

# Identification of a post-transcriptional gene silencing mutant by EMS screen

Rocky Payet

100180993

A thesis submitted for the degree of Doctor of Philosophy (PhD)

School of Biological Sciences,  
University of East Anglia,  
Norwich,  
United Kingdom

December 2020

This copy of the thesis has been supplied on condition that anyone who consults it is understood to recognise that its copyright rests with the author and that use of any information derived therefrom must be in accordance with current UK Copyright Law. In addition, any quotation or extract must include full attribution

## Abstract

MicroRNAs (miRNAs) are small, non-coding RNAs which regulate a wide range of processes in plants ranging from developmental regulation to regulation of stress response. Through extensive research their mode of action and biogenesis is well understood, however the mechanisms through which they are degraded are much less so. In order to identify novel genes involved in miRNA decay, an EMS forward mutagenesis screen is performed in *Arabidopsis thaliana*. One mutant, *microRNA stability mutant 1 (msm1)*, is identified through this screen and analysed for general miRNA phenotypes using qPCR, northern blots and small RNA sequencing. Unfortunately, this does not appear to be a miRNA decay mutant, however it still presents an interesting post-transcriptional gene silencing phenotype. This background is then analysed for alternative splicing phenotypes, using RNA sequencing. Bioinformatic methods are used to identify the frequency of alternative splicing events, the types of alternative splicing events and the gene ontologies of the genes undergoing alternative splicing. This RNA sequencing is then used to identify potential causative mutations in the background using a bioinformatics pipeline, followed by PCR and sanger sequencing of backcrossed segregants. One of the likely causative genes identified in this approach is *cp13*, which is a C terminal phosphatase like gene responsible for dephosphorylating RNA pol II. This background is compared against published literature on the same gene using publicly available sRNA and mRNA sequencing. A large number of sRNAs mapping to transposable elements were discovered in this background which were not identified in published data on the same gene.

## **Access Condition and Agreement**

Each deposit in UEA Digital Repository is protected by copyright and other intellectual property rights, and duplication or sale of all or part of any of the Data Collections is not permitted, except that material may be duplicated by you for your research use or for educational purposes in electronic or print form. You must obtain permission from the copyright holder, usually the author, for any other use. Exceptions only apply where a deposit may be explicitly provided under a stated licence, such as a Creative Commons licence or Open Government licence.

Electronic or print copies may not be offered, whether for sale or otherwise to anyone, unless explicitly stated under a Creative Commons or Open Government license. Unauthorised reproduction, editing or reformatting for resale purposes is explicitly prohibited (except where approved by the copyright holder themselves) and UEA reserves the right to take immediate 'take down' action on behalf of the copyright and/or rights holder if this Access condition of the UEA Digital Repository is breached. Any material in this database has been supplied on the understanding that it is copyright material and that no quotation from the material may be published without proper acknowledgement.

# Contents

Abstract.....	2
Abbreviations.....	8
Acknowledgements.....	11
Chapter 1: Introduction.....	12
What is a miRNA?.....	12
Differences between plant and animal miRNA.....	12
Role of plant miRNAs in development.....	16
miRNA in plant stress response.....	19
Regulatory mechanisms of miRNAs.....	24
Mechanisms of turnover of miRNA.....	25
Exoribonucleolytic degradation of miRNAs – miRNA trimming.....	26
Non-templated nucleotide addition to miRNA – miRNA tailing.....	27
Interplay between truncation and tailing of miRNA.....	29
Target Directed miRNA Decay.....	31
Potential miRNA turnover candidate genes.....	31
Alternative Splicing.....	35
Other forms of PTGS.....	38
siRNAs.....	38
Virus derived siRNAs.....	40
Heterochromatic siRNAs.....	41
Trans-acting siRNAs.....	42
RNA quality control siRNAs.....	44
Ribosomal RNA-derived siRNAs.....	49
Piwi-interacting RNAs.....	49
CPLs.....	50
Aims of the project.....	51
Objectives:.....	52
Chapter 2: Materials and methods.....	53
Plant tissue culture.....	53
Normal conditions.....	53
Low Sulphate Media.....	53
Low Phosphate Media.....	53
ABA exposure.....	53

Seed Sterilisation .....	54
Mutagenesis Screening .....	54
Root and Silique Measurements.....	55
Phyllochron measurements .....	56
DNA Extractions.....	56
RNA Extractions .....	57
sRNA Northern Blots .....	57
RNA size separation by urea polyacrylamide gel electrophoresis .....	57
Transfer of RNA to nylon membrane.....	58
Chemical cross linking.....	58
Hybridisation .....	59
Membrane Stripping.....	60
Data interpretation .....	60
DNase Treatment .....	61
Reverse Transcription.....	61
qPCR.....	61
Primer Testing .....	62
Data interpretation .....	62
Sequencing of backcrosses.....	63
Backcrosses .....	63
sRNA library construction and sequencing.....	64
sRNA library total RNA cleanup .....	64
3' Adapter Adenylation.....	65
sRNA adapter ligation and PCR.....	65
Library size selection .....	65
Library normalisation.....	66
sRNA library Sequencing .....	66
mRNA Sequencing.....	66
Bioinformatics analysis.....	67
Mapping of Reads to Reference Sequences .....	67
Differential Expression Analysis.....	67
Alternative Splicing.....	68
Coverage plots.....	69
miTRATA analysis.....	69
SIMPLE Analysis.....	70

Coverage Plots.....	70
Gene Ontology Analysis.....	70
Venn Diagrams.....	71
Folding Analysis .....	71
Strand Distribution analysis.....	71
Chapter 3: Identification of putative mutants in miRNA decay pathway components by EMS screen .....	72
Introduction.....	72
Results.....	75
EMS Screening.....	75
Silencing of the transgene in the next generation.....	78
Phenotypic validation in the next generation.....	80
Discussion .....	83
Variance between EMS populations and control .....	83
Types of mutants.....	84
Improving the screen .....	85
Validating the phenotype by qPCR.....	85
Conclusion.....	86
Chapter 4: Characterisation of a miRNA phenotype in <i>msm1</i> .....	87
Introduction.....	87
Results.....	89
<i>Msm1</i> morphology .....	89
<i>Msm1</i> miRNA accumulation: .....	91
sRNA sequencing.....	94
Differential Truncation and Tailing analysis: .....	97
Discussion .....	100
Chapter 5: Identifying alternative splicing phenotypes in <i>msm1</i> .....	105
Introduction.....	105
Results: .....	107
ABA phenotype of <i>msm1</i> : .....	107
Differential expression analysis: .....	109
GO Enrichment of differentially expressed genes:.....	111
SplAdder analysis of <i>msm1</i> .....	117
Coverage Plots and counts supporting alternative splicing events.....	120
.....	123

Gene ontology of alternatively spliced genes.....	123
Discussion:.....	125
<i>Msm1</i> has an altered transcriptome profile .....	125
Splicing profiles in <i>msm1</i> roots and shoots .....	127
Functional consequences of alternative splicing in <i>msm1</i> .....	129
Limitations.....	131
Conclusions.....	132
Chapter 6: Identification of the causative SNP(s) in <i>msm1</i> .....	133
Introduction.....	133
Results.....	134
Identifying putative causative mutations .....	134
Backcross genotyping.....	136
Evidence for <i>cp13</i> mutation.....	138
Evidence of OPS mutation .....	140
Discussion .....	141
CPL3 and OPS are the most likely causative mutations.....	141
Limitations.....	143
Conclusion.....	144
Chapter 7: Investigation of <i>cp13</i> through <i>msm1</i> .....	145
Introduction: .....	145
Results.....	146
<i>msm1</i> has populations of transcripts which produce a significantly different number of siRNAs.....	146
Many siRNAs produced from endogenous transcripts in <i>msm1</i> are unlikely to be degradation debris .....	152
Transcripts producing a greater number of sRNAs in <i>msm1</i> are stress related .....	156
sRNA distribution across progenitor transcripts is even implicating an RDR activity.....	160
<i>msm1</i> has a significantly increased level of TE silencing .....	165
Discussion .....	169
Limitations.....	169
24mers in <i>msm1</i> and their consequence.....	171
Purpose of the sRNAs? .....	173
TE silencing in <i>msm1</i> .....	174
Conclusion.....	176

Chapter 8: Discussion and conclusions .....	177
miRNA and splicing phenotype in <i>msm1</i> .....	177
Understanding CPL3 through <i>msm1</i> .....	177
Main limitations of this study.....	179
Future work .....	180
Wider Relevance of this thesis .....	181
Chapter 9: References .....	182

## Abbreviations

ABA – Abscisic acid

ABI – Abscisic Acid Insensitive

AGO – Agonate

APR – Adenosine-5'-phosphosulphate reductase

APS – Adenosine-5'-phosphosulphate

ATP – Adenosine Triphosphate

ATPS – Adenosine Triphosphate Sulfurylase

BLAST - Basic Local Alignment Search Tool

CDS – Coding Sequence

CPL3 – C Terminal Phosphatase-like 3

DCL – Dicer-Like

DCP – Decapping protein

DIS3L2 – Dis3-like 3'-5' Exoribonuclease

DNA – Deoxyribonucleic Acid

EMS – Ethyl Methanesulfonate

FLM – Flowering Locus M

GFP – Green Fluorescent Protein

GO – Gene Ontology

HC-siRNA – Heterochromatic Small Interfering RNA

HEN1 – Hua Enhancer 1

HESO1 – Hen1 Suppressor 1

HST – HASTY

IPS1 – Induced by Phosphate Starvation 1

LNA – Locked Nucleic Acid

MAC7 – MOS4 Associated Complex 7

MFE – Minimum Free Energy

miRNA / miR[number] – microRNA

miTRATA – MicroRNA Truncation and Tailing Analysis

mRNA – Messenger Ribonucleic Acid

MSM1 – MicroRNA Stability Mutant 1  
NGS – Next generation sequencing  
NMD – Nonsense Mediated Decay  
OPS - Octopus  
PARE – Parallel Analysis of cDNA Ends  
PCR – Polymerase chain reaction  
piRNA – Piwi-interacting RNA  
Pre-miRNA – Precursor microRNA  
Pri-miRNA – Primary microRNA  
PSI – Percentage Spliced-In  
PTGS – Posttranscriptional Gene Silencing  
RACE – Rapid Amplification of cDNA Ends  
RD29 – Response to Desiccation 29  
RdDM – RNA-directed DNA Methylation  
RDR – RNA-Dependent RNA polymerase  
RISC – RNA Induced Silencing Complex  
RNA – Ribonucleic acid  
RNA-pol – RNA polymerase  
RRP6 – Ribosomal RNA-processing protein 6  
RT-qPCR – Reverse transcription quantitative polymerase chain reaction  
RQC-siRNA – RNA-quality control small interfering RNA  
risiRNAs – Ribosomal RNA-derived Small Interfering RNAs  
SDN – Small-RNA Degrading Nuclease  
SGI – Suppressor of Growth Inhibition  
SGS3 – Suppressor of Gene Silencing 3  
SIMPLE – Simple Mapping Pipeline  
sRNA – Small RNA  
siRNA – Small Interfering RNA  
SLIM1 – Sulphur limitation 1  
SNP – Single Nucleotide Polymorphism  
SOV – Suppressor of Varicose

S-PTGS – Sense Posttranscriptional Gene Silencing  
SR Proteins – Serine/Arginine Rich Proteins  
SUC2 – Sucrose Transport Protein 2  
SULTR2;1 – Sulphur Transporter 2;1  
TAIR – The Arabidopsis Information Resource (website)  
TAS – Trans-acting Small Interfering RNA producing gene  
tasiRNA – Trans-acting Small Interfering RNA  
TDMD – Target Directed MicroRNA Degradation  
T-DNA – Transfer Deoxyribonucleic Acid  
TE – Transposable Element  
TGS – Transcriptional Gene Silencing  
TUTase – Terminal Uridyltransferase  
UBC – Ubiquitin Conjugating Enzyme  
URT1 – Uridyltransferase 1  
UTR – Untranslated Region  
VCS – Varicose  
WT – Wild-type

## Acknowledgements

First and foremost, I would like to thank Tamas Dalmay for giving me the opportunity to carry out a PhD at UEA. The years that I have spent here as a PhD student will likely shape the rest of my life, and I will look back on them warmly. I would also like to thank my other academic supervisor, Ben Miller, who has made substantial contributions to both my academic and personal development, and provided many pieces of sage advice over the years. I thank also Afroditi Tsaballa, Martina Billmeier, Ping Xu, Maria-Elena Mannarelli, and Zahara Medina, for helping me acclimatise to life in the Dalmay lab, and teaching me a wide range of techniques. Additional thanks must also be given to Afroditi for looking after me in the first year of my PhD, and helping me make the transition from masters' student to PhD student. I would also like to thank Yvonne Ridge for her support with some of my larger experiments, and for helping me with undergraduate supervision when I had too much to do. Thanks is also due to Firas Louis, who helped generate some of the qPCR data for the putative mutants.

I would like to thank Simon Moxon, Josh Thody and Thomas Bradley for inducting me into the field of bioinformatics. All three of them answered countless questions from me without a word of complaint, and were pivotal in my learning of bioinformatics. I am especially grateful for this as it greatly shaped the course of my PhD and opened up many doors to me which would otherwise have been closed. Of all the skills I learned over the course of my PhD, I believe this one is perhaps the one I am proudest of.

None of the achievements of my PhD would have been possible without a strong support network behind me. To that effect, I would firstly like to thank my partner, Sean, who has always been supportive of me and listened to me ramble without complaint. Secondly, I would like to thank my family, all of whom have motivated and empowered me since day one of my PhD, and have always shown (or pretended to show) interest in what I am doing in the lab. Thirdly, I would like to thank my closest friends, Lucy, Louise and Leo, for always calling, sending me things and making me laugh. I find it hard to imagine completing this without their continued support.

# Chapter 1: Introduction

## What is a miRNA?

MicroRNAs (miRNA), first characterised in *Caenorhabditis elegans* in 2001 (Lau et al., 2001, Lee and Ambros, 2001), are ~22 nucleotide RNAs which function as regulators of gene expression through posttranscriptional repression (Bartel, 2009). MiRNA are transcribed by RNA polymerase II (RNA Pol II) as much larger transcripts (Lee et al., 2004) which then undergo a series of processing steps to yield mature, single-stranded miRNAs. These mature miRNA are loaded into ARGONAUTE1 (AGO1) which together form the guide-functionality of the RNA Induced Silencing Complex (RISC) (Pratt and MacRae, 2009). The single-stranded miRNA can anneal to another RNA molecule with a complementary sequence, called a miRNA target site, which guides RISC to the target transcript and results in silencing of those transcripts (Pratt and MacRae, 2009).

## Differences between plant and animal miRNA

It is salient that there are differences between plant and animal miRNAs. The first of these differences can be found in the biogenesis of miRNAs. In animals, long primary miRNA transcripts known as pri-miRNA undergo processing in the nucleus to form approximately 70 nucleotide stem-loop precursors known as pre-miRNA; this processing is carried out by the RNase III, Drosha (Lee et al., 2003). The pre-miRNA is then trafficked out of the nucleus by Exportin-5 (Lund et al., 2004) following which mature miRNAs are excised from the pre-miRNA by Dicer (Zhang et al., 2004). In plants, Drosha is absent, however there are four Dicer-like enzymes (DCL1-4) (Schauer et al., 2002). DCL1 is of particular interest here as it is demonstrably involved in miRNA accumulation (Reinhart et al., 2002) in plants. DCL1 has also been credited with the processing of pri-miRNA to pre-miRNA and then a second cleavage which converts pre-miRNA to an even shorter pre-miRNA (Kurihara and Watanabe, 2004) and it is highly likely that DCL1 is also responsible for the excision of mature miRNA from these pre-miRNA precursors (Chen, 2005). Unlike in animals, it appears that miRNA processing in plants takes place entirely in the nucleus, without a cytoplasmic component, owing to nuclear localisation sequences on DCL1 (Papp et al., 2003). There is a homolog of Exportin-5 in *Arabidopsis* called HASTY (Bollman

et al., 2003), however this is proposed to export the miRNA/miRNA\* duplex rather than the pre-miRNA. miRNA\* are antisense sequences of mature miRNAs which base-pair with them when the hairpin foldback structure characteristic of miRNAs is formed (Chen, 2005). A summary of miRNA biogenesis is depicted in Figure 1.

