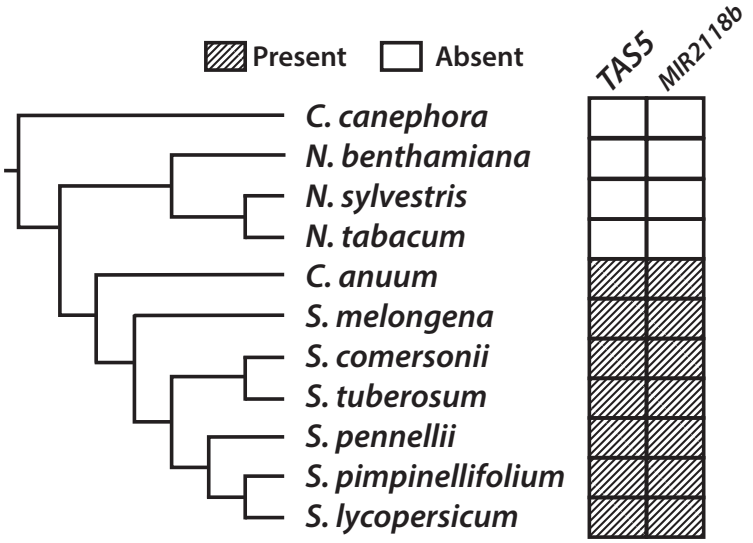

Fig. 3.8 ***TAS5* and miR2118b are only present in *Solanum* species** . Diagram summarising the presence or absence of genomic sequences matching *TAS5* and MIR2118b in Solanaceae and a close relative.The closest sequenced relative of Solanaceae, *Coffea canephora*, was included as an outgroup.

3.2.8 Expression profile of miR482/2118 during reproductive development

The members of miR2118, the only miR482/2118 clade present in monocots, have been described in this clade to have a developmental role during initiation of meiosis in panicles [Johnson et al., 2009; Zhai et al., 2015]. Whether this process is specific to monocots or is present in other angiosperms, remains largely unknown, as they have not been investigated in dicots (Figure 3.2). Most of the work has been done in *A. thaliana* and this organism does not contains any miR2118 members. Shivaprasad et al. [2012] showed that members of the miR482 clade are expressed at similar levels in every tissue, but no reference was made to the miR2118 group.

To assess any differences in expression of the members miR2118 clade in tomato, I performed northern blot analysis of miRNA expression in different developmental stages of reproductive development. Stable expression of miR482b throughout all the tissues tested validated previous observations [Shivaprasad et al., 2012]. Out of the two miR2118 members of tomato, miR2118b accumulation remained generally unchanged in all stages. Interestingly, miR2118a showed significantly higher accumulation during flower development than in leaf, peaking at inflorescence. This pattern of expression coincides with those observed in maize [Zhai et al., 2015] and suggests again that differences between these two clades could have functional significance in tomato beyond ubiquitous control of NLR expression.

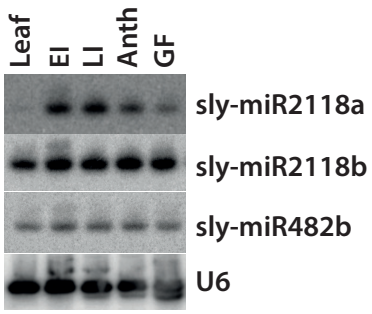

Fig. 3.9 **Expression pattern between different miR482/2118 members in tomato.** RNA gel blot analysis of tomato miRNAs in various tissues of plant development. The lower image shows the same blot hybridized with U6, as a loading control. EI: early inflorescence; LI: Late inflorescence; Anth: Anthesis; GF: Green fruit.

3.3 Discussion

The results described in this chapter provide new evidence into the sub-functionalization of members of the miR482/2118 family in tomato based on previously ignored structural differences. They also support and extend most observations made by Shivaprasad et al. [2012] in relation to the regulation of NLRs by the miR482 clade. Sequence analysis of the individual members of this family in tomato demonstrated the presence of two major structural variants (Figure 3.3). Two of the members that were initially annotated as miR482 variants were a clear product of mis-annotation and I proposed a revision of the current nomenclature of these miRNA family in tomato. The relevance of this proposed change became evident when among the 5 members of the family, just miR482b and miR482e accounted for the majority of pNLRs (Table 3.3), which was the initial role described for this family in tomato [Shivaprasad et al., 2012]. It is important to state again that the ability of a miRNA to produce phasiRNAs is intrinsic to the miRNA structure [Manavella et al., 2012]. Therefore, it is safe to assume that any real targets that were being expressed at an mRNA level in our samples, should contain sRNA counts in our sequencing datasets. What I hypothesize is that only the members of the miR482 clade in tomato account for the production of phasiRNA from pNLRs for the regulation of NLRs.

One of the members of the miR2118 clade, miR2118b, targeted predominantly a non-coding transcript known as *TAS5*. This transcript proved to be responsible for a great proportion of phasiRNAs produced in tomato (Figure 3.4). Additionally, out of the pNLRs, only few were targeted by miR2118b. Generally, all these pNLRs were also predicted for the miR482 members, and in most cases with better scores than for miR2118b. This would suggest that miR2118b would not be a predominant player in the production of phasiRNAs from NLRs. *TAS5* appeared to have a great degree of sequence similarity to key domains of CNL and TNL types of NLRs on both the sense and antisense orientations. Therefore,

any phasiRNA produced from *TAS5* contained sufficient similarity to NLRs and could potentially target these genes. On an evolutionary note, phylogenetic analysis showed that the presence of both *TAS5* and MIR2118b was restricted to *Solanum* ssp. of the Solanaceae (Figure 3.8). De Felippes et al. [2008] suggested that genic regions of a multi-copy miRNA family were not under strong selective pressure and if in any of these cases a newly evolved miRNA fortuitously guides cleavage of an mRNA, this interaction could become the subject of selection. This potential route of miRNA/target coevolution would be similar to what has been suggested for transcription factor binding sites, where a high proportion of binding sites are not functional across lineages and have a high turnover rate [Dermitzakis and Clark, 2002]. This idea was also supported by Xia et al. [2015] when studying phasiRNA loci in gymnosperms. I hypothesize that co-evolution might have played a role in the conservation of this TAS gene and its miRNA trigger. Once one of the two is lost, no selective pressure remains to retain the other. This would explain the conjoint evolutionary history of the pair.

Intriguingly, miR2118a showed little affinity to targeting NLRs compared to all the other members (Table 3.2). It also presented a unique and dynamic expression profile and its accumulation peaked during reproductive development (Figure 3.9). In monocots, studies in wheat, maize, and rice have failed to identify phasiRNAs from NLRs or members of the miR482 clade. All the monocot members belong to the miR2118 clade, accumulate in panicles, and are involved in the production of phasiRNA from intergenic non-coding loci [Johnson et al., 2009; Zhai et al., 2015]. The function of these other phasiRNAs remains a mystery and they have only been described in monocots and some gymnosperms (Figure 3.2). My evidence for reproductive expression and the lack of NLRs as predicted targets of miR2118a in tomato resembles greatly the monocot findings. This pattern suggests the possibility that this microRNA would function in the same fashion as the monocot miR2118, as part of an

evolutionary conserved pathway of yet unknown function.

Shivaprasad et al. [2012] suggested that miR482/2118 functioned as buffers of resistance gene expression in times where there is no pathogen challenge. They would reduce any fitness costs of R gene mis-expression, while retaining the capability to rapidly respond to biotic attacks. An additional hypothesis is that negative regulation of NLRs may also contribute to their evolution [Fei et al., 2013], as miR482/2118 may buffer spontaneous mutations in NLRs to prohibit direct deleterious effects such as constitutive induction of immune responses. These hypotheses are not mutually exclusive, and still apply to my findings in tomato. However, my findings suggest that this might only apply to the miR482 clade of the family. One of the other members, miR2118b, appears to be involved in the regulation of NLRs through the action of a non-coding RNA. This non-coding RNA would be the first reported TAS gene involved in the defence response in the whole plant kingdom. The function of the last member, miR2118a, remains elusive and a matter of speculation. This example of diversification of functions inside a miRNA family should also encourage the small RNA field to re-evaluate conclusions based on exclusively *in silico* predictions of miRNA targets.

3.4 Acknowledgements

Dr. Bruno Santos provided valuable advice on target prediction analysis.

Chapter 4

Dissecting miR482 and miR2118b functions in tomato

4.1 Introduction

The previous chapter describes, based on *in silico* predictions, small RNA sequencing and degradome data, the direct miR482/2118 targets. The suggested model was that miR482-clade members directly control the expression NLR targets, while miR2118b influences the expression of defence genes indirectly through the action of *TAS5*, a non-coding locus in tomato. However, while I showed that the production of of phasiRNA was correlated with the target predictions, I did not have data from a functional test. In order to address this deficiency, I developed a reverse genetic approach, in which I disrupted the targeting of miR482 family members and miR2118b.

Our current understanding of genes and molecular traits has been largely based on studies using mutational approaches. The analysis of collections of mutants by forward and reverse genetics is one the most common techniques

in recent research and, correspondingly, second most cited paper of the last 20 years in plant sciences describes the *A. thaliana* collection of mapped T-DNA mutant lines [Alonso et al., 2003]. Unfortunately, unlike *A. thaliana*, a mapped collection of mutants does not exist for tomato. The sole existing large collection has only been visually phenotyped in the field [Menda et al., 2004] and only a few mutants have been mapped and characterized. From these, the two that are relevant for this study are knock-out alleles of *RDR6* and *DCL* genes [Yifhar et al., 2012]. However, further study of these mutants will only address the general participation of the phasiRNA pathway in the regulation of defence response as has been done in *A. thaliana* [Boccara et al., 2014] and not the specific effects of the miR482/2118 family. Since no mutants of the tomato miR482/2118 family have been described in tomato, one approach would be the use of novel CRISPR-Cas9 techniques of gene editing. This technique has already been successfully applied in tomato [Brooks et al., 2014]. However, the multi-copy nature of the miR482/2118 family, the technical restriction of finding a protospacer adjacent motif (PAM) for the Cas9 targeting in the generally short sequence of a miRNA gene, and the time-frame restrictions of this project, made it inappropriate for this study. Therefore, other approaches were necessary for addressing the specific effects of the miR482/2118b family.

In *A. thaliana*, Franco-Zorrilla et al. [2007] discovered that the non–coding gene *IPS1* (*INDUCED BY PHOSPHATE STARVATION1*) contains a motif with sequence complementarity to the phosphate starvation–induced miRNA miR399. However, the pairing in this motif is interrupted by a mismatched 3-nt loop at the expected miRNA cleavage site. This characteristic loop inhibited the cleavage of *IPS1* RNA and instead sequestered the miR399. Sequestration of miR399 resulted in increased accumulation of the miR399 targets. Yan et al. [2012] later adapted these finding into a synthetic non-coding RNA containing short tandem target mimics that effectively sequestered the desired miRNAs. These short tandem target mimic are composed of two short sequences mimick-

ing small RNA target sites, separated by a linker of an empirically determined optimal size. Yan et al. [2012] also showed that the sequestration leads to the degradation of targeted miRNAs by small RNA degrading nuclease proteins.

In this chapter, I explore the use of target mimics for studying miR482/2118 function in tomato in relation to defence response. The aim was to create miRNA knock-downs to study miR482/2118 members involved in NLR regulation. For that purpose, I first designed target mimics constructs for the specific and simultaneous sequestration of different miR482/2118 members and transiently tested the efficacy of these mimics. I then generated stable transgenic lines in tomato that successfully sequestered the desired miRNAs. Finally I addressed the effect of the miRNA sequestration on the downstream production of phasiRNAs.

4.2 Results

4.2.1 Construct design for target mimic assays

To achieve a time-efficient approach, miRNA target mimic constructs were tested using a synthetic transient expression assay in *Nicotiana benthamiana* leaves based on GFP expression. This assay provided a preliminary screening method prior to the creation of stable transgenic tomato lines. The mimic assay was designed so that an initial vector was introduced containing a GFP with a miRNA target site, initially allowing for GFP expression (Figure 4.1). If the miRNA of interest successfully cleaved their targets, GFP production was blocked (Figure 4.1). Finally, a target mimic construct complementarity to the miRNAs of interest was introduced, thus binding the miRNA and evading the silencing of the GFP transcript (Figure 4.1).

Short tandem mimics for the miRNAs were designed following the instructions of Yan et al. [2012]. In summary, each construct consisted of two mimic sites, separated by a spacer sequence. Each binding site was complementary to the miRNA of interest, but had a three nucleotide insertion between the eleventh and twelfth nucleotides to create a wobble that prevented cleavage. The first construct, *MIM482*, contained one binding site complementary to miR482b and one site complementary to miR482e. The second construct, *MIM2118b*, contained two binding sites complementary to miR2118b (Figure 4.2). Based on my previous analysis, miR2118a was not investigated as its role appeared to be linked to reproductive development (Figure 3.9), similarly to its homologs in monocots [Johnson et al., 2009; Zhai et al., 2015].

The GFP sensors were designed by cloning a *GFP* sequence containing three binding sites at the 3' end, separated by spacer regions. *GFP-T482* contained complementary binding sites for miR482b, miR482c and miR482e,

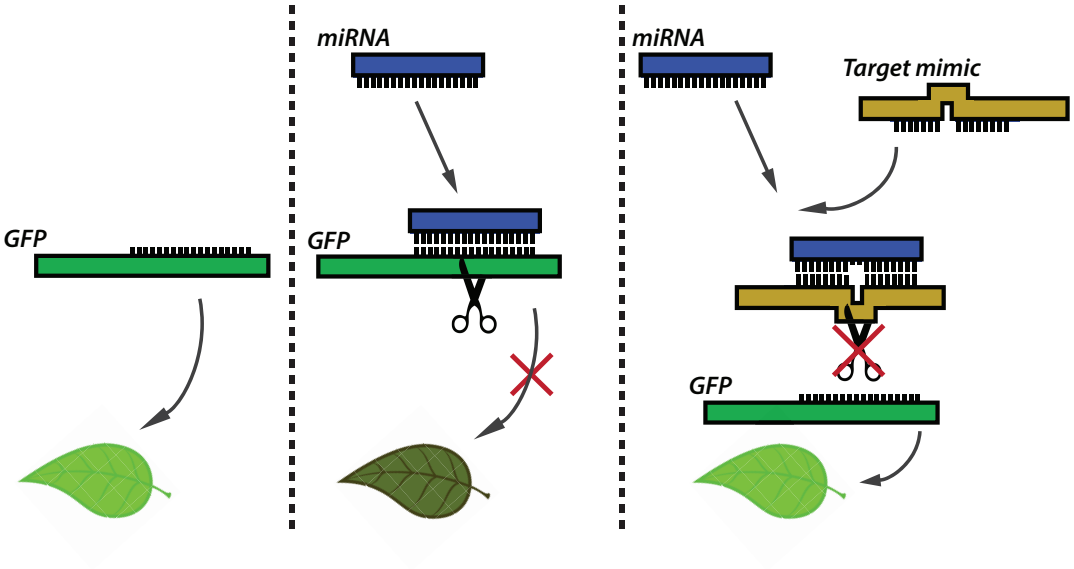

Fig. 4.1 **Target mimic model and assay design.** Experimental concept of target mimics for investigating miR482/2118 function. (Left panel) A GFP transcript containing a miRNA target site will allow GFP protein accumulation. (Middle panel) Introduction of miRNA which will bind to the target site and cleave the transcript, ultimately preventing GFP protein accumulation. (Right panel) Addition of a target mimic which will bind to miRNA. A wobble position at the cleavage position will inhibit cleavage. Sequestration of the miRNA will allow GFP transcripts accumulation and trigger GFP protein levels will be restored.

while *GFP-T2118b* contained three complementary binding sites for miR2118b (Figure 4.2).

The sequences of the most highly expressed member of the miR482 family (miR482e) and of miR2118b were cloned. The correct genomic coordinates of each miRNA precursor were established using the current tomato genome assembly and RNA sequencing data from leaf tissue elaborated by previous members of the Baulcombe lab. Precursors were then cloned under the constitutive over-expression promoter *CaMV 35S*. These constructs are referred to as *MIR482ox* and *MIR2118box* (Figure 4.2).

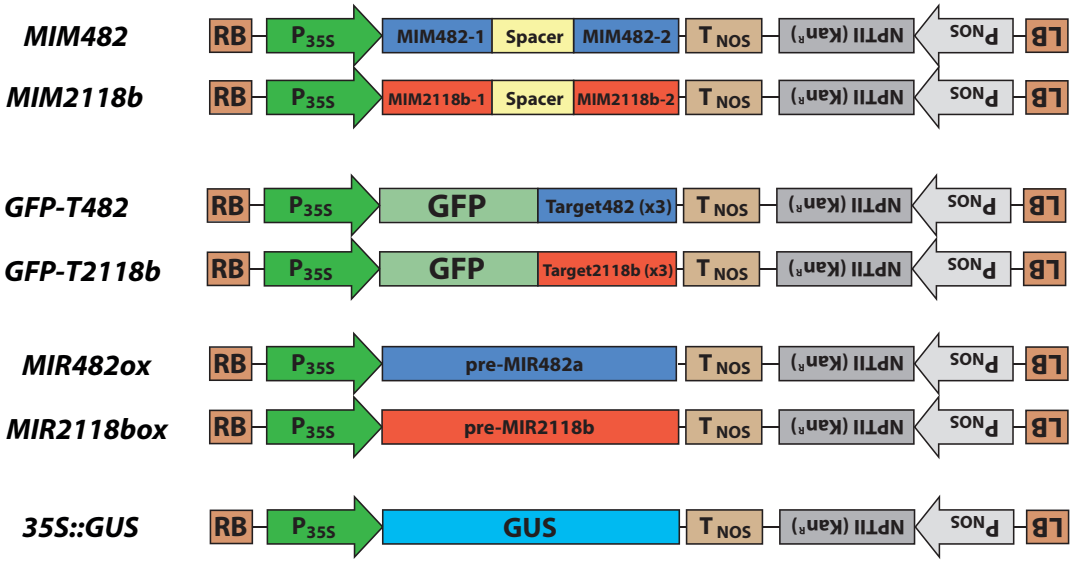

Fig. 4.2 **Constructs for target mimic assay.** Diagram indicating the characteristics of each construct used in this work. All relevant elements are expressed under a constitutive promoter. *MIM* constructs consist of two mimic sites complementary to the miRNA of interest, but had a three nucleotide insertion between the eleventh and twelfth nucleotides to create a wobble that prevented cleavage, and separated by a spacer sequence. *GFP-T* constructs include the *GFP* sequence with three miRNA binding sites at the 3' end, separated by spacer regions. miR-ox contructs include the cloned genomic sequence of each miRNA precursor. A *35S::GUS* construct is added as a control of no relevant function to the assay.

4.2.2 Transiently testing target mimics in *Nicotiana benthamiana*

The experimental design involved five different treatments per condition (miRNA) to test the target mimics (Table 4.1):

1. Plant were infiltrated with the GFP sensor to address basal levels of GFP expression.
2. Alongside the GFP sensor, the correct miRNA over-expression constructs was added and silencing of GFP was assessed. This treatment tested miRNA functionality.
3. An incorrect miRNA was added instead, as a control to ensure the specificity of the silencing effect.
4. The correct target mimic construct was added to the GFP sensor and the miRNA. If the target mimic worked, the miRNA would be sequestered and the GFP expression restored. All three together addressed the functionality of the target mimic.
5. An incorrect target mimic was added to assess the specificity of the sequestration.

Since different treatments contained variable number of constructs, a *35S::GUS* control vector of no relevant function was used so that the bacterial load of each treatment remained constant.

I proceeded to transiently express and test all construct combinations in *N. benthamina* plants via agroinfiltration. The *Agrobacterium tumefaciens* suspension carrying the vectors of interest was introduced into a plant leaf by direct injection. GFP targets showed sufficient level of basal fluorescence to be assessed visually (Figure 4.3). Overexpression of the correct miRNA silenced the GFP transgene; and expression of the incorrect miRNA did not

GFP-Target	miRNAox	MIMIC	Predicted fluorescence	Experiment objective
482 2118b	- -	- -	+++ +++	Establish functionality and basal level of GFP expression
482 2118b	482 2118b	- -	+ +	Test functionality of miRNAs
482 2118b	2118b 482	- -	+++ +++	Test specificity of miRNAs
482 2118b	482 2118b	482 2118b	+++ +++	Test effectiveness of target mimics
482 2118b	482 2118b	2118b 482	+ +	Test specificity of target mimics

Table 4.1 **Construct combination scheme and predictions.** Summary of the experimental design for transient testing of target mimics. Controls with incorrect components to check specificity are indicated in red. Predicted high GFP expression is indicated in green.

result in silencing of the transcript, demonstrating the specificity of the silencing. Target mimics successfully released the silencing of the GFP transcript allowing for GFP protein accumulation. The inability of unmatched mimics to release the silencing demonstrated the precision of mimics. Visual phenotypes were validated in a quantitative manner using Western blot analysis of protein extracts from agroinfiltrated leaves (Figure 4.3). In summary, target mimics successfully and specifically blocked miRNA function in this transient assay. Since the results showed that the target mimic constructs designed in this study were functional, I proceeded with generating of stable transgenic tomato lines carrying the target mimic constructs. The transient assay presented here also proved to be an efficient tool here for rapid validation of target mimic constructs.

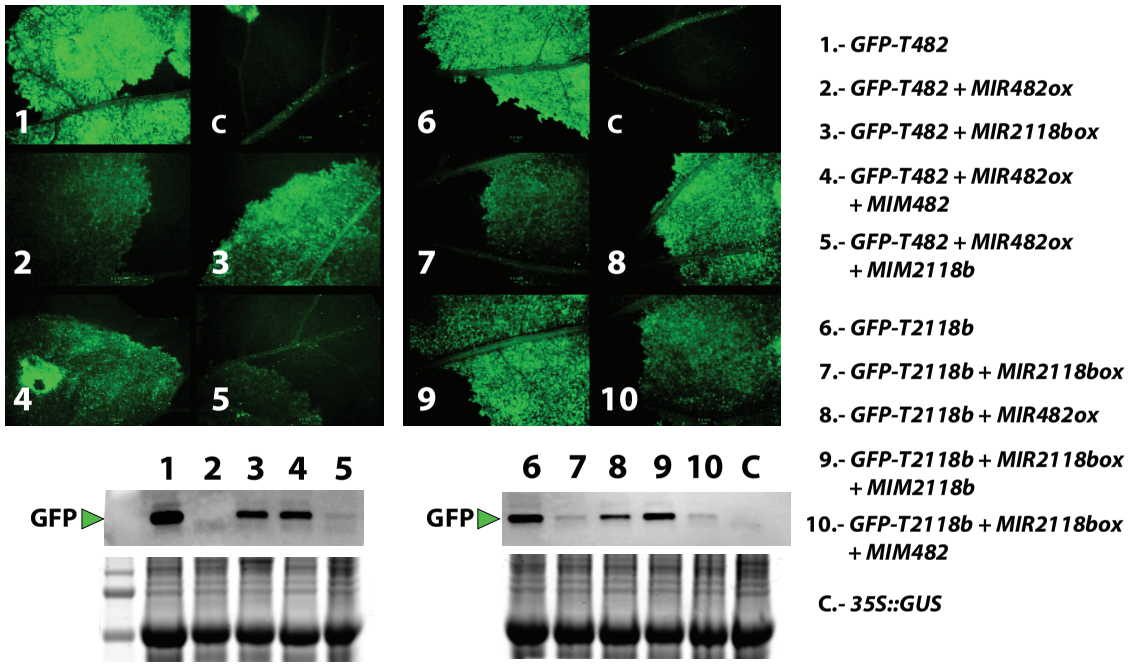

Fig. 4.3 **Transiently testing target mimics in *Nicotiana benthamiana* leaves.** GFP expression estimated (Upper panel) *in vivo* under UV light and (Lower panel) via Western blot analysis of leaf protein extracts using an antibody against GFP. Vector combinations for each number are indicated on the side.

4.2.3 Characterization of tomato transgenic lines expressing target mimics

The majority of plant stable transformation, especially in crop species, is achieved via regeneration of adult plants through tissue culture. In only few species, including plant model *A. thaliana*, methods to generate stable lines '*in natural conditions*' have been achieved. Unfortunately, such approaches have not been developed for tomato. Tissue culture transformation of tomato presents overt challenges. The lengthy regeneration in order to obtain stably transformed tomato plants is still a hurdle that can take up several months with a success rate lower than 10 % per explant [Eck et al., 2006]. Additionally, tomato transformation is prone to generate lines with low expression of the vector of interest. This is probably due to the large size of the tomato genome and the high chances of insertion within transposon silenced sequences.

To secure a sufficient amount of transformation events, approximately two hundred explants per construct were transformed via tissue culture. Transformed calli were grown under selective pressure using the antibiotic kanamycin. Resistance to kanamycin was conferred by the *NPTII* gene present in the construct. Any regenerated plant was PCR-screened for the presence of the transgene.

In 16 validated T0 lines (8 per construct), sequestration of the target miRNA was analysed via Northern blot 4.4. Due to the close similarity of sequences, both miRNA probes were designed using highly specific locked nucleic acid (LNA)-modified probes. LNA is a high-affinity RNA analogue with high mismatch discrimination and thermal stability [Válóczi et al., 2004]. This allows to distinguish highly similar nucleotide sequences. Total miRNA signal was normalized to U6. Overall, a wide range of levels of miRNA sequestration was observed (Figure 4.4). For each construct, at least 3 independent lines had a

significant reduction on the levels of their target miRNAs. These results prove that target mimics are functional in tomato. Further analysis of the lines is presented in the following sections.

Out of all stable lines, the two lines with highest sequestration level per construct were selected and brought to the next generation (lines #1 and #2 for *MIM482*, lines #3 and #4 for *MIM2118*). T1 plants from these parents were grown and used for any subsequent experiments.

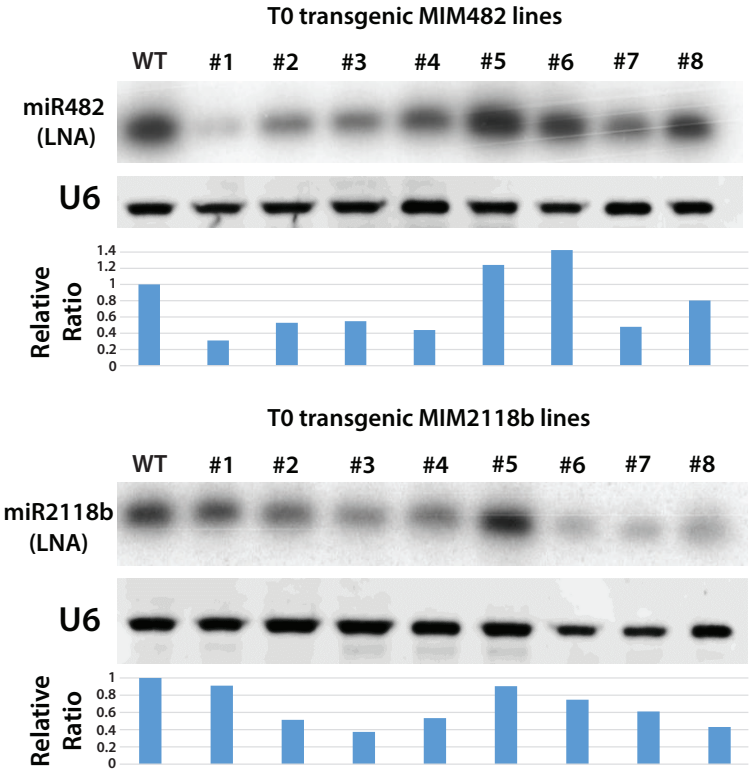

Fig. 4.4 **Northern blots of transgenic mimic lines.** RNA gel blot analysis of tomato transgenic mimic lines. Upper line shows miRNA blotted with highly specific locked nucleic acid (LNA) probes. Lower image shows the same blot hybridized with U6, as a loading control. Barplot indicates relative ratio of miRNA signal vs U6 signal.

4.2.4 Target mimic action is highly specific.

To obtain further insight into the action of my target mimics, I generated high-throughput sequencing libraries of sRNAs from 4-weeks old leaves. I selected descendants of the lines with highest sequestration level from both mimic constructs (hereafter, *MIM482*, and *MIM2118b* lines) and included a wild type line as a control. Libraries were generated using two independent lines per condition and three biological replicates per line. A total of 18 libraries were made, multiplexed, pooled and sequenced.

The primary question I aimed to address was the specificity of the target mimics, particularly among the miR482/2118 family members. To address this issue, reads aligning to all known tomato miRNA sequences (miRBase) were quantified and compared between conditions. In *MIM482* lines, all miR482-clade members presented an approximate 10-fold reduction compared to wild type. All other miRNAs, including members of the miR2118 clade, were within the same range as wild type (Figure 4.5). In *MIM2118b* lines, only miR2118b presented an approximate 10-fold reduction while all other miRNA remained unchaged compared to wild type (Figure 4.5). These results confirmed the high level of specificity of the target mimic RNAs.

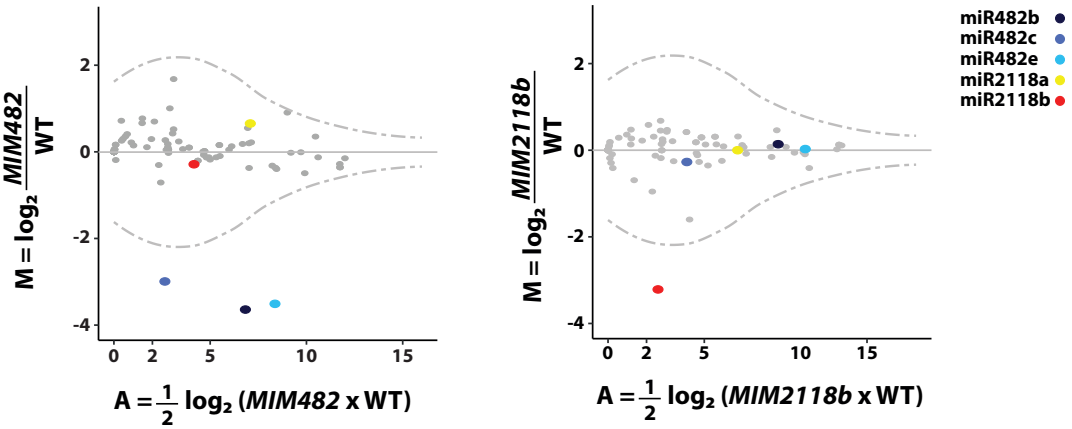

Fig. 4.5 **Target mimics sequester specific miRNAs.** MA plot showing fold changes of miRNAs in mimic lines. Coloured dots indicate miRNAs of the miR2118/482 family; grey indicate other miRNAs. The dotted line represents a Poisson distribution with 1 % significance values at the top and bottom of the range, applying the 0 correction (if nreads=0;+1). sRNA reads are normalized to the whole library with reads per million (nRPM) and presented as the mean from three biological replicates.

4.2.5 Target mimics disrupt phasiRNA production from NLRs and *TAS5*

Once the specific sequestration of target miRNAs was confirmed in my mimic lines (Figure 4.5), I proceeded to investigate the effect of mimic expression on the sRNA population and phasiRNA production. In a *Medicago truncatula* study using transgenic hairy roots over-expressing 22-nt miRNAs, Fei et al. [2015] showed that the abundance of phasiRNAs correlates with the levels of both miRNA triggers and the target transcript. This effect has also been observed in tomato, where DCL1 knock-downs hampered the production of miRNAs, which ultimately led to a global negative effect on phasiRNA production [Kravchik et al., 2013]. Therefore, obstructing the production of phasiRNA-triggering miRNAs should have a direct negative effect on downstream phasiRNA accumulation.

Genome wide siRNA studies have usually focused on heterochromatic RNA-directed DNA methylation (RdDM). In these cases, correctly identifying the regions of the genome associated with sRNA has always been a major concern. Historically, small RNA loci have been vaguely defined as regions which have a high density of tags aligning to them. Since phasiRNA requires PolII transcription, I decided to restrict my loci to those overlapping with genes (based on ITAG annotation 3.2; The Tomato Genome Consortium [2012]). I performed a statistical analysis using total sRNA counts to identify differential sRNA loci (DSLs), based on the same principle of differential expression used when analysing RNAseq datasets. DSLs in *MIM482* line showed a high enrichment of pNLRs, and almost all of these were predicted targets of miR482 members. None of the non-NLR DSLs were predicted targets of any miR482. In the case of *MIM2118b*, exclusively *TAS5* appeared as a DSL (Table 4.2).

Differential sRNA loci between MIM482 and wild type

Gene ID	Annotation	Log2(FC)	Condition	adj. p-value
Solyc11g065780	Disease resistance protein (NLR class) family	1.596	WT > MIM482	2.3E-06
Solyc09g064610	Disease resistance protein (NLR class) family	3.255	WT > MIM482	3.2E-05
Solyc01g008790	Non specific phospholipase C	2.730	WT > MIM482	2.3E-04
Solyc01g067165	Disease resistance protein (NLR class) family	1.971	WT > MIM482	3.4E-04
Solyc01g008800	Disease resistance protein (NLR class) family	2.508	WT > MIM482	4.5E-04
Solyc04g017620	F-box family protein	2.065	WT > MIM482	5.4E-04
Solyc12g044190	Disease resistance protein (NLR class) family	1.924	WT > MIM482	9.4E-04
Solyc07g005770	Disease resistance protein (NLR class) family	2.787	WT > MIM482	1.4E-03
Solyc08g076000	Disease resistance protein (NLR class) family	2.336	WT > MIM482	1.8E-03
Solyc02g036270	Disease resistance protein (NLR class) family	1.921	WT > MIM482	2.3E-03
Solyc02g032650	Disease resistance protein (NLR class) family	2.140	WT > MIM482	3.6E-03
Solyc04g005550	Disease resistance protein (NLR class) family	2.306	WT > MIM482	7.1E-03
Solyc04g005540	Disease resistance protein (NLR class) family	1.751	WT > MIM482	1.7E-02
Solyc01g067147	Asterix-like protein	2.230	WT > MIM482	3.2E-02
Solyc01g100380	Calreticulin	1.322	WT < MIM482	1.2E-02
Solyc03g112330	U-box domain-containing kinase family protein	1.684	WT < MIM482	2.5E-02
Solyc03g112335	O-acyltransferase (WSD1-like) family protein	1.637	WT < MIM482	4.0E-02
Solyc09g097780	Glycine-rich protein	1.499	WT < MIM482	4.8E-02

Differential sRNA loci between MIM2118b and wild type

Gene ID	Annotation	Log2(FC)	Condition	adj. p-value
Solyc06g005410	TAS5	2.168	WT > MIM2118b	8.09E-05

Table 4.2 **Most significant differential sRNA loci in MIMIC lines.** Genetic loci with differential accumulation of sRNAs, with their gene id, annotation, log2 fold changes, direction of the change, and adjusted p-value (cut-off of 0.05). Colour code indicates when the gene is a predicted to be a preferential target of (blue) miR482 or (red) miR2118.

Subsequently, I decided to validate whether these siRNA changes correlated with those triggered by miR482/2118 members. To avoid any biases from other sRNA pathways such as the RdDM-dependent siRNAs, I focused exclusively on 21-nt sRNAs for the subsequent analysis since phasiRNAs are mostly this size class [Axtell, 2013]. I calculated the phasiRNA distribution in three phased loci: *LRR2* (*Solyc04g005540*), a well studied CNL gene targeted by miR482 [Shivaprasad et al., 2012], and two TAS genes (*TAS5* and *TAS3-1*). All these loci showed a clear siRNA accumulation pattern originating from their expected miRNA cleavage sites (Figure 4.6). *MIM482* lines showed a 4-fold reduction of 21-nt siRNAs compared to wild type in *LRR2* (Figure 4.6). This effect was not observed in *MIM2118b* lines. In concordance to previous results, *MIM2118b* showed a 5-fold reduction in *TAS5*-derived siRNAs, while such effect was not

seen for *MIM482* lines (Figure 4.6). I also investigated *TAS3-1* as a control as it is a phasiRNA locus, but unrelated to the miR482/2118 family. None of the mimic lines had an altered accumulation of TAS3-derived siRNAs compared to wild type (Figure 4.6). In summary, this demonstrated that the effect of the target mimics correlated with phasiRNA production and was specific to the miRNAs for each target mimic.

Upon closer inspection of the NLRs, the total *NLR*-derived siRNA population was halved in *MIM482* lines compared to wild type, while remained unaffected in *MIM2118b* (Figure 4.7A). As expected, out of all NLRs, pNRLs give rise to the great majority (>90%) of siRNAs (Figure 4.7A). Even among pNLRs, a large proportion of these siRNAs arise from just a handful of highly productive pNRLs.