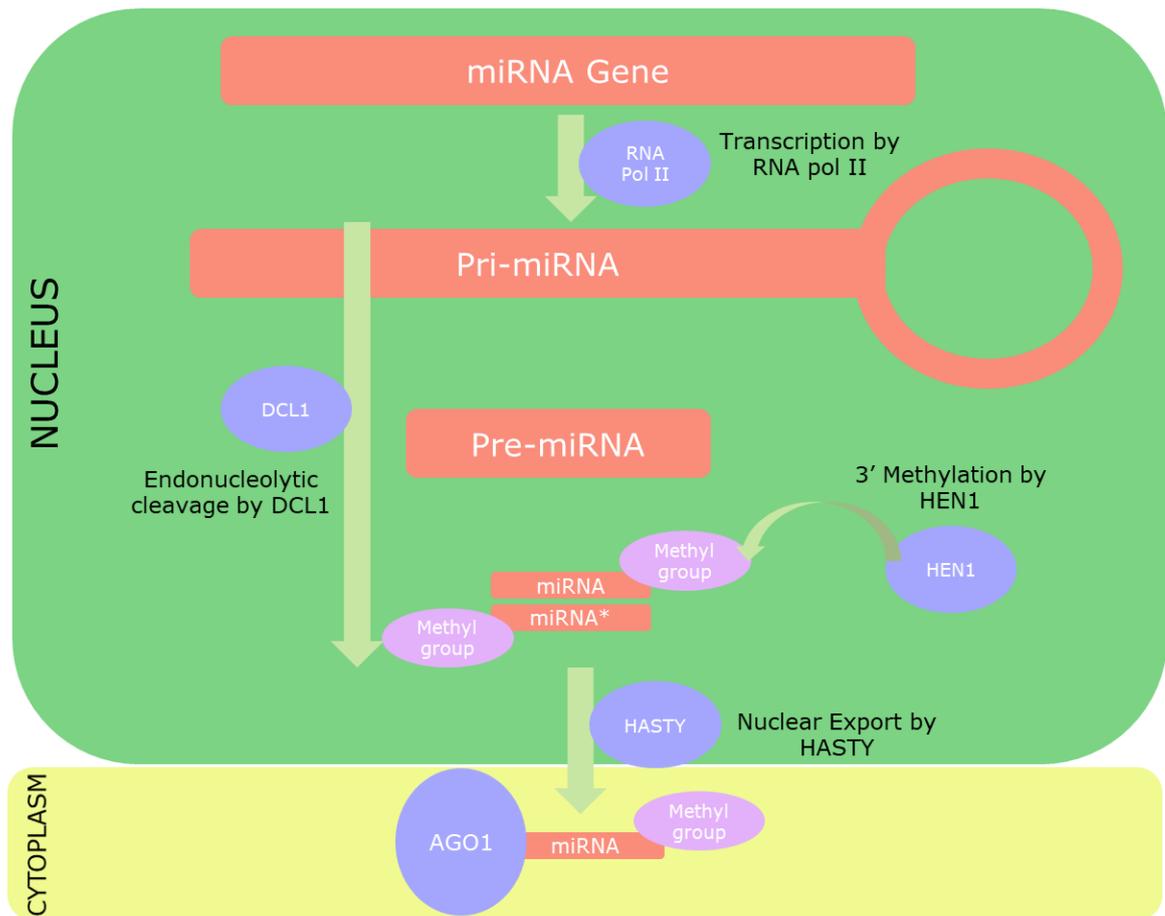
MiRNA target sites of plant and animal miRNAs are typically located on different regions of their target transcripts. Animal miRNA target sites are generally located within the 3' untranslated region (UTR) of the transcript (Enright et al., 2003). In addition to this, miRNA control in animals appears to be the result of a combinatorial effect of multiple miRNAs, as many animal mRNAs have more than one target site for the same miRNA, such as the *lin-14* gene which has seven *lin-4* target sites (Lee et al., 1993). Plant miRNAs differ in this respect also as their target sites are almost invariably found within the open-reading frames (ORFs) and only one target site is generally found per mRNA (Millar and Waterhouse, 2005). It should be noted that these rules are not unyielding, as some miRNA target prediction studies in plants have identified potential sites in 5' or 3'-UTRs of plant mRNAs (Sunkar and Zhu, 2004).

There are also differences between plant and animal miRNA target sites and their complementarity to their cognate miRNAs. Animal miRNAs rarely have perfect complementarity to their target sites (Dalmay, 2013), however the degree of complementarity is not uniform across the miRNA. There are distinct, highly conserved regions between nucleotides 2-8 of the miRNA, called the seed sequence or `miRNA seed` which, typically, have perfect complementarity (Lewis et al., 2003). Based around this seed sequence, three classes of target site have been identified. These are 5' dominant canonical target sites, which show no mismatching in the seed sequence and a high degree of base pairing to the rest of the site; 5'dominant seed only sites in which the seed sequence is perfectly complementary but the rest of the site does not show very much complementarity to the miRNA; 3'-compensatory target sites, where the seed sequence contains mismatches but this is compensated for by high base pairing with the 3' end of the miRNA (Dalmay, 2013). These animal miRNA target sites contrast quite notably with plant miRNA target sites, which usually exhibit

extensive complementarity to their cognate miRNA – there is generally no more than a maximum of 5 mismatches (Parizotto et al., 2004), (Liu et al., 2014).

The mechanism of action of plant and animal miRNAs differs in some respects also. miRNAs have two clear mechanisms of silencing; cleavage of the mRNA transcript and repression of translation of the mRNA transcript. The degree of complementarity between target site and miRNA determines which route is taken. If there is high complementarity then cleavage is favoured, and if there is low complementarity then repression of translation is favoured (Zeng et al., 2003), (Doench et al., 2003). Therefore, predictably, in plants the predominant method of miRNA regulation is through cleavage (Millar and Waterhouse, 2005). This cleavage usually occurs between nucleotides 10 and 11 of the cognate miRNA sequence, relative to the 5' end (Llave et al., 2002). In animals, the dominant mechanism is translational repression, and the presence of multiple miRNA target sites on a single mRNA allows for a graduated response, in which the degree of repression is related to the level of site occupation (Doench et al., 2003). Finally, animals rely on GW182 family proteins for translational repression of transcripts. However, no homologs of these proteins have been found to date in plants, indicating that this is not likely a major mechanism of miRNA mediated silencing in plants (Braun et al., 2013).

Owing to the high degree of complementarity between plant miRNA and their target sites, and the high amount of evolutionary conservation, it is possible to computationally predict miRNA targets. Putative targets are then verified experimentally using a range of techniques such as genetic analysis, 5'-rapid amplification of cDNA ends (5'-RACE) to detect the cleavage or RNA degradome analysis, sometimes called parallel analysis of cDNA ends (PARE) (Sun et al., 2011). As plant miRNAs appear to have near-perfect complementarity to their sites, little has been done to identify if targets exist with less specific pairing. Interestingly, investigation into *Arabidopsis* degradome libraries identified potentially 4000 sites which could function as miRNA target sites, suggesting that there may be even more plant miRNAs yet to be identified (Folkes et al., 2012).



**Figure 1 – Schematic Diagram depicting key stages in plant miRNA Biogenesis**

Diagram depicting the key conceptual stages of miRNA biogenesis in plants. Arrows indicate protein activity, blue elements depict proteins, purple elements depict protective modifications, and red elements depict nucleotide sequences.

MiRNA gene is transcribed by RNA pol II as a pri-miRNA transcript. This transcript folds into a hairpin structure and is then processed by DCL1 into a smaller pre-miRNA, and then further cleaved releasing the miRNA/miRNA\* duplex. This duplex is methylated at the 3' ends of each strand by HEN1, and then exported into the cytoplasm by HASTY. One miRNA strand is selected by AGO1 and loaded which can then go on to silence.

## Role of plant miRNAs in development

The first miRNAs characterised were shown to function as key regulators of development in *C. elegans* (Lagos-Quintana et al., 2001), and so initial emphasis on the study of miRNAs in other organisms was also placed on their role in development. One way to delineate the role of miRNAs in development of the organism is by looking at lines with mutation(s) in proteins involved with miRNA accumulation and function. This is a very broad and non-specific approach which has served to show the importance of miRNA action in plant development.

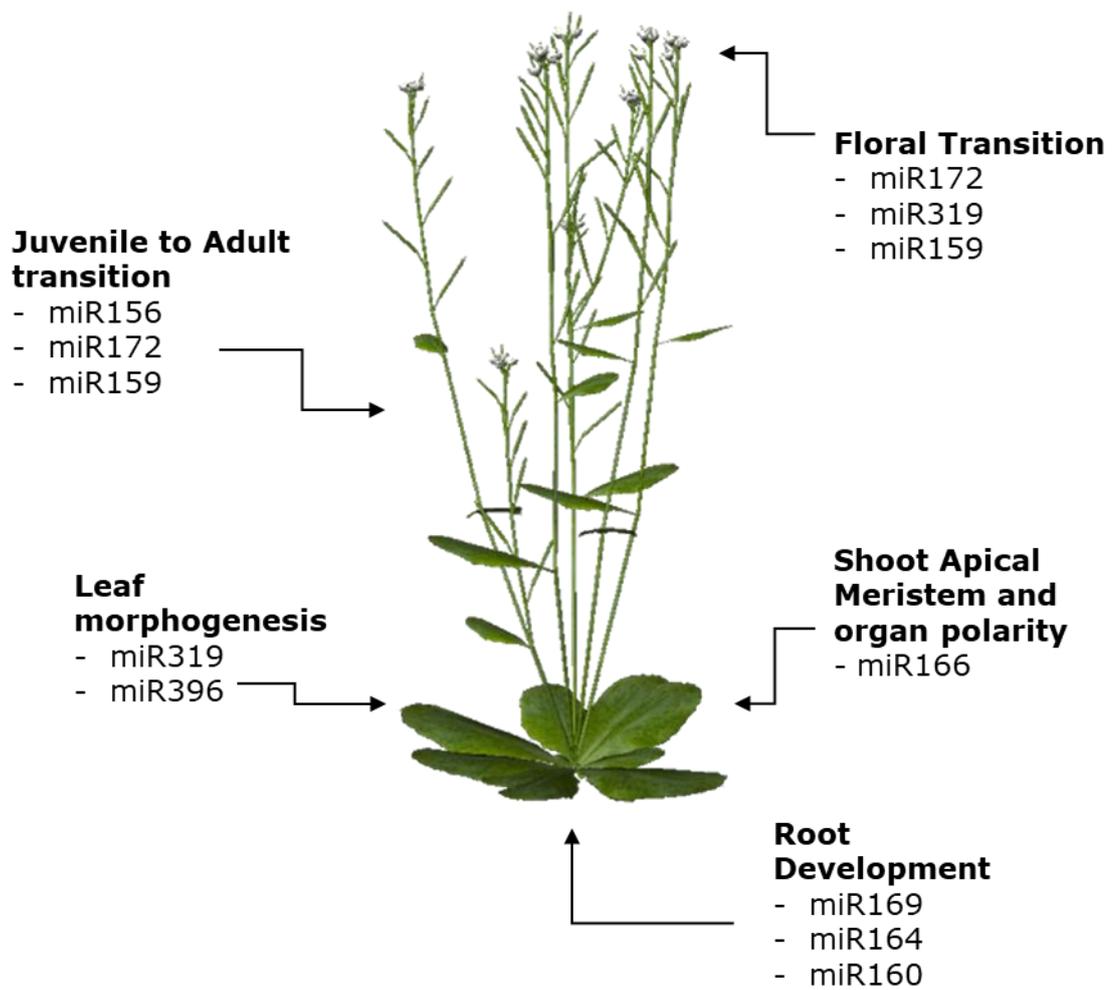
One mutant which has been of particular interest in this respect is the *Arabidopsis carpel factory* mutant, with CARPEL FACTORY (CAF) being a Dicer-1 homolog. This mutant has a truncated DLC1 protein attributable to a T-DNA insert into its 3' end – called *caf|dcl1-9*. This mutant produces many stamen whorls and an indefinite number of carpels, and also exhibits pleiotropic phenotypes (Jacobsen et al., 1999). Complete loss of DCL1 activity results in embryo lethality (Schauer et al., 2002), and the T-DNA insert in *caf|dcl1-9* is towards the end of the transcript, so it is reasonable to surmise that *caf|dcl1-9* protein has partial functionality. As DCL1 null mutants are embryo lethal, there is clearly a lack of redundancy between *Arabidopsis* DCL genes. In addition to this, it has been shown that DCL1 is required for miRNA production, and, consistent with this in *caf|dcl1-9* there is greatly reduced miRNA accumulation (Park et al., 2002). It was later shown that Caf|DCL1-9 did not compromise posttranscriptional gene silencing in any way, and so the severe developmental phenotypes seen in *caf|dcl1-9* mutants are likely attributable to miRNA levels rather than function (Finnegan et al., 2003). Mutants defective in another protein called HEN1 also display similar pleiotropic phenotypes to the *caf|dcl1-9* mutants and, on interrogation, appears to also be defective in miRNA accumulation, further substantiating the role of miRNA accumulation in development.

A good example of a mutation in a gene encoding a protein involved in miRNA function rather than level would be *Arabidopsis ago1* mutants. There are

numerous mutant lines available generated both by EMS and by T-DNA insertions in the *Ago1* gene (Bohmert et al., 1998) which forms an important part of RISC (Pratt and MacRae, 2009). Mutants in this gene show pleiotropic shoot architecture defects, further demonstrating the importance of effective miRNA activity during development. It should be noted however that miRNA accumulation has not been investigated in response to AGO family mutants (Pratt and MacRae, 2009). A summary of some of the key miRNA involved in plant development are illustrated in Figure 2.

As miRNAs target transcripts on a sequence basis, different individual miRNAs regulate different developmental pathways in varied ways, depending on which specific transcripts they repress. These effects may be as a direct result of repression of the transcript, or they may be an indirect effect, such as repression of a repressor.

A good example of miRNA-mediated developmental regulation is miR160. This miRNA targets an auxin-response factor called ARF17. Auxin is a critical plant hormone which regulates development. Studies performed using a miR160 resistant ARF17 transgene, which lacks the miRNA target site, have shown a plethora of developmental defects. These ranged from embryo asymmetry, premature inflorescence, sterility and root growth defects (Mallory et al., 2005). Another example of miRNAs demonstrating a key role in development is that of miR156. This miRNA targets SQUAMOSA PROMOTER BINDING PROTEIN-LIKE (SPLs) family of transcription factors in *Arabidopsis*. Transgenic plants which constitutively express miR156 displayed prolonged expression of juvenile traits and had severely delayed flowering. This phenotype could be rescued by expressing a miR156 resistant SPL3 transcript, indicating that the miRNAs regulation of this transcript was responsible for the phenotype (Wu and Poethig, 2006).



**Figure 2 – Schematic Diagram depicting the developmental stages miRNA regulate in *Arabidopsis***

Diagram depicting key developmental stages that miRNA regulate in *Arabidopsis thaliana*.

## miRNA in plant stress response

Whilst it is clear that miRNAs are fundamental for plant development, their activity does not end with development. There is a rapidly growing body of literature which demonstrates a considerable role for miRNAs in plant stress responses (Sunkar et al., 2012). First evidenced *in silico*, computational analysis of plant miRNA and their targets showed that some miRNAs targeted stress response genes (Jones-Rhoades and Bartel, 2004). This was followed by identification of novel miRNAs cloned from *Arabidopsis* grown under stress conditions which had not previously been attempted (Sunkar and Zhu, 2004).

One particularly well studied example of a plant stress response miRNA is miRNA395, first described in 2004 (Jones-Rhoades and Bartel, 2004). This miRNA was shown to be intimately involved with the sulphate starvation response, being undetectable in plants grown in normal or high level sulphate containing media, and readily detectable in plants grown in low sulphate media (Jones-Rhoades and Bartel, 2004). MiRNA395 was also found in the same study to be complementary to the mRNA of ATP sulphurylase (ATPS) proteins; ATPS being the first enzyme in the sulphur assimilation pathway (Kawashima et al., 2011).

In plants, the sulphur assimilation pathway is a demand driven process in which inorganic sulphate is taken up and sequentially reduced into sulphide, which is then used to produce the amino acid cysteine (Kopriva, 2006). In response to a deficiency of sulphur in the plant, both uptake of sulphur from the environment and rate of reduction of sulphate is increased by an associated induction of sulphate transporter mRNA and adenosine-5'-phosphosulphate (APS) reductase (APR) mRNA levels, which is the crucial enzyme of the pathway (Nikiforova et al., 2003). The molecular mechanisms that underpin the sulphate starvation response are still largely unknown, however to date, several components have been identified which operate in the sulphate `regulatory circuit`. The first of these components is the transcription factor SULPHUR LIMITATION 1 (SLIM1), which increases sulphate transporter expression and also upregulates several

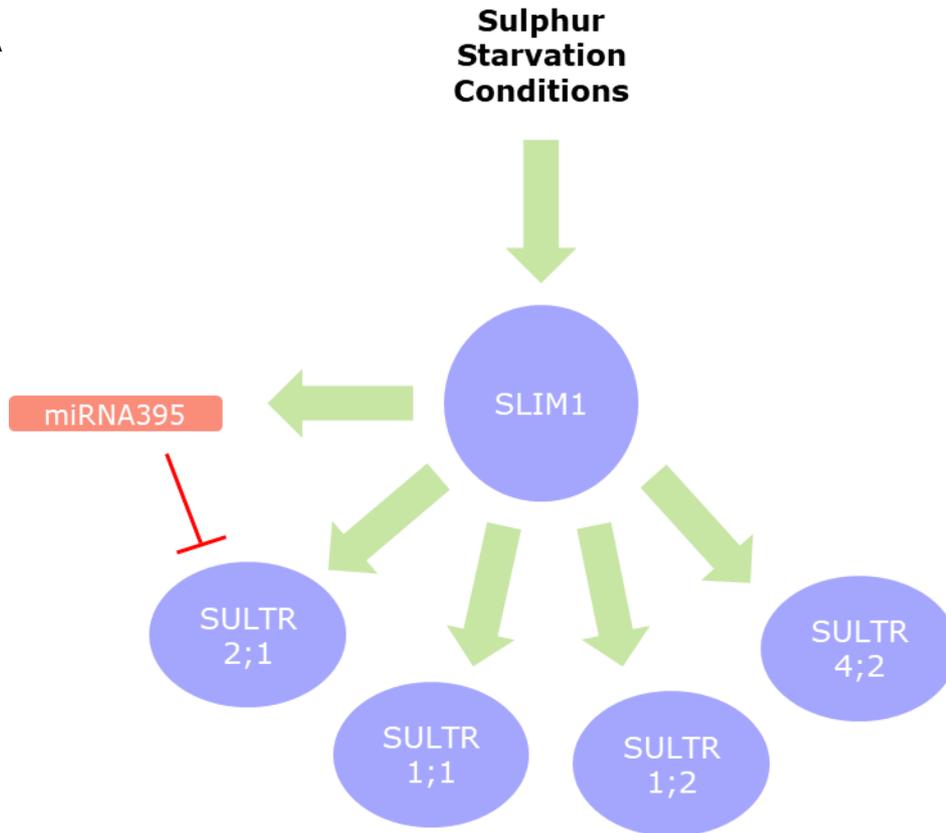
other genes involved in the sulphur starvation response (Maruyama-Nakashita et al., 2006).

MiRNA395 expression is controlled by SLIM1 (Kawashima et al., 2009). In addition to targeting ATPS, miRNA395 also appears to target the low affinity Sulphur transporter SULTR2;1, and yet paradoxically, miRNA395 expression is strongly induced in sulphur starvation conditions (Jones-Rhoades and Bartel, 2004). Additionally, the mRNA levels of SULTR2;1 were found to increase in the roots in sulphate starvation, which was also unexpected given that of the 6 miRNA395 loci (MIR395a-f), MIR395c and MIR395e in particular were also found to be strongly induced in the roots (Kawashima et al., 2009). To investigate this, (Kawashima et al., 2009) analysed SULTR2;1 expression in wild type *Arabidopsis* and *Slim1* mutants, which are unable to induce SLIM1 and by extension unable to upregulate miRNA395. In the wild type they found that when they transferred the seedlings from a sulphur rich environment (S+) to a sulphur deficient environment (S-), there was no significant change in SULTR2;1 expression in the leaves but there was a considerable increase in SULTR2;1 expression in the roots. In the *Slim1* mutants however, SULTR2;1 levels in the leaves on transfer increased as there was no miRNA395 expression, and the SULTR2;1 upregulation in the roots was even more dramatic than in the wild type (Kawashima et al., 2009).