In order to account for any bias introduced by these highly productive loci when looking at total abundance of *pNLR*-derived siRNAs, I also investigated the collective changes in individual *pNLRs*. *MIM482* lines presented a highly significant reduction of siRNA production from individual *pNLRs* compared to all genes, while no change was observed for *MIM482* lines (Figure 4.7B). In summary, *MIM482* reduced the total production of phasiRNAs from NLRs, and this reduction was due to cumulative changes in the majority of pNLRs and not just a handful of the most productive ones.

In conclusion, target mimics were highly specific sequestering either miR482-clade members (*MIM482*) or miR2118b (*MIM2118b*). The production of *pNLR*-derived siRNAs was significantly reduced exclusively in *MIM482* (Figure 4.7). In *MIM2118b*, only the TAS5-derived siRNAs were significantly decrease (5-fold reduction; Figure 4.6 & Table 4.3). These results provide further evidence for the sub-functionalization hypothesis presented in the previous chapter regarding

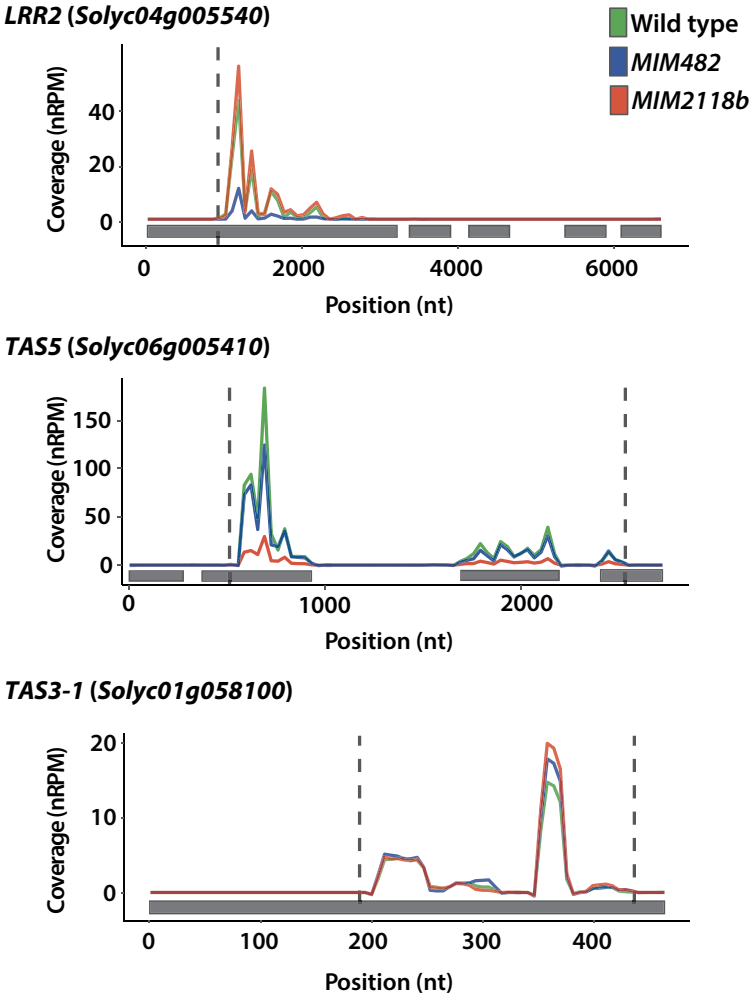

Fig. 4.6 **Target mimics specifically inhibit production of phasiRNAs at the sequestered miRNA targets.** Abundance of 21-nt siRNAs along a CNL, *TAS5*, and *TAS3-1* locus. Positions corresponding to the cleavage sites of their respective miRNA triggers (miR482, miR2118b, and miR390 respectively) are indicated by dotted lines. Grey boxes indicate the position of exons. sRNA reads are normalized to the whole library with reads per million (nRPM) and presented as the mean from three biological replicates.

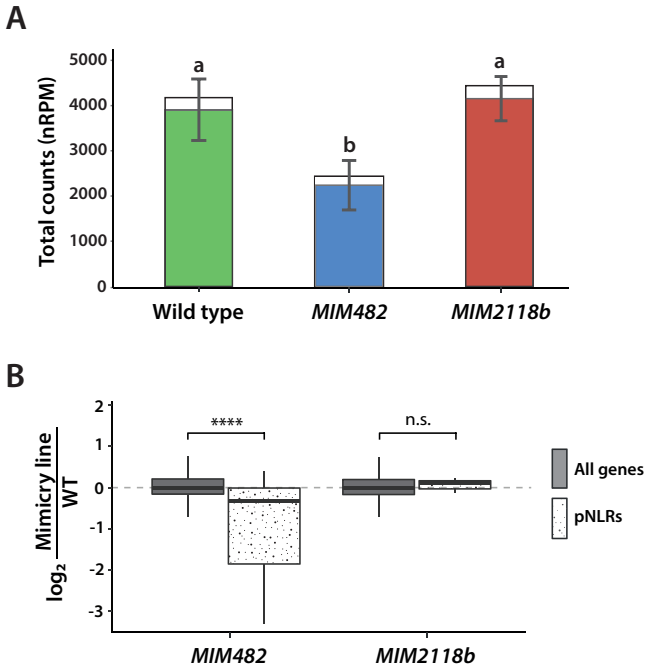

Fig. 4.7 **Small RNA production of *pNLRs* is decreased in *MIM482* lines.** (A) Bar-plot of total 21-nt siRNA counts in NLRs in WT and mimic lines. Coloured fraction correspond to pNLRs. Statistically significant differences were found using one-way ANOVA test followed by Tukey HSD at 95% confidence limits. (B) Box-plots of 21nt siRNA difference between each mimic line and WT at individual loci. Grey-boxes represent siRNAs mapping to NLRs that contain the canonical target site for miR482s, patterned-boxes represent all genes in the tomato genome. Welch Two Sample t-test were performed (p-values of 1.173e-09 and 0.8867, respectively). sRNA reads are normalized to the whole library with reads per million (nRPM) and presented as the mean from three biological replicates.

miR2118b and miR482 and their different targets.

General information		small RNA Analysis					Target Predictions	
		Total 21nt counts			Log2 difference			
Gene_ID	NLR class	WT	M482	M2118b	M4/W	M2/W	miR482	miR2118
Solyc11g065780	CNL	1159.4	420.5	1225.5	-1.463	0.080	be c	-
Solyc05g008070	CNL	569.4	578.4	577.9	0.023	0.021	**be**	b
Solyc02g036270	CNL	514.6	124.2	545.1	-2.050	0.083	b ec	b
Solyc04g005540	CNL	303.2	78.5	349.7	-1.949	0.206	bec	b
Solyc11g071995	CNL	259.4	338.2	291.3	0.382	0.167	be	-
Solyc04g005550	CNL	129.6	24.1	150.8	-2.429	0.219	bec	b
Solyc10g051050	CNL	128.0	131.0	146.7	0.034	0.197	b e	-
Solyc09g064610	CNL	77.7	7.9	83.9	-3.296	0.111	-	-
Solyc01g008800	TNL	74.7	11.2	82.8	-2.738	0.148	e	a
Solyc08g007630	CNL	70.0	48.9	78.6	-0.518	0.167	-	-
Solyc05g009630	CNL	66.9	67.5	71.9	0.013	0.103	-	-
Solyc11g069990	CNL	65.3	56.2	61.3	-0.217	-0.091	be	-
Solyc11g069620	CNL	64.9	36.1	72.4	-0.845	0.158	be	-
Solyc11g069925	CNL	58.4	47.0	58.9	-0.314	0.012	**be**	-
Solyc05g005330	CNL	48.2	47.1	38.3	-0.032	-0.329	-	-
Solyc11g068360	CNL	34.4	24.0	33.6	-0.523	-0.036	be	-
Solyc07g049700	CNL	25.0	22.8	28.7	-0.129	0.200	b e	-
Solyc11g071410	CNL	23.3	27.1	26.6	0.215	0.186	-	-
Solyc10g085460	CNL	19.9	18.4	13.7	-0.114	-0.532	-	-
Solyc11g011350	TNL	18.2	21.7	25.8	0.252	0.504	e	a
Solyc11g006640	CNL	17.8	14.4	19.5	-0.305	0.132	b e	-
Solyc12g044190	CNL	17.3	4.4	17.9	-1.974	0.047	be	-
Solyc11g020100	CNL	17.3	16.1	15.8	-0.108	-0.128	**be**	-
Solyc12g044200	CNL	17.2	5.1	17.3	-1.758	0.012	be	-
Solyc07g005770	CNL	15.5	2.3	15.2	-2.780	-0.028	b e	b
Solyc09g018220	CNL	14.7	16.8	6.4	0.200	-1.196	-	-
Solyc11g064770	CNL	14.3	5.9	16.6	-1.284	0.215	-	-
Solyc02g032650	TNL	14.3	3.1	14.1	-2.214	-0.023	bec	**b**
Solyc12g006040	CNL	13.7	11.5	5.7	-0.258	-1.261	b e	b
Solyc08g076000	CNL	12.5	2.2	12.2	-2.526	-0.033	**c**	**b**
Solyc11g069660	CNL	12.0	6.1	13.2	-0.977	0.132	-	-
Solyc09g098130	CNL	11.1	11.9	12.0	0.101	0.108	-	-
Solyc06g005410	TAS5	431.2	342.2	83.8	-0.334	-2.363	-	**b**

Table 4.3 **siRNA production at *pNLRs* is reduced in mimic lines.** Summary of *pNLRs*, with their gene id, class of NLR protein based on the phylogenetic analysis of Andolfo et al. [2014]. Total counts for 21-nt sRNAs (nRPM) in wild type (WT), *MIM482* and *MIM2118b* lines. Log2 fold changes between WT and MIMIC lines (intensity of colour indicates stronger reduction). Summary of target prediction, with letters indicating the predicted targeting miRNA. *TAS5* (bottom) is added for additional reference.

4.3 Discussion

The *in silico* analysis in the previous chapter suggested distinct functions for the different clades of the miR482/2118 family, and was also supported by degradome and sRNA sequencing data. The results of this chapter provide conclusive evidence of the distinct molecular functions of the miR482/2118 members.

Taking advantage of these target mimics, I have shown that miR482 members account for most of the production of *pNLR*-derived siRNAs. Sequestration of miR482 members resulted in reduction by half of the total amount of *pNLR*-derived siRNAs. While sequestration of miR2118b had an effect on a couple of lowly-producing *pNLRs*, its overall contribution to the *pNLR*-derived siRNA population was insignificant. Sequestration of miR2118b however, resulted in a major reduction of *TAS5*-derived siRNAs, while remaining unaffected in *MIM482* lines. This observation allowed me to conclude that miR2118b regulates exclusively and primarily *TAS5*.

At the time of conception of this project, no use of tandem mimics in tomato had been reported. Therefore, an initial concern was the applicability of target mimics in this species. Over the past couple of years, other studies have used short tandem target mimics to knock down the action of tomato miRNAs [Cao et al., 2016; Damodharan et al., 2016; Jia et al., 2015]. My initial observation of a strong reduction of target miRNAs and the data from these other studies have proven the suitability of target mimics for investigating miRNA function. However, all previous studies focused on the action of a single miRNA. With my approach, I also demonstrated the highly specific action of target mimics. This is the first time target mimics have been utilised to dissect the molecular function of miRNAs of the same family.

By studying miRNA over-expression lines, Fei et al. [2015] showed that levels of miRNA have a linear effect on the production of phasiRNAs. This was similar to what I have observed for *TAS5* in *MIM2118b*, but not the case in *MIM482* datasets, where I showed that a 10-fold reduction in my miRNA-triggers only halved the total population of NLR-derived siRNAs. One explanation for this discrepancy is based on the possibility that redundant mechanisms co-exist. The tomato genome contains several other 22-nt miRNA families that target different sequences conserved in defence genes (e.g. miR5300, miR5301, miR6022, miR6024, miR6026, and miR6027) [Karlova et al., 2013]. Some of those have already been identified as phasiRNA-producing miRNAs [Kravchik et al., 2013; Shivaprasad et al., 2012]. It is likely that the miR482-clade, while accounting for a great proportion of NLR-derived phasiRNAs, might not be the exclusive trigger on an individual NLR. Redundancy could be a key factor in avoiding high fitness costs of R gene mis-expression. Redundancy also reduces selective pressure over the miRNA sequences, resulting in faster evolutionary rates [De Felippes et al., 2008; de Vries et al., 2015]. This is a feature typical of other genetic components of the defence response [Kuang et al., 2004; Zhang et al., 2016a]. A higher evolutionary rate in a miRNA family would ultimately facilitate diversification within the family and could lead to new targeting preferences.

One remaining question is what would be the effect of manipulation of the miR482/NLR and miR2118b/TAS5 silencing pathways on the overall defence response. As indicated previously, Shivaprasad et al. [2012] proposed that the miR482/2118 family regulates NLRs in a manner that confers control over resistance defence response by reducing the expression of these NLRs until a pathogen is encountered. Upon infection, pathogen effectors would release the triggering the suppression of miRNA function and releasing the NLR action. In my target mimic lines, the action of these miRNA is suppressed and therefore an up-regulation of the basal level of resistance is expected. I hypothesize that these transgenic lines will present a higher basal level of resistance and therefore

be less susceptible to biotic stresses. The following chapter will address this hypothesis.

4.4 Acknowledgements

MIM482 related constructs for transient testing in *Nicotiana benthamiana* were originally designed by Dr. Adrian Valli and Mr. Will Summers (unpublished data). Dr. Adrian Valli provided valuable advice for the design of MIM2118b constructs and all the transient assays in *N. benthamiana*.

Chapter 5

Consequences of miR482/2118b depletion

5.1 Introduction

My previous findings indicate that miR482/2118 members are responsible for the production of a large number of secondary siRNAs. It has been suggested that miR482/2118-derived silencing could regulate, directly or through the action of their secondary siRNAs, a large proportion of NLR repertoire [Shivaprasad et al., 2012]. NLRs play critical roles in an organism's health in both plants and animals (reviewed in Jones et al. [2016]). When activated, NLRs often induce the 'hypersensitive response' (HR) which leads to localized cell death, changes to the cell wall, production of active oxygen species and strong activation of *pathogenesis-related* (*PR*) genes [Jones and Dangl, 2006].

The first stage in the negative regulation of NLRs through miR482/2118 involves an interaction between the mature miRNA and a sequence motif in the corresponding mRNA. The second stage – production of the phased siRNAs (phasiRNAs) – would also result in silencing of *NLR* mRNAs, either at genomic, transcriptional or translational level [Fei et al., 2015]. Most phasiRNA targets

would be close homologues of the primary miRNA target. The net effect of these two stages would be a silencing cascade affecting multiple *NLRs* [Halter and Navarro, 2015].

Argonaute (AGO) proteins are the binding partners of sRNAs and catalysts of silencing [Baulcombe, 2004]. Specific endogenous AGO proteins from plants and animals are essential for immunity because they bind infection-derived sRNAs and silence transcripts of both host and pathogen, including the genomes of RNA viruses [Carbonell and Carrington, 2015; Ding, 2010; Li et al., 2013b]. Consequently, pathogens have evolved diverse mechanisms to avoid silencing, most notably through the expression of proteins that act as suppressors of RNA silencing [Pumplin and Voinnet, 2013]. The silencing suppressor activity of these proteins blocks host plant silencing pathways. As a 'counter-defence' mechanism, failure in the silencing pathway caused by the silencing suppressors also impairs the action of the endogenous negative regulators of defence response, ultimately releasing the defence response (Figure 3.1). Therefore, the reduction of these negative regulators, as in my mimic lines, should potentially reduce the silencing of NLR and enhance the basal level of expression of resistance genes.

Spontaneous NLR activation can lead to autoimmune conditions that compromise reproductive fitness. This effect has been shown in transgenic lines overexpressing NLRs [Oldroyd and Staskawicz, 1998; Tao et al., 2000; Zhang et al., 2004], and in gain-of-function mutations like *suppressor of npr1-1, constitutive 1* (*snc1*) [Li et al., 2001] or *suppressor of salicylic acid insensisitve 4* (*ssi4*) [Shirano et al., 2002]. One suggestion is that in normal conditions the negative regulators, including miR482/2118 family members, mitigate the potential cost of NLR gene amplification and diversification.

The results of this chapter explore the transcriptional status of target genes in my target mimic lines. As reported originally in Shivaprasad et al. [2012],

the initial hypothesis is that the miR482/2118 pathway could be involved in negatively regulating a broad range of immune receptors against pests and diseases in tomato and other species. To test this hypothesis, I also explored the effects of infecting my target mimic lines with different classes of pathogens.

5.2 Results

5.2.1 Silencing of miRNA targets is suppressed in target mimic lines

I initially addressed the effect of the target mimic RNAs on the transcript accumulation of targets of miR482/2118 members and their derived phasiRNA. For that purpose, I investigated the transcriptional status of several NLR genes using quantitative RT-PCR in 4 week old leaf tissue of first-generation transgenic plants and wild type controls. I evaluated the levels of expression of confirmed direct targets of each conditions, and three other NLRs not targeted by miR482/2118b but potential targets of their derived phasiRNAs.

For direct targets, I selected the genes *LRR2* and *TAS5* for the *MIM482* and *MIM2118b* lines, respectively, since both have been confirmed in my data (Table 3.5) and previously published work [Li et al., 2012a; Shivaprasad et al., 2012]. For the *MIM2118b* lines, I also assessed the TNL transcript encoding for *Bs4*, as it was originally reported to be an indirect target of a TAS-derived siRNA [Li et al., 2012a]. Finally, I included in the analysis *NRC1* and *Prf*, two NLR-encoding genes that are not among the predicted targets of any miR482/2118 members.

NRC1, which stands for 'NLR Required for Cell death 1', belongs to a larger family of helper CNLs in Solanaceae that are proposed to act downstream of several canonical CNL receptors [Gabriëls et al., 2007; Sueldo et al., 2015; Wu et al., 2017, 2016]. However, the exact mechanism of action of NRC1 remains controversial. Recently, Wu et al. [2016] described how some of the previously suggested interactions of this protein with other NLRs were based on experimentally flawed conditions and that other members of this family were responsible for effects initially attributed to NRC1. In *N. benthamiana*, members

of the helper NRC family NRC2, NRC3 and NRC4 redundantly contribute to immunity mediated by various sensor NLRs [Wu et al., 2017]. A role of NRC1 in immunity can still be expected, since an auto-active version of NRC1 was capable of triggering an elicitor-indepedent HR when transiently expressed in *N. benthamiana* and *N. tabacum* [Gabriëls et al., 2007; Sueldo et al., 2015]. According to my analysis, *NRC1* is not directly targeted by any miR482/2118 members (Tables S.1-5). However, *NRC1* still presents sufficient sequence similarity to other miR482-targeted CNLs and the reversed *pseudoNB-ARC* region of *TAS5* (Figure 3.7), which would make it a potential target of secondary siRNAs derived from pNLRs and *TAS5*. All together, I considered the *NRC1* gene an ideal NLR candidate to preliminarily assess the transcriptional or post-transcriptional effect of my mimicry lines.

Prf is a CNL known for its role in bacterial immunity to *P. syringae* [Gutierrez et al., 2010]. Prf acts in a molecular complex with the Ser/Thr kinase Pto, and mediated by the Pto direct interaction with *P. syringae* effectors AvrPto and AvrPto, Prf confers resistance to the bacteria [Mucyn et al., 2006]. Several other Pto-like kinases also interact with Prf, suggesting that Prf could be involved in the recognition of other effector proteins and have a role in immunity beyond *P. syringae* detection [Gutierrez et al., 2010]. While still an NLR, it is not a predicted target of any miR482/2118 members (Tables S.1-5) and it is not highly similar to regions of *TAS5* (Figure 3.7) or the most prolific pNLRs targeted by miR482 [Andolfo et al., 2014].

The results of the qRT-PCR analysis revealed a significant increase in mRNA levels of *LRR2* in *MIM482*, and *TAS5* in *MIM2118b* lines, compared to wild type (Figure 5.1). The level of up-regulation of these direct targets directly correlated with levels of miRNA sequestration on each of the target mimic lines (Figure 4.4). *NRC1* and *Prf* were also slightly induced in target mimic lines. *Bs4*, the previously predicted target of a TAS5-derived phasiRNA, remained

unchanged (Figure 5.1). From these observations, I concluded that miR482- and miR2118b-mediated silencing is relieved in *MIM482* and *MIM2118b* lines, respectively. Additionally, *NRC1* and *Prf* increased slightly in the target mimic lines, suggesting that miR482 and miR2118b negatively regulate some non-target NLRs. This indirect regulation could occur trough the action of secondary siRNAs or as an indirect consequence of the up-regulation of other NLRs.

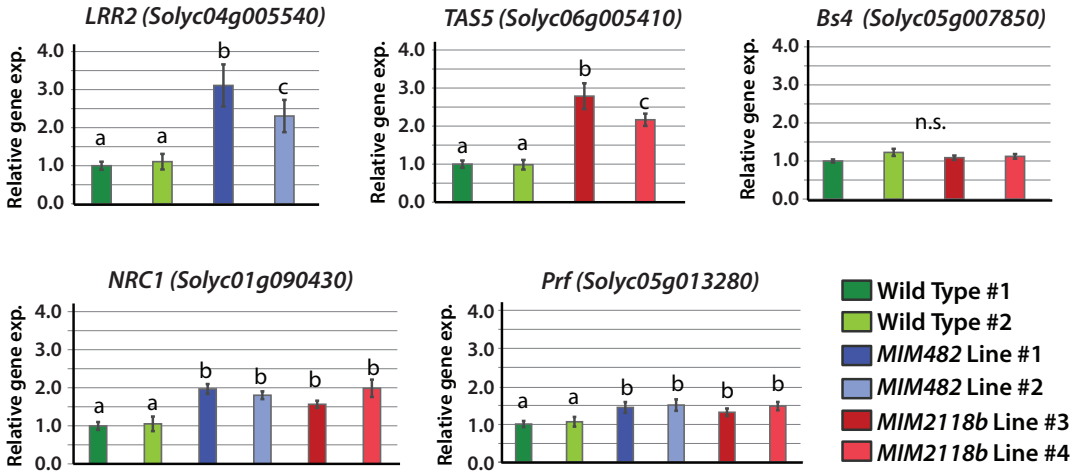

Fig. 5.1 **Direct targets and other defence genes are upregulated in target mimic lines.** Quantitative PCR analysis for the abundance of target mRNAs *LRR2*, *TAS5*, *Bs4*, *Prf* and *NRC1* in 4 week old leaf tissue. Expression values were adjusted to tomato housekeeping gene *EXP* and shown relatively to WT values. Error bars indicate SD (n=4). Statistically significant differences were explored using one-way ANOVA test followed by Tukey HSD at 95% confidence limits.

5.2.2 RNAseq failed to provide any valuable information due to high variability between biological replicates

Previous qRT-PCR results provided enough evidence to hypothesize a transcriptional activation of a broad range of NLRs in the target mimics lines. To investigate this hypothesis, I proceeded to analyse the transcriptional status of tomato NLRs genome-wide using high-throughput RNA sequencing techniques.

I performed mRNA sequencing library preparation using four biological replicates of second generation transgenic plants carrying the target mimic construct and four wild type plants as control. All biological samples were collected and processed at the same time. Libraries from the samples were multiplexed, pooled and sequenced in the same run. Reads were pre-processed and mapped to the latest genome and transcriptome of tomato (ITAG 3.2) [The Tomato Genome Consortium, 2012]. The quality of reads and their mapping was optimal (Figure S.2). However, differential expression analysis using the edgeR package did not yield many significant results. Out of all the tomato genes, only 13 in *MIM482* and 2 in *MIM2118b* showed significant changes in their expression between the control and the target mimic lines(Figure 5.2). Closer inspection of these results revealed that none of the up-regulated RNAs were predicted targets of miR482/2118 members or were directly related to the plant immune system (Figure S.6). These results contradicted my previous qRT-PCR expression data and urged for a thorough investigation of the RNAseq datasets.

Surprisingly, most genes with fold-changes greater than two were not identified by edgeR as having significant differential expression (Figure 5.2). This behaviour was similar when using other programs used for quantification and estimation of RNAseq datasets (tested DESeq2, baySeq and Sleuth; data not

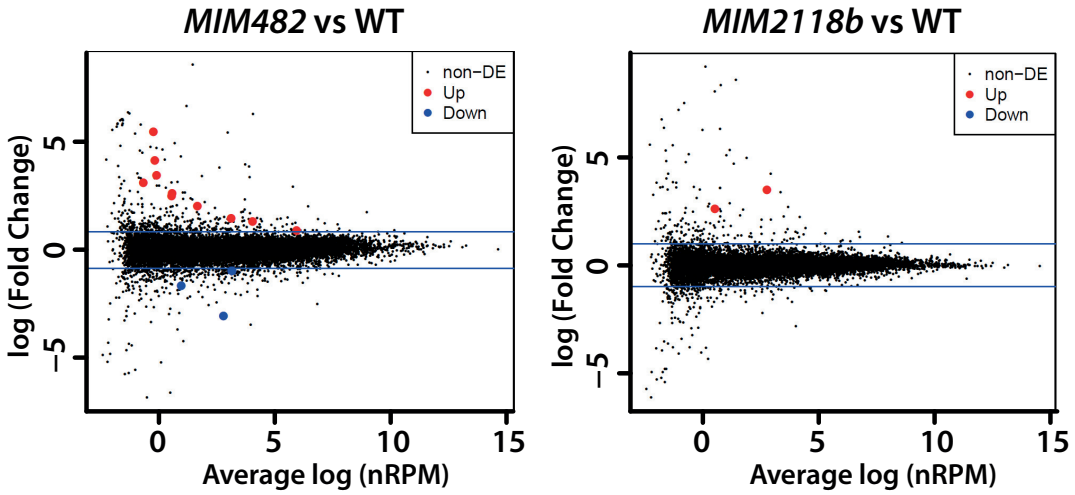

Fig. 5.2 **A small number of differentially expressed genes are detected in my RNAseq datasets**. MA plots of differential expression of genes using edgeR. Red and blue dots indicate up and down-regulated genes, respectively, in target mimic lines. Black represents non-differentially expressed genes. (FDR cut-off = 0.05).

shown). Examination of the variance between replicates showed unexpected high levels of variation across gene expression in each replicate set (Figure 5.3A). In fact, biological replicates failed to cluster together when investigating the heterogeneity of the dataset either by plotting the typical log2 fold changes between the samples or by studying their level of correlation (Figure 5.3B & S.5).

To further characterize sample variance, I calculated the biological coefficient of variation (BCV). BCV represents the coefficient of variation that would remain between biological replicates after removing the technical variation (if sequencing depth could be increased indefinitely). Experiments in which biological replicates share genetically identical backgrounds, such as this, should yield around 10% BCV [McCarthy et al., 2012]. An average BCV of 23.4% in my datasets indicated an unusually high variability between replicates, confirming previous observations. Therefore, I deemed the results of this RNAseq as inconclusive due to experimental noise in the datasets. The genome-wide expression

profile of genes should be re-addressed in further experiments.

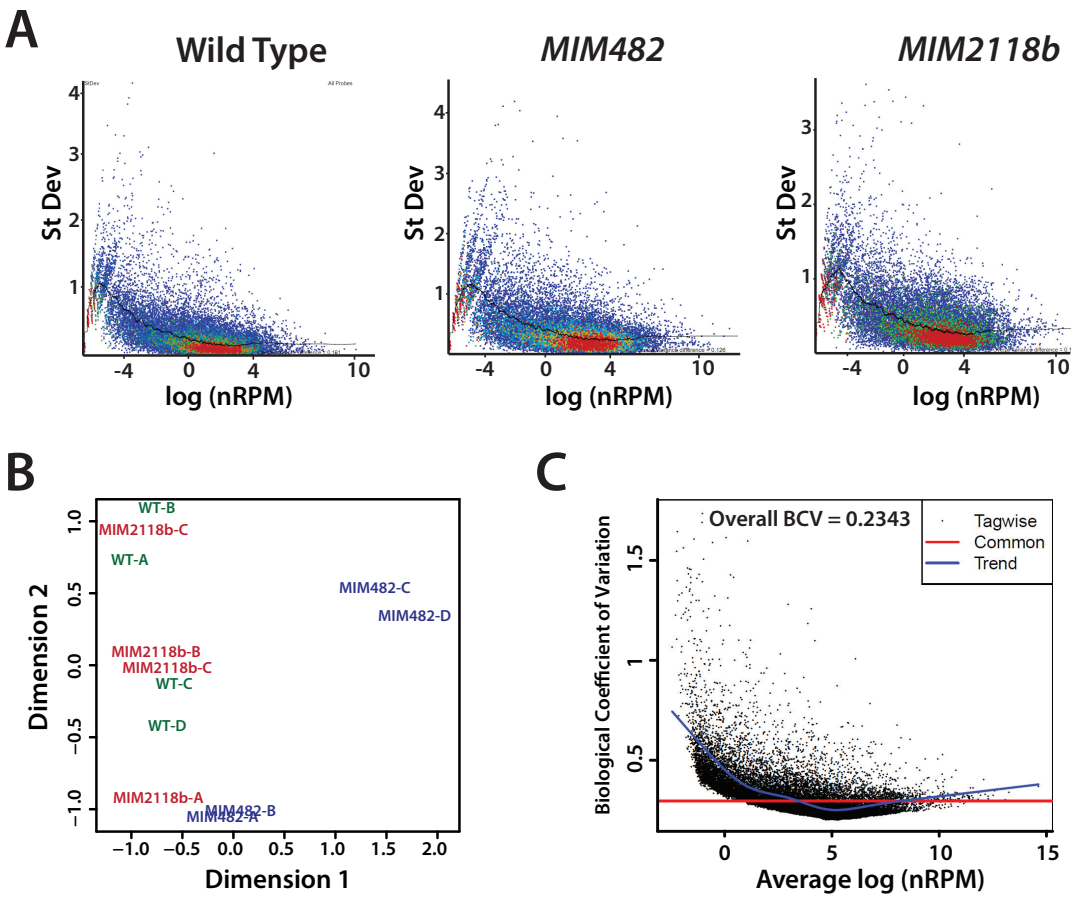

Fig. 5.3 **Variance between biological replicates in this dataset hinders any differential expression analysis**. (A) Scatterplot showing the standard deviation of gene expression between biological replicates against their mean expression. (B) Multidimensional scaling plot of between gene expression profiles where the distances on the plot approximate the typical log2 fold changes between the samples. Replicate libraries are colour coded. (C) Dot-plot of the gene-wise biological coefficient of variation (BCV) against gene abundance (in log2 counts per million). The BCV is the square root of the negative binomial dispersion. Lines display the common and trended BCV estimates by the edgeR package.

5.2.3 Sequestration of miR482 and miR2118b reduces susceptibility to oomycete *Phytophthora infestans*

To investigate the effect of the miR2118/482 cascade on disease resistance, I infected transgenic plants producing the mimic RNAs with the oomycete pathogen *Phytophthora infestans*. *P. infestans* causes the potato late blight, one of the most destructive diseases in crops worldwide [Kamoun, 2001]. Oomycetes secrete effectors to promote infection and colonization of plant tissue [Bozkurt et al., 2012]. Among these effectors, several act as suppressors of gene silencing [de Vries et al., 2017; Qiao et al., 2013; Xiong et al., 2014]. PSR2 is one such suppressor, and is known to interfere with phasiRNA production derived from the action of 22-nt miRNAs [Qiao et al., 2013]. The presence and function of PSR2 are conserved across many species of *Phytophthora* [Xiong et al., 2014]. *PSR2*-silenced *Phytophthora* strains uniformly exhibited significantly decreased virulence [Qiao et al., 2013]. This suggest a relevant interplay between the general phasiRNA machinery of the host and the infection of *Phytophthora*.

I inoculated second generation transgenic plants with zoospore droplets of the *P. infestans* isolate *Pi 88069*. The inoculations were performed on detached leaves of two independently transformed lines of *MIM482* (lines #1 and #2) and *MIM2118* (lines #3 and #4) (Figure 4.4) and a wild type control. The tomato cultivar M82*, from which all the plants in this study are derived, is highly susceptible to *P. infestans* infections [Akhtar et al., 2016]. Since *P. infestans* is a hemibiotroph, I measured the progress of necrotic lesions during the infection to investigate differences between my mimic lines and normal conditions. All mimic lines were less susceptible than the non transgenic control plants, visible through a significantly reduced lesion size at 3 days post infection (dpi) (Figure 5.4). Measurements of lesion area at 3 dpi showed strong statistically significant differences to wild type (Figure 5.5). Subtle differences between independent

*M82 (LA3475) is a processing variety originated in the USA.

lines correlated with levels of sequestration of the target miRNA (Figure 4.4), although these differences were not statistically significant (Figure 5.5).

Fig. 5.4 ***MIM482* and *MIM2118b* are less susceptible to *P. infestans* infection.** Visual phenotype of target mimic lines and wild type under blue light at 3 days post infection of detached leaves with *P. infestans Pi 88069*. Dark zones represent necrotic tissue.

To further assess the specificity of these phenotypes, I infected in the same fashion a transgenic line expressing a target mimic construct designed to sequester miR171. This miRNA affects the expression of GRAS family transcription factors, such as SCARECROW (SCR), that participate in some developmental processes [Hwang et al., 2011; Llave et al., 2002], but have not been reported to play a role in plant defence. Measurements of lesion area at 3 dpi showed no difference between *MIM171* lines and wild type controls (Figure 5.6), suggesting that the effect of my *MIM482* and *MIM2118b* lines was specific to the particular action of miR482 and miR2118b.

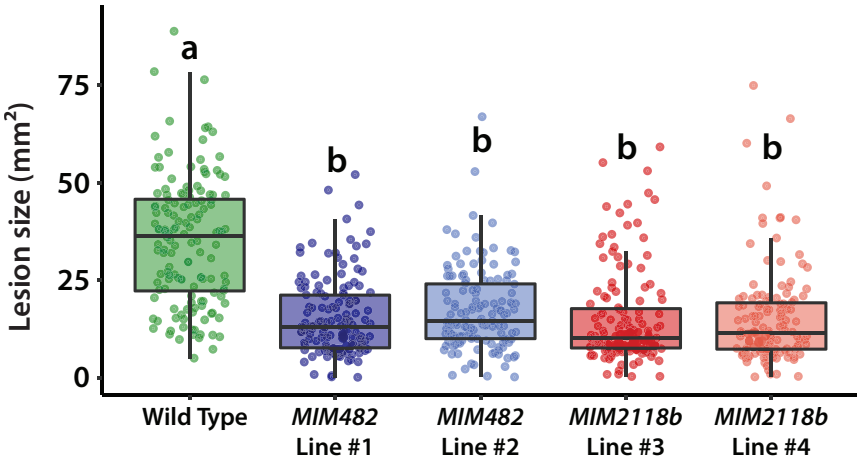

Fig. 5.5 **Sequestration of miR482 and miR2118b reduces susceptibility to *P. infestans*.** Measurements of lesion area 3 days post infection of detached leaves with *P. infestans Pi 88069*. Statistically significant differences were found using one-way ANOVA test followed by Tukey HSD at 95% confidence limits.

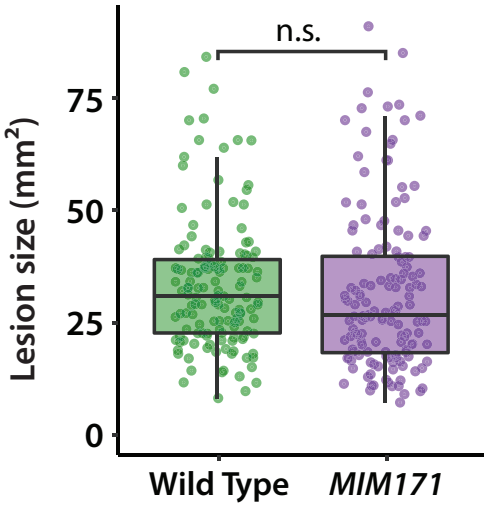

Fig. 5.6 **Non-related target mimics do not affect resistance to *P. infestans*.** Measurements of lesion area 3 days post infection of detached leaves with *P. infestans Pi 88069*. Differences failed a Welch Two Sample t-test.