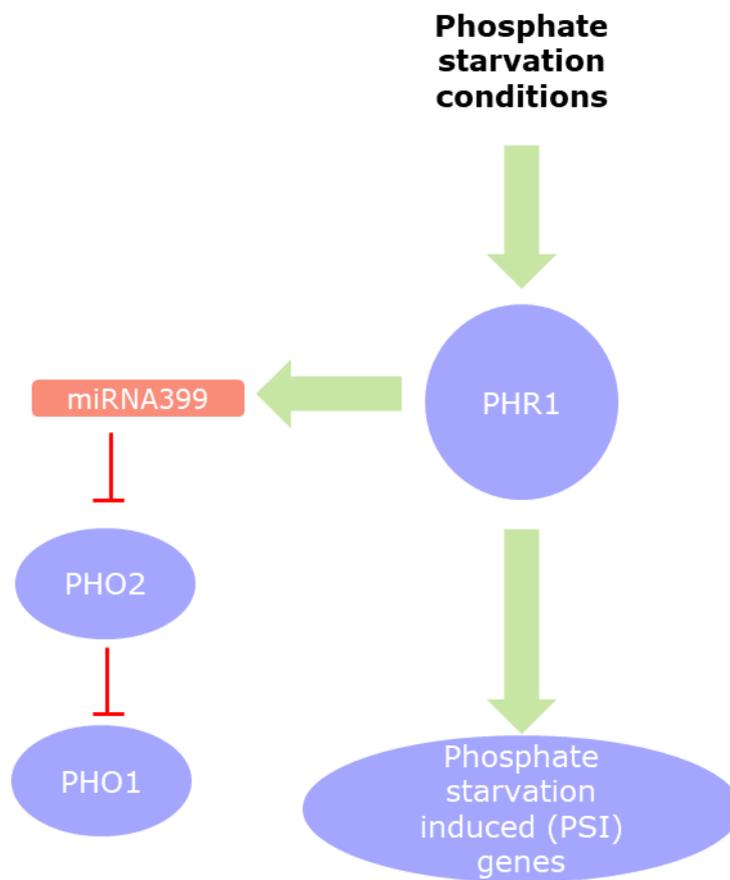
On closer investigation, whilst there is a positive temporal correlation between miRNA395 and SULTR2;1, it appears that spatially they are different. In the roots, most miRNA395 expression occurs in the phloem companion cells, yet the SULTR2;1 is mainly expressed in xylem parenchyma cells. There are however, some cleaved SULTR2;1 fragments in the roots indicating that there is some overlap between these spatial expression patterns. With the available evidence a model has been suggested in which SULTR2;1 has low level expression in the phloem companion cells, and miRNA395 silences this expression in response to S- conditions in order to restrict SULTR2;1 expression to the xylem (Kawashima et al., 2009). This illustrates an elegant and non-canonical method of miRNA-mediated gene regulation in plant stress response.

There exist other plant miRNA which regulate response to nutrient deficiency stresses, such as miRNA399. This miRNA, similar to miRNA395, is upregulated in response to low levels of inorganic phosphate ( $P_i$ ), however in contrast to miRNA395 it exerts its regulatory effects in a canonical way. As miRNA399 levels increase, its target transcript, that of a ubiquitin-conjugating enzyme (UBC) involved in response to low  $P_i$  levels, decreases (Fujii et al., 2005). These two pathways are illustrated in Figure 3. In addition to nutrient starvation responses such as sulphate, phosphate, nitrogen, copper etc., there are many other plant miRNAs which are involved in other plant stress responses such as drought and salinity; cold, heat and other abiotic stresses; biotic stresses (Sunkar et al., 2012).

**A**



**B**



**Figure 3 - Schematic representation of miR395 and miR399 induction by Sulphur and Phosphor starvation respectively**

**A:** On sulphur starvation, the master transcription factor SLIM1 is upregulated. This in turn upregulates the production of sulphur transporters SULTR1;2 – SULTR4;2, as well as miR395. However, miR395 is expressed only in the roots. Therefore, on sulphur uptake, root cells have reduced sulphur incorporation due to this negative regulation.

**B:** On phosphate starvation, the master transcription factor PHR1 is upregulated, which in turn upregulates a suite of phosphate starvation induced (PSI) genes. MiR399 is also upregulated, which targets a ubiquitin conjugating enzyme called PHO2, which normally induces degradation of PHO1.

## Regulatory mechanisms of miRNAs

MiRNAs have the capacity to regulate a plethora of biological processes and, as such, need to be highly regulated themselves. Temporal regulation of miRNAs is of particular importance to ensure the appropriate response is achieved, especially during development. There are three main strata of control of miRNA steady-state levels; transcription of miRNA genes (*MIR*), processing of miRNA precursors, and turnover of miRNAs. As miRNAs are encoded by their own individual genes with their own TATA boxes (Xie et al., 2005a) and transcription binding motifs (Megraw et al., 2006), miRNA gene expression is subject to similar transcriptional regulation to protein coding genes.

MiRNA also have the ability to regulate their own processing and activity through feedback mechanisms. Two good examples of this are miR162 and miR168. DCL1 transcript levels, required for miRNA processing in plants (Reinhart et al., 2002), were found to accumulate in *dcl1* mutant plants and also in miRNA-defective *hen1* mutant plants. Computational analysis also predicted a miR162 target site close to the centre of the DCL1 mRNA, which was then verified with the identification of a DCL1-derived RNA with the properties expected of a miR162-guided cleavage product (Xie et al., 2003). AGO1, a crucial part of RISC (Pratt and MacRae, 2009) also has its expression repressed by a miRNA mediated mechanism – in this instance miRNA168 (Vaucheret et al., 2006). This regulation was also demonstrated to be critical by (Vaucheret et al., 2004) who showed that *Arabidopsis* mutants which expressed miRNA168 resistant AGO1 had developmental defects (Vaucheret et al., 2004).

In the cell, most RNAs created by RNA pol II receive a 5' cap structure, which is added to the nascent transcript and followed by the binding of the heterodimeric cap-binding complex (CBC). This serves to induce splicing, polyadenylation of the transcript and nuclear export by exportin5 (HASTY in *Arabidopsis*). A polyA tail is also added to the 3' end of the transcript, and together with the 5' cap structure serve to protect the transcript from exonucleolytic decay pathways (Orphanides and Reinberg, 2002). Whilst pri-miRNA do have these structures, mature miRNA do not (Bartel, 2004), and so other mechanisms exist to protect them from degradation.

The first of these protective mechanisms is methylation at the 3' end of the transcript. The methyltransferase enzyme HUA Enhancer 1 (HEN1), methylates miRNA/miRNA\* duplexes in plants (Huang et al., 2009). It has been suggested that this modification protects the miRNA from exonucleolytic degradation from the 3' end (Yu et al., 2005) which is consistent with observations in *hen1* mutants, which have reduced miRNA accumulation (Park et al., 2002). A majority of plant miRNAs and other small RNA species have 2'-O methylation marks (Yu et al., 2005).

In addition to methylation, miRNA also appear to be protected by the addition of adenylic acid residues to their 3' ends. *In vitro*, plant extracts which had undergone adenylation were found to degrade slower (Lu et al., 2009). The fact that this is not the case for miRNA suggests that other mechanisms exist to degrade mature miRNAs; this is further supported by the observation that exosome complex depletion does not affect miRNA levels (Chekanova et al., 2007). Isolation of truncated 3' and 5' miRNA products suggests that this mechanism may operate at least in part through 3' to 5' exonucleolytic degradation, however it is unlikely to be 5' to 3' as the greater the number of nucleotides trimmed from the 3' end, the greater the percentage of adenylated miRNAs, however this does not occur at the 5' end (Lu et al., 2009).

The *in vivo* relevance of this is not yet certain. Indeed, it presents a more complex picture of miRNA turnover as it appears the non-templated addition of these adenylic acid residues occurs on truncated miRNAs lacking a 3' methyl group, yet our current understanding of methylation of plant miRNAs suggests that removal of this group enters the miRNA into turnover pathways. There does, however, appear to be a tissue dependent component to this adenylation, as RNA extracts from different tissues have different overall percentages of adenylation. It is also worth noting that these data were produced from *Populus trichocarpa*, and not *Arabidopsis*.

## Mechanisms of turnover of miRNA

MiRNAs are regarded as inherently stable molecules; on average they are ten times more stable than mRNAs (Gantier et al., 2011). However, there is a growing body of evidence which shows that individual miRNAs, or miRNAs in

specific environments, undergo dramatically accelerated decay which alters their levels and activity (reviewed in (Rüegger and Großhans, 2012). These miRNAs are not to be confused with miRNA\*s, which have been shown to exhibit fast or `ultra-fast` turnover kinetics in cell lines (Guo et al., 2015). Rather, these are mature miRNAs which undergo accelerated turnover that correlates with specific cellular events such as progression through the cell cycle (Rissland et al., 2011) or neuron-mediated environmental responses in animals such as changing of light conditions (Krol et al., 2010). Specific mechanisms were also suspected to be involved in the degradation of mature miRNAs different to standard RNA decay machinery, as depletion of the exosome complex does not affect miRNA levels (Chekanova et al., 2007).

### Exoribonucleolytic degradation of miRNAs – miRNA trimming

Active degradation of miRNAs was first reported in *Arabidopsis*, and was shown to be mediated by a family of 3'- 5' exoribonucleases called the small RNA degrading nucleases (SDNs) (Ramachandran and Chen, 2008). Simultaneous knockdown of all three SDN family members (SDN1-3) resulted in increased miRNA accumulation and pleiotropic developmental defects. SDN1 has been singled out as the most crucial of the family members for miRNA degradation, as it was found to act specifically on single stranded miRNAs, yielding degradation products of 8 – 9 nucleotides (Ramachandran and Chen, 2008).

This is far from a complete story of miRNA turnover, however. In the same study by Ramachandran and Chen, it was found using miR173 that SDN1 mediated degradation is attenuated by the addition of the 2'-O methyl mark which is present on almost all plant miRNAs ever characterised (Li et al., 2005). Additionally, miRNAs that had lost their 2'-O methyl group and had undergone 3'-end uridylation also had substantially reduced degradation by SDN1, leading the authors to conclude that 3' uridylation could actually have a protective role against exonucleolytic degradation by SDN1 (Ramachandran and Chen, 2008). As such it appears that a different and uncharacterised enzyme or set of enzymes are responsible for the degradation of 3' uridylated miRNAs. Additionally, for SDN1 to play a role in fast miRNA turnover, it is highly likely that the 2'-O methyl mark would need to be removed as SDN1 mediated degradation is significantly slowed by its presence. This would call for a

mechanism in which the 2'-O methyl group is removed, but no 3'-end uridylation occurs as this also attenuates SDN1 mediated degradation. More recent work has verified that, in at least 15 of the 43 most abundant miRNAs found in Arabidopsis AGO1 immuno-precipitates, incubation with SDN1 resulted in an increase in 3' truncation compared to a non-catalytic SDN1 mutant (Yu et al., 2017), which implies other exonucleases besides the SDNs must act to degrade other miRNAs, or that the SDNs are not fully redundant.

ATRIMMER2 is another gene which has been shown to degrade miRNAs. However, unlike the SDNs, this exonuclease operates on unmethylated miRNAs and miRNA\*s, and does not appear to have much effect on total mature miRNA levels. However, what is interesting about *atrimmer2* plants is that many miRNA\*s have increased abundance, potentially implicating ATRIMMER2 in miRNA\* degradation. It also localises with AGO1, which is consistent with what would be expected of a miRNA degrader (Wang et al., 2018). It is not yet clear where exactly ATRIMMER2 fits into the miRNA decay pathway.

#### Non-templated nucleotide addition to miRNA – miRNA tailing

As well as protective modifications such as adenylation, a miRNA can also undergo degradation following non-templated nucleotide addition. In *hen1* mutants, in which miRNA accumulation is substantially reduced, extracted miRNAs were extensively found to be uridylated at their 3' ends (Li et al., 2005). The *HEN1 SUPPRESSOR1 (HESO1)* was later found to be responsible for this uridylation using the same *hen1 Arabidopsis* mutants (Zhao et al., 2012a). In a *hen1* background overexpression of HESO1 resulted in a further reduction in miRNA accumulation and more severe developmental defects (Ren et al., 2012). When HESO1 was mutated to a loss of function variant (*heso1-1*) in a *hen1* background miRNA accumulation increased and there was a partial rescue of the phenotype seen in *hen1* (Zhao et al., 2012a). As the *hen1* phenotype was only partially rescued, and there were still 3'-uridylated miRNAs present in *heso1* loss of function mutants (Zhao et al., 2012a), it indicated that there was at least one other Terminal Uridylyl Transferase (TUTase) in *Arabidopsis* which uridylates miRNAs.

This was validated with the discovery of another TUTase; UTP:RNA uridylyltransferase (URT1), which is a functional paralog of HESO1 (X. Wang et al., 2015). This appears to be the only other TUTase which acts on miRNA as miRNA uridylation is globally abolished in *hen1 heso1 urt1* triple mutants (X. Wang et al., 2015). Uridylation of miRNAs by both HESO1 and URT1 is completely inhibited by methylation at the 2'-O position (Li et al., 2005), however as most plant miRNAs appear to be 3'-methylated (Yu et al., 2005) it is unclear what the significance of uridylation-mediated miRNA turnover in a wild-type background is. URT1 appears to act cooperatively with HESO1 *in vivo*. Tu et al showed that URT1 and HESO1 have distinct effects on different forms of the same miRNA by quantifying the tailing of full length miRNAs and truncated miRNAs in null *urt1-1* and *heso1-1* mutant backgrounds (Tu et al., 2015). Using miR158 as an example, they found that only HESO1 would tail full-length miRNAs and URT1 could add one nucleotide to one-nucleotide truncated miR158, which would then undergo further tailing by HESO1 (Tu et al., 2015). MiR158 is only partially methylated in the wild type (Zhai et al., 2013), and it was found that HESO1 and URT1 behave the same way in the wild type background (Tu et al., 2015). Both HESO1 and URT1 were shown to act on AGO1-bound miRNAs and have a considerable preference for UTP over other nucleotides, however the two enzymes have different substrate specificities, with URT1 preferring miRNAs ending with A and HESO1 preferring U ending miRNAs (Tu et al., 2015).

When comparing *heso1-1* and *urt1-1* in *hen1-8* backgrounds, HESO1 appeared to have a much more dramatic effect on a wider range of miRNA targets than URT1, which was further substantiated by the observation that whilst *heso1-1* mutation could partially rescue the *hen1-8* phenotype, *urt1-1* could not (Tu et al., 2015). In addition to differences in substrate preference, it is likely that URT1 and HESO1 function at different levels *in vivo*. As HESO1 adds one U to the 3' end of a miRNA, that miRNA becomes a preferred substrate of HESO1 irrespective of its terminal 3' nucleotide. This led Tu et al to conclude that HESO1 likely holds onto the product of this first U addition and uses it as a substrate, suggesting HESO1 could be a processive enzyme. This is not true for URT1, which prefers miRNAs which end in A at their 3' end. A miRNA ending in U is no longer a good substrate for URT1, and so it likely releases the product of its

reaction. This suggested that URT1 is likely not a processive enzyme (Tu et al., 2015). In the *Arabidopsis* genome, there are 10 total putative nucleotidyl transferases including HESO1 and URT1; it is not yet known if any of the other 8 can act on methylated miRNAs where the two TUTases cannot (Tu et al., 2015).

### Interplay between truncation and tailing of miRNA

In addition to the extensive uridylation at the 3' end of miRNAs in *hen1* mutant backgrounds, it was also observed that the miRNAs were heterogeneous in size (Li et al., 2005). Closer interrogation of the system showed these 3' modifications differed significantly between different miRNA families, as well showing that in many cases, heavy truncation of miRNAs had occurred from the 3' end (Zhai et al., 2013). Zhai et al's observations also suggested that this 3' truncation occurs before tailing, as the addition of uridine to the 3' end of the miRNAs did not occur until the miRNA was shortened past a certain length. Whilst these observations were made in a *hen1* background, it was shown that miRNAs which are only partially methylated in the wild type such as miR158 displayed substantially reduced tailing in *heso1-1* mutants (Zhai et al., 2013). This order of events, in which truncation occurs before tailing, was further substantiated by Yu et al. Using published sRNA libraries from *hen1* backgrounds, and *hen1 heso1* backgrounds, along with novel sRNA libraries from *hen1 sdn1 sdn2* triple mutant backgrounds, and then looking at ratios of truncation and tailing in these libraries, they observed that in the *hen1 heso1* libraries, in cases in which tailing was compromised, the level of truncated and tailed miRNAs decreased, however the levels of truncated-only sRNAs increased (Yu et al., 2017). Additionally, 9 of the 10 miRNAs which showed a significant reduction in 3' truncation in the *hen1 sdn1 sdn2* background also showed an increase in truncated only miRNAs in the *hen1 heso1* libraries. This further suggested that 3' truncation occurs before tailing.

*Ago1* mutants have shown that both uridylation and truncation are suppressed in the absence of functional AGO1, but not in AGO1 slicer defective mutant backgrounds, suggesting that these processes occur after AGO1 loading (Zhai et al., 2013). It should be noted that these mutations were in a *hen1* background, and so the miRNA would have lacked 2'-O methylation. In the

absence of *hen1* mutation, conventional wisdom has AGO loading conferring protection to miRNAs from degradation (Winter and Diederichs, 2011), Ji and Chen, 2012), yet it is clearly also required for controlled decay. A diagram of this model of miRNA degradation when AGO1 loaded is depicted in Figure 4.

Investigation into miR166 and miR163 which showed differential truncation followed by uridylation back to their original length, suggests that the pattern of truncation and uridylation is determined by the miRNA sequence. This also looked to be true of miRNAs longer than the general 21nt plant miRNAs, such as the 24nt miRNA163, which is truncated, and then tailed back to 24nt with uridine residues. Despite these observations, it is not completely clear if miRNAs are tailed back to their full length, and not tailed to a greater length, as the method by which the sRNA libraries were generated involves size selection on a PAGE gel, which introduces the potential for biasing of size. It also appeared to be the 3' end of the sequence specifically which determines which modifications the miRNA will undergo (Zhai et al., 2013).