5.2.4 Sequestration of miR482 and miR2118b reduces susceptibility to bacterial pathogen *Pseudomonas syringae*

Testing whether the miR482/2118 pathways could be involved in the broad spectrum regulation of defence against pests and diseases in tomato and other species required extending my initial observations to a different class of pathogen. Shivaprasad et al. [2012] demonstrated that tomato miR482 accumulation is reduced in response to infection bacterial pathogen *Pseudomonas syringae*. Shivaprasad et al. [2012] also showed that miR482-direct targets are up-regulated upon infection. These results already suggested that RNA silencing may play a role in regulating antibacterial immunity in tomato. In addition, during plant infection, *P. syringae* has been shown to produce at least three suppressors of silencing that inhibit multiple steps of silencing pathways [Navarro et al., 2008]. During infection of *A. thaliana rdr6* and *miR472*^{m} (a mutant of an *A. thaliana* miRNA gene encoding for a miRNA similar to miR482) mutants, growth of *P. syringae* was reduced [Boccara et al., 2014; Katiyar-Agarwal et al., 2006]. Therefore, I decided to test the susceptibility of my target mimic lines to *P. syringae*.

To further investigate the effect of the miR482/2118 members on antibacterial immunity, I infiltrated a suspension of *Pseudomonas syringae* pv. *tomato* DC3000 into attached leaves from second generation transgenic plants and a wild type control. The M82 cultivar does not contain the introgressed Pto allele from *Solanum pimpinellifolium* that confers resistance to *P. syringae* and therefore remains susceptible to infection.

I measured bacterial titres at 0 dpi, to control for even bacterial loading, and at 3 dpi. Bacterial titres of leaves were lower in *MIM482* and *MIM2118b* lines than in the control at 3dpi (Figure 5.7). Similarly to *P. infestans* experiments,

subtle differences between independent lines correlated with levels of sequestration of the target miRNA (Figure 4.4), although these differences were not statistically significant (Figure 5.5). These results showed that sequestration of miR482 and miR2118b leads to a reduced susceptibility to *P. syringae*, and suggest a role of these miRNAs in the regulation of anti-bacterial immunity in tomato.

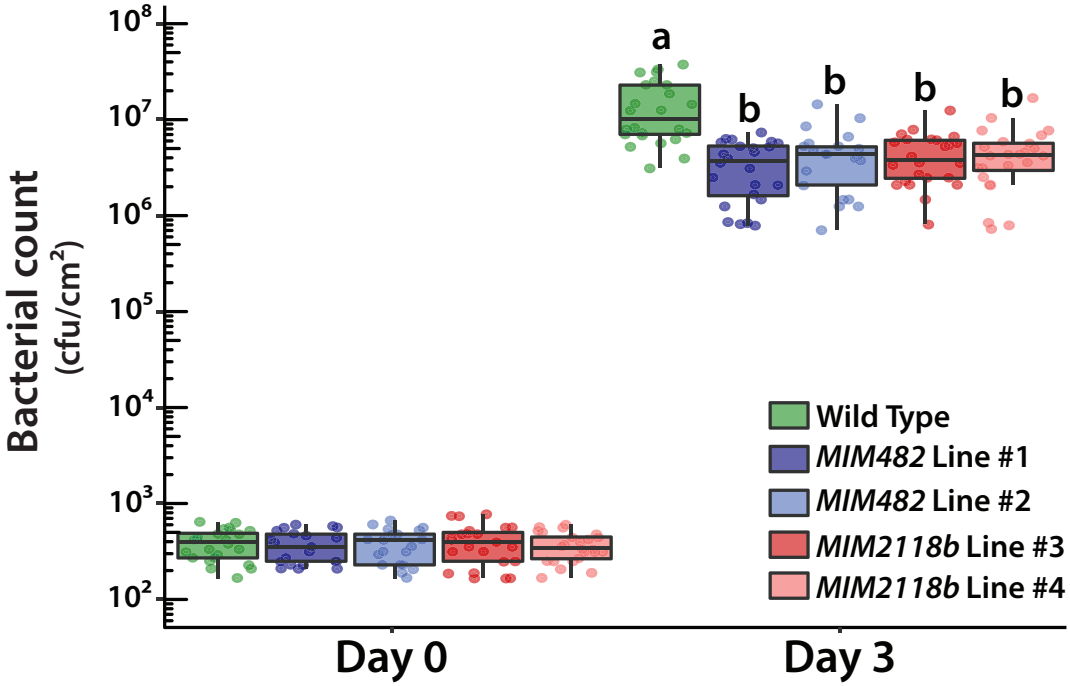

Fig. 5.7 **Sequestration of miR482 and miR2118b reduces susceptibility to *Pseudomonas syringae*.** Boxplot of bacterial population in WT and mimicry lines leaves infected with *Pseudomonas syringae* pv. *tomato* DC3000. Bacterial counts at 0 and 3 days post leaf infiltration. Statistically significant differences were determined using ANCOVA test followed by Tukey HSD at 95% confidence limits.

5.2.5 MIMIC lines do not present any obvious developmental defects

As mentioned previously, mis-regulation of NLRs can lead to deleterious effects on the plant. A plant would normally avoid autoimmunity effects by tightly controlling transcription, translation, and degradation of its defence proteins (reviewed in Huot et al. [2014]). One main remaining question is whether the expression of NLRs, while conferring higher resistance, is regulated under a cost-benefit balance. One hypothesis is that the boost in immunity provided by *MIM482* and *MIM2118b* could limit fitness by negatively affecting plant growth and reproduction. Counter-intuitively, T2 transgenic plants (once the effects of regeneration through tissue culture were washed off) presented no obvious developmental phenotypes in terms of morphology (Figure 5.8A), although closer inspection would be required. Preliminary data of general leaf morphology supports this observation (Figure 5.8B).

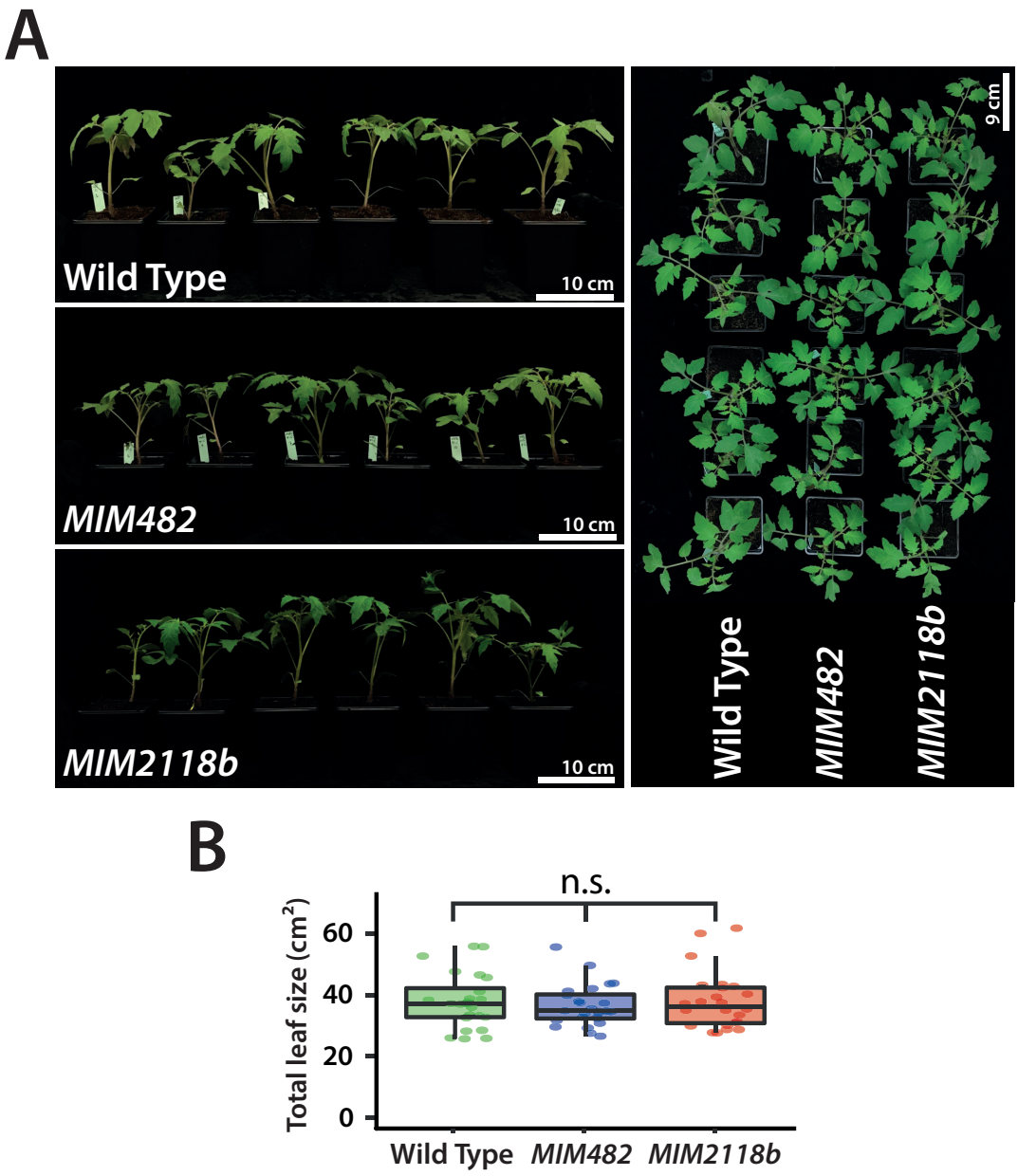

Fig. 5.8 **Transgenic plants do not present any obvious developmental defects.** (A) Lateral and aerial images of 4 weeks old plants. (B) Total leaf size remains unaffected in mimic lines. Boxplot representing total area of main leaflets from first and second true leaves. Differences failed a one-way ANOVA test.

5.3 Discussion

The experiments presented in this chapter explore the effects of my target mimic lines on transcriptional de-regulation of NLRs and its effects on the responses to biotic stresses. Sequestration of miR482 and miR2118b led to a major transcriptional upregulation of direct targets *LRR2* and *TAS5*, respectively (Figure 5.1). This up-regulation was reminiscent of the anti-viral immune response in tomato. During viral infection (possibly through the action of the virus SSs) levels of miR482 are reduced, leading to an increase in expression of *LRR2* [Shivaprasad et al., 2012].

NRC1 and *Prf*, two CNLs that are not directly targeted by miR482/2118 members, showed mild up-regulation in both target mimic lines (Figure 5.1). However TNL *Bs4* was not affected in *MIM2118b*, suggesting that TAS5-derived phasiRNAs do not participate in its transcriptional regulation. This observation contradicts a previous model based on *in silico* analyses [Li et al., 2012a], where a single *TAS5* phasiRNA was predicted to target *Bs4*. Overall, the observations suggest that not all NLRs are transcriptionally regulated by the miR482/2118, and different levels exist for the ones that are. It is possible that several factors influence the level of this relationship, such as quantity of phasiRNAs capable of recognising the transcript, accessibility of the target sites, or co-localization of the transcripts and the components of the silencing machinery. Unfortunately, a genome-wide transcriptional analysis was unsuccessful and could not be repeated due to the time restrictions of this work. It should therefore be re-addressed in the future.

MIM482 and *MIM2118b* lines were significantly less susceptible to the filamentous pathogen *P. infestans* and bacterial pathogen *P. syringae* (Figures 5.5 & 5.7). The effect was specific to the action of the target mimics on the members of the miR482/2118 family (Figure 5.6). These results support the

hypothesis that the actions of the miR482 and miR2118b are involved in the regulation of broad spectrum defence, although the mechanism of action is still unclear. The simplest interpretation is that one or more NLRs conferring weak recognition of *P. infestans* or *P. syringae* are miR482/2118 dependent and therefore up-regulated in the MIMIC lines. Another scenario is that the up-regulation of NLR pathways antagonizes the normal progress of the pathogen life cycle.

In the case of *Phytophthora*, secondary siRNAs are induced during the early infection stages in soybean [Wong et al., 2014]. Late during *P. infestans* infection however, *PSR2* is upregulated in the transition from biotrophic to necrotic phase [de Vries et al., 2017; Qiao et al., 2013], which would result in a down-regulation of phasiRNA production, and up-regulation of NLRs and other defence genes. Expression of *PSR2* would therefore release the silencing on *NLRs*, which would be intuitively disadvantageous. However, this is inconsistent with the observed virulence benefit demonstrated for this suppressor and since silencing of *PsPSR2* leads to largely reduced virulence of *P. sojae* [de Vries et al., 2017; Qiao et al., 2013]. Still, in addition to targeting defence genes, targets of phasiRNAs are essential for developmental and metabolic processes, (such as nutrient and auxin signalling during flower and leaf development) [Allen et al., 2005; Hsieh et al., 2009; Kravchik et al., 2013; Lin et al., 2013; Zhai et al., 2015; Zheng et al., 2015]. In conclusion, the advantages of tampering with all the phasiRNAs simultaneously by disrupting their biogenesis probably outweigh the cost of releasing defence genes from negative regulation. The most likely scenario that my data indicates is that increased basal level of expression of resistance genes in the target mimic lines has a quantitative effect and only slows down the infection.

Another potential explanation is that *Phytophthora* uses the silencing suppressors to hijack the HR mechanism and trigger the death of the host cell upon entering necrotrophic growth of the pathogen. Such mechanism could

act complementary to the action of other pathogen effectors that alter the host for pathogen invasion. Such is the case of PexRD54, another *P. infestans* RXLR-type effector, which has evolved to bind host autophagy proteins to stimulate autophagosome formation [Dagdas et al., 2016]. Via the conjoint action of all these effectors, *Phytophthora* would ultimately control the timing of the defence response to serve its own developmental cycle. Therefore, a 'default' release of NLR silencing in my target mimic lines prior to the transition to the necrotic phase could compromise the virulence of *P. infestans*.

In the case of bacterial pathogen *P. syringae*, although there is enough evidence to support the involvement of RNA silencing in antibacterial immunity, the role of phasiRNAs remains elusive. Silencing mutants in *A. thaliana* are susceptible to infection with multiple RNA viruses and fungi [Deleris et al., 2006; Ellendorff et al., 2008; Garcia-Ruiz et al., 2010; Qu et al., 2008]. In *P. syringae* infections however, a variety of phenotypes arise from these mutants. In the case of *rdr6*, resistance is increased [Boccara et al., 2014]; mutants of *drb4* (the protein that drives dsRNA recognition for phasiRNA production) are highly susceptible [Zhu et al., 2013]; mutants of *dcl4* showed unchanged resistance compared to wild type [Katiyar-Agarwal et al., 2006; Navarro et al., 2008]. Such apparently contradictory data requires further investigation. What it is clear is that miR482 and miR2118b, much like their *A. thaliana* counterpart miR472, are part of a key regulatory control of disease resistance against bacteria (Figure 5.7 & Boccara et al. [2014]).

Sequestration of these miRNAs did not cause any apparent developmental defects, although further analysis is required (Figure 5.8). This was initially unexpected, as defence activation is generally thought to be at the expense of the plant's fitness. However, observations for fitness cost are mostly based on self-activated mutants, ectopic overexpression assays, and the comparison between resistant and susceptible alleles of defence genes [Heidel et al., 2004;

Kempel et al., 2011; Tian et al., 2003]. While there is abundant evidence supporting the fitness–defence trade-off concept (reviewed in Huot et al. [2014]), such findings do not necessarily clash with my observations. The target mimics remove a repressive factor of NLR expression, but it is possible that full activation of NLRs still requires the presence of pathogen-derived molecules or other catalysts. The action of the miR482/2118 members could be part of a multilayer 'fail-safe' mechanism.

The results also represent the first attempt to use short tandem mimics for improving plant defence. Previous use of short tandem mimics for manipulating miRNAs have lead to important improvements in different agronomic traits [Jia et al., 2015; Zhang et al., 2017a]. These results provide additional applications for these target mimics, valuable not only in studying the function of a miRNA family, but also as a potentially useful tool in agriculture and pest-control strategies.

5.4 Acknowledments

Dr. Sebastian Schornack assisted with *Phytophthora infestans* experiments. Dr. Zhengming Wang provided the *MIM171* tomato line used as a control.

Chapter 6

Small RNAs and ARGONAUTE 5 during germline formation

6.1 Introduction

Small RNAs (sRNAs) are the specificity components of silencing machinery and therefore determine the targets, but it is through the action of their associated proteins, that the silencing ultimately occurs. As mentioned in the introduction, ARGONAUTE (AGO) proteins are one of such kind.

Some AGOs are ubiquitously expressed throughout the plant, such as AGO1 or AGO4, while others are restricted to precise cell-types or conditions [Zhang et al., 2015]. Similarly to animals, several AGO proteins have been reported to be expressed during sexual development [Nonomura et al., 2007; Olmedo-Monfil et al., 2010; Singh et al., 2011; Tucker et al., 2012; Zhai et al., 2015]. Out of all of these cell-type specific AGOs, the AGO5 sub-clade has been repeatedly reported to participate in germline formation and meiosis in both monocots and dicots [Nonomura et al., 2007; Tucker et al., 2012; Zhai et al., 2015].

The specific role of AGO5 is still unclear and subject to speculation. In rice, AGO5-class MEIOSIS ARRESTED AT LEPTOTENE1 (MEL1, also named OsAGO5c) has been shown to regulate gender-independent cell division of pre-meiotic germ cells [Nonomura et al., 2007]. In *mel1* mutants, defects in pre-meiotic mitosis and meiotic homolog synapsis ultimately lead to plant sterility [Komiya et al., 2014; Nonomura et al., 2007]. MEL1 preferentially binds siRNAs with a 5' cytosine and derived from phasiRNA-producing loci (PHAS) [Komiya et al., 2014]. These PHAS loci are triggered by miR2118 and are conserved in several monocots [Johnson et al., 2009; Kakrana et al., 2017; Zhai et al., 2015]. Chromatin condensation is also affected in *mel1* mutants, which has led to the suggestion that MEL1 might be involved in epigenetic regulation during meiotic events. *A. thaliana* AGO5 shows the same preference for 5' cytosine siRNAs that are derived from intergenic sequences [Mi et al., 2008], although pre-meiotic PHAS loci have not been reported and its genome does not contain a copy of miR2118. Interestingly, a dominant gain-of-function allele (*ago5-4*) is defective in the initiation of megagametogenesis [Tucker et al., 2012].

Like *A. thaliana*, tomato has been suggested to only contain a single copy of AGO5 (Figure 6.1) [Bai et al., 2012]. Differently to *A. thaliana* but similar to grasses, tomato also contains a copy of miR2118 that is predominantly expressed in flower development (miR2118a; Figure 3.9). An outstanding question remains if the role of AGO5 is conserved between monocots and dicots. While in monocots, its involvement in epigenetic regulation is clear, *A. thaliana* knock-out *ago5* alleles do not present any meiotic defects [Oliver et al., 2014]. However, epigenetic mutants in *A. thaliana* generally do not present developmental abnormalities, possibly due to the low repetitive content of its genome [Lisch, 2009]. In tomato, mutants of those same epigenetic genes usually present major developmental defects [Gouil, 2016]. Thus, in this chapter I set to investigate the role of AGO5 in tomato, and the effect of knock-out mutations in germline formation.

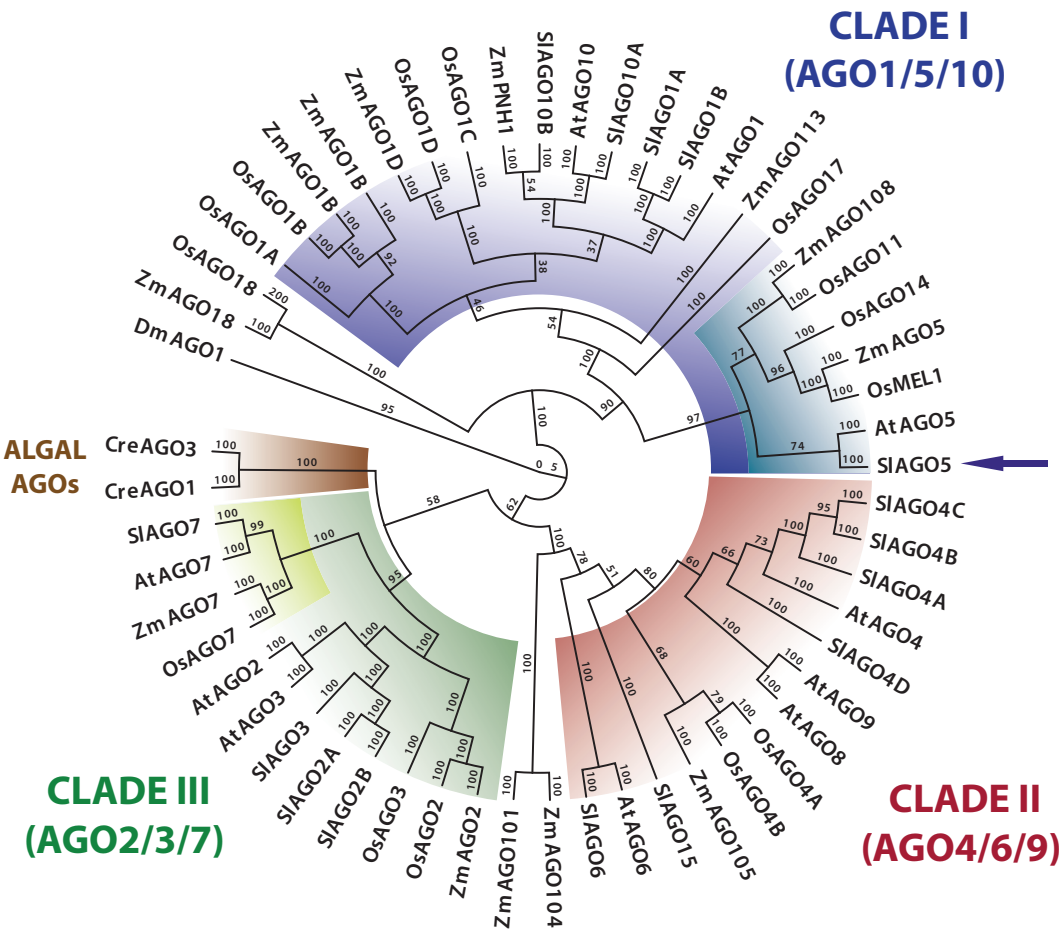

Fig. 6.1 **Phylogenetic tree of AGO proteins in plants.** Colours represent particular clades and sub-clades. Neighbor joining-based phylogenetic analysis was done on full length proteins using Mega 5 (bootstrap = 1000) [Tamura et al., 2011]. Arrow indicates tomato AGO5. DmAGO1 was chosen as an outgroup to determine the root. Abreviations indicate species as follows: At: *Arabidopsis thaliana*; Sl: *Solanum lycopersicum*; Zm: *Zea mays*; Os: *Oryza sativa*; Cre: *Chlamydomonas reinhardtii*; Dm: *Drosophila melanogaster*.

6.2 Results

6.2.1 Expression profile of tomato *ARGONAUTE 5*

In order to investigate the possible roles of tomato AGO5-like members in development, I inspected the genome using TBLASTN (protein query against a translated nucleotide database) in order to identify orthologues of AtAGO5. Using the protein sequence of AtAGO5 as an input against the tomato genome (SL3.0; The Tomato Genome Consortium [2012]) yielded only a single candidate: *Solyc06g074730*, in agreement with a previous study [Bai et al., 2012].

Taking advantage of publicly available RNA expression data of tomato tissue and cell-types, I investigated the expression profile of *Solyc06g074730* (hereafter *AGO5*) throughout development. In a dataset that included all major organs, *AGO5* transcripts dramatically accumulated in developing inflorescences exclusively (Figure 6.2A)*. In order to get a better understanding of the distinct expression of *AGO5*, I explored the Tomato Atlas of Expression, an RNAseq dataset of different stages tomato fruit development coupled with laser capture microdissection (LCM) of different tissue types [Fernandez-Pozo et al., 2017]. In agreement with the previous dataset, *AGO5* transcripts were primarily detected in the influorescence at 0 days post anthesis. within that developmental stage, expression of the gene occurred primarily at the ovules, and also at the surrounding placenta (Figure 6.2B)†. This expression profile broadly coincides with orthologues of AGO5 in other plant species [Nonomura et al., 2007; Tucker et al., 2012]. Regrettably, no data was available in the Tomato Expression Atlas for earlier stages of inflorescence development.

*Tomato eFP browser (URL: http://bar.utoronto.ca/efp_tomato/cgi-bin/efpWeb.cgi). Datasets used were presented in Matas et al. [2011].

†Tomato Expression Atlas (URL: http://tea.solgenomics.net/). Datasets used was originally presented in Richard et al. [2015]

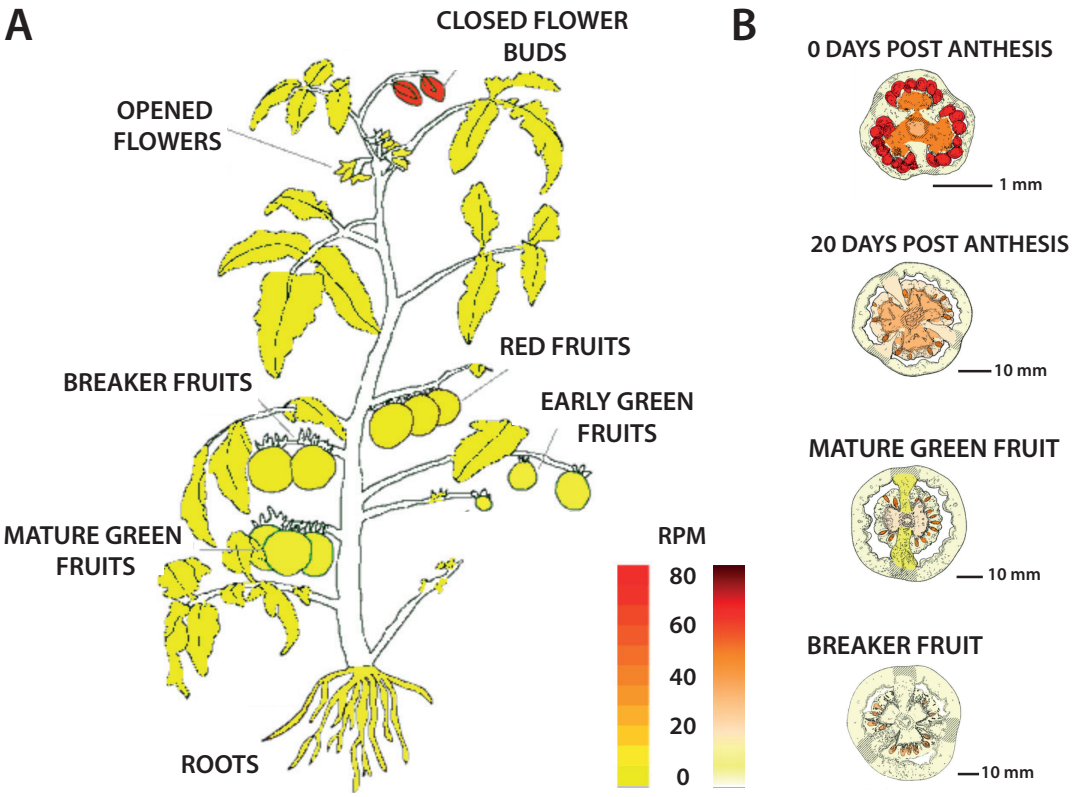

Fig. 6.2 ***AGO5* (*Solyc06g074730*) is most expressed in tomato flowers.** Tomato expression profile of *AGO5* using publicly available data of (A) general organs in and (B) cell types during fruit development. Colour scales were adjusted to show the same range of expression in reads per million (RPM). Data was obtained from the tomato eFP browser [Winter et al., 2007] and the tomato expression atlas [Fernandez-Pozo et al., 2017].

AGO5 proteins in other plants have been reported to accumulate primarily in early flower development stages, corresponding with the occurrence of gamete formation and meiosis [Nonomura et al., 2007; Tucker et al., 2012; Zhai et al., 2015]. To further dissect the expression of *AGO5* in tomato, I performed quantitative RT-PCR on different stages of flower development, and included young leaves as a control. Expression of *AGO5* peaked at inflorescence stages, experiencing an overall 10-fold increase at the later stage which coincides with meiotic events and gamete differentiations (Figure 6.3). This observation would again suggest a conserved role between AGO5 orthologues across plant species based on its common expression profile.

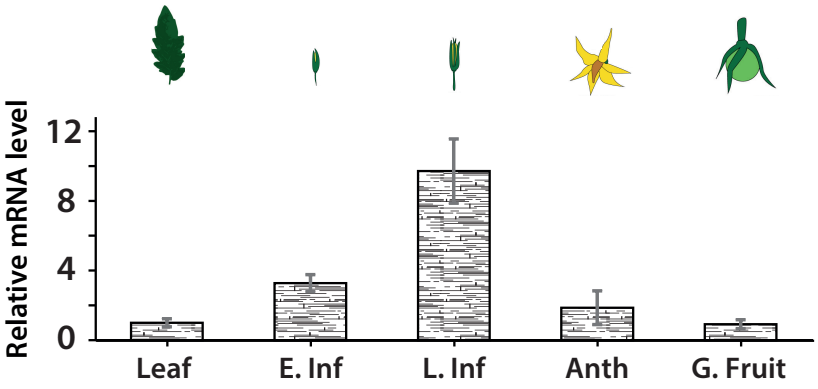

Fig. 6.3 ***AGO5* expression peaks during inflorescence development**. Quantitative analysis for the abundance of *AGO5* mRNA in various tissues of tomato development. Expression values were adjusted to tomato housekeeping gene *EXP* and shown relatively to leaf values. Error bars indicate SD (n=3). E. Inf: early inflorescence; L. Inf: Late inflorescence; Anth: Anthesis; G. Fruit: Green fruit.

6.2.2 Generation *ago5* mutant alleles using genome editing

As mention in previous chapters, a mapped collection of mutants does not exist for tomato. Only few mapped mutants exist, and the only member of the Argonaute family for which a mapped mutant exists is AGO7, involved in leaf development through the regulation of miR390/TAS3 [Yifhar et al., 2012]. Therefore, in order to study the role of AGO5 through a reverse genetic approach, I decided to generate mutant alleles taking advantage of the novel technologies of directed mutagenesis.

Several strategies for genome editing currently exist, ranging from the use of zinc-finger nucleases (ZFNs), transcription activator-like effector nucleases (TALENs), to the Clustered regularly interspaced short palindromic repeats-(CRISPR-) CRISPR associated protein 9 (Cas9) system. ZFNs were directly discarded as they remain difficult to engineer and prone to failure [Ramirez et al., 2008]. TALENs were originally derived from the TAL effector proteins of tomato pathogen *Xanthomonas*, and reduced transformation efficiencies have raised concerns about potential TALEN cytotoxicity [Christian et al., 2013]. CRISPR-Cas9, per contra, has been repeatedly used in plants with virtually none of the presented drawbacks [Feng et al., 2013; Nekrasov et al., 2013; Shan et al., 2013]. Recently, it has even been applied in tomato to mutagenize promoters and create a continuum of expression profiles [Rodriguez-Leal et al., 2017].

To create different mutant alleles of *AGO5*, I designed pairs of small guide RNAs (sgRNAs) spaced by approximately 100 bp. The sgRNAs were designed to recognise a DNA sequence in the genome containing a 3' protospacer adjacent motif (PAM), which could ultimately be targeted by the Cas9 (Figure 6.4A). Cleavage occurs around 3 bp upstream of the PAM domain, where those DNA breaks are later repaired with deletions of nucleotides. Depending on the

success of targeting, the mutagenic event would range from small deletions at the sgRNA targeting site, to larger deletions between the two sgRNA targeting sites. I employed two different pairs of sgRNA to ensure the success of having mutagenic events, and designed then in the 5' region of the gene to ensure that complete knock-out alleles were created.

To secure a sufficient amount of transformation and mutagenic events, approximately a hundred explants per construct were transformed via tissue culture with Agrobaterium tumefaciens carrying the Cas9 construct and one of the sgRNA pairs (Figure 6.4B).

Four independent transformants per sgRNA pair were PCR screened for mutated alleles, and a wide of combinations of different sized bands indicated the presence multiple new alleles in various degrees of combinations (Figure 6.4C). PCR products were cloned into pGEM-T vectors and sequenced. A total of 6 mutated alleles were isolated, ranging from small 1-3 bp, to large 68-80 bp deletions (Figure 6.4D). Generally, most plants appeared to have at least one wild-type allele. Therefore, plants were self-pollinated and brought to the next generation to study the possible effects of mutated alleles in homozygous backgrounds with the Cas9 cassette segregated out. Although some regenerated lines presented low levels of seed setting (data not shown), all first-generation plants were capable of setting fruit.

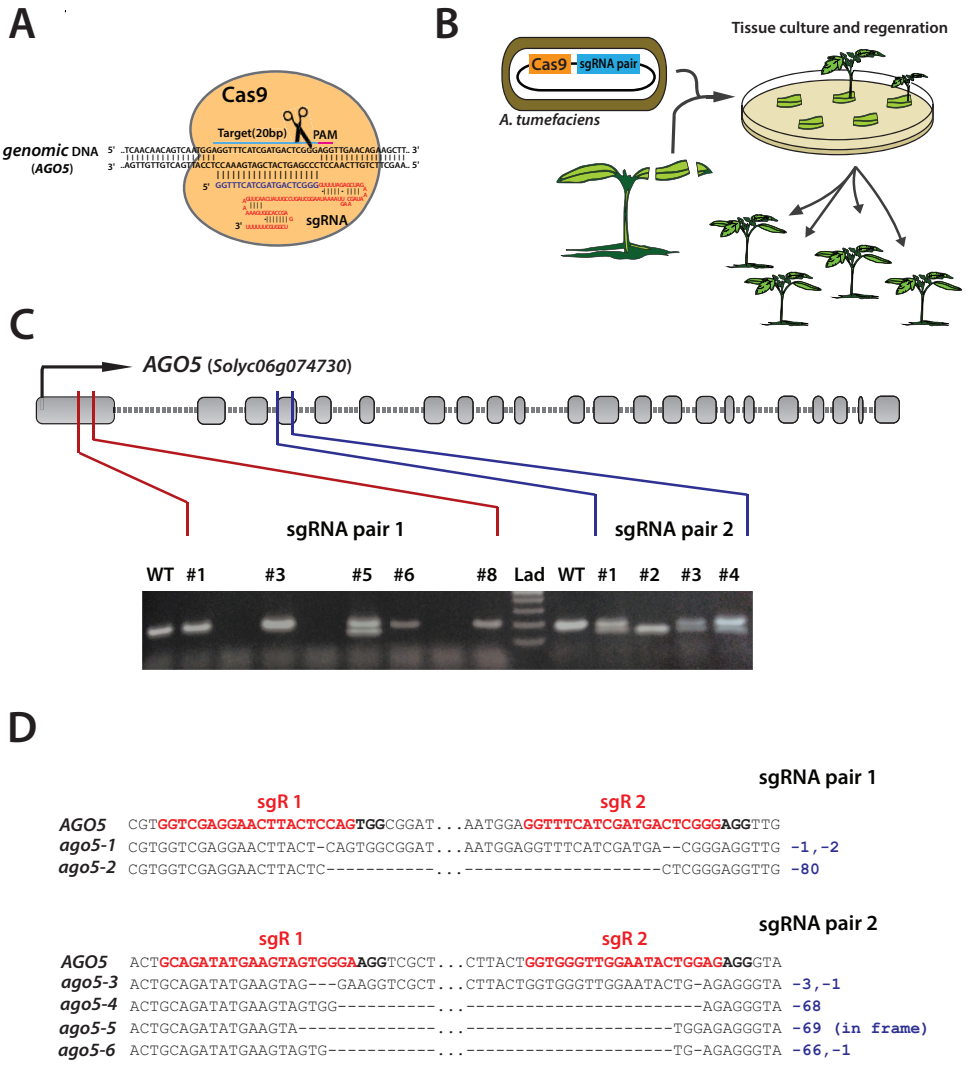

Fig. 6.4 **Directed mutagenesis of *AGO5* using gene editing**. (A) Schematic representation of Cas9-sgRNA interactions with genomic DNA sequences. (B) Generation of tomato transformants using Agrobacterium-mediated transformation and tissue culture. (C) Diagram of the genomic structure of the AGO5 gene. Two sgRNA pairs were designed to target exons 1 and 4, respectively. PCR screening validated the presence of large deletion in some mutagenized alleles. (D) Sequences of the different mutated alleles generated by the Cas9-sgRNA machinery. Complementary regions to the sgRNAs are indicated in red, and ultimate result of the deletions are indicated in blue (unless indicated, mutations result in knock-out alleles). Figure sections A and B are inspired on Gouil [2016].

6.2.3 Mutations of *ago5* decrease pollen fertility in tomato

Several descendants of first-generation transformants were screened for homozygous mutated alleles, and only those that carried a single homozygous mutant allele were further characterized. In order to preliminary asses general defects in gamete formation, pollen viability was calculated using an Alexander staining assay [Peterson et al., 2010].