Despite common machinery, different populations of miRNAs have been shown to have different turnover kinetics (Guo et al., 2015). Whilst it is likely that this difference is attributable to sequence, it could be the result of either different sequences taking longer to degrade by the common machinery, or by different miRNAs being entered into the decay pathways differentially. One example of this latter scenario has been demonstrated by Yu et al. with miR165/6 (Yu et al., 2017).

AGO10, which is the AGO protein most closely related to AGO1 (Ji et al., 2011), had previously been shown to repress miR165/6 accumulation by sequestration (Zhu et al., 2011). This repression was also shown to be important for the maintenance of the shoot apical meristem and floral meristem (Yu et al., 2017). Yu et al. showed that AGO10 competes with AGO1 for a subset of miRNAs, with highest affinity for miR165/6. They subsequently showed that in addition to this competition and sequestration of miRNAs, miRNAs that were bound by AGO10 were more susceptible to SDN mediated 3' truncation (Yu et al., 2017) than AGO1 bound miRNAs. Therefore, this represents one particular scenario

whereby a miRNA is differentially entered into decay pathways relative to other miRNAs, which therefore affects the overall kinetics of the miRNA.

### Target Directed miRNA Decay

Another emerging mechanism by which miRNAs may be specifically entered into miRNA decay pathways is known as Target-Directed miRNA Degradation (TDMD). This phenomenon occurs when AGO-bound miRNAs bind to highly complementary target sites. This results in trimming and tailing of the miRNA (Ghini et al., 2018), made possible by the release of the 3' end by AGO which then is accessible to enzymatic processing (Sheu-Gruttadauria et al., 2019). This mechanism of miRNA decay has predominantly been observed in animals. This is likely due to the fact that animal miRNAs have much lower sequence complementarity than plant miRNAs, and additionally lack the protective 2' O methyl group that plant miRNAs possess (Fuchs Wightman et al., 2018). However, there are instances of this method of decay being functionally relevant in plants as well, for example with the Induced by Phosphate Starvation1 (IPS1) gene which acts as a regulator of P<sub>i</sub> homeostasis.

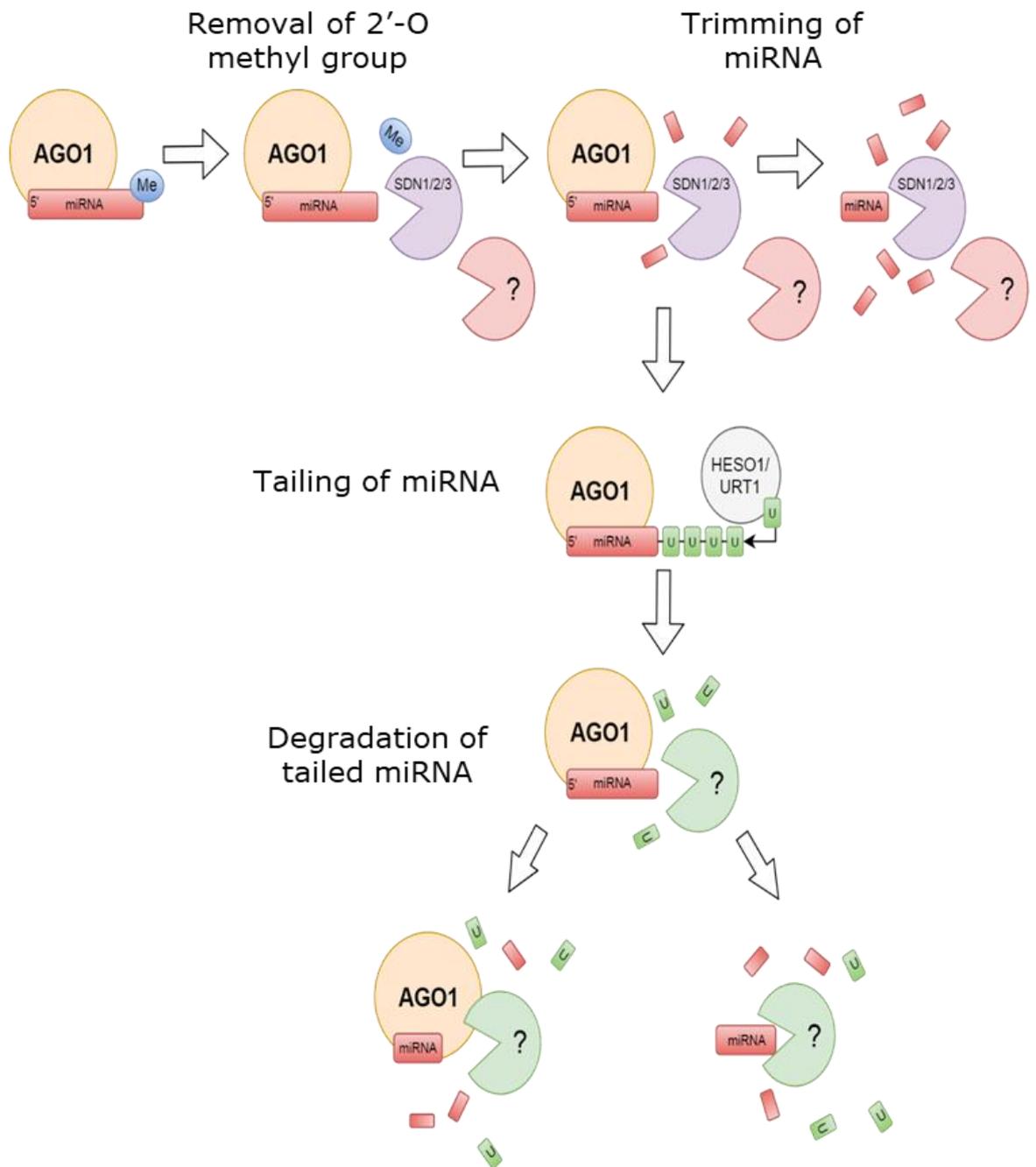
IPS1 RNA contains a target site for miR399 which is upregulated in phosphate starvation and targets the UBC, PHO2 (Fujii et al., 2005). However, the IPS1 miR399 target site contains a three-nucleotide bulge between the 10<sup>th</sup> and 11<sup>th</sup> position of miRNA399 which prevents cleavage. This target site competes with the PHO2 miR399 target site however, as IPS1 RNA cannot be cleaved, it sequesters the miRNA and results in an increased accumulation of PHO2 mRNA (Franco-Zorrilla et al., 2007). This leads to a decrease in miR399 levels.

### Potential miRNA turnover candidate genes

Looking at miRNA turnover in other kingdoms of life has suggested several potential genes which may play a role in uridylylated miRNA turnover in plants. In mammals, DIS3-like exonuclease 2 (DIS3L2) is a 3'-5' exoribonuclease which degrades oligouridylylated pre-let-7 *in vivo*, and has an orthologue in *Arabidopsis*. It should be noted however that this enzyme in mammals targets pre-miRNA; in animals pre-miRNA is processed in the cytoplasm by Dicer (Zhang et al., 2004), however in plants the pri-miRNA and pre-miRNA are all

processed in the nucleus by DCL1 (Kurihara and Watanabe, 2004). Recent work has demonstrated that, in mammals, DIS3L2 can indeed degrade certain populations of uridylated miRNAs (Yang et al., 2020). This gene has a homologue in plants, which is known as Suppressor of Varicose (SOV) (Zhang et al., 2010). However, despite possessing many of the characteristics required of a degrader of uridylated miRNAs, in Col-0 SOV is non-functional, which makes it unlikely to be a core component of the miRNA degradation machinery. Additionally, research performed during the course of this PhD into this gene was not able to show any differences in the accumulation of mature miRNAs in Col-0 backgrounds complemented with functional SOV (data not shown).

In *Chlamydomonas* a subunit of the exosome complex, Ribosomal RNA-processing protein 6 (RRP6), degrades uridylated small RNAs *in vivo*; when knocked down miRNA and siRNA accumulation increases (Ibrahim et al., 2010). Whilst this degradation pathway was suggested by the authors of the paper to be a quality control mechanism to remove damaged or dysfunctional small RNA, *Arabidopsis* has three RRP6 genes which could function in a similar way in miRNA turnover.



**Figure 4 – Schematic diagram showing known and unknown stages of mature miRNA decay**

Diagram depicting potential degradation pathways of mature, AGO1 loaded miRNAs. Question marks depict unknown enzymes.

Mature miRNAs are AGO1 loaded in plants. These are protected by a 2'-O methyl group at the 3' end. This mark is removed by SDN family proteins and potentially other unknown enzymes. The exonuclease continues to degrade the miRNA until it reaches the portion protected by the AGO protein. At this point, some miRNAs may be released by the AGO1 and degraded completely by the exonucleases. In other cases, the exonuclease may drop off when it encounters AGO1. Following this, as the methyl group has been removed, non-templated nucleotide addition occurs by the action of TUTases such as HESO1 and URT1. These newly tailed miRNAs then undergo degradation by an as yet undescribed enzyme either in association with AGO1 or released.

## Alternative Splicing

A growing body of evidence demonstrates an involvement between miRNA biogenesis and alternative splicing. The term `alternative splicing` is used to describe splicing when a decision is made between a multitude of splice sites within a transcript (Reddy et al., 2013). This results in multiple transcript isoforms which derive from the same genomic locus. This differs from constitutive splicing, where no decision is made and the same splice sites are always used, resulting in only one transcript from a gene.

Splicing is itself commonplace in Eukaryotes, with *Arabidopsis* being no exception. In *Arabidopsis* up to 90% of the genes encoded within its genome contain introns, which must be removed in order to produce a functional, protein-coding transcript (Shang et al., 2017). During splicing, regions inherent in the transcript, called splice sites, are recognised by a collection of splicing factors. Most well described of these are small-nuclear RNAs (snRNAs) U1, U4, U5 and U6 (Shang et al., 2017). These sites are defined as 3' splice sites and 5' splice sites, and mark the beginning and end of a pre-mRNA. There also exist non-canonical splice sites which are emerging as RNA sequencing depth increases, and by definition these differ from `canonical splice sites` (Sibley et al., 2016).

In *Arabidopsis*, between 42-61% of genes are known to undergo alternative splicing. This is an important mechanism for increasing the functional diversity of the genome, as each gene can potentially encode multiple transcripts with different functions. There are a number of different types of alternative splicing events, which can have completely different end-results. In some cases, different permutations of functional domains can be assembled to produce proteins with disparate functions. In other cases, it can result in the disruption of reading frames or introduction of premature stop codons, resulting in nonsense mediated decay (NMD), a highly controlled mRNA degradation pathway used to remove transcripts with premature stop codons (Ner-Gaon et al., 2004). Alternative splicing can also open up or liberate transcripts from miRNA based regulation, by including or excluding regions containing miRNA target sites (X. Yang et al., 2012). As many as 45% of alternatively spliced

genes in *Arabidopsis* show evidence of regulation by NMD, indicating that this method of alternative splicing control is widespread (Kalyna et al., 2012).

An example of alternative splicing resulting in varied functional domain permutations, as well as a biological process which relies on alternative splicing for its functionality, is the regulation of flowering locus M (FLM). This is crucial for ambient temperature and vernalisation-mediated flowering (Posé et al., 2013). FLM has two splice isoforms whose proportions alter in response to temperature changes, named FLM- $\delta$  and FLM- $\beta$ . The two isoforms compete for interaction with the floral repressor SVP, which represses flowering – this repressor relies on FLM for DNA binding. However, only one of these isoforms (FLM- $\beta$ ) can interact with DNA; FLM- $\delta$  cannot interact with DNA due to a missing functional domain. At higher temperatures, FLM is alternatively splice so that FLM- $\delta$  is the predominant isoform, which outcompetes FLM- $\beta$  for SVG binding and results in FLM- $\delta$ /SVG complexes which can't bind the DNA. This results in a de-repression of flowering in elevated temperatures (Posé et al., 2013).

There is also a strong association between alternative splicing and stress response in plants. A wide range of studies in various plant models have shown dramatic differences in alternative splicing profiles in response to abiotic stresses, such as cold (Calixto et al., 2018), drought (Chong et al., 2019), salt (Ding et al., 2014) and heat (Keller et al., 2017).

The most common type of alternative splicing in plants is intron retention (Reddy et al., 2013). In this type of alternative splicing, an intron which would usually be spliced out is retained in the resulting transcript. A second type is exon skipping, which is the most common type of alternative splicing in animals; it does still occur in plants but at a lower level (Chaudhary et al., 2019). As the name suggests, this mechanism of alternative splicing involves the skipping of one or more exons in a transcript, such that an exon that is skipped will not be present in the final mRNA transcript. Finally, alternative 3' splice site usage and alternative 5' splice site usage are the last two common types of alternative splicing. These are conceptually the same, but are relative to the 3' splice site or 5' splice site respectively. Either of these can result in transcripts which are

truncated or extended, and involve the masking of some splice sites in favour of others (Y. Wang et al., 2015).

Splice site masking, such as that seen in alternative 3' and 5' splice site usage, is a common way that alternative splicing `decisions` are made in plants. Put simply, this involves the covering of splice sites which would otherwise be used by binding proteins, such as the serine/arginine rich (SR) proteins (Cruz et al., 2014). These have been found to be highly abscisic acid (ABA) responsive, providing a basis for the strong relationship between alternative splicing and ABA which has long been observed (Laloum et al., 2018). ABA and abiotic stress signalling have also long been associated, which provides further mechanistic clues as to how stress related alternative splicing decisions are made (Tuteja, 2007). Because these splice sites are masked by the obstructing protein, the next available splice site is used (Duque, 2011).

However, the splice sites found in introns are both small and degenerate, and not able to convey the necessary information to control these decision-making events. Bioinformatic analyses which focussed on overrepresented sequences in introns and exons thus were used to identify any other sequence motifs which could assist in these decision-making events. These studies identified a number of cis regulatory elements (CREs) which appear to have a hand in conveying the necessary information (Chasin, 2007). There also exist a plethora of RNA binding proteins and accessory proteins which form complexes that can recognise these splice sites and CREs, known collectively as splicing factors (Dvinge, 2018). Many of these factors have been found to be involved in many other processes, and they appear to be at least twice as numerous in plants than in animals (Reddy et al., 2013).

There is also emerging an increasingly intimate connection between alternative splicing, splicing and miRNA biogenesis. In some instances, this can be because some miRNAs are encoded within introns, and therefore rely on splicing machinery to be liberated from the progenitor transcript (G. D. Yang et al., 2012). Another way in which alternative splicing can influence miRNA production and expression is exemplified in miR846 and miR842 in *Arabidopsis*. These miRNAs form a cistronic miRNA pair, which are alternative splicing isoforms of

the same transcript. Therefore, alternative splicing decisions dictate which of these two miRNAs is expressed at a given time, in an ABA-regulated process (Jia and Rock, 2013).

There are also many examples of splicing factors interacting directly with miRNA biogenesis machinery. One such example of this is MAC7, an RNA helicase which forms part of a larger MAC complex. In *mac7*, there are global intron retention defects, supporting that this component is involved in splicing. Additionally, there is a decrease in the accumulation of many miRNAs, as well as a reduction in pri-miRNA levels. However, for these miRNAs, the half-life of the pri-miRNA transcripts is the same as WT, and their promoter activity is also the same (Jia et al., 2017). The authors of this study also found that in *mac7*, HYL1 did not localise to the dicing bodies in the nucleus, suggesting that MAC7 facilitates this event. This links MAC7 to miRNA processing.

MiR400 also represents an example of the involvement of alternative splicing in miRNA processing. This miRNA exists within the intron of a protein coding gene, and when plants undergo heat stress, the intron is retained by the host gene due to a heat specific alternative splicing event. This results in a relative increase in pri-miR400, but a reduction in the levels of mature miR400 (Yan et al., 2012).

## Other forms of PTGS siRNAs

In addition to silencing by miRNAs, there are many other known species of sRNA which can cause silencing. The main overarching class of these sRNAs are siRNAs, which differ from miRNAs in several distinct ways. There is a general difference in the types of targets that miRNAs and siRNAs regulate. Typically, miRNA appear to regulate endogenous genes, and siRNAs appear to protect genomic integrity from foreign elements such as viruses, transgenes and transposons (Carthew and Sontheimer, 2009). These are not without exception, but represent the broader, general specialisation of these sRNA species. In all of these pathways, the sRNA species acts primarily as a guide, ensuring that the corresponding nucleotide sequence is targeted by the process, be it endonucleolytic cleavage or the protection of genome integrity (Vaucheret, 2008). The process itself is determined by the AGO protein that the sRNA is

loaded into; different AGO proteins perform different functions, and therefore the type of regulation a sRNA enacts is determined by which AGO it is loaded into. MiRNAs are typically loaded into AGO1 in *Arabidopsis*, which has slicer activity and therefore enacts endonucleolytic cleavage most commonly (Baumberger and Baulcombe, 2005). SiRNAs are typically loaded into a greater diversity of AGO proteins, such as AGO2, AGO4 and AGO5. The decision as to which of these an siRNA is loaded into is determined primarily by the 5' terminal nucleotide (Mi et al., 2008).

The range of sizes of siRNAs is wider than that of miRNAs; miRNAs are typically 21nt long in plants, whereas siRNAs span from 20 – 25nt in length (Kasschau et al., 2007). The reason for this difference is that there are a number of different sub-species of siRNA which arise from different biogenesis pathways. While miRNAs are typically DCL1 products, whose biochemistry determines their length, siRNAs can be the product of a number of different DCLs, such as DCL3, which produces 24nt species (Kasschau et al., 2007), DCL2 which produces 22nt species, and DCL4 which produces 21nt species (Parent et al., 2015). There are also DCL independent routes of siRNA biogenesis (Ye et al., 2016).

In addition to differences in DCL usage in the biogenesis of siRNAs, these sRNA species are transcribed by a different RNA polymerase than miRNAs. As discussed earlier, miRNAs are RNA pol II products (Lee et al., 2004), however siRNAs are transcribed by RNA pol IV and to a lesser extent RNA pol V (Zhou and Law, 2015). The origin of many of these siRNAs differ from miRNAs in that they can be considered `secondary` transcriptional products. This is because transcripts from which they derive are single stranded, however they must be made double stranded before siRNAs can be produced. One of the ways this is often achieved is via an RDR activity, such as RDR6 coupled with SGS3 (Peragine et al., 2004). MiRNAs do not have this requirement.