Interestingly, lines carrying alleles that resulted in early stop codons generated mostly full fertile pollen grains containing normal cytoplasm like the wild type, which stained magenta (Figure 6.5). Only plants containing homozygous *ago5-5* alleles produced mostly dead pollen grains that were nearly empty, and deformed, and consisting mainly of exine walls, which stained turquoise (Figure 6.5).

The mutant *ago5-5* allele consists of a 69 bp deletion in the fourth exon of *AGO5*. This deletion results in the elimination of a 23 amino-acid sequence motif (...VGRSLFHHTFAGDAGLLTGGLEY...) that is part of the Argonaute linker 1 (L1) domain; also referred to as Domain of Unknown Function 1785 (DUF1785) (Figure S.6). The L1 is a domain found in all argonaute proteins, from prokaryotes to humans. It is a linker region between the N-terminal domain of the protein and the PAZ domain. Crystal structures of bacterial, yeast and human argonautes have shown that it consists of an alpha-helix packed against beta-sheets spanning one face of the adjacent N-terminal and PAZ domains [Elkayam et al., 2012; Nakanishi et al., 2012; Wang et al., 2008]. Together these three regions comprise the PAZ-containing lobe of argonaute, which consists of a compact global fold that ultimately accommodates the small RNA.

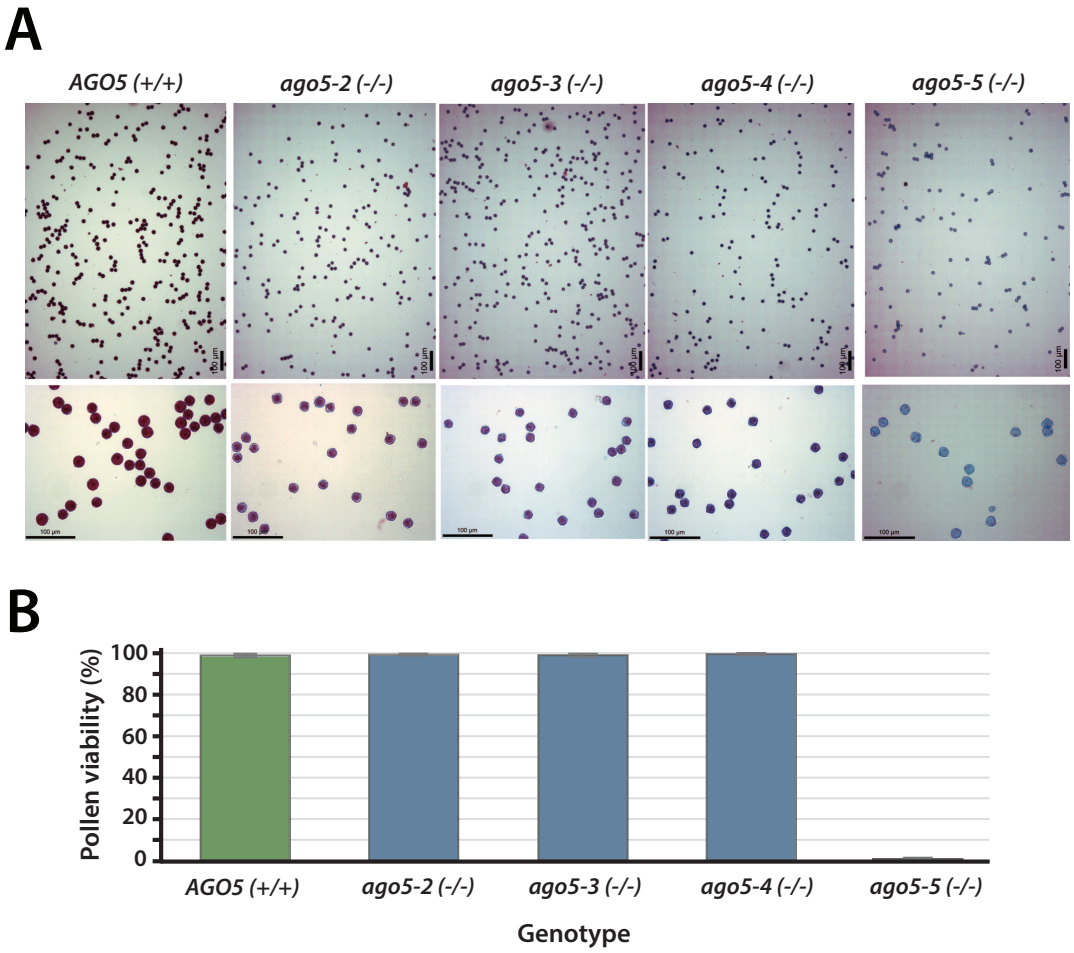

Fig. 6.5 **Pollen viability is only affected in *ago5-5*/- mutants.** (A) Differential staining of aborted and non-aborted pollen grains based on alexander stain assay. Aborted pollen grains stained turquoise. Non-aborted pollen grains stained magenta. (B) Bar-plot of the percentage of magenta (non-aborted) pollen grains. Error bars represent S.D. (n=4-8).

6.3 Discussion

Expression of *Solyc06g074730*, the single member of the *AGO5* clade in tomato, was detected during flower development, coinciding with gamete formation (Figure 6.2&6.3). This expression pattern was reminiscent of other members of this clade in *A. thaliana*, rice, and maize [Nonomura et al., 2007; Tucker et al., 2012; Zhai et al., 2015]. Several *ago5* mutant alleles were generated with CRISPR-Cas9-directed genome editing (Figure 6.4D). Homozygous mutant lines were examined to investigate the potential involvement of AGO5 in gametophyte formation. Alleles that produced early stop codons showed no abnormal pollen fertility (Figure 6.5). Surprisingly, plants carrying the homozygous *ago5-5* allele produced mostly aborted pollen grains. The *ago5-5* allele consisted of a 69 bp deletion that removed 23 a.a. of the L1 domain of the AGO5 protein.

The results presented here are evocative of the findings in A. thaliana *AGO5*. Tucker et al. [2012] showed that while knock-out (amorphic) mutations of *AtAGO5* showed no defects in ovule formation or gametophyte development, a particular dominant allele (*Atago5-4*) presented a substantial (approximately 50 %) decrease in fertility. The transposon insertion in the *Atago5-4* allele gave rise to a truncated AGO5 protein, retaining the N-terminal and PAZ domains but lacking the MID and PIWI domains. Interestingly, the phenotype could be replicated by expression viral silencing suppressors Hc-Pro and P1 under the *AtAGO5* promoter. Tucker et al. [2012] also showed that the promoter of *AtAGO5* restricts its expression to somatic cells surrounding the reproductive cells, suggesting that the required sRNAs pathways for gametogenesis affected in the mutant are actually somatic. Whether both the *A. thaliana* and the tomato *AGO5* function in the same fashion requires further investigation.

A hypothesis for the basis of the mis-function of the tomato *ago5-5* allele arises from structural studies of different AGOs in other eukaryotes and bacteria.

In yeast, conformation changes by the movement of L1 are important for AGO function, which otherwise blocks access to the catalytic pocket [Elkayam et al., 2012]. An amino-acid change (R172A) in L1 domain of bacterial *Thermus thermophilus* AGO impacted the ability of the protein to produce effective cleavage without causing major structural rearrangements, suggesting an important role of the domain in correct AGO functionality [Wang et al., 2008]. Whether this also applies to plant Argonautes is still cause of speculation as no crystal structures of plant AGO proteins have been resolved. However, I hypothesize that the deletion at the L1 domain of AGO5 has likely cause major structural rearrangement, impairing its normal function.

An outstanding question from both the *A. thaliana* observations and my results is what sRNAs pathways are ultimately affected in these gain-of-function mutant alleles. Different plant argonautes have been shown to have distinct functions but can still be partially redundant ([Havecker et al., 2010; Wu et al., 2009]. One possibility is that knock-out mutations of AGO5 do not give phenotypes because its function can be complemented by other AGOs such as AGO1. But when mutations give rise to antimorphs, the normally AGO5-bound sRNAs are now bound to a non-catalytic protein and therefore prevented from any activity or any AGO1 'takeover'. A different explanation is that the defects arise from a neomorphic mutation, where the new proteins are disturbing other sRNA pathways unrelated to normal AGO5 function, and vital for gamete formation.

This analysis, along with previous observations [Tucker et al., 2012], has provided preliminary evidence for the existence of an vital sRNA pathway for gamete formation at cells that express AGO5 in dicots. However, the actual involvement of AGO5 in this pathway requires further investigation. Complementation assays, immuno-precipitation of the proteins and sequencing of their bound sRNAs will be key in further exploratory analysis. My results also remark the capabilities of genome editing using CRISPR-Cas9 in reverse genetic

approaches, as it allowed to create multiple diverse mutations that ultimately enabled these findings.

6.4 Acknowledgments

Dr. Sara Lopez-Gomollon performed the alexander staining assays in pollen.

Chapter 7

Transposable element activation in the shoot apical meristem

7.1 Introduction

Transposable elements (TEs) are major components of eukaryotic genomes. TEs are DNA sequences capable of movement within the genome and are generally considered selfish, harmful genetic elements. The presence and movement of TEs can result in insertional mutagenesis, unequal homologous recombination and genome rearrangements and duplications [Feschotte and Pritham, 2007; Lisch, 2013]. For these reasons, eukaryotes have developed a variety of strategies to limit the activity of TEs. A current consensus suggests that these strategies ensure a balance between maintenance of genome stability and tolerating genome instability as a driving force of evolution [Lisch, 2013].

7.1.1 Transposable elements and their influence on gene expression

In the plant model *Arabidopsis thaliana*, TEs remain transcriptionally inactive in almost every organ [Slotkin et al., 2009]. The primary mechanism of inacti-

vation is by cytosine methylation but the process is not simple and different mechanisms are involved depending on the type of TE, the chromosomal context and other, uncharacterized factors.

Genetic analysis in *A. thaliana* provided some clues about the underlying complexity behind the silencing of TEs. This approach implicated the chromatin remodeler DECREASE IN DNA METHYLATION 1 (DDM1) that interacts with histone H1 and is required for TE methylation [Zemach et al., 2013] in conjunction with the DNA-methyltransferase MET1. In *met1* and *ddm1* mutants there is a loss of symmetric methylation TEs are transcriptionally activated. In this artificially active state a second silencing mechanism comes into play in which the TE transcripts are targeted by miRNAs and siRNAs and subsequently trigger PTGS and secondary siRNA production leading to a third layer of TE silencing [Creasey et al., 2014; Martinez et al., 2017]. In this third layer the secondary siRNAs cause the TEs to become hyper-methylated in the CHH context through the non-canonical RNA-directed DNA methylation (RdDM) pathway that is dependent on RNA-Dependent RNA polymerase 6 (RDR6) and RNA polymerase II (PolII) [Stroud et al., 2013].

Further evidence that a non-canonical RdDM mechanism is involved in TE silencing is from *EVADÉ* (a long-terminal repeat (LTR) that was activated in met1-derived epigenetic inbred lines) [Marí-Ordóñez et al., 2013; Mirouze et al., 2009] or *ATHILA6* (a LTR transposon subfamily activated in *ddm1*) [McCue et al., 2012]. These studies found that transcriptionally active TEs were initially transcribed by PolII and copied by RDR6 to produce dsRNAs. These dsRNAs were processed by different DCLs into 21-22nt siRNAs which, aside from participating in further PTGS, could also initiate de novo DNA methylation in a DRM2- and PolV-dependent manner [Bond and Baulcombe, 2015; Panda and Slotkin, 2013] (Figure 7.1). It has been suggested that this pathway and the phasiRNA pathway described in previous chapters are related

mechanisms and, consistent with that idea there is hyper-methylation in the body of TAS genes [Wu et al., 2012]. How the sRNAs are loaded into one side or the other remains the object of future studies.

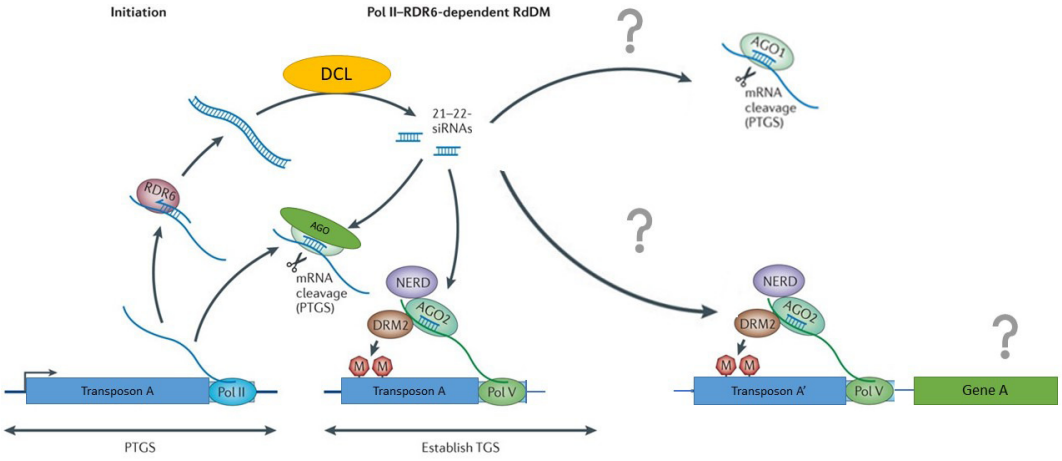

Fig. 7.1 **Non-canonical RdDM simplified model.** Schematic diagram depicting the silencing of a transcriptionally active TE. (Left panel) A TE is initially transcribed by PolII, targeted by a trigger sRNA and transformed into a dsRNA via RDR6. A DCL protein then slices the dsRNA molecule into 21-22nt siRNAs. These siRNAs can then initiate methylation via PolV/DRM2. This process subsequently activates the canonical RdDM pathway, entering a feedback loop reinforcing the newly established cytosine methylation (not showed). Additionally, the same siRNAs could potentially find alternative targets and initiate PTGS or TGS *in trans*, affecting gene expression. Adapted from Matzke and Mosher [2014].

In normal conditions, only few TEs have a basal level of transcription. However, larger TE derepression occurs naturally in plants under certain genetic or environmental conditions. Several reports describe such reactivation in hybrid crosses, and in response to biotic and abiotic stresses (reviewed in Martínez and Slotkin [2012]). There is also evidence for transient relaxation of repression during development. Such is the is case in meiocytes, vegetative nuclei (VN) of pollen, and the central cell of the female gametophyte of both *A. thaliana*, rice, and maize [Martínez and Slotkin, 2012; Slotkin et al., 2009], and also

in the meristem of rice and maize [Vicient, 2010]. The consequence of this relaxation during reproductive development or meristem progression remains largely unknown.

The only natural occurrence of TE activation that has been thoroughly studied so far is during pollen development. DDM1 is down-regulated in the VN, resulting in the expression of LTRs and consequently an increase in 21nt siRNAs and CHH methylation mapping to these loci [Calarco et al., 2012; Ibarra et al., 2012]. These TE-derived siRNA can then accumulate in the sperm cells presumably through a mechanism requiring cell to cell movement. Such accumulation of siRNA reinforces silencing by targeting *de novo* methylation in these cells [Kawashima and Berger, 2014]. In *ddm1*, siRNAs derived from active TEs can act *in trans* by targeting non-neighboring genes that share sequence homology [McCue et al., 2012]. This mechanism has also been observed in animals, where totipotent oocytes exhibit high levels of TE transcription and TE-derived siRNAs that regulate expression of multiple genes [Watanabe et al., 2008]. As such, the presence of TE transcripts in plants may also play a direct role in regulating development by modulating gene expression.

TEs can also be activated under a range of environmental stimuli. Ito et al. [2011], for example, showed that the *ONSEN* TE was transcriptionally activated under high temperature stress but this activation did not correlate with increased genome integration. The lack of transposition suggests that endogenous transcriptional activation of TEs is not sufficient to cause genome instability. Addtionally, a comparative study of TEs in *A. thaliana* and *Arabidopsis lyrata* (a closely related species with 2.5 times more copies of TEs) showed a clear correlation between the presence of TEs and differences in expression of orthologous genes [Hollister et al., 2011]. These effects were partially explained by alterations of the local chromatin structure affected by TE-associated DNA

methylation and histone modifications.

In a recent study, Makarevitch et al. [2015] analysed TEs near stress-responsive genes in maize and found 20 TE families that associated with stress-responsive expression of nearby genes. Expression of these TEs was responsive to the same stress conditions and they appeared to provide enhancers that influenced the expression of nearby genes [Makarevitch et al., 2015]. Taken together, the current model would suggest that epigenetic regulation of TEs might serve additional purposes beyond protecting the host genome integrity, including being co-opted by the host to regulate gene expression.

7.1.2 Tomato vegetative meristem as a new model system for TE regulation studies

A. thaliana has been particularly efficient in restraining TE proliferation. Only 10 % of its genome is comprised of TEs. This small number of TEs may explain why RdDM mutants in *A. thaliana* do not present developmental abnormalities, while species with large numbers of TEs (such as maize or rice) do [Lisch, 2009]. For this reason, *A. thaliana* might not be the ideal model to investigate the effects of TE reactivation during normal development. The crop species tomato, *Solanum lycopersicum*, was chosen to further explore the relationship between TE transcription, sRNA expression and gene expression. The tomato genome is significantly larger than *A. thaliana* (900Mb versus 125Mb), with mobile elements comprising almost two thirds of the genome [The Tomato Genome Consortium, 2012].

In tomato, TE insertions are highly abundant in the promoter sequences of genes [The Tomato Genome Consortium, 2012]. Additionally, siRNAs preferentially map to euchromatic regions with methylated cytosines in the asymmetric

CHH context [The Tomato Genome Consortium, 2012]. The two observations together suggest a higher rate of interplay between the RdDM pathway and the regulation of gene expression. This interplay has also been suggested in a comparative study between tomato and its close wild relative *Solanum pennelli*. Bolger et al. [2014] showed that genetic differences that improve stress tolerance in the *S. pennelli* are tightly linked to the presence of unique TE insertions in regulatory regions of those genes. In addition, fluctuation of methylation has been shown to play an important role during a further developmental process in tomato: fruit development and ripening [Liu et al., 2015a; Zhong et al., 2013].

Genetic regulation of the vegetative meristem has already been described in tomato [Park et al., 2012]. As mentioned previously, natural TE activation has also been reported in the vegetative meristem [Ohtsu et al., 2007a; Tamaki et al., 2015; Vicient, 2010]. In maize, it was also shown that the developmental shift from juvenile growth to an adult reproductive phase in the meristem is associated with activation of the MuDR TE [Li et al., 2010]. In Park et al. [2012] however, TE activation in the tomato meristem was simply not explored.

My aim was to assess whether TE transcriptional activation would occur in tomato meristems and whether this process would be associated with changes in gene expression. TEs have already been reported to be activated in the meristem of species with genomes with large TE content. Due to the colonisation success of the tomato genome by TEs, I hypothesized that this phenomenon is also conserved in this species. Furthermore, I hypothesised an amplification and diversification of the population siRNAs as a consequence of the transcriptional upregulation of certain TEs. These siRNAs derived from transcriptionally-active TEs would account, at least in part, for the changes in gene expression.

7.2 Results

7.2.1 Transposable element transcription upregulation in tomato vegetative meristem

Two developmental programmes appear during the life cycle of flowering plants. First there is a vegetative phase, where the apical meristem produces leaves and lateral shoots. Later in development, floral induction occurs and the apical meristem changes its developmental pattern and initiates the production of flowers. The initial *Solanum lycopersicum* vegetative meristem matures into a floral meristem after the formation of the 6-8th leaf primordia [Park et al., 2012]. To ensure that only vegetative meristems where collected, only plants presenting the 5th leaf primordial where used (Figure 7.2). Meristems were manually dissected following the procedure detailed in Park et al. [2012] explained in chapter 2.

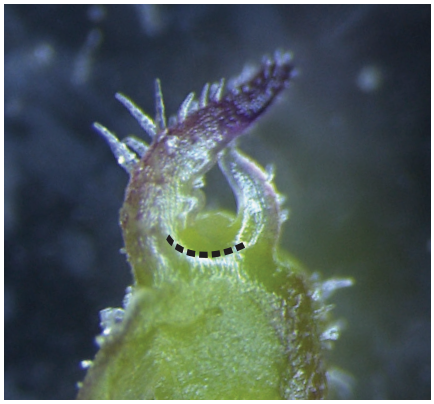

Fig. 7.2 **Microdissection of tomato vegetative meristems.** Solanum lycopersicum var. M82 vegetative meristem under the stereo microscope. Dashed lines indicate dissected tissue line.

High throughput sequencing (HTS) of RNA transcripts was used to compare the expression profile of TEs in the vegetative meristems to those in mature leaves. Sample sets were comprised of four biological replicates. To avoid

artefacts due to miss-mapping of reads, I considered only uniquely mapping reads. A TE was considered 'expressed' if it had over 0.1 reads per kilobase of transcript per million mapped reads (RPKM) in any of the libraries. Given the depth of size of my libraries, this meant that any TE had to have at least 10 non-redundant reads in the library.

Expression of 20% of all TE tags (106,570 of 544,443; ITAG2.4 annotation) could be detected in my datasets (min. RPKM=0.1), with overall higher expression of TEs in the vegetative meristem (Figure 7.3). My results confirm that transcriptional activity of several transposable elements is particularly high in the tomato vegetative meristem, which is consistent with previous observations in maize [Vicient, 2010]. Interestingly, over half of the TEs (61,032, 11.21% of the total amount of TEs) were exclusively detected in the meristem vs 7,560 in the leaf (Figure 7.3), pointing to a major de-regulation of TE silencing in the meristem.

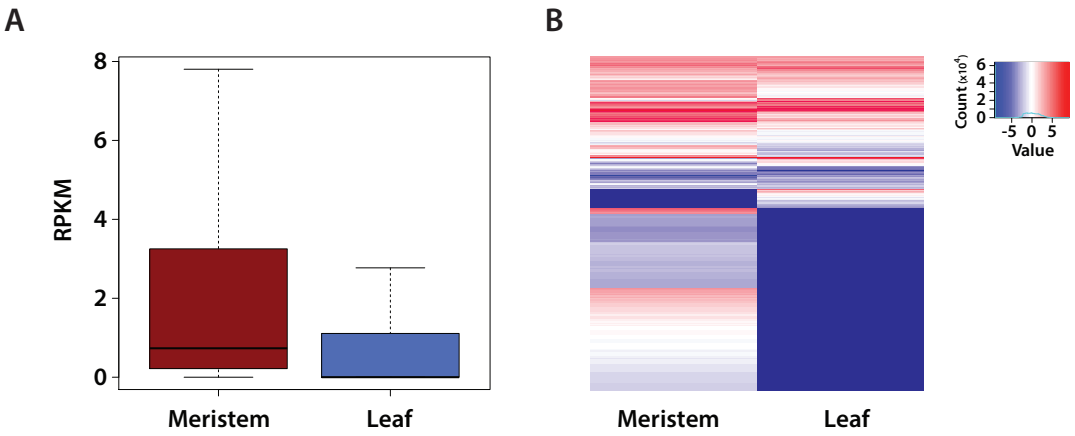

Fig. 7.3 **Transposable elements are upregulated in the tomato vegetative meristem.** (A) Box plot representing relative reads per TE between meristem and leaf libraries (min. RPKM=0.1). (B) Heatmap representation of the log value of expressed TEs.

I performed a differential expression analysis between meristem and leaf samples. Only TEs with a false discovery rate (FDR) smaller or equal to 0.1 were considered in any further analysis. Among the elements upregulated in the meristem, LTRs were significantly overrepresented (Figure 7.4). Since LTRs make up for 95.7 % of the repetitive elements present in tomato genome, I set out to determine if there is any correlation between TE copy number and transcription levels. For that purpose, I used an extensive and well described LTR dataset from a previous study [Xu and Du, 2014]. I did not find a clear correlation ($R^2 = 0.241$) although a trend could be inferred between the number of inserted elements and their level of expression (Figure 7.5A).

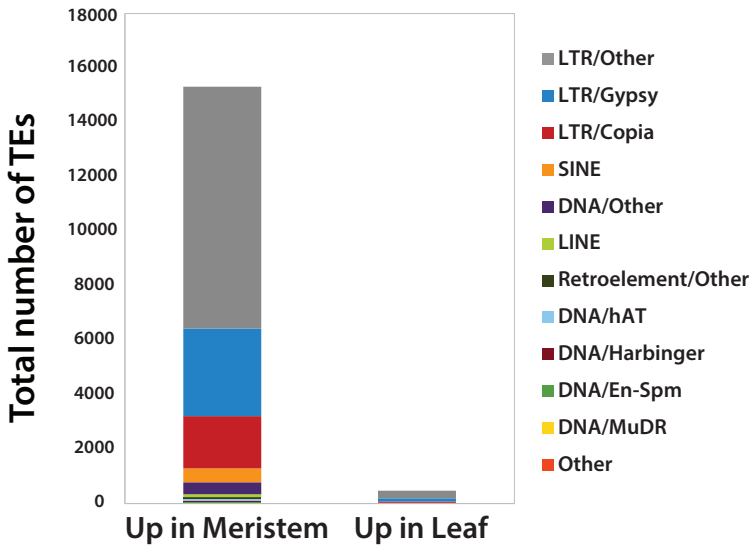

Fig. 7.4 . **Transposable elements are upregulated in the vegetative meristem and young leaves of S. lycopersicum cv. M82.** Table representing the number of elements upregulated in the meristem (left) and leaf (right) based on baySeq results. Elements are grouped by major families. FDR cut-off = 0.1.

The majority of old LTR elements in the tomato genome are located in recombination-suppressed heterochromatin regions, while young LTR elements are preferentially located in the gene-rich euchromatic regions [The Tomato Genome Consortium, 2012; Xu and Du, 2014]. Therefore, I tested whether

the genomic location of LTRs could influence the level of their transcriptional activation in vegetative meristems. My analysis showed no statistical difference between expression levels of elements located in the euchromatin compared to those in the heterochromatin (p-value = 0.3927 in a Welsch two sample t-test) (Figure 7.5B). Alternatively, I tested whether the time of insertion, as estimated by Xu and Du [2014], would influence the transcriptional activation of those elements and no pattern could be inferred (Figure 7.5D). The reactivation of certain LTRs in the vegetative meristem could not be explained by copy number, genomic location or the time of insertion.

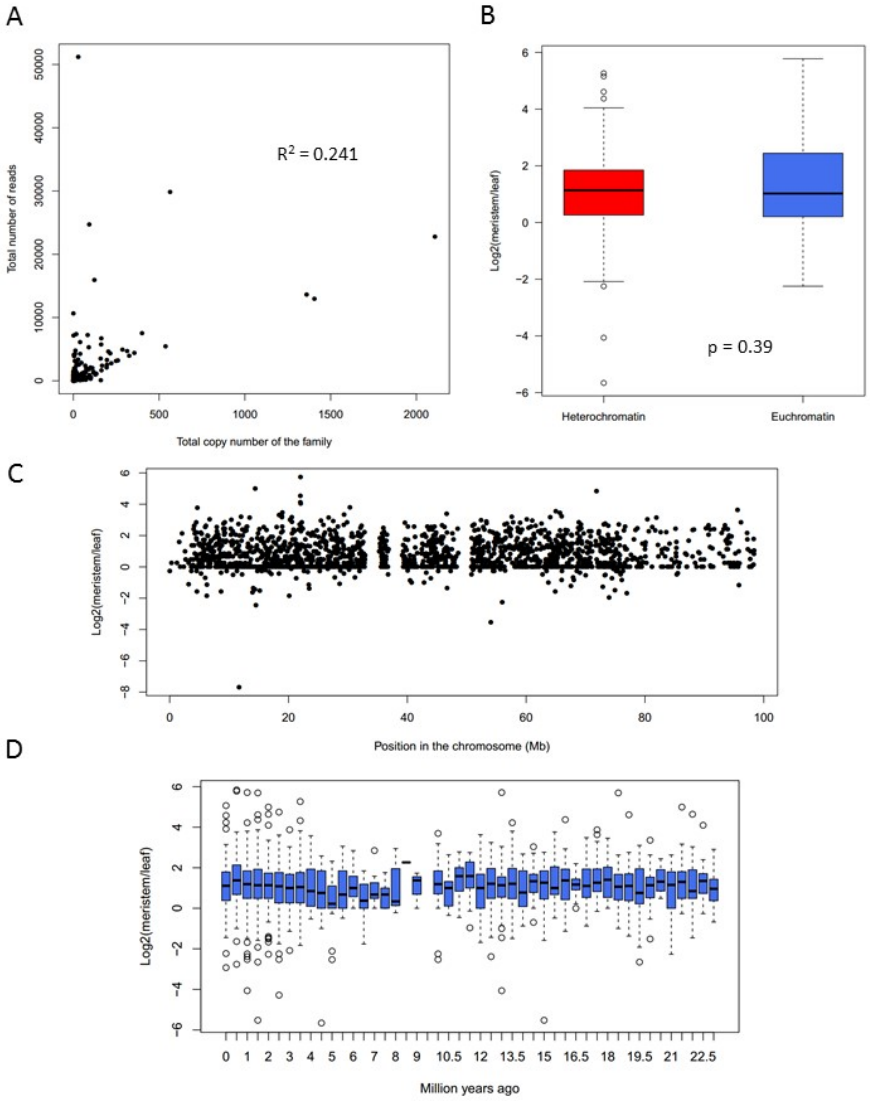

Fig. 7.5 **LTR over-expression in the meristem is not correlated with copy number, genomic location or time of insertion.** (A) Dot-plot of LTR copy number and the number of reads. Each dot represents an LTR subfamily. (B) Boxplot representation of expression change of LTRs elements located in the euchromatin (blue) or heterochromatin (red) (Classification was based on Xu and Du [2014].). (C) Dot-plot representation of expression change of LTRs elements located in chromosome 1. (D) Boxplot representation of expression change of LTRs elements sorted by time of insertion in the genome.

7.2.2 TEs are enriched in the vicinity of genes with dynamic expression

TE insertion in promoter regions may provide novel *cis*-acting regulatory sites and influence transcript production. In *A. thaliana*, insertions of the heat-activated LTR *ONSEN* affect the heat-responsive expression for nearby genes [Ito et al., 2011]. In rice, transposon mPing itself is up-regulated in response to cold stress, and similar changes in regulation of expression has been observed for rice genes located near new mPing insertions [Naito et al., 2009]. I hypothesized that changes in gene expression could be partially conditioned by the presence of TE elements in the vicinity of tissue-specific genes.

First, I performed a differential expression analysis between my meristem and leaf samples. I was able to detect expression of a total of 25,486 and 21,170 genes in meristem and leaf samples, respectively. For the DE analysis, only genes with a FDR smaller or equal to 0.1 were considered as differentially expressed. As quality control of my datasets, I included tomato meristem RNA-seq data from Park et al. [2012]. Around three quarters of the list of total upregulated genes in the meristem were shared between both datasets (Figure 7.6A). Some differences (a quarter of the total in each case) could be expected because of the preparation methods of the datasets. My RNA samples were enriched by ribosomal RNA depletion while Park et al. [2012] were enriched by polyA selection.

From my datasets, I wanted to investigate if TEs insertion near genes was associated with near tissue-specific expression of these genes. Genes with differential expression in any direction in my datasets were clearly enriched for TEs presence in their regulatory regions (Figure 7.6B). The enrichment was higher in the 5' region of the genes, comprised of the promoter region that regulate gene expression. This finding is in line with previous observations in

maize [Makarevitch et al., 2015]. These results suggest that TEs themselves could be a recurrent source of enhancers to influence the expression of genes.

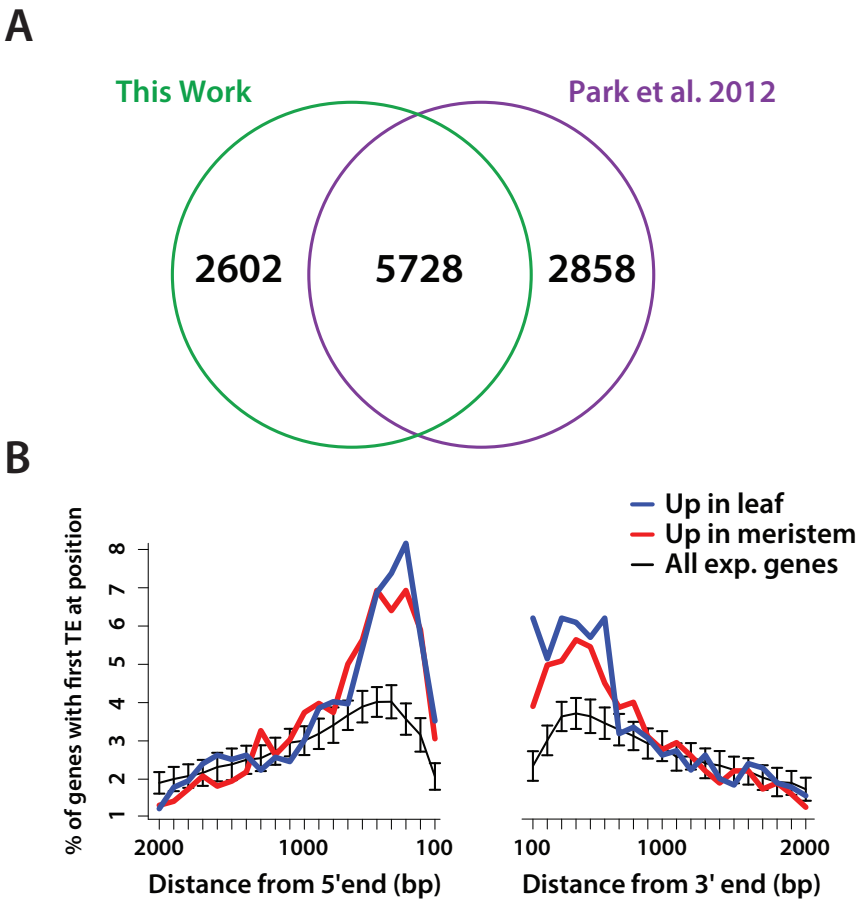

Fig. 7.6 **TEs are enriched in the vicinity of genes with dynamic expression.** (A) Venn diagram of DE genes from meristem samples of this work and the work from Park et al. [2012] vs the leaf samples of this work (B) Line plot of the percentage of genes with the first TE at a certain genetic distance in genes upregulated in the meristem (red) and leaf (blue) compared to all genes expressed in my datasets (black, with variances and mean estimates obtained using bootstrap analysis).

7.2.3 Meristem-specific siRNA profile

The small RNA silencing machinery and the transcriptional regulation of TEs are processes deeply interlinked. For that reason, I decided to investigate the sRNA populations in the vegetative meristem. Overall, sRNA profiles revealed that 21 and 24-nt sRNAs were the predominant size classes in leaf and meristem samples (Figure 7.7A). These two sizes classically represent the majority of miRNAs and siRNA. I further classified sRNAs according to genome annotation; rRNA, miRNA, coding genes, and TE-associated siRNAs. In both size classes and tissues, the largest group of mapped sRNAs were TE-associated (Figure 7.5B). Notably, the proportion of miRNAs observed in the meristem was significantly higher than that in leaf. An effect due primarily to the high abundance of two miRNA species, miR165 and miR166, that regulate the cell fate of vegetative meristem cells.

Due to their intrinsic small nature, mapping sRNAs to repetitive elements proved to be challenging. Out of all possible options, the least biased approach to investigate these datasets was to account only for sRNAs mapping to only one genomic position. Out of all uniquely mapping TE-associated siRNA (rasiRNAs), the majority mapped to LTRs that had been previously identified as expressed in my RNAseq analysis. This was particularly significant for 21-nt sRNAs, where 74.9 % in meristem and 92.5 % in leaf of rasiRNAs were associated with transcriptionally active TEs. Similarly, in *A. thaliana*, activation of TEs is associated with production of 21nt siRNA at LTRs through a non-canonical RdDM pathway [Nuthikattu et al., 2013]. PolII transcripts get targeted by miRNAs or siRNAs, which are used as seeds by RDR6 to produce dsRNAs that will then be processed by a member of the DCL family and amplify the siRNA response [Matzke and Mosher, 2014].

A lower proportion of 24nt than 21nt siRNAs was associated with transcriptionally active TEs (42.4 % in meristem and 71.5 % in leaf). This is consistent with the current model of RdDM where, once DNA methylation is established, polII transcription and RDR6-derived 21nt siRNAs become unnecessary. In this case, transcriptionally silent TEs could still be producing 24nt through the canonical RdDM machinery while remaining transcriptionally inactive.

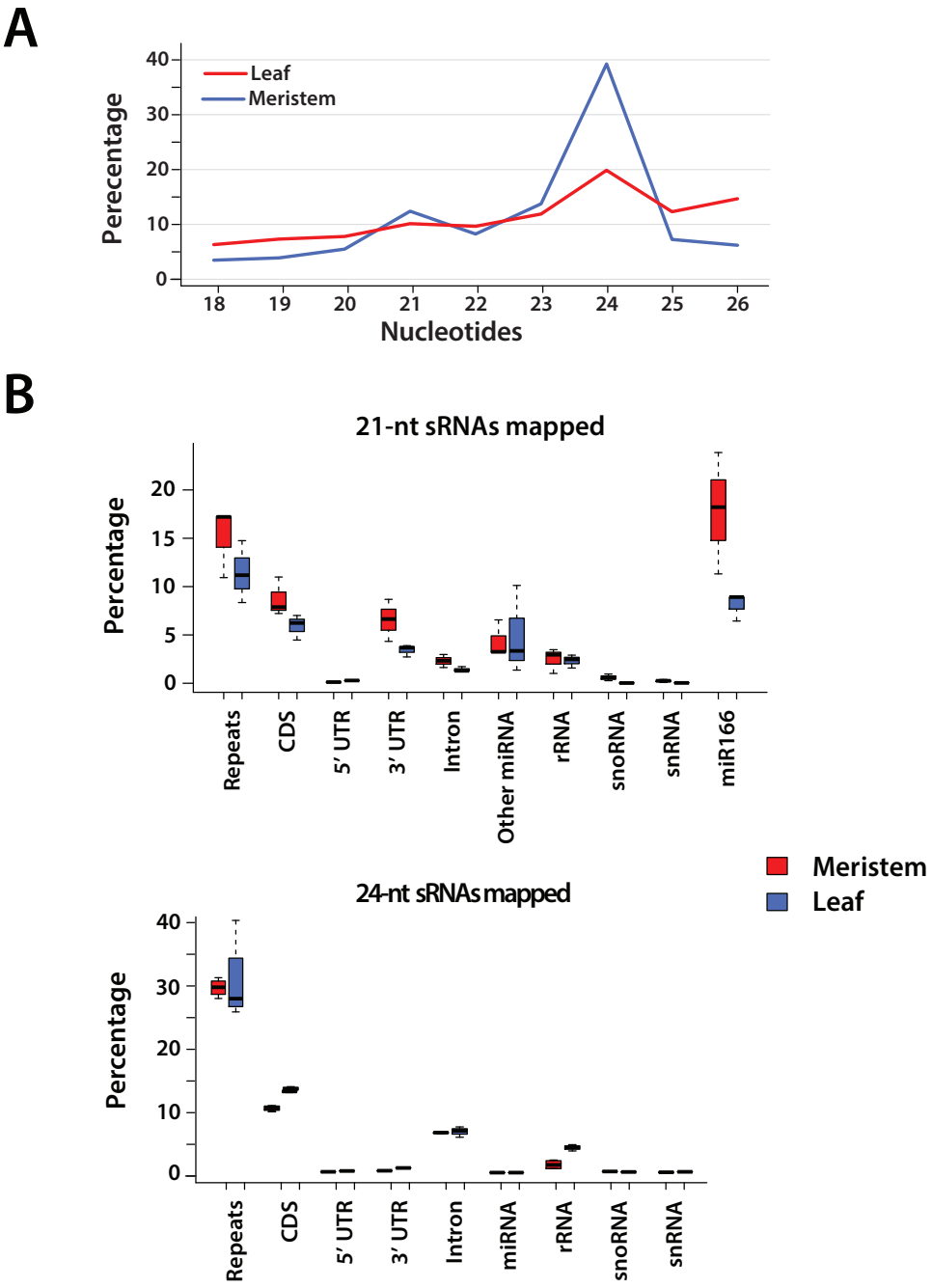

Fig. 7.7 **Meristems accumulates a higher proportion of small RNAs derived from LTRs.** (A) Line plot summarizing the small RNA sizes present in the tomato meristem (red) and leaf (blue) libraries. (B) Box-plots summarizing the annotated proportion of 21-nt and 24-nt small RNAs from (red) meristem and (blue) leaf samples.

7.3 Discussion

The work described in this chapter set out to test the hypothesis that TE transcriptional activation would occur in tomato meristems and that this process would be associated with changes in gene expression. I further hypothesized that the siRNAs associated with activated TEs would account, at least in part, for the changes in gene expression. In this scenario the siRNAs would mediate the 'controlling element' consequence of TEs conceptualized in the work of Barbara McClintock [McClintock, 1953].

The first part of this hypothesis was confirmed by the analysis of RNA-seq datasets from tomato vegetative meristems. The results presented here demonstrated that TEs are transcriptionally activated in the meristem (Figure 7.3). Thousands of TEs, mainly LTR retrotransposons, become upregulated in the vegetative meristem when compared to leaf tissue. This is consistent with previous findings in maize [Ohtsu et al., 2007b; Vicient, 2010].

The vegetative meristem is a complex organ responsible for the development of all above-ground structures, including the germline. It is comprised of several cell types that surround a region of pluripotent stem cells that divide and differentiate, as well as self-replenish. Genetic studies have revealed a complex network of dozens of genes involved in regulating maintenance of stem-cell identity, specification and differentiation, and control of the transition from the vegetative to the reproductive program [Soyars et al., 2016]. Reactivation of TEs might be restricted to a particular cell type that would act as the companion cells of the rapidly dividing pluripotent cells. If this TE activation took place in a particular cell type I envision that the associated siRNAs could move into the meristematic cells.This has already been suggested to occur in the root meristem, where the columella presents major TE de-methylation [Kawakatsu et al., 2016]. This would also be similar to the activation of TEs in the vegetative nuclei (VN)

during pollen formation. This activation coincides with a substantial loss of centromere identity, H3K9 methylation, and heterochromatin decondensation [Calarco et al., 2012; Schoft et al., 2009]. The activation has the specific function of delivering siRNAs into the gametes to reinforce TE silencing [Martínez et al., 2016]. In the vegetative meristem, the main purpose of such a scenario would be aiding the silencing machinery in cells that take part in propagating the plant to the next generation. One way to address the possibility of sRNA movement between cells in the meristem would be to study the location of precursors (TE transcripts) and signals (TE-derived siRNAs) using RNA in-situ hybridization assays of meristem sections.

Transposition is likely to be governed by multiple redundant mechanisms. During certain developmental stages or stresses, some of these mechanisms may be impaired. If the plant presents redundant mechanisms of TE regulation, no transposition will occur even when TEs are transcribed. And since TE regions are not under strong selective pressure, if in any of these cases a newly evolved rasiRNA fortuitously guides the cleavage of an mRNA, this interaction could become the subject of selection. Situations in which RNA-mediated regulation of TEs provides a fitness benefit would be subject to positive selection. Eventually such a process could lead to full domestication of a TE in a manner that is parallel to the way that TE derived genes have given rise to transcription factors [Dermitzakis and Clark, 2002; Lin et al., 2007] or be involved in the evolution of miRNAs [Li et al., 2011; Piriyapongsa and Jordan, 2008; Xia et al., 2015]. In *A. thaliana*, a secondary siRNA from an activated LTR can direct PTGS of a stress related gene [McCue et al., 2012]. I suggest that similar events could be occurring in the meristem, where siRNAs from active TEs could find new targets of PTGS or TGS in genes and their regulatory regions where new insertions have occurred (Figure 7.1). Studying of cleaved RNAs in meristem through degradome analysis and correlate them with abundant TE-derived siRNAs could provide clues as of the existence of such regulatory function of

'domesticated' TEs as a source of regulatory sRNAs.

The possibility, however, that the associated siRNAs would influence gene expression was not confirmed by this analysis. Nevertheless, I also found that upregulated TEs were associated with 21nt and 24nt siRNA loci, while silenced TEs were only associated with 24nt siRNAs. This can be compared to what has been previously reported in *A. thaliana*, where upregulated LTRs in the pollen VN give rise to 21nt siRNA and become CHH hypermethylated by the RDR6 non-canonical pathway. TE-derived 21nt siRNAs have been linked to the initiation of de-novo DNA methylation [Bond and Baulcombe, 2015; Panda and Slotkin, 2013] but their direct role in DNA methylation reprogramming in the plant remains unclear. There is evidence that 24nt siRNAs are not sufficient to initiate silencing in *trans* at least in transgenes [Robinson, 2006] or in endogenous alleles (paramutagenesis) [Hollick, 2012]. It is likely that a receptive chromatin status and/or the presence of 21nt siRNA are the requirements for TGS initiation. Then, the switch to 24nt would boost the signal through the canonical RdDM feedback loop and reinforce the silencing. Whether 21nt siRNAs are the sole initiators of the silencing remains unknown because it is impossible to generate a mutant where their production is completely lost.

Transcriptional activation of TEs in normal development appears to be a common feature in animals and plants [Hollick, 2012; Malone et al., 2009; Vicient, 2010; Watanabe et al., 2008]. One possible explanation is that one of the aims of a transcriptional upregulation of TEs is to amplify and diversify the population of siRNAs. With such a mechanism, a cell may store regulatory information in different TEs that would be accessed at particular times in development. siRNAs derived from transcriptionally-active TEs could trigger PTGS and TGS of developmentally relevant genes. TEs are highly variable and so this proposed information storage could be highly variable between species or even between members of the same species. In the recent years, transcriptional

activation of TEs during reproductive stages has been linked to epigenetic reprogramming, imprinting and TE control in the germline [Calarco et al., 2012; Pignatta et al., 2014], but additional roles cannot be ruled out.

Overall, these results have provided sufficient preliminary evidence to support the use of the tomato vegetative meristem as a new model system for TE regulation. This study has also provided a first glance at the transcriptional and sRNA profile of the vegetative meristem, but further work is required to test the hypotheses presented before. Investigating the epigenetic state of TEs and genes in the meristem should be an immediate priority. Linking sRNAs to repetitive elements turned out to be a herculean task, as much information was lost due to the inability to map many sRNA to unique genomic locations. Methylation analyses are performed on longer strings of information (sequencing reads) and therefore, their association with singular genomic locations is less challenging. Correlation of methylation analysis with my siRNA would reveal any links between epigenetic changes associated with gene regulation, and siRNAs derived from transcriptionally active TEs.

It has been over 60 years since Barbara McClintock used, for the first time, the term 'controling units' (later renamed and popularized as 'controlling elements') [McClintock, 1953]. Her initial findings on how TEs could both move and control gene expression are still very relevant today. My findings build on that idea by accounting for the interplay between TEs and both the RNA silencing and the epigentic machinery.

7.4 Acknowledgements

Prof. Zachary Lippman (Cold Spring Harbor Laboratory, USA) provided valuable advice for the meristem collection technique. Dr. Thomas Hardcastle and Dr. Bruno Santos provided valuable advice on HTS data analysis.

Chapter 8

Discussion

The work presented in this dissertation has included new evidence about the extent that RNA silencing contributes to the regulation of vital processes in plants. Preliminary data about its involvement in reproductive development and TE regulation in the vegetative meristem was discussed extensively in chapters 6 and 7, respectively. Therefore, the following section will be centred around the major focus of this dissertation: the role of RNA silencing in plant immunity.

Prior to this work, there had been numerous reports of NLR regulation through RNA silencing in plants [Arikit et al., 2014; Li et al., 2012b; Liu et al., 2014; Shivaprasad et al., 2012; Zhai et al., 2011]. However, a clear link between plant immunity and these mechanisms was still missing. By studying the miR482/2118 family, one of the most extended miRNA families involved in NLR regulation, this dissertation aimed to shed light on this phenomenon. The results presented here expanded the current knowledge of the miR482/2118 family in tomato and showed that their action quantitatively affects plant immunity.

8.1 The miR482/2118 family and its role in plant immunity

8.1.1 Final remarks on miR482/2118 targets

Based on my findings, expressed NLR RNAs could be classified in three types: (1) NLRs not giving rise to siRNAs, (2) pNLRs that did not change abundance of siRNAs in target mimic lines, and (3) pNLRs that changed abundance of siRNA in target mimic lines.

NLR targets that not give rise to phasiRNAs

One of the biggest current challenges when studying sRNA function is accurately identifying true targets. It was initially estimated that up to 30 % of all tomato NLRs were direct targets of miR482 [Shivaprasad et al., 2012]. Using target mimics, I could only observe 10 % of all NLRs (32 out of 301) giving rise to phasiRNAs (pNLRs), and less than 6 % (17; 10 % of NLRs with evidence of expression based on Andolfo et al. [2014]) seemed to be miR482-dependent (Figure 3.4 & Table 4.3).

Factors other than miRNA targeting or target abundance have been shown to impact phasiRNA production. Fei et al. [2015] indicated that false positives in target predictions might be due to not accounting for target inaccessibility, which is determined by mRNA secondary structure on flanking target sites. They also observed that extensive 3' pairing, different to canonical rules for miRNA–target interactions, is necessary for 22-nt miRNA targeting. In addition, ectopically expressing *Medicago* miR482/2118 members in *A. thaliana* showed that only a few NLRs were capable of phasiRNA production, and this effect was correlated with amount of free energy at the flanking regions of the target site [Fei et al., 2015]. The results are also consistent with prior evidence of

evolutionary selection in the flanking sequence around miRNA–target sites to increase target accessibility [Gu et al., 2012]. These observations would indicate that the particular sequence motif required for miR482/2118 targeting is not required for NLR function, and the selective pressure to keep such motif exists solely as result of a co-evolution of the miRNA and its targets.

One additional explanation for the abundance of false positives is that the phasiRNA machinery could be only co-localized with a subset of "potential" targets in a specific sub-cellular domain. Several studies have indicated that different types of small RNAs are distinctly partitioned between sub-cellular locations such as the nucleus, the endoplasmic reticulum (ER) or cytosol [Li et al., 2016, 2013a; Pontes et al., 2006]. Additionally, protein components components of these pathways are also distinctly localized in the cell (reviewed in Leung [2015]). In the case of phasiRNA production, main effectors such as RDR6, SGS3 or (in the case of TAS3) AGO7, are localized in cytoplasmic siRNA bodies that are linked to the ER/Golgi endomembrane [Jouannet et al., 2012].

In summary, even if a miRNA/transcript interaction was plausible, it would not occur if both component do not meet sub-cellularly. I would hypothesise that this might be the true even in a developmental basis. Different cell types could trigger targeting and phasiRNA formation only under certain developmental conditions or stress responses. I would argue therefore that the concept of miRNA target and the conclusions of silencing studies should be limited to the spatio-temporal characteristics of the observations. I speculate that many previous *in silico* analysis of the regulation potential of the miR482/2118 family, or any other miRNA family, could be based on great overestimation of direct targets of these miRNAs. However, the extent by which these miRNAs could influence regulation through translational repression or through their derived

phasiRNAs is currently unknown.

pNLRs not triggered by miR482

Although the most significant contributors to the pNLR-derived productions were miR482-dependent, almost half of the pNLRs did not change their levels of phasiRNA production in the *MIM482* lines (Table 4.3). This suggest that other miRNAs also participate in this regulation, even though miR482 is probably the most prominent player. This suggestion is consistent with previous observations in tomato and other Solanaceae, where other miRNAs such as miR5300, miR6019, miR6027, miR6024, and miR6026 were also reported to target NLRs and trigger phasiRNAs [Li et al., 2012b; Shivaprasad et al., 2012]. The existence of redundant mechanisms to trigger phasiRNAs at NLR loci reinforces the suggestion that regulation is of great biological relevance.

miR482-triggered pNLRs

When NLR regulation by miR482/2118 was initially discovered, these miRNAs were proposed as 'master regulators' of the expression of the defence genes [Li et al., 2012b; Shivaprasad et al., 2012; Zhai et al., 2011]. As mentioned previously however, my experiments suggest that miR482 interacts with fewer NLRs than previously anticipated [de Vries et al., 2015; Shivaprasad et al., 2012]. One remaining question is whether there is any relationship between the different miR482-dependent pNLRs that could explain why some of them produce phasiRNAs and others do not.

Based on the phylogenetic analysis in Andolfo et al. [2014], my preliminary observation indicates that the majority of miR482-dependent pNLRs belong mostly to two basal CNL clades (Figure S.7). One of these clades, which

contains the first reported targets of miR482 (*LRR1* and *LRR2*; Shivaprasad et al. [2012]), was identified as the helper CC_R sub-class of CNLs in Andolfo et al. [2014]. However, I believe this interpretation to be incorrect, as the reference NLRs that cluster together with this clade are not CC_R NLRs but canonical CNLs such as *A. thaliana* RPS2 or *lettuce* Dm3 (RPGC2B)*. Additionally, a recent report has indicated that tomato only contains two CC_R NLRs (*Solyc02g090380* and *Solyc04g079420*)† [Qian et al., 2017], which is in agreement with other studies that have shown that CC_R NLRs are scarce among all angiosperms [Shao et al., 2016].

Interestingly, all representative members of the pNLR clades (RPS2, RPS5, Dm3, I2, and R3a; Figure S.7) have been reported to have a CC domain that lacks the highly conserved EDVID motif found among most characterized CNLs [Giannakopoulou et al., 2015; Qi et al., 2012; Rairdan et al., 2008]. Overall, this suggests that miR482 targets belong to a very distinct and evolutionary ancient CNL subclade. The representative members of this subclade have been shown to not require the action of any known NLR helpers [Collier et al., 2011; Wu et al., 2017]. It is possible that these CNLs act through different and maybe archaic mechanisms that do not require the action of any helpers. Missing a regulatory bottleneck that would enable downstream regulation of the deleterious effects of autoimmunity would prompt the necessity of direct regulation through miRNA targeting. This would explain why the tomato Prf/Pto guard/guardee pair does not appear to function outside the Solanaceae as it requires NRC helpers, while members of this CNL clade such as *A. thaliana* RPS2/RIN4 and RPS5/PBS1 retain function across plant lineages [Ade et al., 2007; Day et al., 2005]. A more in-depth study of the characteristics of these

*Upon contacting authors of Andolfo et al. [2014], they confirmed the mistake in the interpretation of the data.

†These two tomato CC_R NLRs were not among the pNLRs or the predicted targets of any members of the miR482/2118 family (Tables 4.3 & S1-5).

NLRs could be crucial in the understanding of the miR482-dependent regulation.

8.1.2 Function of miR482-derived phasiRNAs

The widespread conservation of NLR-derived phasiRNAs strongly implies a role in the regulation of NLR expression that has already been discussed [Liu et al., 2014]. My work additionally provides direct evidence of miR482 involvement in immunity (Figures 5.5 & 5.7), although the precise mechanism of action of miR482/pNLRs-derived phasiRNAs remains unknown (Figure 8.1).

Based on the knowledge obtained from other PHAS loci, it has generally been suggested that pNLRs-derived phasiRNAs could act *in trans* to regulate other similar genes [Fei et al., 2013]. However, I could not identify any degradome products that would occur as a consequence of targeting by the most abundant phasiRNAs (data not shown). One explanation is that given the large number and variability in sequence of pNLR-derived phasiRNAs with relative low abundance of individual forms, dosage and repeated targeting is likely to be a determining factor. The nature of this repeated targeting would at same time mask any degradome signatures of a single phasiRNA. However, it also is possible that pNLR-derived phasiRNA might work preferentially through other mechanisms of RNA silencing.

Li et al. [2016] found that phasiRNA biogenesis in *A. thaliana* is associated exclusively to membrane-bound polysomes close to siRNA bodies. This observation suggests that phasiRNA production occurs at the rough ER, where another ER membrane protein, AMP1, has been shown to co-localize and be required for translational repression [Li et al., 2013a]. Other studies have shown that 21-nt phasiRNAs also trigger TGS by initiating DNA cytosine methylation by the non-canonical RdDM pathway [Stroud et al., 2014; Wu et al., 2012]. One

hypothesis would be that 21-nt phasiRNAs might preferentially work through translational repression and/or TGS instead of RNA cleavage and degradation. In order to test this hypothesis, one could investigate any correlations between the transcriptional status of NLRs, and their DNA methylation level and ribosome footprints in both wild type and my mimic lines using Next-generation sequencing techniques (RNA-seq, Bisulphite-seq and Ribo-seq).

8.1.3 Evolutionary significance of miR482-pNLR regulation

One unresolved question is the evolutionary origin and benefits of these miRNAs and pNLRs. Recent work has shown that highly duplicated NLRs are typically targeted by miRNAs [Wei et al., 2014; Zhang et al., 2016a]. Wei et al. [2014] suggested that miRNA targeting of the I2 gene family may have enabled rapid duplication and divergence of this family in the Solanaceae. However this can also be seen from the opposite angle, since highly duplicated NLRs may have a higher chance of generating new miRNAs by genomic rearrangement. Any miRNA generated could then target and regulate the family of NLR genes from which it evolved [Zhang et al., 2015].

A. thaliana contains only two miRNAs (miR472 and miR825*) that target four NLRs (*RPS5*, *RSG1*, *RSG2* and TNL *At5g38850*) [Chen et al., 2010a; Howell et al., 2007]. Interestingly *A. thaliana* miR472 mature sequence is highly similar to miR482, but its precursor sequence suggests that it evolved independently in Brassicae [Zhang et al., 2016a]. The fact that many miRNA families have evolved to target the same encoded motif through distinct, lineage-specific origins suggests a high level of convergent evolution. However, the *A. thaliana* example also shows that the degree of prevalence of this mechanism varies greatly between lineages, with more extreme examples such as rice and other

grasses that contain large numbers of NLRs but lack miRNAs that trigger phasiRNAs at these.

In summary, the evolutionary benefits of pNLRs are yet to be identified. While there is correlation with NLR family expansion and miRNA targeting, a more straightforward explanation would be that miR482 appeared simply to buffer NLR expression levels to reduce the fitness cost of these genes (previously suggested in González et al. [2015]). Broad recognition in NLRs has been associated with higher risk of inappropriate activation [Farnham and Baulcombe, 2006; Harris et al., 2013]. Strikingly, the level of expression of potato CNL of this subclade, *R3a*, conditions its recognition spectrum of several oomycete AVR3a variants (S. Schornack and collaborators; personal communication). An increase in sensitivity and range of detection based on gene expression level might cause a trade-off with plant development for which breeding might have selected against, and in favour of negative regulators such as miR482.

8.2 miR2118b and a TAS gene of novel function

An additional finding in this work was the existence of a *Solanum*-specific TAS gene derived from scrambled NLR sequences and targeted specifically by miR2118b (Figure 3.7). *TAS5* is ncRNA that acts as one of the most prolific phasiRNA loci in tomato. My results have shown that lines with reduced levels miR2118b and *TAS5*-derived phasiRNAs have significantly reduced susceptibility to bacterial and oomycete pathogens (Figures 5.5 & 5.7).

The 'both or none' taxonomical pattern of *TAS5* and miR2118b strongly suggests co-evolution of both genes (Figure 3.8). This is the first reported case

of a TAS gene directly involved in the regulation of defence response. However, I cannot truly conclude that the reduction of susceptibility in *MIM2118b* lines is directly linked to *TAS5* and not to other yet unidentified targets of miR2118b. One way to confirm the role of *TAS5* would be to silence the gene using VIGS against its promoter, which would cause a loss of phasiRNAs without affecting the trigger (miR2118b). Responses in the TAS5-silenced lines to *P. infestans* or *P. syringae* infections would then be measured to check if it reproduces the effects of the *MIM2118b* lines.

The specific targets of *TAS5*-derived phasiRNAs are still unknown and it requires further studying of transcriptional and translational changes occurring in the target mimic lines (similarly to what exposed before for miR482) (Figure 8.1). Potential secondary targets of this miR2118b/*TAS5* mechanism would possibly be highly different to the miR482 pathway, since *TAS5* is composed of sequences similar to TNLs and CNLs of different families to the miR482-dependent pNLRs. It would be of great interest to test whether the effects of both target mimics are additive for NLR regulation and the defence response. Whether other TAS5-like non coding RNAs have appeared in different plant genomes should also be explored.

8.3 Conclusions

This work addressed the extent to which the miR482/2118 family and its derived phasiRNAs regulate NLRs and influence plant immunity in tomato. Taking advantage of small RNA sequencing of target mimic lines, I was able to characterize miR482/2118 targets beyond the *in silico* predictions of previous works, and included the recently discovered *TAS5* gene. These tomato target mimic lines were less susceptible than their non transgenic precursors to pathogens *P. infestans* and *P. syringae*. This analysis extended our understanding of

the miR482/2118 cascade and provided direct evidence of its influence on quantitative disease resistance. These results support the hypothesis that the actions of the miR482 and miR2118b are involved in the regulation of broad spectrum defence, although the mechanism of action is still unclear (Figure 8.1). Transcriptional analysis will provide additional clues to elucidate this silencing cascade. Overall, my results add further support to the idea that RNA silencing may contribute substantially to plant immunity beyond the antiviral response, and should therefore garner careful attention.

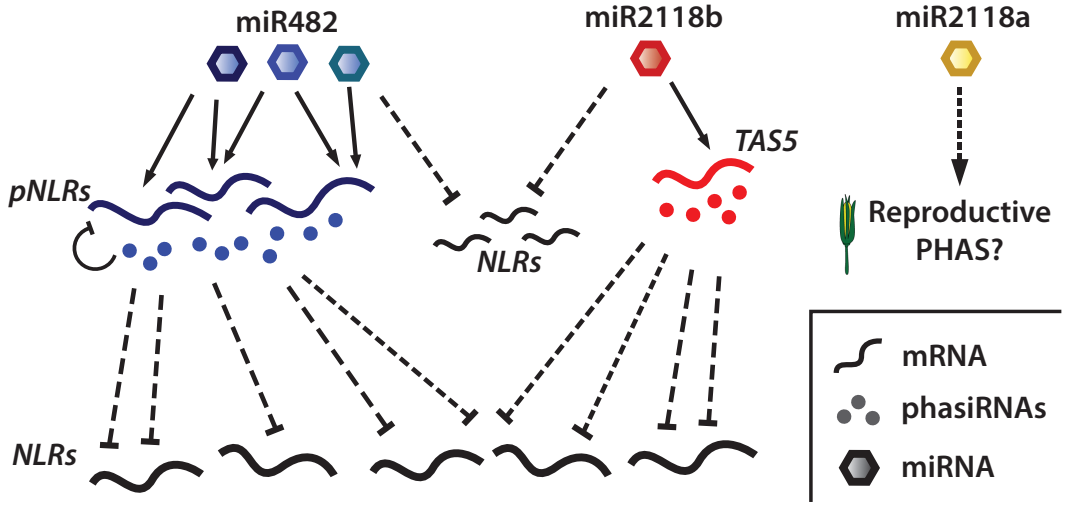

Fig. 8.1 **Proposed model of the miR482/2118 silencing cascade in tomato.** Based on my findings, members of the miR482 family target pNLR mRNAs, triggering phasiRNAs that, in turn may target other NLR mRNAs by suppressing translation or promoting RNA degradation. miR2118b targets *TAS5* that also produces phasiRNAs with the potential to target NLRs. Both miRNAs clades may also promote NLR silencing by directly suppressing translation or promoting RNA degradation of NLR mRNAs. miR2118a is primarily expressed during flower development and is potentially involved in the production of reproductive phasiRNAs of unknown function.

The characterised examples presented in this dissertation, including the study of *ago5* mutants during male gamete development and the transcriptional analysis of vegetative meristems, provide new evidence about the extent that

RNA silencing influences plant development. These findings also explore the use of new biotechnological tools such target mimics against NLR-regulating miRNAs as an approach for enhancing disease resistance in highly bred cultivars. Study and manipulation of RNA silencing mechanisms should bring about significant improvements in crop traits, including yield and resistance to different stresses.

Bibliography

Ade, J., DeYoung, B. J., Golstein, C., and Innes, R. W. (2007). Indirect activation of a plant nucleotide binding site–leucine-rich repeat protein by a bacterial protease. *Proceedings of the National Academy of Sciences*, 104(7):2531–2536.

Adenot, X., Elmayan, T., Lauressergues, D., Boutet, S., Bouché, N., Gasciolli, V., and Vaucheret, H. (2006). Drb4-dependent tas3 trans-acting sirnas control leaf morphology through ago7. *Current Biology*, 16(9):927–932.

Akhtar, K., Saleem, M., Iqbal, Q., Asghar, M., Hameed, A., and Sarwar, N. (2016). Evaluation of tomato genotypes for late blight resistance using low tunnel assay. *Journal of Plant Pathology*, 98(3).

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

Alonso, J. M., Stepanova, A. N., Leisse, T. J., Kim, C. J., Chen, H., Shinn, P., Stevenson, D. K., Zimmerman, J., Barajas, P., Cheuk, R., et al. (2003). Genome-wide insertional mutagenesis of arabidopsis thaliana. *Science*, 301(5633):653–657.

Anandalakshmi, R., Pruss, G. J., Ge, X., Marathe, R., Mallory, A. C., Smith, T. H., and Vance, V. B. (1998). A viral suppressor of gene silencing in plants. *Proceedings of the National Academy of Sciences*, 95(22):13079–13084.

Andolfo, G., Jupe, F., Witek, K., Etherington, G. J., Ercolano, M. R., and Jones, J. D. (2014). Defining the full tomato nb-lrr resistance gene repertoire using genomic and cdna renseq. *BMC plant biology*, 14(1):120.

Arikit, S., Xia, R., Kakrana, A., Huang, K., Zhai, J., Yan, Z., Valdés-López, O., Prince, S., Musket, T. A., Nguyen, H. T., et al. (2014). An atlas of soybean small rnas identifies phased sirnas from hundreds of coding genes. *The Plant Cell*, 26(12):4584–4601.

Axtell, M. J. (2013). Classification and comparison of small RNAs from plants. *Annual review of plant biology*, 64:137–59.

Axtell, M. J., Jan, C., Rajagopalan, R., and Bartel, D. P. (2006). A two-hit trigger for sirna biogenesis in plants. *Cell*, 127(3):565–577.

Axtell, M. J. and Staskawicz, B. J. (2003). Initiation of rps2-specified disease resistance in arabidopsis is coupled to the avrrpt2-directed elimination of rin4. *Cell*, 112(3):369–377.

Bai, M., Yang, G.-S., Chen, W.-T., Mao, Z.-C., Kang, H.-X., Chen, G.-H., Yang, Y.-H., and Xie, B.-Y. (2012). Genome-wide identification of dicer-like, argonaute and rna-dependent rna polymerase gene families and their expression analyses in response to viral infection and abiotic stresses in solanum lycopersicum. *Gene*, 501(1):52–62.

Baulcombe, D. (2004). Rna silencing in plants. *Nature*, 431(7006):356.

Baumberger, N. and Baulcombe, D. (2005). Arabidopsis argonaute1 is an rna slicer that selectively recruits micrornas and short interfering rnas. *Proceedings of the National Academy of Sciences of the United States of America*, 102(33):11928–11933.

Bendahmane, A., Farnham, G., Moffett, P., and Baulcombe, D. C. (2002). Constitutive gain-of-function mutants in a nucleotide binding site–leucine rich repeat protein encoded at the rx locus of potato. *The Plant Journal*, 32(2):195–204.

Bernoux, M., Ve, T., Williams, S., Warren, C., Hatters, D., Valkov, E., Zhang, X., Ellis, J. G., Kobe, B., and Dodds, P. N. (2011). Structural and functional analysis of a plant resistance protein tir domain reveals interfaces for self-association, signaling, and autoregulation. *Cell host & microbe*, 9(3):200–211.

Bhattacharjee, S., Zamora, A., Azhar, M. T., Sacco, M. A., Lambert, L. H., and Moffett, P. (2009). Virus resistance induced by nb–lrr proteins involves argonaute4-dependent translational control. *The Plant Journal*, 58(6):940–951.

Boccara, M., Sarazin, A., Thiebeauld, O., Jay, F., Voinnet, O., Navarro, L., and Colot, V. (2014). The arabidopsis mir472-rdr6 silencing pathway modulates pamp-and effector-triggered immunity through the post-transcriptional control of disease resistance genes. *PLoS pathogens*, 10(1):e1003883.

Bohmert, K., Camus, I., Bellini, C., Bouchez, D., Caboche, M., and Benning, C. (1998). Ago1 defines a novel locus of arabidopsis controlling leaf development. *The EMBO journal*, 17(1):170–180.

Bolger, A., Scossa, F., Bolger, M. E., Lanz, C., Maumus, F., Tohge, T., Quesneville, H., Alseekh, S., Sørensen, I., Lichtenstein, G., Fich, E. A., Conte, M., Keller, H., Schneeberger, K., Schwacke, R., Ofner, I., Vrebalov, J., Xu, Y., Osorio, S., Aflitos, S. A., Schijlen, E., Jiménez-Goméz, J. M., Ryngajllo, M., Kimura, S., Kumar, R., Koenig, D., Headland, L. R., Maloof, J. N., Sinha, N., van Ham, R. C. H. J., Lankhorst, R. K., Mao, L., Vogel, A., Arsova, B., Panstruga, R., Fei, Z., Rose, J. K. C., Zamir, D., Carrari, F.,

Giovannoni, J. J., Weigel, D., Usadel, B., and Fernie, A. R. (2014). The genome of the stress-tolerant wild tomato species Solanum pennellii. *Nature Genetics*, advance on.

Bologna, N. G., Mateos, J. L., Bresso, E. G., and Palatnik, J. F. (2009). A loop-to-base processing mechanism underlies the biogenesis of plant microrna mir319 and mir159. *The EMBO journal*, 28(23):3646–3656.

Bologna, N. G. and Voinnet, O. (2014). The diversity, biogenesis, and activities of endogenous silencing small rnas in arabidopsis. *Annual review of plant biology*, 65:473–503.

Bonardi, V., Tang, S., Stallmann, A., Roberts, M., Cherkis, K., and Dangl, J. L. (2011). Expanded functions for a family of plant intracellular immune receptors beyond specific recognition of pathogen effectors. *Proceedings of the National Academy of Sciences*, 108(39):16463–16468.

Bond, D. M. and Baulcombe, D. C. (2015). Epigenetic transitions leading to heritable, RNA-mediated de novo silencing in Arabidopsis thaliana. *Proceedings of the National Academy of Sciences*, page 201413053.

Bozkurt, T. O., Schornack, S., Banfield, M. J., and Kamoun, S. (2012). Oomycetes, effectors, and all that jazz. *Current opinion in plant biology*, 15(4):483–492.

Bray, N., Pimentel, H., Melsted, P., and Pachter, L. (2016). Near-optimal probabilistic rna-seq quantification. *Nature biotechnology*, 34(5):525.

Brodersen, P., Sakvarelidze-Achard, L., Bruun-Rasmussen, M., Dunoyer, P., Yamamoto, Y. Y., Sieburth, L., and Voinnet, O. (2008). Widespread translational inhibition by plant mirnas and sirnas. *Science*, 320(5880):1185–1190.

Brooks, C., Nekrasov, V., Lippman, Z. B., and Van Eck, J. (2014). Efficient gene editing in tomato in the first generation using the clustered regularly interspaced short palindromic repeats/crispr-associated9 system. *Plant physiology*, 166(3):1292–1297.

Brukhin, V., Hernould, M., Gonzalez, N., Chevalier, C., and Mouras, A. (2003). Flower development schedule in tomato lycopersicon esculentum cv. sweet cherry. *Sexual Plant Reproduction*, 15(6):311–320.

Burch-Smith, T. M., Schiff, M., Caplan, J. L., Tsao, J., Czymmek, K., and Dinesh-Kumar, S. P. (2007). A novel role for the tir domain in association with pathogen-derived elicitors. *PLoS biology*, 5(3):e68.

Calarco, J. P., Borges, F., Donoghue, M. T. A., Van Ex, F., Jullien, P. E., Lopes, T., Gardner, R., Berger, F., Feijó, J. A., Becker, J. D., and Martienssen, R. A. (2012). Reprogramming of DNA methylation in pollen guides epigenetic inheritance via small RNA. *Cell*, 151(1):194–205.

Campo, S., Peris-Peris, C., Siré, C., Moreno, A. B., Donaire, L., Zytnicki, M., Notredame, C., Llave, C., and San Segundo, B. (2013). Identification of a novel microrna (mirna) from rice that targets an alternatively spliced transcript of the nramp6 (natural resistance-associated macrophage protein 6) gene involved in pathogen resistance. *New Phytologist*, 199(1):212–227.

Cao, D., Wang, J., Ju, Z., Liu, Q., Li, S., Tian, H., Fu, D., Zhu, H., Luo, Y., and Zhu, B. (2016). Regulations on growth and development in tomato cotyledon, flower and fruit via destruction of mir396 with short tandem target mimic. *Plant Science*, 247:1–12.

Carbonell, A. and Carrington, J. C. (2015). Antiviral roles of plant argonautes. *Current opinion in plant biology*, 27:111–117.