SiRNAs are far more numerous than miRNAs in plants. This can be inferred from sRNA library size distributions which show the highest peak at 24nts. However, many standard techniques which demonstrate the quantity of sRNA, such as sRNA northern blot, may fail to convey this as where miRNAs are high abundance, single sequences which can be hybridised to a single sRNA probe,

siRNAs can derive from many points along the same transcript or template, and therefore are not a single sequence. Each individual siRNA sequence may recur far fewer times in a cell than a miRNA, but the combination of all of these many sequences far exceeds miRNA levels. (Vazquez and Hohn, 2013).

In addition to the earlier described methylation of miRNA/miRNA\* duplexes by HEN1, siRNAs also undergo 2'-O methylation by HEN1. This was demonstrated in RNA pol IV knockouts in the *hen1* background, which were found to show a partial rescue (Yu et al., 2010). It also appears likely that many of the same components which degrade mature, AGO loaded miRNAs also degraded AGO loaded siRNAs. For example, knockdown of SDN1/2/3 resulted in an observable increase in a number of highly expressed siRNAs in *Arabidopsis* (Ramachandran and Chen, 2008). In studies using combinations of *hen1* and *hen1 heso1* backgrounds there also appeared to be a reduction in the levels of 3' tailing of siRNAs which were examined in *hen1 heso1*, suggesting that HESO1 also tails AGO bound siRNAs as well as miRNAs (Zhao et al., 2012b).

There are a number of significant sub-populations of siRNA which are distinguished by length, biogenesis and function.

### Virus derived siRNAs

SiRNAs also represent a significant arm of plant antiviral response. Indeed the first studies describing antisense sRNAs identified them as part of a defence mechanism against viruses and transgenes (Hamilton and Baulcombe, 1999). In plants, both RDR1 and RDR6 are used to create double-stranded templates of virus derived genes or virus genome. When these are suppressed, which is a strategy employed by some viruses such as Cucumber Mosaic Virus (CMV), there is a substantial reduction in the number of 21, 22 and 24nt long virus derived sRNAs (Diaz-Pendon et al., 2007).

Many of these virus derived siRNAs appear to be DCL4 dependent (Bouché et al., 2006). DCL4 produces 21nt sRNAs (Xie et al., 2004), and as such these are most likely loaded into AGO1 (Mi et al., 2008), and enact silencing by endonucleolytic cleavage or translational repression in the same way as miRNA-

mediated repression. Mutants in AGO1 have reduced virus resistance, which supports this (Morel et al., 2002).

As the years have progressed, and further contributions to the field of study have been made, many other subfamilies of siRNA have been uncovered which serve a myriad of functions.

### Heterochromatic siRNAs

Heterochromatin is one of the two forms of organisation of a genome, the other being euchromatin. While euchromatin contains a majority of the genes, heterochromatin contains a high density of repeat regions and transposable elements. As a result of this, heterochromatin is maintained in a densely packed, silenced state, which relies heavily on hc-siRNAs (heterochromatic siRNAs) (Tamaru, 2010). These are not unique to plants and are found in a variety of other species in different kingdoms of life, such as *Drosophila* (Fagegaltier et al., 2009) and *Schizosaccharomyces pombe* (Kloc and Martienssen, 2008).

These siRNAs are produced when RNA pol IV products are converted into double stranded RNA species by RDR2, which are then processed by DCL3 (Onodera et al., 2005) (Pikaard, 2006). These 24nt hc-siRNAs are then loaded into AGO4, which enacts RNA directed DNA methylation (RdDM); this results in *de novo* methylation at cytosine residues in the DNA regions that the hc-siRNA guides the AGO4 complex to (Zilberman et al., 2003) (Qi et al., 2006). This methylation is not achieved by the siRNA:AGO4 complex alone, but rather by a larger complex containing AGO4. This complex assembles on nascent transcripts – guided by the siRNA – and then further proteins are recruited to enact the methylation, such as Domains Rearranged Methyltransferase 2 (DRM2) (Wierzbicki et al., 2009).

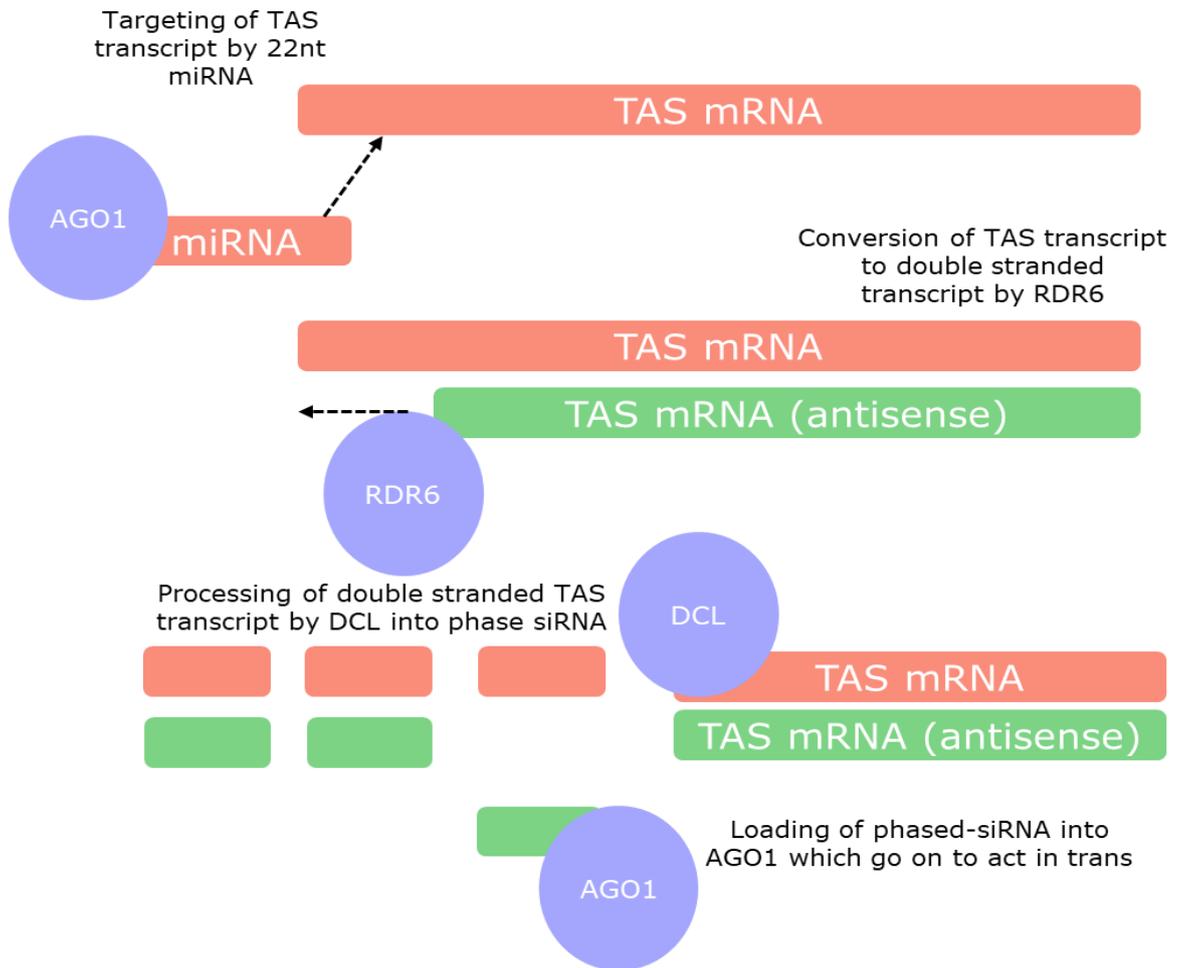
The same studies that describe *de novo* DNA methylation as a result of RdDM also document histone methylation arising as a consequence of this method of regulation. The combined result of these methylation events results in alteration of the chromatin architecture, resulting in tightly packed heterochromatin. This type of siRNA regulation is referred to as Transcriptional Gene Silencing (TGS), rather than PTGS, as the level of regulation occurs at the transcriptional stage.

Many elements such as transposons, or transposable elements (TEs) are found within these silenced, heterochromatic regions. Transposons are genetic parasites which are capable of self-propagation when expressed, followed by re-insertion into the host genome. This re-insertion can have major fitness implications, as it can occur within a gene thereby disrupting it and effectively knocking it out (Underwood et al., 2017). As such, it is of vital importance that these elements are maintained in a transcriptionally silenced state, and justifies the existence of RdDM targeting of these elements.

### Trans-acting siRNAs

Trans-acting siRNAs (tasiRNA) are another category of siRNA. These siRNAs derive from TAS (Trans-acting siRNA) gene transcripts, and work in trans to repress specific mRNAs, much like miRNAs (Vazquez et al., 2004). The production of tasiRNAs from TAS transcripts first requires the targeting of the transcript by a miRNA, such as miR390 (de Felippes et al., 2017). There are three described TAS transcripts in *Arabidopsis*, usefully named TAS1/2/3. Following this miRNA binding event, phased production of 21mer tasiRNAs is initiated, which are dependent on RDR6, SGS3 and DCL4 (Peragine et al., 2004)(Xie et al., 2005b).

These tasiRNAs are involved in a number of different pathways. The best characterised pathway involves repression by TAS3 of Auxin Response Factor 3 (ARF3), which suppresses juvenile-to-adult phase transition in *Arabidopsis*. Loss of tasiRNAs in *rdr6* or *dcl4* backgrounds results in accelerated phase change and patterning defects in leaves and floral organs (Fahlgren et al., 2006) (Hunter et al., 2006). The mechanism of tasi-RNA biogenesis is illustrated in Figure 5.



**Figure 5 Schematic representation of tasi-RNA biogenesis**

Diagram depicting the production of tasi-RNA from TAS transcripts.

TAS transcripts are targeted by 22nt miRNA. This targeting triggers the conversion of the TAS transcript to a double-stranded RNA by RDR6. This double-stranded RNA species is recognised by DCL proteins resulting in the production of phased siRNAs, which are then loaded into AGO1 and go on to regulated other genes *in trans*.

## RNA quality control siRNAs

One subfamily of siRNA of particular interest to this thesis is the RNA quality control siRNA (rq-siRNA / rqc-siRNA) family. Studies into the connection between PTGS and RNA quality control pathways such as NMD first alluded to the existence of these siRNA species. Experiments in this study by Moreno et al. (2013) used NMD mutants *upf1* and *upf2*, both of which are required for NMD to occur (Kashima et al., 2006), polyadenylation mutants PARN and CCR4a/b, and exosome factor mutants such as RRP4/41/44A. These various RNA decay pathway mutants were transformed using a transgene reporter, and the levels of PTGS in them was quantified by counting the level of repression of the transgenes. In these mutants, the levels of PTGS of the transgenes relative to WT backgrounds transformed with only the transgene but without the NMD mutations, increased by as much as 78%. This suggested that many of these different RNA decay pathways inhibit the levels of PTGS, and implied the existence of what they dubbed a `tug of war` between RNA quality control pathways and PTGS (Moreno et al., 2013). They also assayed the levels of the TAS1 and TAS2 loci by RNA gel and found no difference in the levels between the control and mutant backgrounds, suggesting that the nature of this PTGS was unlike the previously described RDR6 SGS pathway (Dalmay et al., 2000) (Yoshikawa et al., 2005). However, this study did not employ the use of sRNA sequencing, and so the nature of this PTGS remained unknown.

Following this, this species of siRNA was first directly described in mRNA decapping mutants such as *vcs* and *dcp2*. Decapping is an irreversible step of mRNA degradation which typically occurs within cytoplasmic foci called P bodies. It is an elaborate process involving a complex of many constituents, and is required for the appropriate 5'-3' decay of mRNAs by XRN4 (Xu et al., 2006). There is some redundancy in this process but when core functionality is lost, postembryonic development is inhibited and the phenotype is lethal.

In *vcs* and *dcp2*, a large number of endogenous transcripts were found to produce more sRNAs than WT backgrounds in sRNA sequencing data produced from them. In *vcs*, around 1250 transcripts did this, and in *dcp2*, 1351. These sRNAs appeared to be the product of an RDR activity, as they mapped both in

sense and antisense to the genes from which they derived in a roughly 50/50 split (Martínez de Alba et al., 2015). These mutants had severe phenotypes which did not pass the two-cotyledon stage, but following introgression of *rdr6* into the backgrounds there was a degree of phenotypic rescue. Accompanying this reduction in severity of phenotype, in both *rdr6/vcs* and *rdr6/dcp2* roughly 1/5<sup>th</sup> of the transcripts producing more sRNAs in the single mutants, no longer did so in the double mutants. This implied that the sRNAs deriving from these transcripts which changed were RDR6 dependent in nature. Additionally, when the size distribution of these siRNAs was analysed the largest peak was at 21nt, which is consistent with RDR6/DCL4 products (Taochy et al., 2017). There is also a second peak at 24nt, however the introgression of *rdr6* does not appear to influence these peaks.

In this study by Martínez de Alba et al. (2015), the authors make several interesting observations with regard to the localisation of these processes. First of all, they demonstrate that VCS and DCP2 localise to cytoplasmic P bodies, which had been previously shown. They also show that RDR6 and SGS3 localise to cytoplasmic siRNA bodies, again which had previously been shown. The novel observation, however, was that these P bodies and siRNA bodies in the cytoplasm were colocalising, which raised the possibility that cross-talk was occurring between these two types of bodies. This was the first paper which described the so-called RNA quality control siRNAs, which they abbreviated to rqc-siRNA. The model proposed by the authors of this study was that decapping prevented the entry of aberrant or dysfunctional RNAs in P bodies into siRNA bodies. They suggest that in decapping mutants, P bodies become saturated with transcripts which can't be efficiently degraded as the 5'-3' degradation pathway is unable to process them. As a result of this, the 'overflow' of transcripts are passed to the siRNA bodies, and deleterious siRNAs are produced from them, which result in the degradation of the transcript. Therefore, in a wild type background, in a situation in which a large number of aberrant transcripts are being produced, for example in viral infection (Garcia et al., 2014), the siRNA bodies act in concert with the P bodies to maximise the degradation capacity. As demonstrated in the study by Moreno et al. (2013), there is also competition between these pathways, as knockdown of RNA decay pathway components results in increases in PTGS.

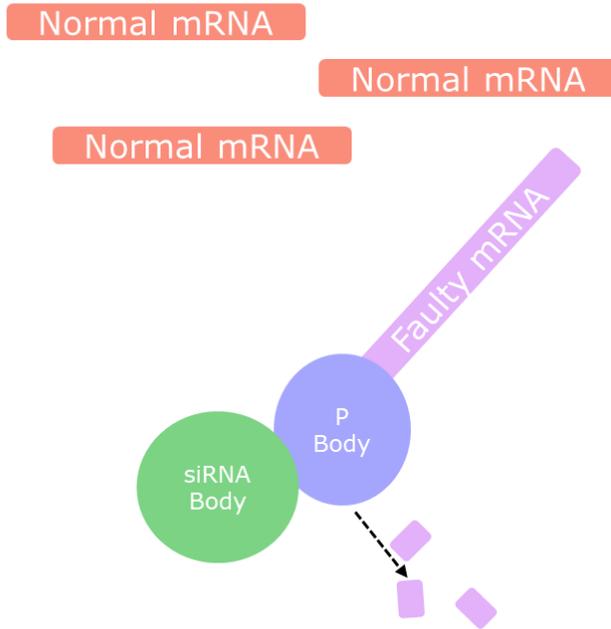
The clear problem with this manner of control is that whilst aberrant transcripts are directed to P bodies for degradation in a specific manner, siRNAs produced are loaded into AGOs which go on to repress further transcripts. The only criterion for a match between the siRNA loaded AGO and the target transcript is the guide siRNA sequence, and so it is not possible for this mechanism of repression to differentiate between aberrant and functional transcripts. Therefore, if endogenous transcripts enter into this pathway, it can potentially result in inappropriate silencing of transcripts simply because some of them have aberrations. A schematic of rqc-siRNA biogenesis and consequences is depicted in Figure 6.

Finally, the most recently described gene involved in the production of these rqc-siRNAs is C-terminal Phosphatase-like 3 (CPL3) (Li et al., 2019). In this study, it was found, through the use of 3' RACE, that transgenes exhibited varying degrees of truncation which resulted in their entry into RNA decay pathways, concomitant with a high density of sRNAs which mapped to the transgenes. These sRNAs also appeared to derive from a number of endogenous transcripts, although at a much lower level than from the transgene.

A majority of the sRNAs which accumulated in *cp13* were 21nt long, with a second, much smaller subpopulation of 24nt sRNAs also observed. Of the many endogenous transcripts which accumulated more sRNAs relative to the control, only 56 appeared to be downregulated by the corresponding sRNAs, suggesting that many of these sRNAs either don't accumulate at levels high enough to enact silencing, or are not AGO loaded. This downregulation was exacerbated when *cp13* plants were exposed to stress conditions. Four genes which demonstrated measurable levels of repression by qPCR in *cp13* under stress were analysed again under the same conditions but in the *cp13/rdr6* background. In these double mutants, the repression of the transcripts appeared to be lost, or reduced when measured by qPCR, indicating that these sRNAs may in part be RDR6 dependent in nature (Li et al., 2019).