Carbonell, A., Fahlgren, N., Garcia-Ruiz, H., Gilbert, K. B., Montgomery, T. A., Nguyen, T., Cuperus, J. T., and Carrington, J. C. (2012). Functional analysis of three arabidopsis argonautes using slicer-defective mutants. *The Plant Cell Online*, 24(9):3613–3629.

Catanzariti, A.-M., Dodds, P. N., Ve, T., Kobe, B., Ellis, J. G., and Staskawicz, B. J. (2010). The avrm effector from flax rust has a structured c-terminal domain and interacts directly with the m resistance protein. *Molecular plant-microbe interactions*, 23(1):49–57.

Cesari, S., Thilliez, G., Ribot, C., Chalvon, V., Michel, C., Jauneau, A., Rivas, S., Alaux, L., Kanzaki, H., Okuyama, Y., et al. (2013). The rice resistance protein pair rga4/rga5 recognizes the magnaporthe oryzae effectors avr-pia and avr1-co39 by direct binding. *The Plant Cell*, 25(4):1463–1481.

Chen, H.-M., Chen, L.-T., Patel, K., Li, Y.-H., Baulcombe, D. C., and Wu, S.-H. (2010a). 22-nucleotide rnas trigger secondary sirna biogenesis in plants. *Proceedings of the National Academy of Sciences*, 107(34):15269–15274.

Chen, Q., Han, Z., Jiang, H., Tian, D., and Yang, S. (2010b). Strong positive selection drives rapid diversification of r-genes in arabidopsis relatives. *Journal of molecular evolution*, 70(2):137–148.

Chen, Y., Liu, Z., and Halterman, D. A. (2012). Molecular determinants of resistance activation and suppression by phytophthora infestans effector ipi-o. *PLoS pathogens*, 8(3):e1002595.

Christian, M., Qi, Y., Zhang, Y., and Voytas, D. F. (2013). Targeted mutagenesis of arabidopsis thaliana using engineered tal effector nucleases. *G3: Genes, Genomes, Genetics*, 3(10):1697–1705.

Collier, S. M., Hamel, L.-P., and Moffett, P. (2011). Cell death mediated by the n-terminal domains of a unique and highly conserved class of nb-lrr protein. *Molecular plant-microbe interactions*, 24(8):918–931.

Couto, D. and Zipfel, C. (2016). Regulation of pattern recognition receptor signalling in plants. *Nature Reviews Immunology*.

Creasey, K. M., Zhai, J., Borges, F., Van Ex, F., Regulski, M., Meyers, B. C., and Martienssen, R. A. (2014). miRNAs trigger widespread epigenetically activated siRNAs from transposons in Arabidopsis. *Nature*.

Cui, H., Tsuda, K., and Parker, J. E. (2015). Effector-triggered immunity: from pathogen perception to robust defense. *Annual review of plant biology*, 66:487–511.

Dagdas, Y. F., Belhaj, K., Maqbool, A., Chaparro-Garcia, A., Pandey, P., Petre, B., Tabassum, N., Cruz-Mireles, N., Hughes, R. K., Sklenar, J., et al. (2016). An effector of the irish potato famine pathogen antagonizes a host autophagy cargo receptor. *Elife*, 5:e10856.

Dai, X. and Zhao, P. X. (2011). psrnatarget: a plant small rna target analysis server. *Nucleic acids research*, 39(suppl_2):W155–W159.

Damodharan, S., Zhao, D., and Arazi, T. (2016). A common mirna160-based mechanism regulates ovary patterning, floral organ abscission and lamina outgrowth in tomato. *The Plant Journal*, 86(6):458–471.

Day, B., Dahlbeck, D., Huang, J., Chisholm, S. T., Li, D., and Staskawicz, B. J. (2005). Molecular basis for the rin4 negative regulation of rps2 disease resistance. *The Plant Cell*, 17(4):1292–1305.

De Felippes, F. F., Schneeberger, K., Dezulian, T., Huson, D. H., and Weigel, D. (2008). Evolution of arabidopsis thaliana microrna from random sequences. *Rna*, 14(12):2455–2459.

de Vries, S., Kloesges, T., and Rose, L. E. (2015). Evolutionarily dynamic, but robust, targeting of resistance genes by the mir482/2118 gene family in the solanaceae. *Genome biology and evolution*, 7(12):3307–3321.

de Vries, S., von Dahlen, J. K., Uhlmann, C., Schnake, A., Kloesges, T., and Rose, L. E. (2017). Signatures of selection and host-adapted gene expression of the phytophthora infestans rna silencing suppressor psr2. *Molecular plant pathology*, 18(1):110–124.

Deleris, A., Gallego-Bartolome, J., Bao, J., Kasschau, K. D., Carrington, J. C., and Voinnet, O. (2006). Hierarchical action and inhibition of plant dicer-like proteins in antiviral defense. *Science*, 313(5783):68–71.

Deng, Y., Zhai, K., Xie, Z., Yang, D., Zhu, X., Liu, J., Wang, X., Qin, P., Yang, Y., Zhang, G., et al. (2017). Epigenetic regulation of antagonistic receptors confers rice blast resistance with yield balance. *Science*, page eaai8898.

Dermitzakis, E. T. and Clark, A. G. (2002). Evolution of transcription factor binding sites in mammalian gene regulatory regions: conservation and turnover. *Molecular biology and evolution*, 19(7):1114–1121.

Dinesh-Kumar, S. and Baker, B. J. (2000). Alternatively spliced n resistance gene transcripts: their possible role in tobacco mosaic virus resistance. *Proceedings of the National Academy of Sciences*, 97(4):1908–1913.

Ding, S.-W. (2010). Rna-based antiviral immunity. *Nature Reviews Immunology*, 10(9):632–644.

Dodds, P. N., Lawrence, G. J., Catanzariti, A.-M., Teh, T., Wang, C.-I., Ayliffe, M. A., Kobe, B., and Ellis, J. G. (2006). Direct protein interaction underlies gene-for-gene specificity and coevolution of the flax resistance genes and flax rust avirulence genes. *Proceedings of the National Academy of Sciences*, 103(23):8888–8893.

Du, J., Verzaux, E., Chaparro-Garcia, A., Bijsterbosch, G., Keizer, L. P., Zhou, J., Liebrand, T. W., Xie, C., Govers, F., Robatzek, S., et al. (2015). Elicitin recognition confers enhanced resistance to phytophthora infestans in potato. *Nature Plants*, 1:15034.

Eck, J. V., Kirk, D. D., and Walmsley, A. M. (2006). Tomato (lycopersicum esculentum). *Agrobacterium Protocols*, pages 459–474.

Elkayam, E., Kuhn, C.-D., Tocilj, A., Haase, A. D., Greene, E. M., Hannon, G. J., and Joshua-Tor, L. (2012). The structure of human argonaute-2 in complex with mir-20a. *Cell*, 150(1):100–110.

Ellendorff, U., Fradin, E. F., De Jonge, R., and Thomma, B. P. (2008). Rna silencing is required for arabidopsis defence against verticillium wilt disease. *Journal of experimental botany*, 60(2):591–602.

Ellis, J. G., Lawrence, G. J., and Dodds, P. N. (2007). Further analysis of gene-for-gene disease resistance specificity in flax. *Molecular plant pathology*, 8(1):103–109.

Fahlgren, N. and Carrington, J. C. (2010). mirna target prediction in plants. *Plant MicroRNAs: Methods and Protocols*, pages 51–57.

Farnham, G. and Baulcombe, D. C. (2006). Artificial evolution extends the spectrum of viruses that are targeted by a disease-resistance gene from potato. *Proceedings of the National Academy of Sciences*, 103(49):18828–18833.

Fei, Q., Li, P., Teng, C., and Meyers, B. C. (2015). Secondary sirnas from medicago nb-lrrs modulated via mirna–target interactions and their abundances. *The Plant Journal*, 83(3):451–465.

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

Fei, Q., Zhang, Y., Xia, R., and Meyers, B. C. (2016). Small rnas add zing to the zig-zag-zig model of plant defenses. *Molecular Plant-Microbe Interactions*, 29(3):165–169.

Feng, Z., Zhang, B., Ding, W., Liu, X., Yang, D.-L., Wei, P., Cao, F., Zhu, S., Zhang, F., Mao, Y., et al. (2013). Efficient genome editing in plants using a crispr/cas system. *Cell research*, 23(10):1229.

Fernandez-Pozo, N., Zheng, Y., Snyder, S. I., Nicolas, P., Shinozaki, Y., Fei, Z., Catala, C., Giovannoni, J. J., Rose, J. K., and Mueller, L. A. (2017). The tomato expression atlas. *Bioinformatics*, page btx190.

Feschotte, C. and Pritham, E. J. (2007). DNA transposons and the evolution of eukaryotic genomes. *Annual review of genetics*, 41:331–68.

Flor, H. (1956). The complementary genic systems in flax and flax rust. *Advances in genetics*, 8:29–54.

Flor, H. H. (1971). Current status of the gene-for-gene concept. *Annual review of phytopathology*, 9(1):275–296.

Franco-Zorrilla, J. M., Valli, A., Todesco, M., Mateos, I., Puga, M. I., Rubio-Somoza, I., Leyva, A., Weigel, D., García, J. A., and Paz-Ares, J. (2007). Target mimicry provides a new mechanism for regulation of microRNA activity. *Nature genetics*, 39(8):1033–7.

Fukuoka, S., Saka, N., Koga, H., Ono, K., Shimizu, T., Ebana, K., Hayashi, N., Takahashi, A., Hirochika, H., Okuno, K., et al. (2009). Loss of function of a proline-containing protein confers durable disease resistance in rice. *Science*, 325(5943):998–1001.

Gabriëls, S. H., Vossen, J. H., Ekengren, S. K., Ooijen, G. v., Abd-El-Haliem, A. M., Berg, G., Rainey, D. Y., Martin, G. B., Takken, F. L., de Wit, P. J., et al. (2007). An nb-lrr protein required for hr signalling mediated by both extra-and intracellular resistance proteins. *The Plant Journal*, 50(1):14–28.

Gao, Z., Liu, H.-L., Daxinger, L., Pontes, O., He, X., Qian, W., Lin, H., Xie, M., Lorkovic, Z. J., Zhang, S., et al. (2010). An rna polymerase ii-and ago4-associated protein acts in rna-directed dna methylation. *Nature*, 465(7294):106–109.

Garcia-Ruiz, H., Takeda, A., Chapman, E. J., Sullivan, C. M., Fahlgren, N., Brempelis, K. J., and Carrington, J. C. (2010). Arabidopsis rna-dependent rna polymerases and dicer-like proteins in antiviral defense and small interfering rna biogenesis during turnip mosaic virus infection. *The Plant Cell*, 22(2):481–496.

Gasciolli, V., Mallory, A. C., Bartel, D. P., and Vaucheret, H. (2005). Partially redundant functions of arabidopsis dicer-like enzymes and a role for dcl4 in producing trans-acting sirnas. *Current Biology*, 15(16):1494–1500.

Giannakopoulou, A., Steele, J. F., Segretin, M. E., Bozkurt, T. O., Zhou, J., Robatzek, S., Banfield, M. J., Pais, M., and Kamoun, S. (2015). Tomato i2 immune receptor can be engineered to confer partial resistance to the oomycete phytophthora infestans in addition to the fungus fusarium oxysporum. *Molecular Plant-Microbe Interactions*, 28(12):1316–1329.

Gloggnitzer, J., Akimcheva, S., Srinivasan, A., Kusenda, B., Riehs, N., Stampfl, H., Bautor, J., Dekrout, B., Jonak, C., Jiménez-Gómez, J. M., et al. (2014). Nonsense-mediated mrna decay modulates immune receptor levels to regulate plant antibacterial defense. *Cell host & microbe*, 16(3):376–390.

González, V. M., Müller, S., Baulcombe, D., and Puigdomènech, P. (2015). Evolution of nbs-lrr gene copies among dicot plants and its regulation by members of the mir482/2118 superfamily of mirnas. *Molecular plant*, 8(2):329–331.

Gouil, Q. (2016). *Epigenetic in tomato: Hybrids and paramutation*. PhD thesis, Department of Plant Sciences; University of Cambrdige.

Gu, W., Wang, X., Zhai, C., Xie, X., and Zhou, T. (2012). Selection on synonymous sites for increased accessibility around mirna binding sites in plants. *Molecular biology and evolution*, 29(10):3037–3044.

Gutierrez, J. R., Balmuth, A. L., Ntoukakis, V., Mucyn, T. S., Gimenez-Ibanez, S., Jones, A. M., and Rathjen, J. P. (2010). Prf immune complexes of tomato are oligomeric and contain multiple pto-like kinases that diversify effector recognition. *The Plant Journal*, 61(3):507–518.

Halter, T. and Navarro, L. (2015). Multilayer and interconnected post-transcriptional and co-transcriptional control of plant nlrs. *Current opinion in plant biology*, 26:127–134.

Hardcastle, T. J. and Kelly, K. A. (2013). Empirical Bayesian analysis of paired high-throughput sequencing data with a beta-binomial distribution. *BMC bioinformatics*, 14(1):135.

Hardcastle, T. J., Kelly, K. A., and Baulcombe, D. C. (2012). Identifying small interfering RNA loci from high-throughput sequencing data. *Bioinformatics (Oxford, England)*, 28(4):457–63.

Harris, C. J., Molnar, A., Müller, S. Y., and Baulcombe, D. C. (2015). Nar breakthrough article: Fdf-page: a powerful technique revealing previously undetected small rnas sequestered by complementary transcripts. *Nucleic Acids Research*, 43(15):7590.

Harris, C. J., Slootweg, E. J., Goverse, A., and Baulcombe, D. C. (2013). Stepwise artificial evolution of a plant disease resistance gene. *Proceedings of the National Academy of Sciences*, 110(52):21189–21194.

Havecker, E. R., Wallbridge, L. M., Hardcastle, T. J., Bush, M. S., Kelly, K. A., Dunn, R. M., Schwach, F., Doonan, J. H., and Baulcombe, D. C. (2010). The Arabidopsis RNA-directed DNA methylation argonautes functionally diverge based on their expression and interaction with target loci. *The Plant cell*, 22(2):321–34.

Heidel, A. J., Clarke, J. D., Antonovics, J., and Dong, X. (2004). Fitness costs of mutations affecting the systemic acquired resistance pathway in arabidopsis thaliana. *Genetics*, 168(4):2197–2206.

Henderson, I. R., Zhang, X., Lu, C., Johnson, L., Meyers, B. C., Green, P. J., and Jacobsen, S. E. (2006). Dissecting arabidopsis thaliana dicer function in small rna processing, gene silencing and dna methylation patterning. *Nature genetics*, 38(6):721–725.

Hollick, J. B. (2012). Paramutation: a trans-homolog interaction affecting heritable gene regulation. *Current opinion in plant biology*, 15(5):536–43.

Hollister, J. D., Smith, L. M., Guo, Y.-L., Ott, F., Weigel, D., and Gaut, B. S. (2011). Transposable elements and small RNAs contribute to gene expression divergence between Arabidopsis thaliana and Arabidopsis lyrata. *Proceedings of the National Academy of Sciences of the United States of America*, 108(6):2322–7.

Howell, M. D., Fahlgren, N., Chapman, E. J., Cumbie, J. S., Sullivan, C. M., Givan, S. A., Kasschau, K. D., and Carrington, J. C. (2007). Genome-wide analysis of the rna-dependent rna polymerase6/dicer-like4 pathway in arabidopsis reveals dependency on mirna-and tasirna-directed targeting. *The Plant Cell*, 19(3):926–942.

Howles, P., Lawrence, G., Finnegan, J., McFadden, H., Ayliffe, M., Dodds, P., and Ellis, J. (2005). Autoactive alleles of the flax l6 rust resistance gene induce non-race-specific rust resistance associated with the hypersensitive response. *Molecular Plant-Microbe Interactions*, 18(6):570–582.

Hsieh, L.-C., Lin, S.-I., Shih, A. C.-C., Chen, J.-W., Lin, W.-Y., Tseng, C.-Y., Li, W.-H., and Chiou, T.-J. (2009). Uncovering small rna-mediated responses to phosphate deficiency in arabidopsis by deep sequencing. *Plant physiology*, 151(4):2120–2132.

Huot, B., Yao, J., Montgomery, B. L., and He, S. Y. (2014). Growth–defense tradeoffs in plants: a balancing act to optimize fitness. *Molecular plant*, 7(8):1267–1287.

Hwang, E.-W., Shin, S.-J., Yu, B.-K., Byun, M.-O., and Kwon, H.-B. (2011). mir171 family members are involved in drought response in solanum tuberosum. *Journal of Plant Biology*, 54(1):43–48.

Ibarra, C. A., Feng, X., Schoft, V. K., Hsieh, T.-F., Uzawa, R., Rodrigues, J. A., Zemach, A., Chumak, N., Machlicova, A., Nishimura, T., Rojas, D., Fischer, R. L., Tamaru, H., and Zilberman, D. (2012). Active DNA demethylation in plant companion cells reinforces transposon methylation in gametes. *Science (New York, N.Y.)*, 337(6100):1360–4.

Ito, H., Gaubert, H., Bucher, E., Mirouze, M., Vaillant, I., and Paszkowski, J. (2011). An siRNA pathway prevents transgenerational retrotransposition in plants subjected to stress. *Nature*, 472(7341):115–9.

Iwakawa, H.-o. and Tomari, Y. (2013). Molecular insights into microrna-mediated translational repression in plants. *Molecular cell*, 52(4):591–601.

Jacob, F., Vernaldi, S., and Maekawa, T. (2013). Evolution and conservation of plant nlr functions. *Frontiers in immunology*, 4.

Jaubert, M., Bhattacharjee, S., Mello, A. F., Perry, K. L., and Moffett, P. (2011). Argonaute2 mediates rna-silencing antiviral defenses against potato virus x in arabidopsis. *Plant physiology*, 156(3):1556–1564.

Jia, X., Shen, J., Liu, H., Li, F., Ding, N., Gao, C., Pattanaik, S., Patra, B., Li, R., and Yuan, L. (2015). Small tandem target mimic-mediated blockage of microrna858 induces anthocyanin accumulation in tomato. *Planta*, 242(1):283–293.

Jia, Y., McAdams, S. A., Bryan, G. T., Hershey, H. P., and Valent, B. (2000). Direct interaction of resistance gene and avirulence gene products confers rice blast resistance. *The EMBO journal*, 19(15):4004–4014.

Johnson, C., Kasprzewska, A., Tennessen, K., Fernandes, J., Nan, G.-L., Walbot, V., Sundaresan, V., Vance, V., and Bowman, L. H. (2009). Clusters and superclusters of phased small rnas in the developing inflorescence of rice. *Genome research*, 19(8):1429–1440.

Jones, J. D. and Dangl, J. L. (2006). The plant immune system. *Nature*, 444(7117):323.

Jones, J. D., Vance, R. E., and Dangl, J. L. (2016). Intracellular innate immune surveillance devices in plants and animals. *Science*, 354(6316):aaf6395.

Jouannet, V., Moreno, A. B., Elmayan, T., Vaucheret, H., Crespi, M. D., and Maizel, A. (2012). Cytoplasmic arabidopsis ago7 accumulates in membrane-associated sirna bodies and is required for ta-sirna biogenesis. *The EMBO journal*, 31(7):1704–1713.

Kakrana, A., Hammond, R., Patel, P., Nakano, M., and Meyers, B. C. (2014). sparta: a parallelized pipeline for integrated analysis of plant mirna and cleaved mrna data sets, including new mirna target-identification software. *Nucleic acids research*, 42(18):e139–e139.

Kakrana, A., Li, P., Patel, P., Hammond, R., Anand, D., Mathioni, S., and Meyers, B. (2017). Phasis: A computational suite for de novo discovery and characterization of phased, sirna-generating loci and their mirna triggers. *bioRxiv*, page 158832.

Kamoun, S. (2001). Nonhost resistance to phytophthora: novel prospects for a classical problem. *Current opinion in plant biology*, 4(4):295–300.

Kanzaki, H., Yoshida, K., Saitoh, H., Fujisaki, K., Hirabuchi, A., Alaux, L., Fournier, E., Tharreau, D., and Terauchi, R. (2012). Arms race co-evolution of magnaporthe oryzae avr-pik and rice pik genes driven by their physical interactions. *The Plant Journal*, 72(6):894–907.

Karasov, T. L., Chae, E., Herman, J. J., and Bergelson, J. (2017). Mechanisms to mitigate the trade-off between growth and defense. *The Plant Cell*, 29(4):666–680.

Karlova, R., van Haarst, J. C., Maliepaard, C., van de Geest, H., Bovy, A. G., Lammers, M., Angenent, G. C., and de Maagd, R. A. (2013). Identification of microrna targets in tomato fruit development using high-throughput sequencing and degradome analysis. *Journal of experimental botany*, 64(7):1863–1878.

Kasschau, K. D., Fahlgren, N., Chapman, E. J., Sullivan, C. M., Cumbie, J. S., Givan, S. A., and Carrington, J. C. (2007). Genome-wide profiling and analysis of arabidopsis sirnas. *PLoS biology*, 5(3):e57.

Katagiri, F., Thilmony, R., and He, S. Y. (2002). The arabidopsis thaliana-pseudomonas syringae interaction. *The Arabidopsis Book*, page e0039.

Katiyar-Agarwal, S., Morgan, R., Dahlbeck, D., Borsani, O., Villegas, A., Zhu, J.-K., Staskawicz, B. J., and Jin, H. (2006). A pathogen-inducible endogenous sirna in plant immunity. *Proceedings of the National Academy of Sciences*, 103(47):18002–18007.

Kawakatsu, T., Stuart, T., Valdes, M., Breakfield, N., Schmitz, R. J., Nery, J. R., Urich, M. A., Han, X., Lister, R., Benfey, P. N., et al. (2016). Unique cell-type-specific patterns of dna methylation in the root meristem. *Nature plants*, 2:16058.

Kawashima, T. and Berger, F. (2014). Epigenetic reprogramming in plant sexual reproduction. *Nature Reviews Genetics*, 15(9):613–624.

Kempel, A., Schädler, M., Chrobock, T., Fischer, M., and van Kleunen, M. (2011). Tradeoffs associated with constitutive and induced plant resistance against herbivory. *Proceedings of the National Academy of Sciences*, 108(14):5685–5689.

Kobe, B. and Kajava, A. V. (2001). The leucine-rich repeat as a protein recognition motif. *Current opinion in structural biology*, 11(6):725–732.

Komiya, R., Ohyanagi, H., Niihama, M., Watanabe, T., Nakano, M., Kurata, N., and Nonomura, K.-I. (2014). Rice Germline-specific Argonaute MEL1 protein binds to phasiRNAs generated from more than 700 lincRNAs. *The Plant journal : for cell and molecular biology*.

Kong, L., Zhang, Y., Ye, Z.-Q., Liu, X.-Q., Zhao, S.-Q., Wei, L., and Gao, G. (2007). Cpc: assess the protein-coding potential of transcripts using sequence features and support vector machine. *Nucleic acids research*, 35(suppl_2):W345–W349.

Kozomara, A. and Griffiths-Jones, S. (2013). mirbase: annotating high confidence micrornas using deep sequencing data. *Nucleic acids research*, 42(D1):D68–D73.

Krasileva, K. V., Dahlbeck, D., and Staskawicz, B. J. (2010). Activation of an arabidopsis resistance protein is specified by the in planta association of its leucine-rich repeat domain with the cognate oomycete effector. *The Plant Cell*, 22(7):2444–2458.

Kravchik, M., Sunkar, R., Damodharan, S., Stav, R., Zohar, M., Isaacson, T., and Arazi, T. (2013). Global and local perturbation of the tomato microrna pathway by a trans-activated dicer-like 1 mutant. *Journal of experimental botany*, 65(2):725–739.

Kuang, H., Woo, S.-S., Meyers, B. C., Nevo, E., and Michelmore, R. W. (2004). Multiple genetic processes result in heterogeneous rates of evolution within the major cluster disease resistance genes in lettuce. *The Plant Cell*, 16(11):2870–2894.

Kumakura, N., Takeda, A., Fujioka, Y., Motose, H., Takano, R., and Watanabe, Y. (2009). Sgs3 and rdr6 interact and colocalize in cytoplasmic sgs3/rdr6-bodies. *FEBS letters*, 583(8):1261–1266.

Langmead, B. and Salzberg, S. L. (2012). Fast gapped-read alignment with bowtie 2. *Nature methods*, 9(4):357–359.

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

Le Roux, C., Huet, G., Jauneau, A., Camborde, L., Trémousaygue, D., Kraut, A., Zhou, B., Levaillant, M., Adachi, H., Yoshioka, H., et al. (2015). A receptor pair with an integrated decoy converts pathogen disabling of transcription factors to immunity. *Cell*, 161(5):1074–1088.

Leipe, D. D., Koonin, E. V., and Aravind, L. (2004). Stand, a class of p-loop ntpases including animal and plant regulators of programmed cell death: multiple, complex domain architectures, unusual phyletic patterns, and evolution by horizontal gene transfer. *Journal of molecular biology*, 343(1):1–28.

Lelandais-Brière, C., Naya, L., Sallet, E., Calenge, F., Frugier, F., Hartmann, C., Gouzy, J., and Crespi, M. (2009). Genome-wide medicago truncatula small rna analysis revealed novel micrornas and isoforms differentially regulated in roots and nodules. *The Plant Cell*, 21(9):2780–2796.

Leung, A. K. (2015). The whereabouts of microrna actions: cytoplasm and beyond. *Trends in cell biology*, 25(10):601–610.

Li, F., Orban, R., and Baker, B. (2012a). Somart: a web server for plant mirna, tasirna and target gene analysis. *The Plant Journal*, 70(5):891–901.

Li, F., Pignatta, D., Bendix, C., Brunkard, J. O., Cohn, M. M., Tung, J., Sun, H., Kumar, P., and Baker, B. (2012b). Microrna regulation of plant innate immune receptors. *Proceedings of the National Academy of Sciences*, 109(5):1790–1795.

Li, H., Freeling, M., and Lisch, D. (2010). Epigenetic reprogramming during vegetative phase change in maize. *Proceedings of the National Academy of Sciences*, 107(51):22184–22189.

Li, S., Le, B., Ma, X., Li, S., You, C., Yu, Y., Zhang, B., Liu, L., Gao, L., Shi, T., et al. (2016). Biogenesis of phased sirnas on membrane-bound polysomes in arabidopsis. *Elife*, 5:e22750.

Li, S., Liu, L., Zhuang, X., Yu, Y., Liu, X., Cui, X., Ji, L., Pan, Z., Cao, X., Mo, B., et al. (2013a). Micrornas inhibit the translation of target mrnas on the endoplasmic reticulum in arabidopsis. *Cell*, 153(3):562–574.

Li, X., Clarke, J. D., Zhang, Y., and Dong, X. (2001). Activation of an eds1-mediated r-gene pathway in the snc1 mutant leads to constitutive, npr1-independent pathogen resistance. *Molecular Plant-Microbe Interactions*, 14(10):1131–1139.

Li, Y., Li, C., Xia, J., and Jin, Y. (2011). Domestication of transposable elements into MicroRNA genes in plants. *PloS one*, 6(5):e19212.

Li, Y., Lu, J., Han, Y., Fan, X., and Ding, S.-W. (2013b). Rna interference functions as an antiviral immunity mechanism in mammals. *Science*, 342(6155):231–234.

Li, Y., Lu, Y.-G., Shi, Y., Wu, L., Xu, Y.-J., Huang, F., Guo, X.-Y., Zhang, Y., Fan, J., Zhao, J.-Q., et al. (2014). Multiple rice micrornas are involved in immunity against the blast fungus magnaporthe oryzae. *Plant physiology*, 164(2):1077–1092.

Lin, D., Yang, Y., Khalil, R., Xian, Z., Hu, G., and Li, Z. (2013). Slmir393 controls the auxin receptor homologous genes expression, and regulates sensitivity to auxin in tomato root growth. *Scientia horticulturae*, 162:90–99.

Lin, R., Ding, L., Casola, C., Ripoll, D. R., Feschotte, C., and Wang, H. (2007). Transposase-derived transcription factors regulate light signaling in Arabidopsis. *Science (New York, N.Y.)*, 318(5854):1302–5.

Lisch, D. (2009). Epigenetic regulation of transposable elements in plants. *Annual review of plant biology*, 60:43–66.

Lisch, D. (2013). How important are transposons for plant evolution? *Nature reviews. Genetics*, 14(1):49–61.

Liu, C., Axtell, M. J., and Fedoroff, N. V. (2012). The helicase and rnaseiiia domains of arabidopsis dicer-like1 modulate catalytic parameters during microrna biogenesis. *Plant physiology*, 159(2):748–758.

Liu, J., Cheng, X., Liu, D., Xu, W., Wise, R., and Shen, Q.-H. (2014). The mir9863 family regulates distinct mla alleles in barley to attenuate nlr receptor-triggered disease resistance and cell-death signaling. *PLoS genetics*, 10(12):e1004755.

Liu, R., How-Kit, A., Stammitti, L., Teyssier, E., Rolin, D., Mortain-Bertrand, A., Halle, S., Liu, M., Kong, J., Wu, C., et al. (2015a). A demeter-like dna demethylase governs tomato fruit ripening. *Proceedings of the National Academy of Sciences*, 112(34):10804–10809.

Liu, X., Sun, Y., Kørner, C. J., Du, X., Vollmer, M. E., and Pajerowska-Mukhtar, K. M. (2015b). Bacterial leaf infiltration assay for fine characterization of plant defense responses using the arabidopsis thaliana-pseudomonas syringae pathosystem. *JoVE (Journal of Visualized Experiments)*, (104):e53364–e53364.

Llave, C., Xie, Z., Kasschau, K. D., and Carrington, J. C. (2002). Cleavage of scarecrow-like mrna targets directed by a class of arabidopsis mirna. *Science*, 297(5589):2053–2056.

Lopez-Gomollon, S., Mohorianu, I., Szittya, G., Moulton, V., and Dalmay, T. (2012). Diverse correlation patterns between micrornas and their targets during tomato fruit development indicates different modes of microrna actions. *Planta*, 236(6):1875–1887.

Mackey, D., Holt, B. F., Wiig, A., and Dangl, J. L. (2002). Rin4 interacts with pseudomonas syringae type iii effector molecules and is required for rpm1-mediated resistance in arabidopsis. *Cell*, 108(6):743–754.

Maekawa, T., Cheng, W., Spiridon, L. N., Töller, A., Lukasik, E., Saijo, Y., Liu, P., Shen, Q.-H., Micluta, M. A., Somssich, I. E., et al. (2011). Coiled-coil domain-dependent homodimerization of intracellular barley immune receptors defines a minimal functional module for triggering cell death. *Cell host & microbe*, 9(3):187–199.

Makarevitch, I., Waters, A. J., West, P. T., Stitzer, M., Hirsch, C. N., Ross-Ibarra, J., and Springer, N. M. (2015). Transposable elements contribute to activation of maize genes in response to abiotic stress. *PLoS genetics*, 11(1):e1004915.

Malone, C. D., Brennecke, J., Dus, M., Stark, A., McCombie, W. R., Sachidanandam, R., and Hannon, G. J. (2009). Specialized piRNA pathways act in germline and somatic tissues of the Drosophila ovary. *Cell*, 137(3):522–35.

Manavella, P. A., Koenig, D., and Weigel, D. (2012). Plant secondary sirna production determined by microrna-duplex structure. *Proceedings of the National Academy of Sciences*, 109(7):2461–2466.

Margis, R., Fusaro, A. F., Smith, N. A., Curtin, S. J., Watson, J. M., Finnegan, E. J., and Waterhouse, P. M. (2006). The evolution and diversification of dicers in plants. *FEBS letters*, 580(10):2442–2450.

Marí-Ordóñez, A., Marchais, A., Etcheverry, M., Martin, A., Colot, V., and Voinnet, O. (2013). Reconstructing de novo silencing of an active plant retrotransposon. *Nature genetics*, 45(9):1029–39.

Martinez, G., Choudury, S. G., and Slotkin, R. K. (2017). trna-derived small rnas target transposable element transcripts. *Nucleic Acids Research*, 45(9):5142–5152.

Martínez, G., Panda, K., Köhler, C., and Slotkin, R. K. (2016). Silencing in sperm cells is directed by rna movement from the surrounding nurse cell. *Nature plants*, 2:16030.

Martínez, G. and Slotkin, R. K. (2012). Developmental relaxation of transposable element silencing in plants: functional or byproduct? *Current opinion in plant biology*, 15(5):496–502.

Matas, A. J., Yeats, T. H., Buda, G. J., Zheng, Y., Chatterjee, S., Tohge, T., Ponnala, L., Adato, A., Aharoni, A., Stark, R., et al. (2011). Tissue-and cell-type specific transcriptome profiling of expanding tomato fruit provides insights into metabolic and regulatory specialization and cuticle formation. *The Plant Cell*, 23(11):3893–3910.

Matzke, M. A. and Mosher, R. A. (2014). RNA-directed DNA methylation: an epigenetic pathway of increasing complexity. *Nature reviews. Genetics*, 15(6):394–408.

McCarthy, D. J., Chen, Y., and Smyth, G. K. (2012). Differential expression analysis of multifactor rna-seq experiments with respect to biological variation. *Nucleic acids research*, 40(10):4288–4297.

McClintock, B. (1953). Induction of instability at selected loci in maize. *Genetics*, 38(6):579.

McCue, A. D., Nuthikattu, S., Reeder, S. H., and Slotkin, R. K. (2012). Gene expression and stress response mediated by the epigenetic regulation of a transposable element small RNA. *PLoS genetics*, 8(2):e1002474.

McCue, A. D., Panda, K., Nuthikattu, S., Choudury, S. G., Thomas, E. N., and Slotkin, R. K. (2014). Argonaute 6 bridges transposable element mrna-derived sirnas to the establishment of dna methylation. *The EMBO journal*, page e201489499.

Menda, N., Semel, Y., Peled, D., Eshed, Y., and Zamir, D. (2004). In silico screening of a saturated mutation library of tomato. *The Plant Journal*, 38(5):861–872.

Meyers, B. C., Kozik, A., Griego, A., Kuang, H., and Michelmore, R. W. (2003). Genome-wide analysis of nbs-lrr–encoding genes in arabidopsis. *The Plant Cell*, 15(4):809–834.

Mi, S., Cai, T., Hu, Y., Chen, Y., Hodges, E., Ni, F., Wu, L., Li, S., Zhou, H., Long, C., et al. (2008). Sorting of small rnas into arabidopsis argonaute complexes is directed by the 5' terminal nucleotide. *Cell*, 133(1):116–127.

Mirouze, M., Reinders, J., Bucher, E., Nishimura, T., Schneeberger, K., Ossowski, S., Cao, J., Weigel, D., Paszkowski, J., and Mathieu, O. (2009). Selective epigenetic control of retrotransposition in Arabidopsis. *Nature*, 461(7262):427–30.

Moore, M. J., Soltis, P. S., Bell, C. D., Burleigh, J. G., and Soltis, D. E. (2010). Phylogenetic analysis of 83 plastid genes further resolves the early diversification of eudicots. *Proceedings of the National Academy of Sciences*, 107(10):4623–4628.

Mucyn, T. S., Clemente, A., Andriotis, V. M., Balmuth, A. L., Oldroyd, G. E., Staskawicz, B. J., and Rathjen, J. P. (2006). The tomato nbarc-lrr protein prf interacts with pto kinase in vivo to regulate specific plant immunity. *The Plant Cell*, 18(10):2792–2806.

Mukhtar, M. S., Carvunis, A.-R., Dreze, M., Epple, P., Steinbrenner, J., Moore, J., Tasan, M., Galli, M., Hao, T., Nishimura, M. T., et al. (2011). Independently evolved virulence effectors converge onto hubs in a plant immune system network. *science*, 333(6042):596–601.

Naito, K., Zhang, F., Tsukiyama, T., Saito, H., Hancock, C. N., Richardson, A. O., Okumoto, Y., Tanisaka, T., and Wessler, S. R. (2009). Unexpected consequences of a sudden and massive transposon amplification on rice gene expression. *Nature*, 461(7267):1130.

Nakanishi, K., Weinberg, D. E., Bartel, D. P., and Patel, D. J. (2012). Structure of yeast argonaute with guide rna. *Nature*, 486(7403):368–374.

Narusaka, M., Shirasu, K., Noutoshi, Y., Kubo, Y., Shiraishi, T., Iwabuchi, M., and Narusaka, Y. (2009). Rrs1 and rps4 provide a dual resistance-gene system against fungal and bacterial pathogens. *The Plant Journal*, 60(2):218–226.

Navarro, L., Dunoyer, P., Jay, F., Arnold, B., Dharmasiri, N., Estelle, M., Voinnet, O., and Jones, J. D. (2006). A plant mirna contributes to antibacterial resistance by repressing auxin signaling. *Science*, 312(5772):436–439.

Navarro, L., Jay, F., Nomura, K., He, S. Y., and Voinnet, O. (2008). Suppression of the microrna pathway by bacterial effector proteins. *Science*, 321(5891):964–967.

Nekrasov, V., Staskawicz, B., Weigel, D., Jones, J. D., and Kamoun, S. (2013). Targeted mutagenesis in the model plant nicotiana benthamiana using cas9 rna-guided endonuclease. *Nature biotechnology*, 31(8):691–693.

Niu, D., Lii, Y. E., Chellappan, P., Lei, L., Peralta, K., Jiang, C., Guo, J., Coaker, G., and Jin, H. (2016). mirna863-3p sequentially targets negative immune regulator arlpks and positive regulator serrate upon bacterial infection. *Nature communications*, 7.

Nonomura, K.-I., Morohoshi, A., Nakano, M., Eiguchi, M., Miyao, A., Hirochika, H., and Kurata, N. (2007). A germ cell–specific gene of the argonaute family is essential for the progression of premeiotic mitosis and meiosis during sporogenesis in rice. *The Plant Cell*, 19(8):2583–2594.

Nuthikattu, S., McCue, A. D., Panda, K., Fultz, D., DeFraia, C., Thomas, E. N., and Slotkin, R. K. (2013). The initiation of epigenetic silencing of active transposable elements is triggered by RDR6 and 21-22 nucleotide small interfering RNAs. *Plant physiology*, 162(1):116–31.

Ohtsu, K., Smith, M. B., Emrich, S. J., Borsuk, L. A., Zhou, R., Chen, T., Zhang, X., Timmermans, M. C., Beck, J., Buckner, B., et al. (2007a). Global gene expression analysis of the shoot apical meristem of maize (zea mays l.). *The Plant Journal*, 52(3):391–404.

Ohtsu, K., Smith, M. B., Emrich, S. J., Borsuk, L. A., Zhou, R., Chen, T., Zhang, X., Timmermans, M. C. P., Beck, J., Buckner, B., Janick-Buckner, D., Nettleton, D., Scanlon, M. J., and Schnable, P. S. (2007b). Global gene expression analysis of the shoot apical meristem of maize (Zea mays L.). *The Plant journal : for cell and molecular biology*, 52(3):391–404.

Oldroyd, G. E. and Staskawicz, B. J. (1998). Genetically engineered broad-spectrum disease resistance in tomato. *Proceedings of the National Academy of Sciences*, 95(17):10300–10305.

Oliver, C., Santos, J. L., and Pradillo, M. (2014). On the role of some argonaute proteins in meiosis and dna repair in arabidopsis thaliana. *Frontiers in plant science*, 5.

Olmedo-Monfil, V., Durán-Figueroa, N., Arteaga-Vázquez, M., Demesa-Arévalo, E., Autran, D., Grimanelli, D., Slotkin, R. K., Martienssen, R. A., and Vielle-Calzada, J.-P. (2010). Control of female gamete formation by a small rna pathway in arabidopsis. *Nature*, 464(7288):628–632.

Ouyang, S., Park, G., Atamian, H. S., Han, C. S., Stajich, J. E., Kaloshian, I., and Borkovich, K. A. (2014). Micrornas suppress nb domain genes in tomato that confer resistance to fusarium oxysporum. *PLoS pathogens*, 10(10):e1004464.

Panda, K. and Slotkin, R. K. (2013). Proposed mechanism for the initiation of transposable element silencing by the RDR6-directed DNA methylation pathway. *Plant signaling & behavior*, 8(8).

Parent, J.-S., Bouteiller, N., Elmayan, T., and Vaucheret, H. (2015). Respective contributions of arabidopsis dcl2 and dcl4 to rna silencing. *The Plant Journal*, 81(2):223–232.

Park, S. J., Jiang, K., Schatz, M. C., and Lippman, Z. B. (2012). Rate of meristem maturation determines inflorescence architecture in tomato. *Proceedings of the National Academy of Sciences of the United States of America*, 109(2):639–44.

Peart, J. R., Mestre, P., Lu, R., Malcuit, I., and Baulcombe, D. C. (2005). Nrg1, a cc-nb-lrr protein, together with n, a tir-nb-lrr protein, mediates resistance against tobacco mosaic virus. *Current biology*, 15(10):968–973.

Peterson, R., Slovin, J. P., and Chen, C. (2010). A simplified method for differential staining of aborted and non-aborted pollen grains. *Int J Plant Biol*, 1:e13.

Pignatta, D., Erdmann, R. M., Scheer, E., Picard, C. L., Bell, G. W., and Gehring, M. (2014). Natural epigenetic polymorphisms lead to intraspecific variation in Arabidopsis gene imprinting. *eLife*, 3:e03198.

Pikaard, C. S., Haag, J. R., Pontes, O. M. F., Blevins, T., and Cocklin, R. (2012). A transcription fork model for Pol IV and Pol V-dependent RNA-directed DNA methylation. *Cold Spring Harbor symposia on quantitative biology*, 77(0):205–12.

Piriyapongsa, J. and Jordan, I. K. (2008). Dual coding of siRNAs and miRNAs by plant transposable elements. *RNA (New York, N.Y.)*, 14(5):814–21.

Pontes, O., Li, C. F., Nunes, P. C., Haag, J., Ream, T., Vitins, A., Jacobsen, S. E., and Pikaard, C. S. (2006). The arabidopsis chromatin-modifying nuclear sirna pathway involves a nucleolar rna processing center. *Cell*, 126(1):79–92.

Pumplin, N., Sarazin, A., Jullien, P. E., Bologna, N. G., Oberlin, S., and Voinnet, O. (2016). Dna methylation influences the expression of dicer-like4 isoforms, which encode proteins of alternative localization and function. *The Plant Cell Online*, 28(11):2786–2804.

Pumplin, N. and Voinnet, O. (2013). Rna silencing suppression by plant pathogens: defence, counter-defence and counter-counter-defence. *Nature Reviews. Microbiology*, 11(11):745.

Qi, D., DeYoung, B. J., and Innes, R. W. (2012). Structure-function analysis of the coiled-coil and leucine-rich repeat domains of the rps5 disease resistance protein. *Plant physiology*, 158(4):1819–1832.

Qian, L.-H., Zhou, G.-C., Sun, X.-Q., Lei, Z., Zhang, Y.-M., Xue, J.-Y., and Hang, Y.-Y. (2017). Distinct patterns of gene gain and loss: Diverse evolutionary modes of nbs-encoding genes in three solanaceae crop species. *G3: Genes, Genomes, Genetics*, 7(5):1577–1585.

Qiao, Y., Liu, L., Xiong, Q., Flores, C., Wong, J., Shi, J., Wang, X., Liu, X., Xiang, Q., Jiang, S., et al. (2013). Oomycete pathogens encode rna silencing suppressors. *Nature genetics*, 45(3):330–333.

Qu, F., Ye, X., and Morris, T. J. (2008). Arabidopsis drb4, ago1, ago7, and rdr6 participate in a dcl4-initiated antiviral rna silencing pathway negatively regulated by dcl1. *Proceedings of the National Academy of Sciences*, 105(38):14732–14737.

Rairdan, G. J., Collier, S. M., Sacco, M. A., Baldwin, T. T., Boettrich, T., and Moffett, P. (2008). The coiled-coil and nucleotide binding domains of the potato rx disease resistance protein function in pathogen recognition and signaling. *The Plant Cell*, 20(3):739–751.

Rairdan, G. J. and Moffett, P. (2006). Distinct domains in the arc region of the potato resistance protein rx mediate lrr binding and inhibition of activation. *The Plant Cell*, 18(8):2082–2093.

Ramirez, C. L., Foley, J. E., Wright, D. A., Müller-Lerch, F., Rahman, S. H., Cornu, T. I., Winfrey, R. J., Sander, J. D., Fu, F., Townsend, J. A., et al. (2008). Unexpected failure rates for modular assembly of engineered zinc fingers. *Nature methods*, 5(5):374–375.

Ravensdale, M., Bernoux, M., Ve, T., Kobe, B., Thrall, P. H., Ellis, J. G., and Dodds, P. N. (2012). Intramolecular interaction influences binding of the flax l5 and l6 resistance proteins to their avrl567 ligands. *PLoS pathogens*, 8(11):e1003004.

Richard, J. P., Csukasi, F., Zheng, Y., Fei, Z., van der Knaap, E., and Catalá, C. (2015). Comprehensive tissue-specific transcriptome analysis reveals distinct regulatory programs during early tomato fruit development. *Plant physiology*, pages pp–00287.

Riehs-Kearnan, N., Gloggnitzer, J., Dekrout, B., Jonak, C., and Riha, K. (2012). Aberrant growth and lethality of arabidopsis deficient in nonsense-mediated rna decay factors is caused by autoimmune-like response. *Nucleic acids research*, 40(12):5615–5624.

Robinson, R. (2006). siRNAs and DNA methylation do a two-step to silence tandem sequences. *PLoS biology*, 4(11):e407.

Rodriguez-Leal, D., Lemmon, Z. H., Man, J., Bartlett, M. E., and Lippman, Z. B. (2017). Engineering quantitative trait variation for crop improvement by genome editing. *Cell*, 171(2):470–480.

Santos, B. A. M. C. (2014). *Small RNAs in gene regulatory networks*. PhD thesis, Department of Plant Sciences; University of Cambrdige.

Sarris, P. F., Duxbury, Z., Huh, S. U., Ma, Y., Segonzac, C., Sklenar, J., Derbyshire, P., Cevik, V., Rallapalli, G., Saucet, S. B., et al. (2015). A plant immune receptor detects pathogen effectors that target wrky transcription factors. *Cell*, 161(5):1089–1100.

Schoft, V. K., Chumak, N., Mosiolek, M., Slusarz, L., Komnenovic, V., Brownfield, L., Twell, D., Kakutani, T., and Tamaru, H. (2009). Induction of RNA-directed DNA methylation upon decondensation of constitutive heterochromatin. *EMBO reports*, 10(9):1015–21.

Schreiber, K. J., Bentham, A., Williams, S. J., Kobe, B., and Staskawicz, B. J. (2016). Multiple domain associations within the arabidopsis immune receptor rpp1 regulate the activation of programmed cell death. *PLoS pathogens*, 12(7):e1005769.

Sela, H., Spiridon, L. N., PETRESCU, A.-J., Akerman, M., MANDEL-GUTFREUND, Y., Nevo, E., Loutre, C., Keller, B., Schulman, A. H., and Fahima, T. (2012). Ancient diversity of splicing motifs and protein surfaces

in the wild emmer wheat (triticum dicoccoides) lr10 coiled coil (cc) and leucine-rich repeat (lrr) domains. *Molecular plant pathology*, 13(3):276–287.

Shan, Q., Wang, Y., Li, J., Zhang, Y., Chen, K., Liang, Z., Zhang, K., Liu, J., Xi, J. J., Qiu, J.-L., et al. (2013). Targeted genome modification of crop plants using a crispr-cas system. *Nature biotechnology*, 31(8):686–688.

Shao, Z.-Q., Xue, J.-Y., Wu, P., Zhang, Y.-M., Wu, Y., Hang, Y.-Y., Wang, B., and Chen, J.-Q. (2016). Large-scale analyses of angiosperm nucleotide-binding site-leucine-rich repeat (nbs-lrr) genes reveal three anciently diverged classes with distinct evolutionary patterns. *Plant physiology*, pages pp–01487.

Shen, Q.-H., Zhou, F., Bieri, S., Haizel, T., Shirasu, K., and Schulze-Lefert, P. (2003). Recognition specificity and rar1/sgt1 dependence in barley mla disease resistance genes to the powdery mildew fungus. *The Plant Cell*, 15(3):732–744.

Shirano, Y., Kachroo, P., Shah, J., and Klessig, D. F. (2002). A gain-of-function mutation in an arabidopsis toll interleukin1 receptor–nucleotide binding site–leucine-rich repeat type r gene triggers defense responses and results in enhanced disease resistance. *The Plant Cell*, 14(12):3149–3162.

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

Singh, M., Goel, S., Meeley, R. B., Dantec, C., Parrinello, H., Michaud, C., Leblanc, O., and Grimanelli, D. (2011). Production of viable gametes without meiosis in maize deficient for an argonaute protein. *The Plant Cell Online*, 23(2):443–458.

Slootweg, E. J., Spiridon, L. N., Roosien, J., Butterbach, P., Pomp, R., Westerhof, L., Wilbers, R., Bakker, E., Bakker, J., Petrescu, A.-J., et al. (2013). Structural determinants at the interface of the arc2 and leucine-rich repeat domains control the activation of the plant immune receptors rx1 and gpa2. *Plant physiology*, 162(3):1510–1528.

Slotkin, R. K., Vaughn, M., Borges, F., Tanurdzić, M., Becker, J. D., Feijó, J. A., and Martienssen, R. A. (2009). Epigenetic reprogramming and small RNA silencing of transposable elements in pollen. *Cell*, 136(3):461–72.

Soto-Suárez, M., Baldrich, P., Weigel, D., Rubio-Somoza, I., and San Segundo, B. (2017). The arabidopsis mir396 mediates pathogen-associated molecular pattern-triggered immune responses against fungal pathogens. *Scientific Reports*, 7:44898.

Soyars, C. L., James, S. R., and Nimchuk, Z. L. (2016). Ready, aim, shoot: stem cell regulation of the shoot apical meristem. *Current opinion in plant biology*, 29:163–168.

Srivastava, P. K., Moturu, T. R., Pandey, P., Baldwin, I. T., and Pandey, S. P. (2014). A comparison of performance of plant mirna target prediction tools and the characterization of features for genome-wide target prediction. *Bmc Genomics*, 15(1):348.

Stroud, H., Do, T., Du, J., Zhong, X., Feng, S., Johnson, L., Patel, D. J., and Jacobsen, S. E. (2014). Non-CG methylation patterns shape the epigenetic landscape in Arabidopsis. *Nature structural & molecular biology*, 21(1):64–72.

Stroud, H., Greenberg, M. V. C., Feng, S., Bernatavichute, Y. V., and Jacobsen, S. E. (2013). Comprehensive analysis of silencing mutants reveals complex regulation of the Arabidopsis methylome. *Cell*, 152(1-2):352–64.

Sueldo, D. J., Shimels, M., Spiridon, L. N., Caldararu, O., Petrescu, A.-J., Joosten, M. H., and Tameling, W. I. (2015). Random mutagenesis of the nucleotide-binding domain of nrc1 (nb-lrr required for hypersensitive response-associated cell death-1), a downstream signalling nucleotide-binding, leucine-rich repeat (nb-lrr) protein, identifies gain-of-function mutations in the nucleotide-binding pocket. *New Phytologist*, 208(1):210–223.

Tai, H. H., Goyer, C., De Koeyer, D., Murphy, A., Uribe, P., Halterman, D., et al. (2013). Decreased defense gene expression in tolerance versus resistance to verticillium dahliae in potato. *Functional & integrative genomics*, 13(3):367–378.

Tamaki, S., Tsuji, H., Matsumoto, A., Fujita, A., Shimatani, Z., Terada, R., Sakamoto, T., Kurata, T., and Shimamoto, K. (2015). Ft-like proteins induce transposon silencing in the shoot apex during floral induction in rice. *Proceedings of the National Academy of Sciences*, 112(8):E901–E910.

Tameling, W. I., Elzinga, S. D., Darmin, P. S., Vossen, J. H., Takken, F. L., Haring, M. A., and Cornelissen, B. J. (2002). The tomato r gene products i-2 and mi-1 are functional atp binding proteins with atpase activity. *The Plant Cell*, 14(11):2929–2939.

Tameling, W. I., Vossen, J. H., Albrecht, M., Lengauer, T., Berden, J. A., Haring, M. A., Cornelissen, B. J., and Takken, F. L. (2006). Mutations in the nb-arc domain of i-2 that impair atp hydrolysis cause autoactivation. *Plant physiology*, 140(4):1233–1245.

Tamura, K., Peterson, D., Peterson, N., Stecher, G., Nei, M., and Kumar, S. (2011). Mega5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. *Molecular biology and evolution*, 28(10):2731–2739.

Tao, Y., Yuan, F., Leister, R. T., Ausubel, F. M., and Katagiri, F. (2000). Mutational analysis of the arabidopsis nucleotide binding site–leucine-rich repeat resistance gene rps2. *The Plant Cell*, 12(12):2541–2554.

The Tomato Genome Consortium (2012). The tomato genome sequence provides insights into fleshy fruit evolution. *Nature*, 485(7400):635–41.

Tian, D., Traw, M., Chen, J., Kreitman, M., and Bergelson, J. (2003). Fitness costs of r-gene-mediated resistance in arabidopsis thaliana. *Nature*, 423(6935):74–77.

Tucker, M. R., Okada, T., Hu, Y., Scholefield, A., Taylor, J. M., and Koltunow, A. M. G. (2012). Somatic small RNA pathways promote the mitotic events of megagametogenesis during female reproductive development in Arabidopsis. *Development (Cambridge, England)*, 139(8):1399–404.

Válóczi, A., Hornyik, C., Varga, N., Burgyán, J., Kauppinen, S., and Havelda, Z. (2004). Sensitive and specific detection of microrna by northern blot analysis using lna-modified oligonucleotide probes. *Nucleic acids research*, 32(22):e175–e175.

Van Eck, J., Kirk, D. D., and Walmsley, A. M. (2006). Tomato (Lycopersicum esculentum). *Methods in molecular biology (Clifton, N.J.)*, 343:459–73.

Vaucheret, H., Vazquez, F., Crété, P., and Bartel, D. P. (2004). The action of argonaute1 in the mirna pathway and its regulation by the mirna pathway are crucial for plant development. *Genes & development*, 18(10):1187–1197.

Vicient, C. M. (2010). Transcriptional activity of transposable elements in maize. *BMC genomics*, 11(1):601.

Wang, Y., Itaya, A., Zhong, X., Wu, Y., Zhang, J., van der Knaap, E., Olmstead, R., Qi, Y., and Ding, B. (2011). Function and evolution of a microrna that regulates a ca2+-atpase and triggers the formation of phased small interfering rnas in tomato reproductive growth. *The Plant Cell*, 23(9):3185–3203.

Wang, Y., Sheng, G., Juranek, S., Tuschl, T., and Patel, D. J. (2008). Structure of the guide-strand-containing argonaute silencing complex. *Nature*, 456(7219):209–213.

Watanabe, T., Totoki, Y., Toyoda, A., Kaneda, M., Kuramochi-Miyagawa, S., Obata, Y., Chiba, H., Kohara, Y., Kono, T., Nakano, T., Surani, M. A., Sakaki, Y., and Sasaki, H. (2008). Endogenous siRNAs from naturally formed dsRNAs regulate transcripts in mouse oocytes. *Nature*, 453(7194):539–43.

Wei, C., Kuang, H., Li, F., and Chen, J. (2014). The i2 resistance gene homologues in solanum have complex evolutionary patterns and are targeted by mirnas. *BMC genomics*, 15(1):743.

Wei, W., Ba, Z., Gao, M., Wu, Y., Ma, Y., Amiard, S., White, C. I., Danielsen, J. M. R., Yang, Y.-G., and Qi, Y. (2012). A role for small rnas in dna double-strand break repair. *Cell*, 149(1):101–112.

Weiberg, A., Wang, M., Lin, F.-M., Zhao, H., Zhang, Z., Kaloshian, I., Huang, H.-D., and Jin, H. (2013). Fungal small rnas suppress plant immunity by hijacking host rna interference pathways. *Science*, 342(6154):118–123.

Weßling, R., Epple, P., Altmann, S., He, Y., Yang, L., Henz, S. R., McDonald, N., Wiley, K., Bader, K. C., Gläßer, C., et al. (2014). Convergent targeting of a common host protein-network by pathogen effectors from three kingdoms of life. *Cell host & microbe*, 16(3):364–375.

Williams, S. J., Sohn, K. H., Wan, L., Bernoux, M., Sarris, P. F., Segonzac, C., Ve, T., Ma, Y., Saucet, S. B., Ericsson, D. J., et al. (2014). Structural basis for assembly and function of a heterodimeric plant immune receptor. *Science*, 344(6181):299–303.

Williams, S. J., Sornaraj, P., deCourcy Ireland, E., Menz, R. I., Kobe, B., Ellis, J. G., Dodds, P. N., and Anderson, P. A. (2011). An autoactive mutant of the m flax rust resistance protein has a preference for binding atp, whereas wild-type m protein binds adp. *Molecular plant-microbe interactions*, 24(8):897–906.

Wingard, S. A. (1928). Hosts and symptoms of ring spot, a virus disease of plants. *J. agric. Res*, 37:127–153.

Winter, D., Vinegar, B., Nahal, H., Ammar, R., Wilson, G. V., and Provart, N. J. (2007). An electronic fluorescent pictograph browser for exploring and analyzing large-scale biological data sets. *PloS one*, 2(8):e718.

Wong, J., Gao, L., Yang, Y., Zhai, J., Arikit, S., Yu, Y., Duan, S., Chan, V., Xiong, Q., Yan, J., et al. (2014). Roles of small rnas in soybean defense against phytophthora sojae infection. *The Plant Journal*, 79(6):928–940.

Wu, C.-H., Abd-El-Haliem, A., Bozkurt, T. O., Belhaj, K., Terauchi, R., Vossen, J. H., and Kamoun, S. (2017). Nlr network mediates immunity to diverse plant pathogens. *Proceedings of the National Academy of Sciences*, page 201702041.

Wu, C.-H., Belhaj, K., Bozkurt, T. O., Birk, M. S., and Kamoun, S. (2016). Helper nlr proteins nrc2a/b and nrc3 but not nrc1 are required for pto-mediated cell death and resistance in nicotiana benthamiana. *New Phytologist*, 209(4):1344–1352.

Wu, L., Mao, L., and Qi, Y. (2012). Roles of dicer-like and argonaute proteins in tas-derived small interfering rna-triggered dna methylation. *Plant physiology*, 160(2):990–999.

Wu, L., Zhang, Q., Zhou, H., Ni, F., Wu, X., and Qi, Y. (2009). Rice microrna effector complexes and targets. *The Plant Cell*, 21(11):3421–3435.

Xia, R., Xu, J., Arikit, S., and Meyers, B. C. (2015). Extensive families of mirnas and phas loci in norway spruce demonstrate the origins of complex phasirna networks in seed plants. *Molecular biology and evolution*, 32(11):2905–2918.

Xia, R., Zhu, H., An, Y.-q., Beers, E. P., and Liu, Z. (2012). Apple mirnas and tasirnas with novel regulatory networks. *Genome biology*, 13(6):R47.

Xie, Z., Allen, E., Wilken, A., and Carrington, J. C. (2005). Dicer-like 4 functions in trans-acting small interfering rna biogenesis and vegetative phase change in arabidopsis thaliana. *Proceedings of the National Academy of Sciences of the United States of America*, 102(36):12984–12989.

Xiong, Q., Ye, W., Choi, D., Wong, J., Qiao, Y., Tao, K., Wang, Y., and Ma, W. (2014). Phytophthora suppressor of rna silencing 2 is a conserved rxlr effector that promotes infection in soybean and arabidopsis thaliana. *Molecular Plant-Microbe Interactions*, 27(12):1379–1389.

Xu, Y. and Du, J. (2014). Young but not Relatively Old Retrotransposons Are Preferentially Located in Gene-rich Euchromatic Regions in Tomato Plants. *The Plant Journal*, 80(4):n/a–n/a.

Yan, J., Gu, Y., Jia, X., Kang, W., Pan, S., Tang, X., Chen, X., and Tang, G. (2012). Effective small rna destruction by the expression of a short tandem target mimic in arabidopsis. *The Plant Cell*, 24(2):415–427.

Yi, X., Zhang, Z., Ling, Y., Xu, W., and Su, Z. (2014). Pnrd: a plant non-coding rna database. *Nucleic acids research*, 43(D1):D982–D989.

Yifhar, T., Pekker, I., Peled, D., Friedlander, G., Pistunov, A., Sabban, M., Wachsman, G., Alvarez, J. P., Amsellem, Z., and Eshed, Y. (2012). Failure of the tomato trans-acting short interfering rna program to regulate auxin response factor3 and arf4 underlies the wiry leaf syndrome. *The Plant Cell*, 24(9):3575–3589.

Yoshikawa, M., Iki, T., Tsutsui, Y., Miyashita, K., Poethig, R. S., Habu, Y., and Ishikawa, M. (2013). 3' fragment of mir173-programmed risc-cleaved rna is protected from degradation in a complex with risc and sgs3. *Proceedings of the National Academy of Sciences*, 110(10):4117–4122.

Yue, J.-X., Meyers, B. C., Chen, J.-Q., Tian, D., and Yang, S. (2012). Tracing the origin and evolutionary history of plant nucleotide-binding site–leucine-rich repeat (nbs-lrr) genes. *New Phytologist*, 193(4):1049–1063.

Zemach, A., Kim, M. Y., Hsieh, P.-H., Coleman-Derr, D., Eshed-Williams, L., Thao, K., Harmer, S. L., and Zilberman, D. (2013). The arabidopsis nucleosome remodeler ddm1 allows dna methyltransferases to access h1-containing heterochromatin. *Cell*, 153(1):193–205.

Zhai, J., Jeong, D.-H., De Paoli, E., Park, S., Rosen, B. D., Li, Y., González, A. J., Yan, Z., Kitto, S. L., Grusak, M. A., et al. (2011). Micrornas as master regulators of the plant nb-lrr defense gene family via the production of phased, trans-acting sirnas. *Genes & development*, 25(23):2540–2553.

Zhai, J., Zhang, H., Arikit, S., Huang, K., Nan, G.-L., Walbot, V., and Meyers, B. C. (2015). Spatiotemporally dynamic, cell-type–dependent premeiotic and meiotic phasirnas in maize anthers. *Proceedings of the National Academy of Sciences*, 112(10):3146–3151.

Zhang, H., Xia, R., Meyers, B. C., and Walbot, V. (2015). Evolution, functions, and mysteries of plant argonaute proteins. *Current opinion in plant biology*, 27:84–90.

Zhang, H., Zhang, J., Yan, J., Gou, F., Mao, Y., Tang, G., Botella, J. R., and Zhu, J.-K. (2017a). Short tandem target mimic rice lines uncover functions of mirnas in regulating important agronomic traits. *Proceedings of the National Academy of Sciences*, page 201703752.

Zhang, J., Li, W., Xiang, T., Liu, Z., Laluk, K., Ding, X., Zou, Y., Gao, M., Zhang, X., Chen, S., et al. (2010). Receptor-like cytoplasmic kinases integrate signaling from multiple plant immune receptors and are targeted by a pseudomonas syringae effector. *Cell host & microbe*, 7(4):290–301.

Zhang, X., Bernoux, M., Bentham, A. R., Newman, T. E., Ve, T., Casey, L. W., Raaymakers, T. M., Hu, J., Croll, T. I., Schreiber, K. J., et al. (2017b). Multiple functional self-association interfaces in plant tir domains. *Proceedings of the National Academy of Sciences*, page 201621248.

Zhang, X., Dodds, P. N., and Bernoux, M. (2017c). What do we know about nod-like receptors in plant immunity? *Annual Review of Phytopathology*, 55(1):205–229. PMID: 28637398.

Zhang, Y., Dorey, S., Swiderski, M., and Jones, J. D. (2004). Expression of rps4 in tobacco induces an avrrps4-independent hr that requires eds1, sgt1 and hsp90. *The Plant Journal*, 40(2):213–224.

Zhang, Y., Xia, R., Kuang, H., and Meyers, B. C. (2016a). The diversification of plant nbs-lrr defense genes directs the evolution of microrna that target them. *Molecular biology and evolution*, 33(10):2692–2705.

Zhang, Z., Liu, X., Guo, X., Wang, X.-J., and Zhang, X. (2016b). Arabidopsis ago3 predominantly recruits 24-nt small rnas to regulate epigenetic silencing. *Nature plants*, 2:16049.

Zheng, Y., Wang, Y., Wu, J., Ding, B., and Fei, Z. (2015). A dynamic evolutionary and functional landscape of plant phased small interfering rnas. *BMC biology*, 13(1):32.

Zhong, S., Fei, Z., Chen, Y.-R., Zheng, Y., Huang, M., Vrebalov, J., McQuinn, R., Gapper, N., Liu, B., Xiang, J., Shao, Y., and Giovannoni, J. J. (2013). Single-base resolution methylomes of tomato fruit development reveal epigenome modifications associated with ripening. *Nature biotechnology*, 31(2):154–9.

Zhu, H., Hu, F., Wang, R., Zhou, X., Sze, S.-H., Liou, L. W., Barefoot, A., Dickman, M., and Zhang, X. (2011). Arabidopsis argonaute10 specifically sequesters mir166/165 to regulate shoot apical meristem development. *Cell*, 145(2):242–256.

Zhu, S., Jeong, R.-D., Lim, G.-H., Yu, K., Wang, C., Chandra-Shekara, A., Navarre, D., Klessig, D. F., Kachroo, A., and Kachroo, P. (2013). Double-stranded rna-binding protein 4 is required for resistance signaling against viral and bacterial pathogens. *Cell reports*, 4(6):1168–1184.

Zong, J., Yao, X., Yin, J., Zhang, D., and Ma, H. (2009). Evolution of the rna-dependent rna polymerase (rdrp) genes: duplications and possible losses before and after the divergence of major eukaryotic groups. *Gene*, 447(1):29–39.

Appendix

Supplemental figures

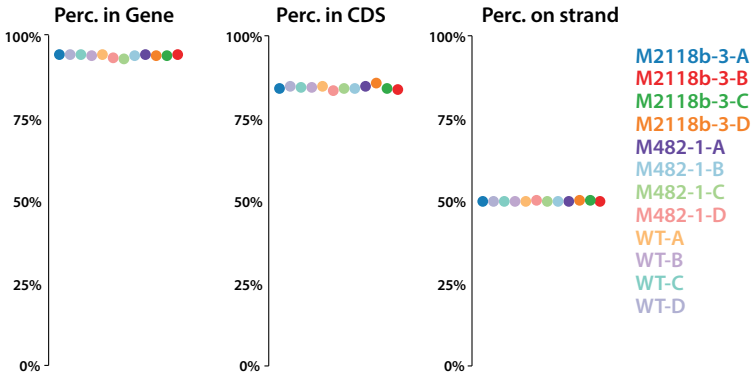

Fig. S.2 **RNAseq quality control does not show any obvious biases regarding mapping of the data.** Percentage of reads that fall into genes, CDS, and in the sense strand of transcripts.

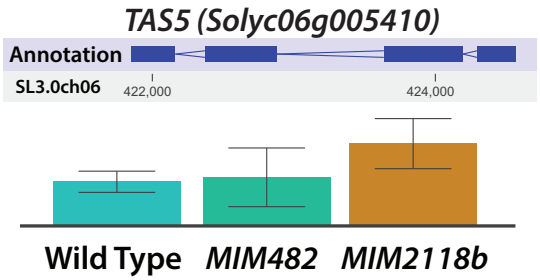

Fig. S.3 **RNAseq expression levels of TAS5.** SeqMonk relative expression levels at the genomic location of TAS5. Error bars represent SD (n=4).

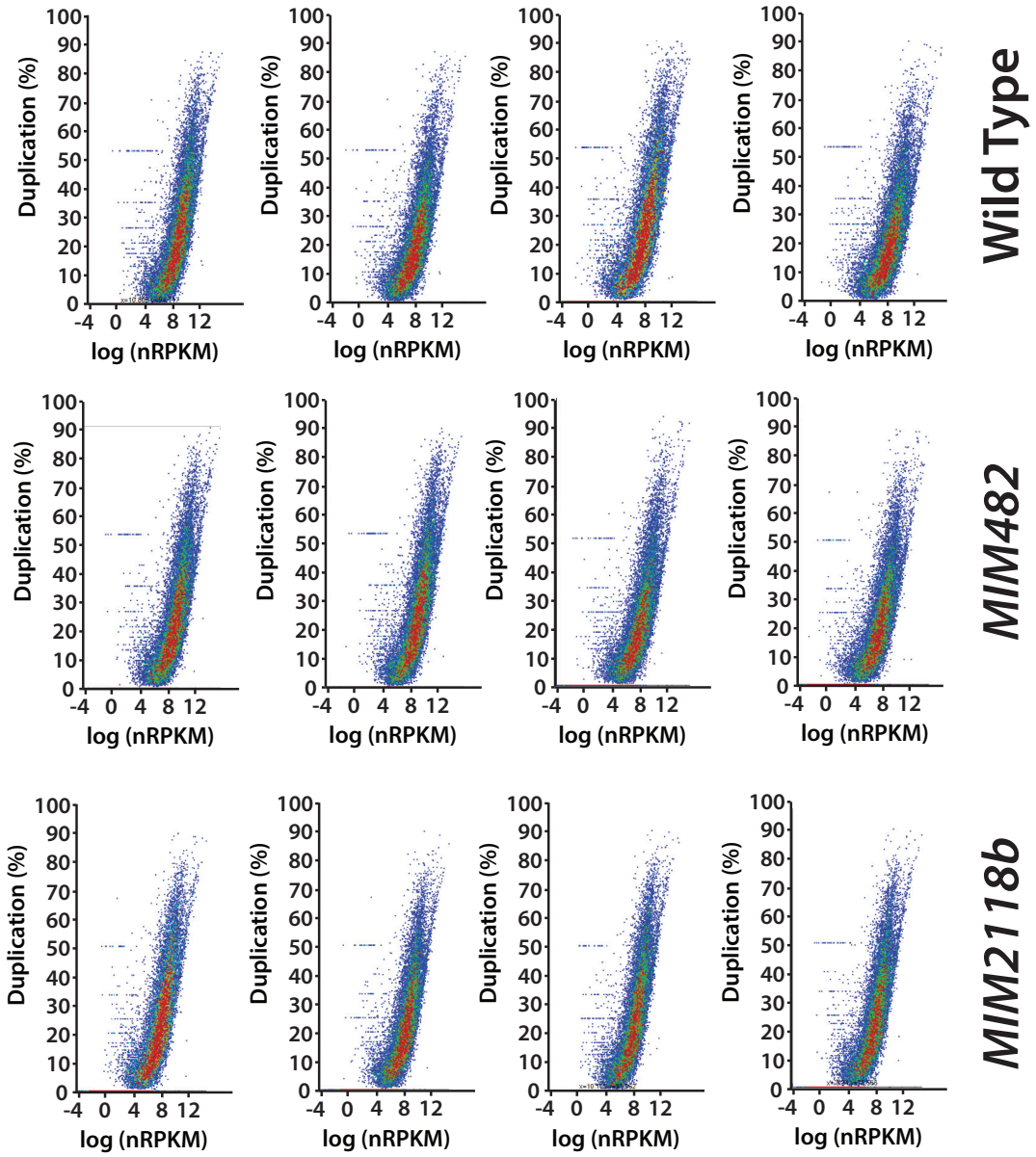

Fig. S.4 **RNAseq quality control does not show overduplication problems in the dataset.** Dot-plots across biological replicates representing the percentage of duplicated reads versus abundance of transcript (log of normalized reads per kilobase of transcripts and million).