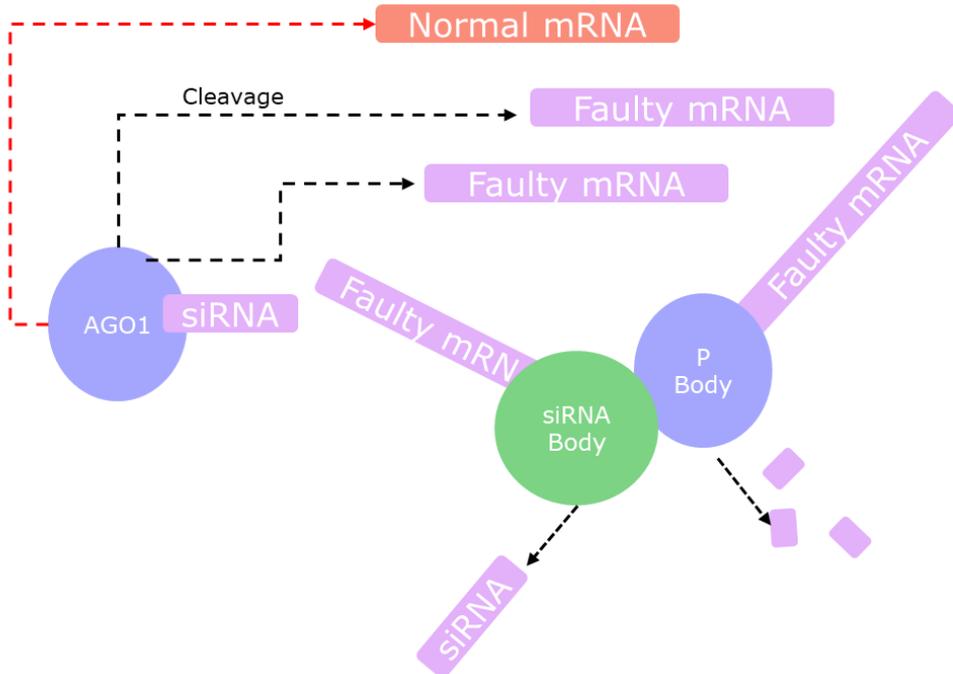
**A**

**Normal Conditions**



**B**

**NMD / Stress Conditions**



**Figure 6 - Schematic representation of production and consequence of rqc-siRNA**

Diagram depicting the production and consequence of rqc-siRNA.

**A:** Under normal conditions, occasional faulty transcripts are sent to the P body, where they undergo controlled and specific degradation.

**B:** Under stress conditions or in backgrounds with NMD phenotypes, there are too many faulty transcripts for the P body to process. Some are consequently fed into the associated siRNA body which produces siRNA. These go on to repress both faulty and non-faulty transcripts in a non-discriminatory way.

## Ribosomal RNA-derived siRNAs

Another recently described category of siRNA are the ribosomal RNA-derived siRNAs (risRNAs). This species of siRNA was discovered in mutants of *FIERY1*, which is involved in 5'-3' RNA degradation (You et al., 2019). In *fiery1*, a number of miRNAs had reduced abundance relative to the WT. However, sRNA sequencing of these mutants showed the presence of elevated peaks of sRNA of length 21nt.

There was a significant overlap between genes producing more sRNAs in *fiery1* with genes producing more sRNAs in the previously described *dcp2* and *vcs* mutants which the rqc-siRNAs were described in. However, there were also many genes in *fiery1* which produced more sRNAs that did not do so in either of these mutants. Indeed, over half of the sRNAs appeared to derive from ribosomal regions, which was not reported for *dcp2* and *vcs*, suggesting that these sRNAs should be classified differently. The authors of this study also observed that these ribosomal RNA-derived sRNAs were AGO1 loaded, which resulted in competition for AGO binding for miRNAs. This most likely explained the reduction in miRNA levels seen in *fr1*, and has not been reported in *vcs* or *dcp2*, further implying that the nature of the sRNAs produced in *fiery1* is different to rqc-siRNAs (You et al., 2019).

## Piwi-interacting RNAs

The third main class of sRNA is that of the piwi-interacting RNA (piRNA). These sRNAs are so named because they interact with PIWI proteins, which exist in animals but not in plants (Thomson and Lin, 2009). Therefore, by definition, piRNAs do not occur in plants and therefore will not be focussed on in this thesis. Briefly, piRNAs have been largely associated with germline development, epigenetic regulation, transposon silencing and translational control (Ku and Lin, 2014).

## CPLs

The previously mentioned CPL3 is part of a larger family of CPLs found in a wide variety of organisms (Archambault et al., 1997). In *Arabidopsis* there are four known CPLs (Koiwa et al., 2002). They dephosphorylate the C terminal domains of RNA polymerase II. All the CPLs in *Arabidopsis* contain domains which have previously been associated with the transcriptional elongation complex. A study by Koiwa et al (2002). demonstrated that CPL1 and to a lesser extent CPL3 reduce the rate of transcriptional elongation. They did this using a luciferase reporter under the control of a cold responsive promoter. In *cp1* and to a lesser extent *cp3*, the intensity of the luminescence was higher under cold conditions than the WT background transformed with the transgene. However, there was no corresponding increase in other cold responsive transcripts in these mutants, and a nuclear run-on analysis demonstrated there was no difference in the rate of transcriptional initiation. The most likely conclusion therefore is that the rate of transcriptional elongation was greater in *cp1* and *cp3* (Koiwa et al., 2002).

Whilst they share common catalytic interactions they have some non-overlapping, distinct *in planta* functions (Bang et al., 2006), such that different CPL family knockouts display distinct phenotypes. CPL1 and CPL3 isoforms have been found to regulate ABA signalling and osmotic stress response (Koiwa et al., 2002). CPL4 appears to be required for growth and development, as complete knockout of this gene is lethal. Suppression of it by RNAi has revealed a wide range of morphological phenotypes, such as incomplete cotyledon expansion, a slow rate of growth, stunted petioles and downward curling leaves (Bang et al., 2006). The same study compared alongside *cp3*, however they found no such lethality in this knockout. However, between the two, only *cp3* appeared to induce ABA hyperactivation.

CPL1 and CPL2 appear to be involved in miRNA processing and strand selection, something that CPL3 and CPL4 do not. RT-qPCR analyses of a number of different miRNAs such as miR156, miR164 and miR319 showed that the accumulation of these miRNAs was lower in *cp1* backgrounds. However, interestingly the same study was not able to corroborate this with RNA gel blots, which showed no clear difference (Manavella et al., 2012). This reduction in

miRNA accumulation in *cp1* was later attributed to a loss of dephosphorylation of HYL1 by CPL1.

A later study into the function of CPL1 demonstrated involvement with nonsense mediated decay in *Arabidopsis*. Specifically, CPL1 was found to interact with two NMD factors, eIF4AIII and UPF3 (2016). In this study, there were a number of alternatively spliced, 5' extended mRNAs which accumulated more highly in *cp1* than the control, which had NMD-eliciting features. This study identified CPL1 through a genetic screen, and as such only CPL1 of the CPL family were analysed in this way. It is not clear if there is any involvement between any of the other CPL family members and NMD or RNA decay. However, in the earlier mentioned study by Li et al. (2019), which demonstrated the accumulation of rqc-siRNAs in *cp3*, the authors were not able to detect any 5' extended transcripts, suggesting that CPL3 is not involved in NMD in the same way that CPL1 is (Li et al., 2019). Li et al. (2019) also demonstrated in their analysis that it is highly likely that CPL3 is involved in polyadenylation, as they found that 67% of their transgene transcripts were non-polyadenylated. 43% of the transgene transcripts were also 3' truncated. The combination of these two factors is used as an explanation as to why a greater proportion of mRNAs enter into mRNA decay pathways in *cp3*.

### Aims of the project

This project aims to expand current understanding of miRNA decay in plants by identifying previously undescribed loci involved in the process. To achieve this, a forward mutagenesis screen will be performed which is designed to be sensitised to changes in the dynamics of miRNA decay. Mutants isolated through this screening will be taken to the next generation, ratified, and then characterised using a combination of northern blot, qPCR and NGS methods. The causative mutation(s) will be identified also by NGS, and compared against the literature – this will largely depend on whether or not any other alleles have previously been described of the gene in question.

## Objectives:

1. Perform a forward mutagenesis screen using the AtSUC2::GFP-miR395 system to identify putative mutants in sRNA decay pathways. This is described in Chapter 3.
2. Take putative mutants to the next generation and ratify them through a combination of Northern blot, qPCR and both sRNA and mRNA sequencing. This is described in Chapter(s) 4 and 5, with Chapter 4 focussing on sRNAs and Chapter 5 focussing on mRNAs.
3. Identify the causative SNP(s) in the mutant background and perform a literature search to determine if anything additional is known about the locus. This is described in Chapter 6.
4. Further characterise the mutant in a literature informed manner, based on the findings of Chapter 6. This is described in Chapter 7.

## Chapter 2: Materials and methods

### Plant tissue culture

#### Normal conditions

*Arabidopsis* were grown under long day conditions (16h light/ 8h dark) at 22°C. Stratification was performed for 2 days in the dark at 4°C prior to transfer to growth chamber. Plants were grown on Murashige and Skoog (MS) medium (Rédei, 2008). Plants grown for EMS screening were grown on 48-well plates (ThermoFisher Scientific, 140675), and plants grown for molecular and morphological phenotyping were grown on 10cm square plates (ThermoFisher Scientific, 166508), spaced at 1cm intervals.

#### Low Sulphate Media

In order to induce sulphur starvation, the MS media recipe was modified to no longer include sulphate. In this recipe, MgSO<sub>4</sub> was substituted for MgCl<sub>2</sub>·6H<sub>2</sub>O (Sigma-Aldrich, M2670), FeSO<sub>4</sub> for FeCl<sub>3</sub>·6H<sub>2</sub>O (Sigma-Aldrich, 236489), MnSO<sub>4</sub> for MnCl<sub>2</sub>·4H<sub>2</sub>O (Sigma-Aldrich, 203734), ZnSO<sub>4</sub> for ZnCl<sub>2</sub> (Sigma-Aldrich, 208086), and CuSO<sub>4</sub> for CuCl<sub>2</sub>·2H<sub>2</sub>O (Sigma-Aldrich, 307483).

#### Low Phosphate Media

In order to induce phosphate starvation, the MS media recipe was modified to no longer include phosphate. In this recipe, KH<sub>2</sub>PO<sub>4</sub> was substituted for MES (Sigma-Aldrich, M3671).

#### ABA exposure

A liquid stock of 10mg/ml abscisic acid (ABA) was prepared by dissolving 10mg of solid +/- Abscisic acid (Sigma, A1049) in 1ml of methanol. *Abi-4* seedlings were purchased from the Nottingham Arabidopsis Stock Centre (NASC).

MS media containing plates were prepared in the usual way. When the molten MS media was still molten but cool enough to hold, for each concentration of ABA a corresponding amount of the liquid stock was added and then thoroughly mixed. Plants were grown for 8 days under long day conditions, and then the number of seedlings which had germinated were counted.

Germination counts were tested between backgrounds at each ABA concentration by Kruskal-Wallis test. For the concentrations at which this reported a significant difference, Dunn test was carried out to determine which the significantly different backgrounds were, using  $p = 0.05$ .

### Seed Sterilisation

Seeds to be grown on sterile culture media were sterilised by incubating for 13 minutes in a 20% sodium hypochlorite (Sigma-Aldrich, 28-3100), 0.1% Triton™ (Sigma-Aldrich, T8787) solution. They were centrifuged to gather the seeds in the bottom, and the sterilisation solution was then removed. The seeds were then washed with sterilised water six times.

### Mutagenesis Screening

Sterilised EMS treated seedlings were germinated on low sulphate media in 48-well plates (ThermoFisher Scientific, 140675). For each round of screening, one plate of non-EMS treated seedlings were also germinated and used as a control. Plates were placed vertically in growth chambers, so that the root grew to span the height of the well. On day 6 and 7 of growth, seedlings were viewed under the Leica MZ16 F fluorescence microscope to assess background levels of GFP fluorescence. As the seedlings were on low sulphur media, and GFP should not be visible, any seedlings which had grade 2 or higher fluorescence as defined in table 1 were discarded from the rest of screening process.

After background screening on day 7, 150 $\mu$ l of autoclaved 30mM MgSO<sub>4</sub> (Sigma-Aldrich, 7487-88-9) was added to each well of each plate. Plates were then returned to the growth chamber, but placed flat instead of vertically. On days 6, 7, 10, 12, 14 and 17 of growth, plants were viewed under the microscope

and GFP was scored using a numeric scoring system. This system is depicted in table 1.

**Table 1 – GFP fluorescence grading system**

Table showing the GFP fluorescence grading system with the numeric grade on the left and the criteria for that grade on the right.

<b>Grade</b>	<b>Description</b>
0	No apparent fluorescence.
1	Very faint fluorescence, but not defined so may not be GFP.
2	Faint GFP fluorescence localised to small regions.
3	Fluorescence clearly in phloem.
4	Bright fluorescence in phloem clearly spanning most if not all of the phloem.
5	Brightest fluorescence spanning the entirety of the phloem.

When a seedling was graded 3 or higher, it was considered to have passed the recovery threshold. Seedlings in the EMS treated plates which passed grade 3 before any of the control seedlings had passed grade 3 were taken as putative early response mutants and taken to the next generation.

Most control seedlings had passed the recovery threshold by day 16. EMS treated seedlings which passed the recovery threshold on day 19 of growth, but not on day 16, were taken as putative late response mutants and taken to the next generation.

### Root and Silique Measurements

Images were taken of roots on 10cm square plates (ThermoFisher Scientific, 166508) *in situ*. For silique images, mature siliques which had not yet browned were cut from plants and placed on a white background with a 1cm marking. Images were then loaded into FIJI software (Schindelin et al., 2012) and

measurements were calibrated using the 1cm spacing in square plates for roots and 1cm marker on the white background for siliques.

Tracks were then drawn following the path of the root, or from tip to tip of siliques, and the length of the track was taken relative to the 1cm calibration. Lengths were then tested between backgrounds using two-sample t tests, as they were normally distributed.

## Phyllochron measurements

Seedlings were germinated on MS media. At days 7, 10, 14 and 18 the number of leaves on each seedling were counted. Statistical significance between backgrounds at each time point was then calculated using a Kruskal-Wallis test followed by a Dunn post hoc test as with root and silique measurement data.

## DNA Extractions

Prior to extraction, plant materials were snap frozen in liquid nitrogen for 5 minutes. A 200mM Tris HCl pH 7.5, 250mM NaCl, 25mM EDTA, 0.5% SDS DNA extraction buffer was prepared.

Snap frozen plant samples were ground up in a liquid nitrogen environment using sterilised mortar and pestles. For every 1mg of tissue, 500 $\mu$ l of DNA extraction buffer was then added to the ground powder. The samples were heated in a water bath at 60°C for 30 minutes, and were vortexed halfway through. An equal volume of chloroform (Fisher Scientific) was then added to each sample. The samples were then centrifuged for 5 minutes at 15000rcf, at 4°C, after which there was a clear phase separation.

The top, aqueous phase of each sample was transferred to a new tube, and an equal volume of isopropanol (Sigma-Aldrich, I9516) was added. This was followed by a 30-minute precipitation step at -20°C. After precipitation, the samples were centrifuged again for 5 minutes at 15000rcf at 4°C, resulting in a pellet of solid material at the bottom of the tube.

The supernatant was removed and the pellet was washed 80% ethanol. After removal of the ethanol, the sample was resuspended in dH<sub>2</sub>O. This was then

centrifuged once again at 15000rcf for 5 minutes, and kept on ice for 2 minutes, in order to collect as much starch and polysaccharides as possible in a pellet in bottom of the tube. The water was then transferred into a final tube, and quantified by Nanodrop 8000 spectrophotometer (Thermo Fisher Scientific).

## RNA Extractions

Prior to extraction, plant materials were snap frozen in liquid nitrogen for 5 minutes.

Total RNA was extracted using Tri Reagent Solution (ThermoFisher Scientific, AM9738). The manufacturer's instructions were followed, except for the precipitation step which was substituted from a 5 – 10-minute room temperature incubation with isopropanol, to a 2-hour incubation with 2 volumes of 100% ethanol.

RNA concentration was measured using the Nanodrop 8000 spectrophotometer (Thermo Fisher Scientific). Integrity of the RNA was subsequently tested by gel electrophoresis; 1µg of RNA was run on a 1.5% (w/v) agarose gel prepared in 0.5 X Tris-Borate-EDTA Buffer (TBE). RNA gels were then stained with 10mg/ml ethidium bromide (Fisher Scientific) and imaged using the Typhoon FLA 9500 (GE Healthcare Life Sciences).

## sRNA Northern Blots

### RNA size separation by urea polyacrylamide gel electrophoresis

Prior to electrophoresis, 1 volume of Gel Loading Buffer II (ThermoFisher Scientific, AM8546G) was added to each total RNA sample, and secondary structures were denatured by heating at 70°C for 2 minutes.

A 15% urea polyacrylamide gel was prepared. For one gel, 2.1g of urea (Fisher Scientific, 10142740) was added to 1.25ml of deionised water and 0.5ml of 5X TBE, and was then heated for 20 seconds in the microwave to dissolve it. Once this solution had cooled, 1.8ml of 19:1 acrylamide/bis solution (Bio-Rad, 1610144) was added, followed by 2.5µl of Tetramethylethylenediamine

(TEMED) (Sigma-Aldrich, 110-18-9) and 50 $\mu$ l of 10% Ammonium Persulfate solution (Thermo-Fisher, 17874). Gel was poured between 1mm glass plates (Bio-Rad, 1651824). A 1mm comb (Bio-Rad, 1653359) was then added and the gel was allowed to set.

Following denaturation, RNA was run in a Mini-Protean<sup>R</sup> Tetra Cell tank (Bio-Rad, 185-8000) on a 15% Urea polyacrylamide gel until the dye front reached the bottom of the gel. 0.5X TBE was used as a buffer for the gel running.

After the gel was run, it was stained with ethidium bromide (Fisher Scientific, 1239-45-8) for 5 minutes, and then imaged using the Typhoon FLA 9500 (GE Healthcare Life Sciences).

### Transfer of RNA to nylon membrane

Whatman paper (ThermoFisher Scientific, 3030-335) was cut into 9cm x 7cm squares and pre-soaked in water. Amersham Hybond-NX membrane (GE Healthcare Life Sciences) was also cut into one 9cm x 7cm square for each gel. This was also pre-soaked in water. Three soaked pieces of Whatman paper were placed in a stack on the base plate of the Semidry apparatus (Fisher), and a sterile serological pipette was rolled across to remove air bubbles. On top of this stack, the soaked Hybond-NX membrane was placed, followed by the urea polyacrylamide gel from the first step. Three more soaked pieces of Whatman paper were placed on top of the gel, and the pipette rolling was repeated in order to eliminate any further air bubbles.

The top plate was placed on the Semidry apparatus, and a constant voltage of 20v was applied to the apparatus for 90 minutes. After this time had elapsed, the apparatus was dismantled, and the urea polyacrylamide gel was re-imaged using the Typhoon FLA 9500. If RNA was no longer visible on the gel, then the transfer was considered a success.