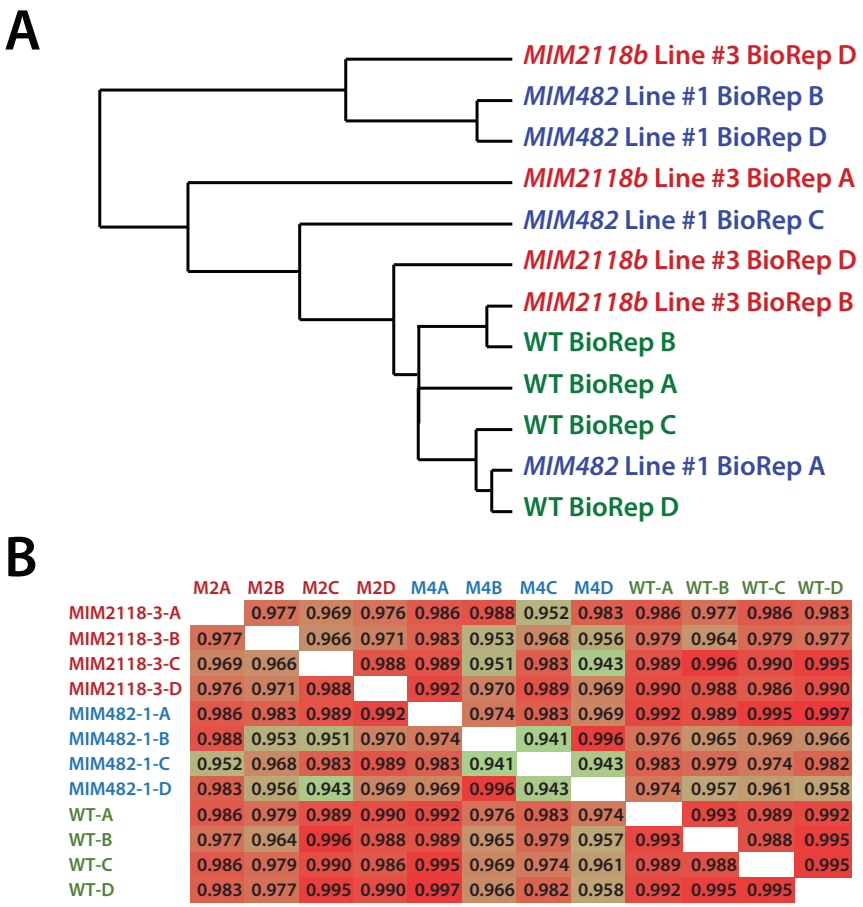

Fig. S.5 **Biological replicates do not have a higher degree of relatedness compared to other conditions.** (A) Datastore tree and (B) correlation matrix of RNAseq datasets

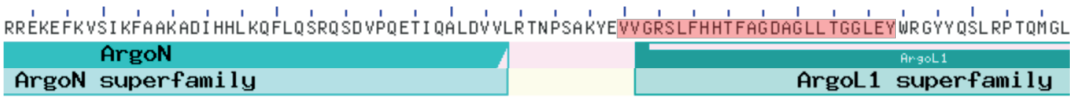

Fig. S.6 ***ago5-5* allele mutation deletes several amino-acids of the Linker 1 (L1) domain.** Domain sequence distribution around the mutated region in *ago5-5*. Red box indicate deleted amino-acids.

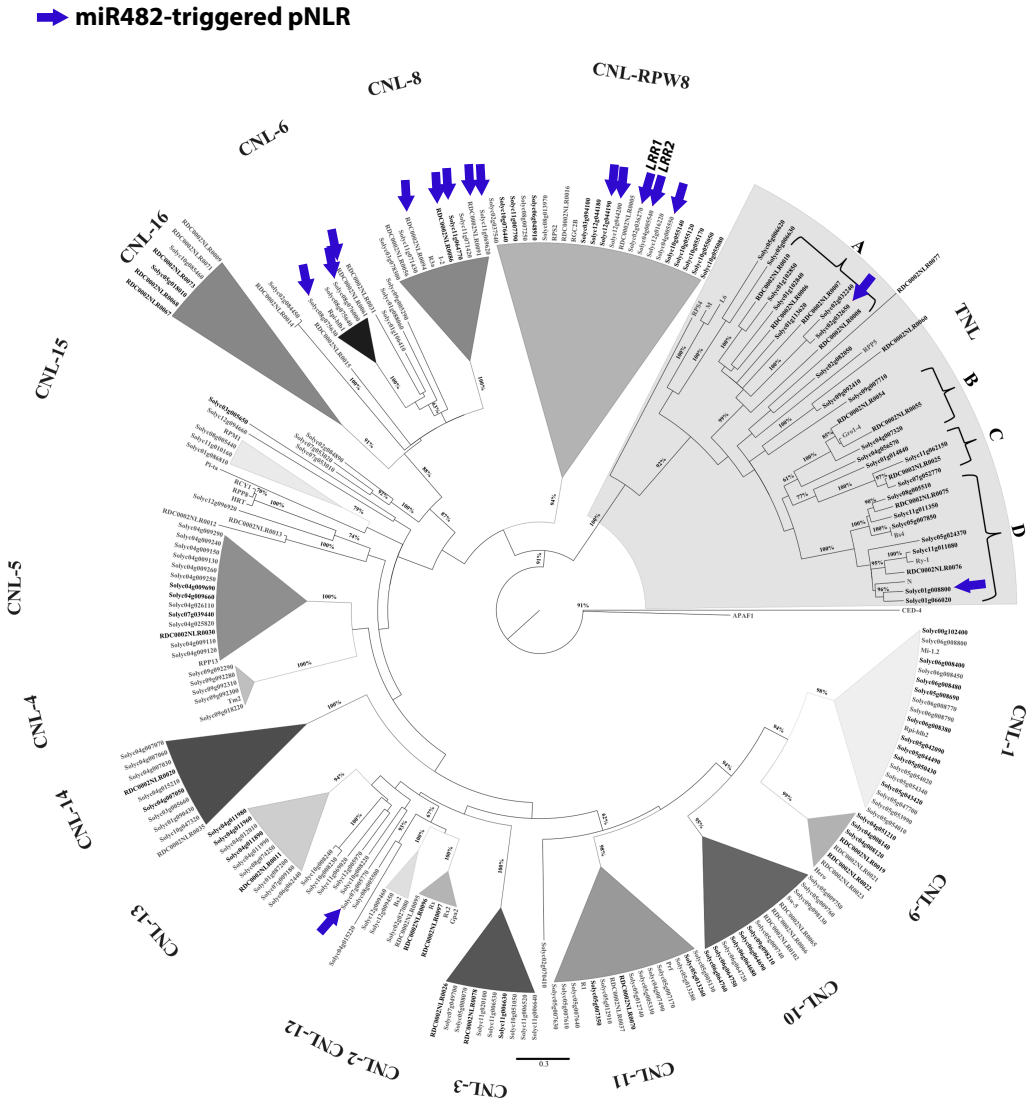

Fig. S.7 **Targets of miR482 in tomato belong to very distinct CNL clades.** NLR phylogeny adapted from Andolfo et al. [2014]. Blue arrows mark pNLRs targeted by miR482. LRR1 and LRR2, originally characterized in Shivaprasad et al. [2012], are indicated as a reference.

Supplemental tables

miRNA	Target	Score	UPE	Posit	Target description
mi2118a	Solyc11g008140	1	11.101	1224	Pectate lyase
mi2118a	Solyc01g102920	2	14.713	635	Disease resistance protein (TIR-NBS-LRR class)
mi2118a	Solyc04g007320	2	21.199	1110	Disease resistance protein (TIR-NBS-LRR class)
mi2118a	Solyc01g020371	2.5	20.058	349	GRF zinc finger family protein
mi2118a	Solyc03g116360	2.5	15.966	2405	Regulator of chromosome condensation (RCC1) family protein
mi2118a	Solyc04g024950	2.5	15.862	4	MATH domain/coiled-coil protein
mi2118a	Solyc04g049780	2.5	13.961	70	Retrovirus-related Pol polyprotein from transposon TNT 1-94
mi2118a	Solyc04g053070	2.5	17.989	278	DNA topoisomerase
mi2118a	Solyc06g009533	2.5	18.082	77	Kinase family protein
mi2118a	Solyc06g062440	2.5	22.729	593	Disease resistance protein
mi2118a	Solyc10g007065	2.5	17.785	580	Phenylalanyl-tRNA synthetase alpha chain
mi2118a	Solyc01g066020	3	25.484	656	disease resistance protein (TIR-NBS-LRR class)
mi2118a	Solyc01g087200	3	19.95	515	Disease resistance protein
mi2118a	Solyc01g090860	3	8.404	782	Nucleotidyltransferase family protein
mi2118a	Solyc02g030100	3	17.13	3745	Vacuolar protein sorting-associated protein 54
mi2118a	Solyc02g030105	3	17.13	2139	Vacuolar protein sorting-associated protein 54
mi2118a	Solyc02g091890	3	6.22	4704	myb-like protein X
mi2118a	Solyc03g083130	3	16.198	1745	gamma-irradiation and mitomycin c induced 1
mi2118a	Solyc05g009750	3	17.541	164	NBS-LRR resistance protein-like protein
mi2118a	Solyc08g005510	3	18.753	1512	disease resistance protein (TIR-NBS-LRR class)
mi2118a	Solyc09g090390	3	18.271	1709	2-oxoglutarate-dependent dioxygenase AOP2
mi2118a	Solyc10g050115	3	5.699	891	Transposon Ty3-I Gag-Pol polyprotein
mi2118a	Solyc11g011350	3	19.87	1381	disease resistance protein (TIR-NBS-LRR class)
mi2118a	Solyc12g009450	3	19.542	727	Disease resistance protein (CC-NBS-LRR class) family
mi2118a	Solyc12g056490	3	23.311	324	WD40 repeat-containing protein
mi2118a	Solyc01g094520	3.5	20.797	978	F-box/kelch-repeat protein
mi2118a	Solyc01g112260	3.5	30.089	316	Phosphoenolpyruvate carboxylase
mi2118a	Solyc02g064680	3.5	19.849	1125	Calcium-transporting ATPase
mi2118a	Solyc02g090860	3.5	19.613	1387	Phenylalanyl-tRNA synthetase alpha chain
mi2118a	Solyc03g025190	3.5	21.744	803	anthocyanin permease
mi2118a	Solyc03g112630	3.5	16.495	3497	Sec14p-like phosphatidylinositol transfer family protein
mi2118a	Solyc04g011960	3.5	15.189	521	Disease resistance protein (CC-NBS-LRR class) family
mi2118a	Solyc04g011980	3.5	16.151	521	Disease resistance protein (CC-NBS-LRR class) family
mi2118a	Solyc04g011990	3.5	15.575	1077	Disease resistance protein (NBS-LRR class) family
mi2118a	Solyc04g012000	3.5	17.228	221	NBS-coding resistance gene analog
mi2118a	Solyc04g012010	3.5	18.466	541	Disease resistance protein (NBS-LRR class) family
mi2118a	Solyc05g010240	3.5	11.485	3748	Chaperonin-60 beta subunit
mi2118a	Solyc06g061215	3.5	20.712	170	Proteinase inhibitor II
mi2118a	Solyc06g069390	3.5	24.527	1442	D-aminoacyl-tRNA deacylase
mi2118a	Solyc07g063430	3.5	13.261	643	Peroxisomal membrane (Mpv17/PMP22) family protein
mi2118a	Solyc08g066500	3.5	13.834	554	Homeobox leucine-zipper protein
mi2118a	Solyc09g005290	3.5	20.812	668	Nbs-lrr resistance protein
mi2118a	Solyc09g091990	3.5	18.213	665	Kinase family protein
mi2118a	Solyc10g087013	3.5	18.136	838	Cytochrome P450
mi2118a	Solyc12g056960	3.5	13.791	484	Glucan 1
mi2118a	Solyc01g008800	4	14.87	1561	disease resistance protein (TIR-NBS-LRR class)
mi2118a	Solyc01g111160	4	17.893	181	far-red elongated hypocotyls 3
mi2118a	Solyc02g081870	4	16.621	267	Pleiotropic drug resistance ABC transporter
mi2118a	Solyc02g093340	4	15.426	1078	RNA-binding (RRM/RBD/RNP motifs) family protein
mi2118a	Solyc03g111140	4	18.294	766	Malate synthase
mi2118a	Solyc03g115740	4	10.662	2443	Xyloglucan alpha-1
mi2118a	Solyc05g008340	4	23.996	791	Core-2/I-branching beta-1
mi2118a	Solyc05g009470	4	14.478	2663	Alpha-glucosidase
mi2118a	Solyc06g009533	4	20.01	1085	Kinase family protein
mi2118a	Solyc06g065820	4	14.049	582	Ethylene Response Factor H.1
mi2118a	Solyc06g068700	4	21.249	2627	Calreticulin/calnexin
mi2118a	Solyc07g008950	4	20.753	2523	Methionyl-tRNA synthetase family protein
mi2118a	Solyc07g008955	4	20.753	2631	Unknown protein
mi2118a	Solyc07g041030	4	22.468	277	DNA topoisomerase
mi2118a	Solyc08g013900	4	19.494	4686	Plant regulator RWP-RK family protein
mi2118a	Solyc08g082000	4	17.749	789	Homeobox-leucine zipper HOX24
mi2118a	Solyc09g007710	4	26.176	986	Disease resistance protein (TIR-NBS-LRR class) family
mi2118a	Solyc09g075010	4	4.321	683	HSP20-like chaperones superfamily protein
mi2118a	Solyc11g045350	4	19.839	3532	Plant regulator RWP-RK family protein
mi2118a	Solyc11g062220	4	19.961	5265	Zinc finger CCCH domain-containing protein 44

Table S.1 **Summary of all predicted targets of miR2118a.**

miRNA	Target	Score	UPE	Posit	Target description
mi2118b	Solyc06g005410	1	15.878	592	TAS5
mi2118b	Solyc06g005410	1	15.569	1639	TAS5
mi2118b	Solyc02g032650	2	15.284	796	Disease resistance protein (TIR-NBS-LRR class)
mi2118b	Solyc01g105340	2.5	20.142	810	Chaperone protein DnaJ
mi2118b	Solyc01g113620	2.5	16.895	806	Disease resistance protein (TIR-NBS-LRR class) family
mi2118b	Solyc04g009110	2.5	18.119	572	Nbs-lrr resistance protein
mi2118b	Solyc04g009130	2.5	20.332	584	Nbs-lrr resistance protein
mi2118b	Solyc04g009290	2.5	18.898	572	Disease resistance protein
mi2118b	Solyc04g026110	2.5	22.415	491	Disease resistance family protein
mi2118b	Solyc05g009750	2.5	17.541	164	NBS-LRR resistance protein-like protein
mi2118b	Solyc08g075630	2.5	26.52	743	NBS-LRR resistance protein
mi2118b	Solyc08g076000	2.5	25.842	851	NBS-LRR resistance protein
mi2118b	Solyc01g102920	3	14.713	635	Disease resistance protein (TIR-NBS-LRR class)
mi2118b	Solyc02g030290	3	19.46	281	Nbs-lrr resistance protein
mi2118b	Solyc02g037540	3	23.811	479	Disease resistance protein
mi2118b	Solyc06g062440	3	22.729	592	Disease resistance protein
mi2118b	Solyc08g075640	3	24.051	836	NBS-LRR resistance protein
mi2118b	Solyc09g075010	3	16.92	611	HSP20-like chaperones superfamily protein
mi2118b	Solyc09g098100	3	18.265	2059	CC-NBS-LRR_Solyc09g098100
mi2118b	Solyc01g110000	3.5	16.414	2336	Beta-galactosidase
mi2118b	Solyc02g036270	3.5	16.852	679	Disease resistance protein (NBS-LRR class) family
mi2118b	Solyc03g123630	3.5	23.333	1670	pectin methylesterase
mi2118b	Solyc04g009120	3.5	22.379	599	Nbs-lrr resistance protein
mi2118b	Solyc04g025820	3.5	20.87	312	Disease resistance protein
mi2118b	Solyc04g025840	3.5	20.843	491	Disease resistance family protein
mi2118b	Solyc05g008070	3.5	21.429	530	Disease resistance protein
mi2118b	Solyc05g014030	3.5	21.401	738	Regulator of chromosome condensation (RCC1) family protein
mi2118b	Solyc06g007780	3.5	21.021	2776	Nuclear transport factor 2 (NTF2)
mi2118b	Solyc07g039400	3.5	22.502	491	Disease resistance protein
mi2118b	Solyc07g039420	3.5	22.458	566	Disease resistance protein (NBS-LRR class) family
mi2118b	Solyc08g067060	3.5	24.14	674	Pentatricopeptide repeat superfamily protein
mi2118b	Solyc08g074250	3.5	15.378	545	Disease resistance protein (CC-NBS-LRR class) family
mi2118b	Solyc11g008140	3.5	11.101	1224	Pectate lyase
mi2118b	Solyc12g006040	3.5	21.275	662	NBS-LRR protein
mi2118b	Solyc01g105775	4	23.141	221	Carbonic anhydrase
mi2118b	Solyc01g111100	4	18.433	635	Neutral invertase
mi2118b	Solyc02g079310	4	16.167	387	Equilibrative nucleoside transporter family protein
mi2118b	Solyc02g079350	4	14.312	2490	Equilibrative nucleoside transporter family protein
mi2118b	Solyc02g083960	4	19.892	3334	2-oxoglutarate and Fe-dependent oxygenase-like protein
mi2118b	Solyc03g007330	4	22.962	983	ATP-dependent zinc metalloprotease FTSH protein
mi2118b	Solyc03g083430	4	21.732	1551	Splicing factor 3A subunit 3
mi2118b	Solyc04g005540	4	17.383	850	Disease resistance protein (NBS-LRR class) family
mi2118b	Solyc04g005550	4	21.364	868	Disease resistance protein (NBS-LRR class) family
mi2118b	Solyc04g071260	4	21.507	2316	Actin
mi2118b	Solyc05g014760	4	21.583	3126	Kinase family protein
mi2118b	Solyc05g018720	4	17.342	62	NBS-coding resistance protein
mi2118b	Solyc07g005770	4	24.094	539	Disease resistance protein
mi2118b	Solyc07g064700	4	25.463	1362	Bromodomain-containing protein
mi2118b	Solyc09g005120	4	18.535	483	DnaJ domain-containing protein
mi2118b	Solyc09g076010	4	8.678	370	Acyl-CoA N-acyltransferase
mi2118b	Solyc10g008230	4	19.11	548	Disease resistance protein
mi2118b	Solyc11g069020	4	20.947	497	Disease resistance protein
mi2118b	Solyc12g017800	4	23.624	1102	NBS-LRR class disease resistance protein
mi2118b	Solyc12g099060	4	23.521	644	Disease resistance protein
mi2118b	Solyc12g099940	4	15.941	2435	Acyl-CoA N-acyltransferase

Table S.2 **Summary of all predicted targets of miR2118b.**

miRNA	Target	Score	UPE	Posit	Target description
miR482b	Solyc04g009070	0	25.758	211	Disease resistance family protein
miR482b	Solyc02g036270	1.5	12.267	681	Disease resistance protein (NBS-LRR class) family
miR482b	Solyc04g009120	1.5	21.183	601	Nbs-lrr resistance protein
miR482b	Solyc05g008070	1.5	21.303	532	Disease resistance protein
miR482b	Solyc04g025820	2	19.14	314	Disease resistance protein
miR482b	Solyc04g025840	2	19.226	493	Disease resistance family protein
miR482b	Solyc07g039420	2	23.494	568	Disease resistance protein (NBS-LRR class) family
miR482b	Solyc11g065780	2	21.191	786	Disease resistance protein
miR482b	Solyc12g017800	2	24.453	1104	NBS-LRR class disease resistance protein
miR482b	Solyc01g067165	2.5	20.018	529	Disease resistance protein (CC-NBS-LRR class) family protein
miR482b	Solyc04g009130	2.5	20.794	586	Nbs-lrr resistance protein
miR482b	Solyc04g009290	2.5	23.41	574	Disease resistance protein
miR482b	Solyc07g039400	2.5	22.497	493	Disease resistance protein
miR482b	Solyc10g054970	2.5	19.563	538	CCNBS gene
miR482b	Solyc10g054990	2.5	16.38	520	Disease resistance protein (NBS-LRR class) family
miR482b	Solyc10g055170	2.5	21.742	46	Disease resistance protein (CC-NBS-LRR class) family protein
miR482b	Solyc11g006530	2.5	18.804	526	Disease resistance protein
miR482b	Solyc11g006630	2.5	21.57	532	Disease resistance protein
miR482b	Solyc04g009240	3	27.611	565	Nbs-lrr resistance protein
miR482b	Solyc04g009250	3	28.375	577	Nbs-lrr resistance protein
miR482b	Solyc04g009660	3	28.833	550	Nbs-lrr resistance protein
miR482b	Solyc04g009690	3	27.03	562	Nbs-lrr resistance protein
miR482b	Solyc07g005770	3	23.779	541	Disease resistance protein
miR482b	Solyc08g005440	3	21.814	627	NBS-LRR disease resistance protein
miR482b	Solyc10g051050	3	21.243	929	Disease resistance protein
miR482b	Solyc11g020090	3	20.197	103	Disease resistance protein
miR482b	Solyc11g020100	3	23.736	1110	Disease resistance protein
miR482b	Solyc11g069925	3	24.835	622	Disease resistance protein
miR482b	Solyc01g113620	3.5	16.593	808	Disease resistance protein (TIR-NBS-LRR class) family
miR482b	Solyc02g032650	3.5	15.288	798	Disease resistance protein (TIR-NBS-LRR class)
miR482b	Solyc02g070730	3.5	25.498	412	NBS-LRR resistance protein
miR482b	Solyc04g005540	3.5	15.547	852	Disease resistance protein (NBS-LRR class) family
miR482b	Solyc04g005550	3.5	18.848	870	Disease resistance protein (NBS-LRR class) family
miR482b	Solyc04g009110	3.5	17.68	574	Nbs-lrr resistance protein
miR482b	Solyc04g056746	3.5	16.29	687	Pentatricopeptide repeat-containing protein
miR482b	Solyc05g006630	3.5	29.979	1034	disease resistance protein (TIR-NBS-LRR class)
miR482b	Solyc05g007170	3.5	22.989	5882	Disease resistance protein
miR482b	Solyc06g072000	3.5	20.074	307	P-loop containing nucleoside triphosphate hydrolases superfamily protein
miR482b	Solyc07g044790	3.5	21.329	1223	Pvr4
miR482b	Solyc07g044797	3.5	21.329	286	CC-NBS-LRR disease resistance protein
miR482b	Solyc07g049700	3.5	24.47	544	Disease resistance protein
miR482b	Solyc09g065560	3.5	14.452	1794	Sulfate transporter
miR482b	Solyc11g006520	3.5	23.607	819	Disease resistance protein
miR482b	Solyc11g006640	3.5	26.45	532	Disease resistance protein
miR482b	Solyc11g068360	3.5	22.83	622	Disease resistance protein
miR482b	Solyc11g069620	3.5	21.64	706	Disease resistance protein
miR482b	Solyc12g044180	3.5	19.352	526	CC-NBS-LRR disease resistance protein
miR482b	Solyc12g044190	3.5	18.858	526	Disease resistance protein (CC-NBS-LRR class) family protein
miR482b	Solyc12g044200	3.5	18.45	1186	Disease resistance protein (CC-NBS-LRR class) family protein
miR482b	Solyc02g037540	4	23.271	481	Disease resistance protein
miR482b	Solyc03g078300	4	20.201	726	Disease resistance protein
miR482b	Solyc04g009150	4	25.145	565	Nbs-lrr resistance protein
miR482b	Solyc04g026110	4	21.107	493	Disease resistance family protein
miR482b	Solyc04g048920	4	15.927	46	CC-NBS-LRR disease resistance protein
miR482b	Solyc05g012740	4	20.929	1585	Disease resistance protein
miR482b	Solyc07g053010	4	28.026	1171	NBS-LRR type disease resistance protein
miR482b	Solyc10g045050	4	14.193	268	Enolase
miR482b	Solyc10g051170	4	20.714	673	Disease resistance protein
miR482b	Solyc10g055050	4	17.252	379	CC-NBS-LRR disease resistance protein
miR482b	Solyc11g069990	4	23.911	1218	I2C5
miR482b	Solyc11g070020	4	23.911	1990	Disease resistance protein
miR482b	Solyc11g071423	4	25.798	229	Disease resistance protein
miR482b	Solyc11g071995	4	26.508	622	Disease resistance protein
miR482b	Solyc12g006040	4	20.75	664	NBS-LRR protein

Table S.3 **Summary of all predicted targets of miR482b.**

miRNA	Target	Score	UPE	Posit	Target description
miR482c	Solyc08g075630	1.5	24.851	745	NBS-LRR resistance protein
miR482c	Solyc08g076000	1.5	24.722	853	NBS-LRR resistance protein
miR482c	Solyc02g021140	2.5	7.72	4	Superoxide dismutase
miR482c	Solyc02g078280	3	11.484	1597	DNA ligase-like protein
miR482c	Solyc05g006630	3	29.979	1034	disease resistance protein (TIR-NBS-LRR class)
miR482c	Solyc06g076350	3	15.598	392	LePCL1
miR482c	Solyc11g011560	3	19.304	1085	PHD finger protein family
miR482c	Solyc11g065780	3	21.191	786	Disease resistance protein
miR482c	Solyc00g006530	3.5	25.591	771	Calmodulin-binding protein
miR482c	Solyc01g102660	3.5	18.543	218	Glutathione S-transferase
miR482c	Solyc02g032650	3.5	15.288	798	Disease resistance protein (TIR-NBS-LRR class)
miR482c	Solyc02g036270	3.5	12.267	681	Disease resistance protein (NBS-LRR class) family
miR482c	Solyc04g005540	3.5	15.547	852	Disease resistance protein (NBS-LRR class) family
miR482c	Solyc04g005550	3.5	18.848	870	Disease resistance protein (NBS-LRR class) family
miR482c	Solyc04g009110	3.5	17.68	574	Nbs-lrr resistance protein
miR482c	Solyc04g009150	3.5	25.145	565	Nbs-lrr resistance protein
miR482c	Solyc04g025160	3.5	19.754	340	ATPase
miR482c	Solyc04g026110	3.5	21.107	493	Disease resistance family protein
miR482c	Solyc04g080590	3.5	25.393	748	Proteasome subunit alpha type
miR482c	Solyc06g060360	3.5	21.646	1010	Adenine nucleotide alpha hydrolases-like superfamily protein
miR482c	Solyc06g083875	3.5	17.908	247	pollen Ole e I family allergen protein
miR482c	Solyc07g053200	3.5	12.314	923	Adenine nucleotide alpha hydrolases-like superfamily protein
miR482c	Solyc08g065740	3.5	23.005	475	Vacuolar processing enzyme
miR482c	Solyc08g068040	3.5	19.304	2071	zinc finger FYVE domain protein
miR482c	Solyc10g007200	3.5	11.289	503	Hexosyltransferase
miR482c	Solyc10g076440	3.5	18.053	475	Disease resistance protein
miR482c	Solyc11g017370	3.5	19.701	1910	Pentatricopeptide repeat-containing protein
miR482c	Solyc11g062010	3.5	20.533	9655	Chromatin structure-remodeling complex subunit snf21
miR482c	Solyc12g005230	3.5	13.933	4336	Breast carcinoma-amplified sequence 3
miR482c	Solyc12g019144	3.5	23.587	1538	RING/U-box superfamily protein
miR482c	Solyc01g097390	4	20.058	1141	NAD(P)-linked oxidoreductase superfamily protein
miR482c	Solyc01g103450	4	30.773	415	Heat shock protein 70
miR482c	Solyc02g005180	4	25.396	273	Sugar facilitator protein 2
miR482c	Solyc02g070730	4	25.498	412	NBS-LRR resistance protein
miR482c	Solyc02g078790	4	14.982	1601	Transcription factor jumonji domain protein
miR482c	Solyc02g080960	4	18.235	39	transmembrane protein
miR482c	Solyc03g097980	4	17.739	550	Guanine nucleotide-binding alpha-2 subunit
miR482c	Solyc03g113620	4	28.565	230	MYB transcription factor
miR482c	Solyc05g005460	4	23.644	2043	Quinone oxidoreductase-like protein
miR482c	Solyc05g018370	4	5.408	808	Leguminosin group485 secreted peptide
miR482c	Solyc06g062440	4	22.498	595	Disease resistance protein
miR482c	Solyc06g068210	4	6.629	2252	Protein FAR1-RELATED SEQUENCE 8
miR482c	Solyc07g052760	4	16.463	540	DNA-binding storekeeper protein-related transcriptional regulator
miR482c	Solyc07g055380	4	19.526	1969	Disease resistance protein (TIR-NBS-LRR class)
miR482c	Solyc07g055610	4	20.758	553	Disease resistance protein (TIR-NBS-LRR class)
miR482c	Solyc09g007830	4	14.32	265	Cytokinin riboside 5'-monophosphate phosphoribohydrolase
miR482c	Solyc11g010660	4	12.438	988	protein SGT1
miR482c	Solyc11g062010	4	21.236	9856	Chromatin structure-remodeling complex subunit snf21
miR482c	Solyc11g069830	4	21.579	309	ATPase ASNA1

Table S.4 **Summary of all predicted targets of miR482c.**

miRNA	Target	Score	UPE	Posit	Target description
miR482e	Solyc11g006530	1.5	18.804	526	Disease resistance protein
miR482e	Solyc11g006630	1.5	21.57	532	Disease resistance protein
miR482e	Solyc07g049700	2	24.47	544	Disease resistance protein
miR482e	Solyc01g067165	2.5	20.018	529	Disease resistance protein (CC-NBS-LRR class) family protein
miR482e	Solyc04g009070	2.5	25.758	211	Disease resistance family protein
miR482e	Solyc05g008070	2.5	21.303	532	Disease resistance protein
miR482e	Solyc06g074760	2.5	15.026	1529	Ring/U-Box superfamily protein
miR482e	Solyc10g054970	2.5	19.563	538	CCNBS gene
miR482e	Solyc10g054990	2.5	16.38	520	Disease resistance protein (NBS-LRR class) family
miR482e	Solyc10g055170	2.5	21.742	46	Disease resistance protein (CC-NBS-LRR class) family protein
miR482e	Solyc11g020100	2.5	23.736	1110	Disease resistance protein
miR482e	Solyc12g009450	2.5	19.509	729	Disease resistance protein (CC-NBS-LRR class) family
miR482e	Solyc12g017800	2.5	24.453	1104	NBS-LRR class disease resistance protein
miR482e	Solyc01g014840	3	16.059	649	disease resistance protein (TIR-NBS-LRR class)
miR482e	Solyc01g108460	3	19.975	5259	Carboxypeptidase
miR482e	Solyc04g009250	3	28.375	577	Nbs-lrr resistance protein
miR482e	Solyc04g009660	3	28.833	550	Nbs-lrr resistance protein
miR482e	Solyc04g009690	3	27.03	562	Nbs-lrr resistance protein
miR482e	Solyc05g032850	3	17.25	2848	evolutionarily conserved C-terminal region 2
miR482e	Solyc07g027020	3	15.263	914	Protein kinase family protein
miR482e	Solyc09g091050	3	15.779	833	Calcium-dependent lipid-binding (CaLB domain) family protein
miR482e	Solyc11g006520	3	23.607	819	Disease resistance protein
miR482e	Solyc11g006640	3	26.45	532	Disease resistance protein
miR482e	Solyc11g069925	3	24.835	622	Disease resistance protein
miR482e	Solyc12g006040	3	20.75	664	NBS-LRR protein
miR482e	Solyc01g008800	3.5	16.457	1564	disease resistance protein (TIR-NBS-LRR class)
miR482e	Solyc01g066020	3.5	26.023	658	disease resistance protein (TIR-NBS-LRR class)
miR482e	Solyc02g032200	3.5	18.488	769	Disease resistance protein (TIR-NBS-LRR class) family
miR482e	Solyc02g036270	3.5	12.267	681	Disease resistance protein (NBS-LRR class) family
miR482e	Solyc03g046207	3.5	24.478	703	Disease resistance protein (CC-NBS-LRR class) family protein
miR482e	Solyc04g005540	3.5	15.547	852	Disease resistance protein (NBS-LRR class) family
miR482e	Solyc04g005550	3.5	18.848	870	Disease resistance protein (NBS-LRR class) family
miR482e	Solyc04g009120	3.5	21.183	601	Nbs-lrr resistance protein
miR482e	Solyc04g074865	3.5	16.828	1400	Retrovirus-related Pol polyprotein from transposon TNT 1-94
miR482e	Solyc05g006630	3.5	29.979	1034	disease resistance protein (TIR-NBS-LRR class)
miR482e	Solyc07g005770	3.5	23.779	541	Disease resistance protein
miR482e	Solyc07g044790	3.5	21.329	1223	Pvr4
miR482e	Solyc07g044797	3.5	21.329	286	CC-NBS-LRR disease resistance protein
miR482e	Solyc10g051050	3.5	21.243	928	Disease resistance protein
miR482e	Solyc11g011350	3.5	20.579	1383	disease resistance protein (TIR-NBS-LRR class)
miR482e	Solyc11g020090	3.5	20.197	103	Disease resistance protein
miR482e	Solyc11g068360	3.5	22.83	622	Disease resistance protein
miR482e	Solyc11g069300	3.5	25.111	2436	Kinase family protein
miR482e	Solyc11g069620	3.5	21.64	706	Disease resistance protein
miR482e	Solyc11g069990	3.5	23.911	1218	I2C5
miR482e	Solyc11g070020	3.5	23.911	1990	Disease resistance protein
miR482e	Solyc11g071423	3.5	25.798	229	Disease resistance protein
miR482e	Solyc11g071995	3.5	26.508	622	Disease resistance protein
miR482e	Solyc12g005970	3.5	15.572	493	Disease resistance protein (CC-NBS-LRR class) family
miR482e	Solyc12g044180	3.5	19.352	526	CC-NBS-LRR disease resistance protein
miR482e	Solyc12g044190	3.5	18.858	526	Disease resistance protein (CC-NBS-LRR class) family protein
miR482e	Solyc12g044200	3.5	18.45	1186	Disease resistance protein (CC-NBS-LRR class) family protein
miR482e	Solyc01g087200	4	19.546	517	Disease resistance protein
miR482e	Solyc01g100310	4	14.841	947	Calmodulin-binding protein
miR482e	Solyc02g032650	4	15.288	798	Disease resistance protein (TIR-NBS-LRR class)
miR482e	Solyc02g073574	4	17.186	583	Disease resistance protein
miR482e	Solyc02g084450	4	17.186	2110	Disease resistance protein
miR482e	Solyc03g078300	4	20.201	727	Disease resistance protein
miR482e	Solyc04g011590	4	30.314	622	Amino acid transporter
miR482e	Solyc04g011960	4	15.415	523	Disease resistance protein (CC-NBS-LRR class) family
miR482e	Solyc04g011980	4	16.137	523	Disease resistance protein (CC-NBS-LRR class) family
miR482e	Solyc04g011990	4	15.741	1079	Disease resistance protein (NBS-LRR class) family
miR482e	Solyc04g012000	4	17.872	223	NBS-coding resistance gene analog
miR482e	Solyc04g012010	4	18.376	543	Disease resistance protein (NBS-LRR class) family
miR482e	Solyc04g048920	4	15.927	46	CC-NBS-LRR disease resistance protein
miR482e	Solyc08g079150	4	16.89	22	SAUR-like auxin-responsive protein family
miR482e	Solyc09g007710	4	25.33	989	Disease resistance protein (TIR-NBS-LRR class) family
miR482e	Solyc10g055050	4	17.252	379	CC-NBS-LRR disease resistance protein
miR482e	Solyc11g065780	4	21.191	786	Disease resistance protein
miR482e	Solyc12g005520	4	23.107	121	Disease resistance protein (CC-NBS-LRR class) family
miR482e	Solyc12g096920	4	17.72	544	Disease resistance protein (CC-NBS-LRR class) family protein

Table S.5 **Summary of all predicted targets of miR482e.**

Differentially expressed genes between MIM482 and wild type (FDR < 0.05)

Gene ID	Log2(FC)	Log (mean)	PValue	FDR	Annotation
Solyc01g066790	2.944	0.531	2.58E-08	5.71E-04	LOW QUALITY:Diacylglycerol O-acyltransferase 2
Solyc11g050938	4.096	-0.116	1.58E-07	1.33E-03	Peroxisomal (S)-2-hydroxy-acid oxidase GLO1
Solyc01g081250	3.105	0.551	1.80E-07	1.33E-03	Glutathione s-transferase
Solyc08g083500	6.470	-0.257	1.39E-06	7.68E-03	Cytochrome P450 family protein
Solyc07g053525	-1.152	3.134	3.13E-06	1.31E-02	Glyceraldehyde-3-phosphate dehydrogenase-like family protein
Solyc03g006860	-1.951	0.939	4.63E-06	1.31E-02	Fructokinase
Solyc06g068190	2.414	1.645	4.77E-06	1.31E-02	Ubiquitin system component Cue protein
Solyc08g044280	1.727	3.097	5.08E-06	1.31E-02	LOW QUALITY:Adenine nucleotide alpha hydrolase-like domain kinase
Solyc05g015847	3.673	-0.686	5.32E-06	1.31E-02	BED zinc finger
Solyc04g074850	-3.629	2.765	9.16E-06	2.01E-02	Ripening regulated protein DDTFR18
Solyc08g065265	4.917	-0.191	1.00E-05	2.01E-02	BED zinc finger
Solyc04g081700	1.549	4.022	1.20E-05	2.21E-02	Ribosomal protein S5/Elongation factor G/III/V family protein
Solyc09g011030	1.071	5.915	1.78E-05	3.03E-02	Hsp70-binding protein 1

Differentially expressed genes between MIM2118b and wild type (FDR < 0.05)

Gene ID	Log2(FC)	Log (mean)	PValue	FDR	Annotation
Solyc07g063320	3.497	2.784	3.54E-09	7.84E-05	LanC-like protein 2
Solyc01g066790	2.630	0.531	9.31E-08	1.03E-03	LOW QUALITY:Diacylglycerol O-acyltransferase 2

Table S.6 **Summary of differentially expressed genes.** .