### Chemical cross linking

In a 50ml falcon tube, 10ml of dH<sub>2</sub>O, 122.5 $\mu$ l of 12.5M 1-methylimidazole (Sigma, M50834) and 10 $\mu$ l of concentrated HCl (Sigma-Aldrich, H1758) were

combined. In a separate, 15ml falcon tube 0.373g of 1-ethyl-3-(3-dimethylaminopropyl)carbodiimide hydrochloride (EDC) (Thermo-Fisher, 22980) was measured. 10ml of the previous solution was added to the 15ml tube, and was shaken until the EDC went into solution. The volume of this cross-linking solution was then brought up to 12ml.

A 9cm x 7cm cut piece of Whattman paper was then placed on a sheet of saran wrap. This was then soaked with approximately 5ml of the cross-linking solution. The Hybond-NX membrane from the transfer stage was then placed on top of the cross-linking solution soaked Whattman paper, so that the side of the Hybond-NX membrane that was facing the gel in the transfer step, was facing up. The saran wrap was used to seal the soaked Whattman paper and Hybond-NX membrane assembly, and the crosslinking was performed by placing this in a 60°C hybridisation oven (Thermo) for 90 minutes.

After 90 minutes of cross-linking had elapsed, the saran wrap was opened, and the Hybond-NX membrane was cut in the top left corner in order to delineate the orientation. The membrane was then placed in dH<sub>2</sub>O and rocked on MACHINE for 10 minutes.

## Hybridisation

For each hybridisation to be performed, 2µl of 10µM sRNA specific DNA oligo (Sigma) without 5' phosphate were labelled with 2µl γ-<sup>32</sup>P (Perkin Elmer), using T4 polynucleotide kinase (PNK) (New England Biolabs, M0201S) according to manufacturers recommended volumes. This reaction was performed at 37°C for 60 minutes. All probes used are listed in supplemental table S1.

Cross-linked membranes were incubated with 5ml of ULTRAhyb<sup>TM</sup>-Oligo buffer (Thermo-Fisher, AM8663) per membrane at 37°C with spinning in a rotary hybridisation oven (Thermo). The membranes were placed such that the RNA side of the membrane faced inwards in the tube. Following this, 30µl of dH<sub>2</sub>O were added to the labelled probe solution, in order to bring the volume up to 50µl. This 50µl solution was then added to the membrane and Prehybridisation Solution containing tubes, with care being taken not to get any probe solution directly on the membranes themselves. The membranes were then incubated at 37°C with spinning overnight.

A 0.2% saline-sodium citrate (SSC), 0.1% sodium dodecyl sulphate (SDS) wash buffer was prepared for the membrane wash stages. The membranes were washed four times with 50ml of wash buffer, incubating for 20 minutes each time at 37°C with spinning. The hybridised membranes were then placed on a Fujifilm plate (Fujifilm) and exposed for a minimum of 30 minutes at 4°C. Following exposure, the Fujifilm was imaged using the Typhoon FLA 9500.

### Membrane Stripping

To strip membranes for re-probing, a 0.1% SDS strip solution was prepared. Membranes to be stripped were incubated in this solution for 60 – 120 minutes at 85°C with spinning. Following this incubation, the membranes were checked for successful stripping by running a Geiger counter across the membrane and ensuring they did not register above background.

### Data interpretation

After imaging each hybridisation, ImageQuant (GE Healthcare) software was used to quantify the number of pixels present in each band. One sample was selected to be normalised against – usually the control – and its band intensity was arbitrarily set as 1. All other bands were then normalised to this, so that their intensity was expressed as a ratio of the normalising sample.

U6 was used as a loading control. The intensity of U6 was measured using ImageQuant in the same way, and for each membrane, the same sample which was assigned the value of 1 in the previous hybridisation, was assigned 1 in the U6 hybridisation.

The normalised intensity value for each band in the first hybridisation was divided by the corresponding normalised U6 intensity value for the same sample. These resulting values were the values presented on each northern blot figure.

## DNase Treatment

For total RNA samples which were to be used for qPCR, TURBO™ DNase (ThermoFisher Scientific, AM2238) was used according to manufacturer's instructions. Successful DNase treatment was verified by PCR of atMON1 gene using primers which flanked an intron, in which genomic DNA produced bands 100bp larger than cDNA.

## Reverse Transcription

Superscript™ IV Reverse Transcriptase (ThermoFisher Scientific, 18090010) was used for reverse transcription of total RNA according to manufacturer's instructions for all cDNA samples prepared, except those produced for sRNA library preparation, for which MMLV Reverse Transcriptase (Lucigen, MM070150) was used.

For priming of the reverse transcription, a variety of methods were used. OligoDT primer (Sigma) was used to capture mRNAs in cDNA libraries. In order to capture sRNAs, two-tailed hemiprobcs were designed as described in as described in the publication describing the method for the detection of mature miRNAs by two-tailed hemiprobcs (Androvic et al., 2017) and synthesised (Sigma). All miRNA hemiprobcs used are listed in supplemental table S1.

Multiplex RT priming mixes were prepared to contain a mixture of two-tailed hemiprobcs at 20µM each and oligoDT at 10µM. This priming strategy was used for all Superscript™ cDNA samples prepared.

## qPCR

For qPCR of mRNA transcripts, primers were designed using Primer-BLAST (Ye et al., 2012), ensuring the T<sub>m</sub> was 60°C, and that the amplicon size was between 100 – 200 base pairs. For miRNA qPCRs, primers were designed as described in Androvic et al, 2017. Where possible, these also had a T<sub>m</sub> of 60°C. qPCR BIO SyGreen Mix Lo-ROX (PCR Biosystems, PB20.11-01) was used for qPCR reactions, at 10µl final volumes per reaction – four technical replicates per

sample were performed to account for increased technical variation. 5ng of cDNA was used as template for the reactions, with the exception of pri-miRNA qPCRs, for which 10ng of template were used to account for low abundance.

For cycling conditions, polymerase activation was performed at 95°C for 2 minutes, followed by 40 cycles of 95°C for 5 seconds, 60°C for 30 seconds. All qPCR reactions were performed in the 7500 Real-time PCR system (Applied Biosystems).

## Primer Testing

In order to calculate the efficiency of each primer pair, serial dilutions between 50ng and 0.625ng of cDNA were prepared. qPCR was then run on each cDNA dilution using each primer pair. The Ct values of each dilution were plotted against the log of quantity, and then the gradient of the line was used to calculate the primer efficiency of the pair.

Melt curves of each primer pair were also visualised, and any primer set which produced multiple peaks were not used for data collection.

## Data interpretation

To interpret the Ct values produced from qPCR reactions, the Pfaffl method (Pfaffl, 2001) for relative quantification was used. After excluding outliers, the mean of technical replicates for each biological replicate was calculated, and then the Pfaffl method was performed using these numbers, and the primer efficiencies calculated for each primer set.

In order to calculate significant differential expression between two conditions, normality of the data was tested using the Shapiro-Wilk test (Ghasemi and Zahediasl, 2012). If the data were normally distributed, a two-sample t-test was performed, using a significance threshold of 5%. If the data were not normally distributed after a log transformation, Mann-Whitney U test was performed using the same significance threshold. R was used to perform all these analyses (R Core Team, 2018).

## Sequencing of backcrosses

In order to genotype F2 backcrosses, primer pairs were designed to flank the regions containing SNPs, as determined from mRNA sequencing data. A third, sequencing primer was designed upstream of the SNP, within each amplicon. PCR amplification was performed using GoTaq™ Flexi DNA polymerase. 1µl of PCR product was then added to 14µl of dH<sub>2</sub>O, and 2 µl of the respective sequencing primer was added. Each was then added to a tube from Eurofins Mix2Seq kit (Eurofins) and sent off for sanger sequencing. All primers used are listed in supplemental table S1.

## Backcrosses

pSUC2:GFP:395 control and *msm1* seedlings were grown in soil, under long day conditions until they began to flower. Forceps and scissors were sterilised using 80% ethanol prior to use. Floral buds which had not yet opened were selected on mother plants; buds which were too small or that had already opened were removed. The sepals and immature stamen were removed from these selected floral buds using sterilised forceps, under Leica MZ16 F fluorescence microscope. The plants were then returned to grow for 2 – 3 days.

After 2 -3 days, open flowers were then selected from the father plant, again under the Leica microscope. Mature stamen were removed from these flowers, and touched on the tip of stigmas of the previously prepared mother plant. Plants were then returned to grow and the stigmas of the mother plant were monitored. When they began to swell, the cross was determined to have taken.

When the siliques formed were close to shattering, they were transferred to a 0.5ml tube (Eppendorf, 0030121023), and left to dry out. When dry, these F1 seeds were then sown, and all seeds were collected from these plants to obtain the segregating F2 population.

## sRNA library construction and sequencing

### sRNA library total RNA cleanup

Prior to sRNA library preparation, total RNA samples were cleaned using the miRVANA miRNA isolation kit (Ambion, AM1560) following a slightly modified protocol. Firstly, total RNA samples were brought up to a volume of 50 $\mu$ l. 5 volumes of lysis/binding buffer were added, followed by 1/10<sup>th</sup> volume of miRNA Homogenate Additive. The samples were mixed and left on ice for 10 minutes.

After 10 minutes has elapsed, 1.25 volumes of 100% ethanol were added and then mixed thoroughly. The lysate and ethanol mixture were then passed through one of the miRVana filter cartridges by centrifuging for 30 seconds at 10,000rcf. The flow-through was discarded and 700 $\mu$ l of miRNA wash solution 1 was passed through the column as before. The column was then washed twice, passing 500 $\mu$ l at a time of miRNA Wash Solution 2/3 through it, followed by spinning the empty column for 1 minute at 10,000rcf to remove all traces of the wash buffers.

The column was transferred to a fresh collection tube and 50 $\mu$ l of Elution Solution which had been pre-heated to 95°C was added. This was left to soak in for 2 minutes, and then the column was centrifuged at full speed for 30 seconds. This was repeated with another 50 $\mu$ l of pre-heated Elution Solution, so that the final elute volume was 100 $\mu$ l. After this, 10 $\mu$ l of 3M NaOAc (Sigma-Aldrich, S2889) and 3 volumes of 100% ethanol was added to the 100 $\mu$ l of purified RNA. This was precipitated in the -80°C for 4 hours.

After precipitation, the RNA was pelleted by centrifugation at 4°C, full speed for 30 minutes. The supernatant was then removed, and replaced with 80% ethanol to wash the pellet. After centrifuging one last time at 10,000rcf, the ethanol was removed and the pellet was air dried. 15 $\mu$ l of nuclease free water was then used to resuspend the total RNA, and the sample was quantified again using the Nanodrop 8000 spectrophotometer (ThermoFisher Scientific).

### 3' Adapter Adenylation

sRNA libraries were prepared as outlined in the published protocol by Xu et al. (Xu et al., 2015). As described in this protocol, 3' and 5' `HD` adapters, with 4 random nucleotides facing towards the insert, were used. The 3' adapter was purchased pre-phosphorylated, however it still required adenylation.

Adenylation of the 3' adapter was performed using the 5' DNA Adenylation kit from New England Biolabs (NEB, E2610L) as directed by the manufacturer. After this reaction had been performed, it was cleaned up using Oligo clean and Concentrator™ (Zymo-Research, D4061). The eluted adapter was then quantified using Nanodrop 8000 spectrophotometer (Thermo Fisher Scientific), and nuclease free water was added to bring the concentration to 10µM. The adenylation was then verified by running the adenylated adapter on a 16% UREA polyacrylamide gel, alongside non-adenylated adapter. If the adenylated adapter sample appeared heavier than the non-adenylated adapter on the gel, then it was considered to have been successfully adenylated.

### sRNA adapter ligation and PCR

The adenylated 3' adapter and 5' adapter were ligated sequentially to 2µg of miRVANA purified total RNA, with a RecJ exonuclease (New England Biolabs, M0264S) digestion step in between to reduce adapter/adapter ligation products, as described (Xu et al., 2015). In this protocol, following adapter ligation samples are amplified using a number of different PCR programs which differ in the number of cycles of amplification. For these libraries, 12, 14 and 16 cycles were selected, and amplification was carried out using Phusion High-Fidelity DNA Polymerase (Thermo Scientific™, F530L).

### Library size selection

Library size selection was performed as described in P.Xu et al (2015).

## Library normalisation

In order to normalise the library quantities, so that the same amount of each was loaded into the sequencer, 1µl from all completed libraries was combined with Novex™ Hi-Density TBE Sample Buffer (Invitrogen™, LC6678) and run on an 8% polyacrylamide gel. Following staining with SYBR™ Gold Nucleic Acid Gel Stain (Invitrogen™, S11494), gels were imaged using the Typhoon FLA 9500 (GE Healthcare Life Sciences). Library intensity was quantified using ImageQuant (GE Healthcare) and normalised to one library, as described in the sRNA Northern Blot protocol. The value produced was used to adjust the 1µl loaded, and the new adjusted volume for each library was one again combined with the Novex™ sample buffer. These adjusted volumes were run again on another 8% polyacrylamide gel and imaged as before, and if the libraries appeared to be of a similar intensity, the remaining the volumes were pooled and sent for sequencing.

## sRNA library Sequencing

Normalised sRNA libraries were sequenced using the Illumina NextSeq 550 at a depth of approximately 33million reads each, single end.

## mRNA Sequencing

The same RNA used to produce the sRNA libraries was used for mRNA sequencing. RNA was DNase treated as previously described, and then cleaned up using RNA Clean & Concentrator-25 (Zymo Research, R1017).

mRNA library preparation and sequencing was performed by Novogene (UK) Company Limited. Paired-end sequencing was carried out to a depth of 20 million reads.

## Bioinformatics analysis

### Mapping of Reads to Reference Sequences

Note: here will talk about both HISAT2 and BISMARCK which will use for bisulfite seq analysis.

Trim Galore was used to remove adapter sequences from fastq files generated by the sequencer (Krueger, 2020). For sRNA libraries generated using HD adapters, a further 4 nucleotides were removed from either end.

HISAT2 was used for the alignment of trimmed reads to reference sequences (Kim et al., 2015). HISAT2 was used according to its default settings. For mRNA sequencing, trimmed reads were aligned to The *Arabidopsis thaliana* Information Resource (TAIR10) genome ("TAIR10 - Genome - Assembly - NCBI," n.d.). Reads which did not align to the reference genome were discarded, and the genome mapped reads were used for the subsequent steps of the analysis.

For sRNA sequencing data, trimmed sRNA reads were also genome aligned, and the non-mapping reads were discarded. These reads were then aligned to a ribosomal RNA database obtained from the SILVA rRNA database project (Quast et al., 2013). Reads which aligned to the ribosomal RNAs were discarded, and the non-aligning reads were then mapped to an *Arabidopsis* tRNA database (Chan and Lowe, 2009), and a fasta list of transposable elements in the TAIR10 genome assembly ("TAIR - Download - TAIR10 transposable elements," n.d.). After each alignment to these databases, the aligning reads were discarded, and the non-aligning reads were taken forward. For the miRNA differential abundance analysis, these reads were used. For all analyses of sRNA abundance that were not miRNAs, the reads generated in the previous step, were aligned to miRbase (Griffiths-Jones et al., 2006), and the aligning reads were discarded. The non-aligning reads were then used.

### Differential Expression Analysis

For transcript quantification, the Salmon tool was used (Patro et al., 2017). For mRNA sequencing, Salmon was run on the genome aligned mRNA reads in

mapping mode. A Salmon transcriptome index was built using the TAIR10 reference cDNA, and the quantitative alignment was performed.

For miRNA differential expression, genome aligned reads which did not align to ribosomal RNA, tRNAs, or transposable elements were quantified with Salmon in mapping mode. The transcriptome index was built using *Arabidopsis thaliana* miRbase sequences (Griffiths-Jones et al., 2006).

For non-miRNA sRNA differential expression, genome aligned reads which did not align to ribosomal RNA, tRNAs, transposable elements or miRbase were quantified with Salmon in mapping mode. The transcriptome index was built using the TAIR10 genome assembly.

For differential expression of mRNA sequencing data and sRNA sequencing data aligned to miRbase, 20 counts were added to each estimated count from Salmon in order to offset the fold change and rule out inflated log<sub>2</sub> fold changes which can arise as a result of low counts. For sRNA sequencing data not aligned to miRbase, 5 counts were added as the total counts were much lower.

Statistically significant fold change differences in sRNA sequencing and mRNA sequencing were determined using DESeq2 R package (Love et al., 2014). The number of aligning reads taken from Salmon was used as the input for DESeq2. For mRNA sequencing analysis, the threshold of significance was set to  $p = 0.01$ . For sRNA sequencing analyses, this was set to  $p = 0.05$ .

## Alternative Splicing

For the alternative splicing analysis, the SplAdder tool was used (Kahles et al., 2016). mRNA sequencing data was used as input, and the TAIR10 *Arabidopsis thaliana* gff3 file, sourced from ensemble (Howe et al., 2020), was used as an annotation file. After events had been identified and quantified, SplAdder was run in the test mode using its default statistical model. The threshold of significance was set to  $p = 0.05$ .

This tool takes genome aligned mRNA sequencing data and an annotation file as input. The annotation file is used by the algorithm to inform a gene-model like graph of each gene, and then evidence to support each intron junction is extracted from the provided mRNA sequencing data (Kahles et al.,

2016). Once all of the evidenced splice sites have been identified, from each graph the individual alternative splicing events are extracted, and then the provided read alignment file is used to quantify the events (Kahles et al., 2016).

### Coverage plots

To create the *Arabidopsis* exon bed file, exon coordinates were extracted from the gff3 file used in the SplAdder run. Intron coordinates were also inferred from this file. These subset gff3 files were converted into bed format using the BEDOPS command line toolkit (Neph et al., 2012).

For quantitative mapping to exons and introns, bedtools was used to obtain a fasta file containing *Arabidopsis* exon sequences, and a separate file containing intron sequences, using the bed formatted files produced in the earlier step (Quinlan and Hall, 2010). Genome aligned mRNA reads were quantitatively mapped to these fasta files using Salmon as earlier described.

For each gene, the number of reads mapping to each intron was divided by the number of reads mapping to each exon, to generate an intron/exon ratio value. This was performed on each library, so that each gene had three intron/exon ratios. The three intron/exon ratios for each gene were then tested against the same gene in the contrary condition using a two-sample t test.

For each library, the total number of intron mapping reads across all genes were then divided by the total number of exon mapping reads across all genes, in order to generate a global intron/exon ratio. These were again tested between contrary conditions using a two-sample t test.

### miTRATA analysis

To quantify truncated or tailed isoforms of miRNAs, the sRNA sequencing files in tag-count format were provided to the miTRATA tool (Patel et al., 2016). MiRBase was used as the reference. From the summary output produced by this tool, the tailing ratio was extracted for each miRNA in each library. This tailing ratio represented the sum of the abundance of the isoforms divided by the total abundance of the mature miRNA.

For each miRNA in each condition, the three tailing ratios were compared to those of the corresponding miRNA in the contrary condition and tested using the non-parametric Mann-Whitney test. The tailing ratios of each miRNA in each library were then summed to produce a global tailing ratio. The global tailing ratios were then compared between opposing conditions and again tested using the Mann-Whitney test.

## SIMPLE Analysis

In order to identify potential phenotypic SNPs in *msm1*, the SIMPLE pipeline was used (Wachsman et al., 2017). This tool was run on each set of *msm1* libraries using the default settings, and then the SNPs which were present in all of the predicted phenotypic SNPs list, which appeared to be homozygous in the sequencing data, were carried forward as candidates.

## Coverage Plots

For each gene plotted, a bed file was manually created using the gene coordinates found on TAIR. Coverage of genome-aligned reads across this bed file was computed using the bedtools coverage tool. The numeric output from this tool was then visualised in R using the ggplot2 package (Wickham, 2009).

## Gene Ontology Analysis

All Gene Ontology enrichment analyses were performed using ShinyGO v0.61 (Ge et al., 2020). Data were exported from this analysis, and `Gene Ratios` were calculated by dividing the number of genes represented of a given annotation divided by the total number of genes in the database with that annotation, in order to normalise for misrepresentation caused by highly annotated GO groups vs sparsely annotated GO groups. Graphs of these data were then plotting using GGplot2 (Wickham, 2009).

## Venn Diagrams

Venn diagrams were constructed using InteractiVenn (Heberle et al., 2015).

## Folding Analysis

MFE of transcripts was calculated by running cDNA sequence obtained from TAIR through RNAfold in the ViennaRNA package. These MFE values were then normalised for transcript size by dividing the MFE by the total length of the transcript in nucleotides. The normalised values were plotted against each other using GGplot2.

## Strand Distribution analysis

Cleaned sRNA reads were mapped against transcripts which produced more 21mers and transcripts which produced more 24mers separately using HISAT2. Samtools was then used to count the number of reads aligning to the forward and reverse strand of each of the alignment files produced, and these data were tabulated. Mean percentages and graphs were then produced using R.

## Chapter 3: Identification of putative mutants in miRNA decay pathway components by EMS screen

### Introduction

One particularly well studied example of a plant stress responsive miRNA is miR395, first described in 2004 (Jones-Rhoades and Bartel, 2004). This miRNA was shown to be intimately involved with the sulphate starvation response, being undetectable in plants grown in normal or high-level sulphate containing media, and readily detectable in plants grown in low sulphate media (Jones-Rhoades and Bartel, 2004). MiR395 was also found in the same study to be complementary to the mRNA of ATP sulphurylase (ATPS) proteins; ATPS being the first enzyme in the sulphur assimilation pathway (Kawashima et al., 2011). In plants, the sulphur assimilation pathway is a demand driven process in which inorganic sulphate is taken up and sequentially reduced into sulphide, which is then used to produce the amino acid cysteine (Kopriva, 2006). In response to a deficiency of sulphur in the plant, both uptake of sulphur from the environment and rate of reduction of sulphate is increased by an associated induction of sulphate transporter mRNA and adenosine-5'-phosphosulphate (APS) reductase (APR) mRNA levels, which is the crucial enzyme of the pathway (Nikiforova et al., 2003). The molecular mechanisms that underpin the sulphate starvation response are still largely unknown, however to date, several components have been identified which operate in the sulphate `regulatory circuit`. The first of these components is the transcription factor SULPHUR LIMITATION 1 (SLIM1), which increases sulphate transporter expression and also upregulates several other genes involved in the sulphur starvation response (Maruyama-Nakashita et al., 2006).

MiR395 expression is controlled by SLIM1 (Kawashima et al., 2009). In addition to targeting ATPS, miR395 also appears to target the low affinity sulphur transporter SULTR2;1, and yet paradoxically, miR395 expression is strongly induced in sulphur starvation conditions (Jones-Rhoades and Bartel, 2004). Additionally, the mRNA levels of SULTR2;1 were found to increase in the roots in sulphate starvation, which was also unexpected given that of the 6 miR395 loci (MIR395a-f), MIR395c and MIR395e in particular were also found to be strongly induced in the roots (Kawashima et al., 2009). To investigate

this, Kawashima *et al* analysed SULTR2;1 expression in wild type *Arabidopsis* and *slim1* mutants, which are unable to induce SLIM1 and by extension unable to upregulate miR395. In the wild type they found that when they transferred the seedlings from a sulphur rich environment (S+) to a sulphur deficient environment (S-), there was no significant change in SULTR2;1 expression in the leaves but there was a considerable increase in SULTR2;1 expression in the roots. In the *Slim1* mutants however, SULTR2;1 levels in the leaves on transfer increased as there was no miR395 expression, and the SULTR2;1 upregulation in the roots was even more dramatic than in the wild type (Kawashima *et al.*, 2009).

On closer investigation, whilst there is a positive temporal correlation between miR395 and SULTR2;1, it appears that spatially they are different. In the roots, most miR395 expression occurs in the phloem companion cells, yet the SULTR2;1 is mainly expressed in xylem parenchyma cells. There are however, some cleaved SULTR2;1 fragments in the roots indicating that there is some overlap between these spatial expression patterns. With the available evidence a model has been suggested in which SULTR2;1 has low level expression in the phloem companion cells, and miR395 silences this expression in response to S-conditions in order to restrict SULTR2;1 expression to the xylem (Kawashima *et al.*, 2009). This illustrates an elegant and non-canonical method of miRNA-mediated gene regulation in plant stress response.

Forward genetic screens are powerful tools for the *de novo* identification of genes involved in processes of interest. Through combination of intelligent reporter system design and random mutagenesis, they represent a non-biased approach to gene discovery, as they do not presuppose the involvement of any particular gene in a process. A common mutagenizing agent used in these screens is ethyl methanesulfonate (EMS). EMS is an alkylating agent used to induce chemical modification of nucleotides; alkylation of guanine results in O<sup>6</sup>-ethylguanine, which can no longer base pair with cytosine, but rather thymine. DNA repair mechanisms then act to repair the damaged DNA, and this results in G/C pairs being replaced with A/T pairs (Kim *et al.*, 2006). EMS treatment results in randomly distributed mutations throughout the genome, and twice as many heterozygotes to homozygotes (Greene *et al.*, 2003). When screening is performed in the M2 generation, both dominant and recessive traits can be screened for.

The Dalmy group previously generated a sulphate responsive assay using GFP and miR395 in *Arabidopsis thaliana*. To generate this system, an RDR6-null line (*rdr6-15*) of *Arabidopsis* was transformed with a GFP transgene containing a miR395 target site. This transgene was put under the control of the *Arabidopsis* SUC2 (AtSUC2) promoter, which directs expression specifically to the phloem of *Arabidopsis* (Truernit and Sauer, 1995). In these AtSUC2::GFP-miR395 seedlings, the GFP transgene is expressed under normal conditions in the phloem companion cells of the plant. On sulphate starvation miR395 is expressed and accumulates (Jones-Rhoades and Bartel, 2004), which results in degradation of the GFP transgene mRNA because it contains a miR395 target site, and therefore loss of fluorescence. Following the addition of sulphate to the system, expression of miR395 is switched off so no new miR395 is produced. The only miR395 present in the system is that which has already been expressed and has not been turned over. Therefore, the rate at which fluorescence recovers in seedlings which have been exposed to low sulphate conditions is indicative of the rate of turnover of miR395, as the two have an inverse relationship.

In order to delineate genes which control this turnover, the AtSUC2::GFP-miR395 seedlings were segregated into pools, and treated with three respective treatments of ethyl methanesulfonate (EMS). Different concentrations of EMS result in different frequencies of mutation, however it is not a linear relationship as higher concentrations can actually result in decreased mutation frequencies (Zhu et al., 1995). As such, the purpose of three pools of EMS treated seedlings was to increase the likelihood of being within the appropriate range of EMS treatment, and to provide three groups with varying levels of mutagenesis.

EMS:AtSUC2::GFP-miR395 seedlings which exhibited faster or slower GFP recovery than the control AtSUC2::GFP-miR395 were to be selected and allowed to progress to produce seeds. M3 plants would then be similarly screened to see if progeny exhibit the same recovery rate as parent, which would be a good indicator of whether or not the observed rate is genetic in nature. There may still be segregation in progeny, however, as EMS generates heterozygotes at twice the frequency of homozygotes.

The rationale of this screening process is that the faster or slower EMS:AtSUC2::GFP-miR395 seedlings display these variances in recovery

of fluorescence from the control because they have mutation in genes which control the turnover of miRNA. The fast recovering individuals could contain loss of function mutations in genes which normally inhibit or suppress controlled miRNA turnover or, less likely, gain of function mutations in genes actively involved in turnover which undergo an increase in activity. Slow turnover mutants could possess mutations in genes which are actively involved in the turnover of miRNAs or in regulatory mechanisms of the turnover. However, whilst genetic screens of this nature are powerful techniques for identifying and characterising new genes (Page and Grossniklaus, 2002), there is the potential rediscovery of already characterised genes. In addition to that, there is also the potential for false positives, for example genes involved in the sulphate starvation response could theoretically influence the rate of fluorescence recovery; there is therefore also scope here for negative results for individuals with mutation in genes involved in miRNA turnover however the statistical likelihood of this is much lower and there is no way of knowing when this happens.

## Results

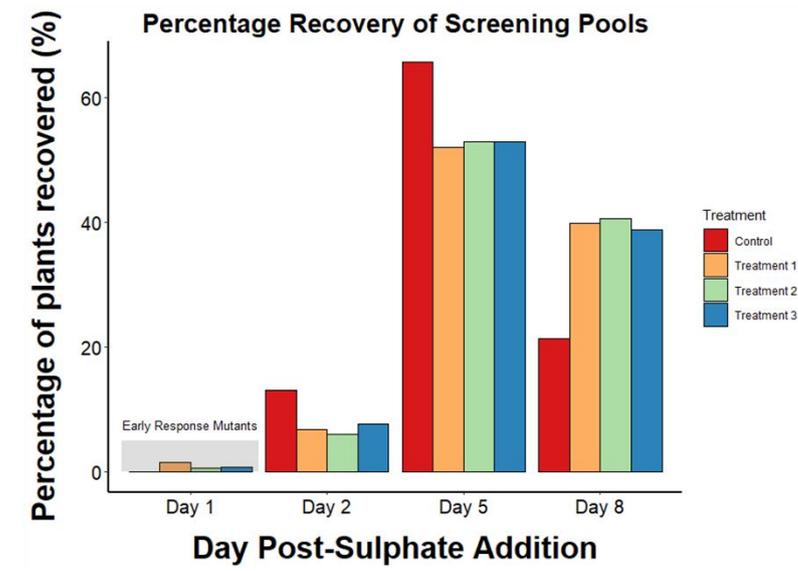
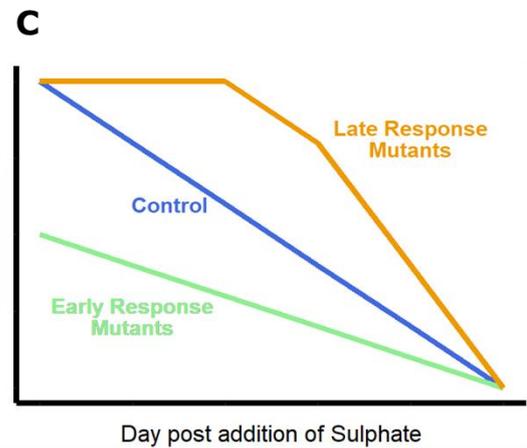
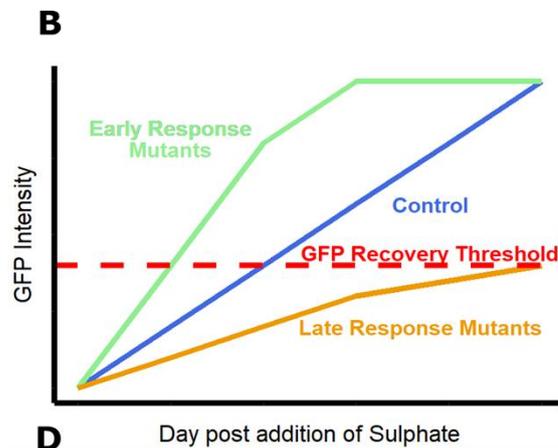
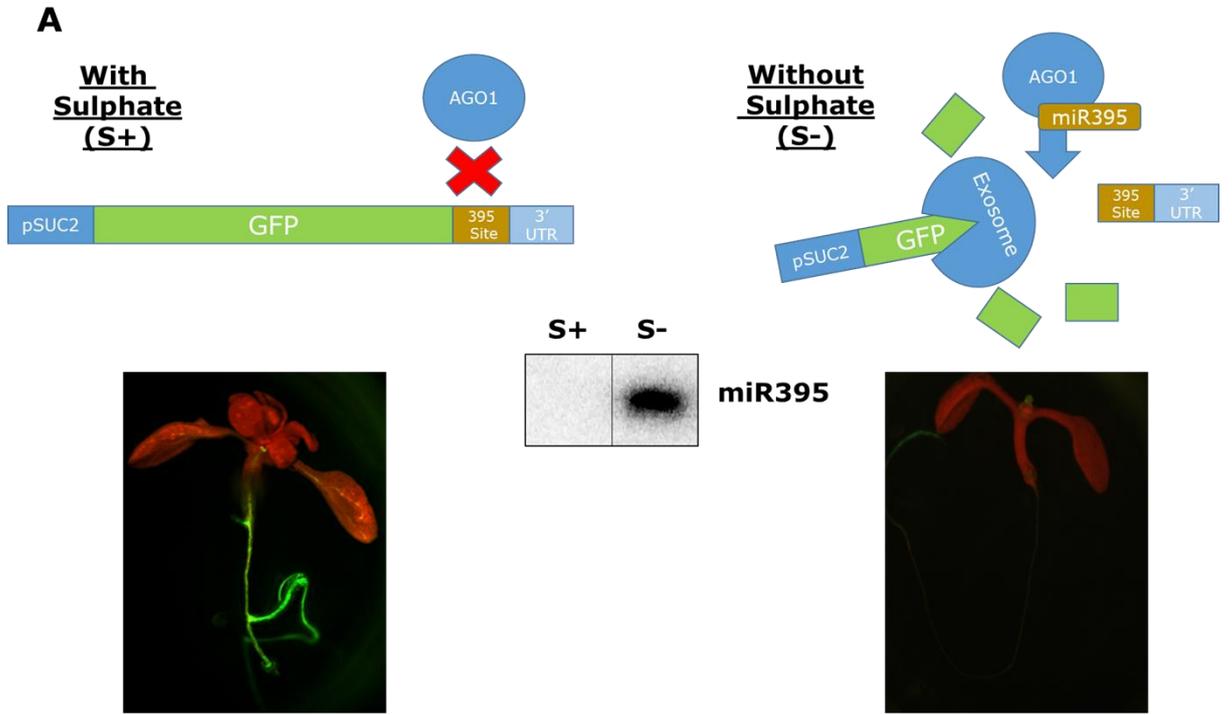
### EMS Screening

Figure 7.1A shows a graphical depiction of the screen described. Sterilised EMS treated and non-EMS treated AtSUC2::GFP-miR395 seedlings were germinated on low sulphate media to induce miR395 expression, and repress GFP expression. Once germinated, MgSO<sub>4</sub> was added to switch off miR395 expression, and seedlings were screened for GFP under UV on days 1, 2, 5, 7, 10 and 13 post addition of sulphate. The GFP intensity of the EMS treated seedlings was compared to that of the non-EMS treated seedlings, and those that recovered outside of the window of the non-EMS treated seedlings were taken for further study. Figure 7.1B and 2.1C show the expected change in GFP and miR395 levels. The decision was made to carry out the screening at the M2 stage, as this allowed for the identification of both dominant and recessive mutant phenotypes.

In total, 14,000 plants were screened in the described method. Of these, roughly 10,000 were EMS treated and 4,000 were control, non-treated lines. Figure 7.1D

depicts the percentage of the total of plants that recovered GFP for each day. This figure represents data taken from 5000 seedlings, as, unfortunately, the rest of the data was lost. Of 1147 control plants screened, none of them recovered GFP on Day 1. Therefore, all of the individuals in the EMS pools that recovered on Day 1 were retained as putative early response mutants.

Late response mutants were individuals which recovered GFP after the control had recovered GFP, but had not recovered GFP within the window of control GFP recovery. In total, 12 putative early response mutants and 15 putative late response mutants were isolated through this approach.



## **Figure 7.1 - Schematic and model of the screen with recovery data**

**A:** Visual depiction of the background in which the EMS screen was produced. The left-hand side shows the RNA transcript produced from the transgene, which is not degraded in the presence of sulphate. This is accompanied by a northern blot image showing no detectable miR395 and a microscopy image showing the GFP fluorescence in the phloem. The right-hand side shows the transcript being degraded by AGO1 loaded miR395, along with a northern blot image showing detectable levels of miR395 in the absence of sulphate, and a microscopy image using the same exposure time showing no observable GFP fluorescence.

**B:** Model showing predicted increase in GFP intensity over time in early response mutants, late response mutants, and non-mutagenised seedlings following relief from sulphate starvation. Red line indicates the point at which GFP fluorescence passes the defined `recovery threshold`.

**C:** Model showing predicted decrease in miR395 levels in early response mutants, late response mutants and non-mutagenised seedlings following relief from sulphate starvation.

**D:** Percentage of seedlings which passed the defined GFP recovery threshold on each day after relief of sulphate starvation. The grey box represents EMS treated seedlings which recovered earlier than the control, and were therefore taken further. Treatment refers to the intensity of EMS treatment each line received in earlier generations.

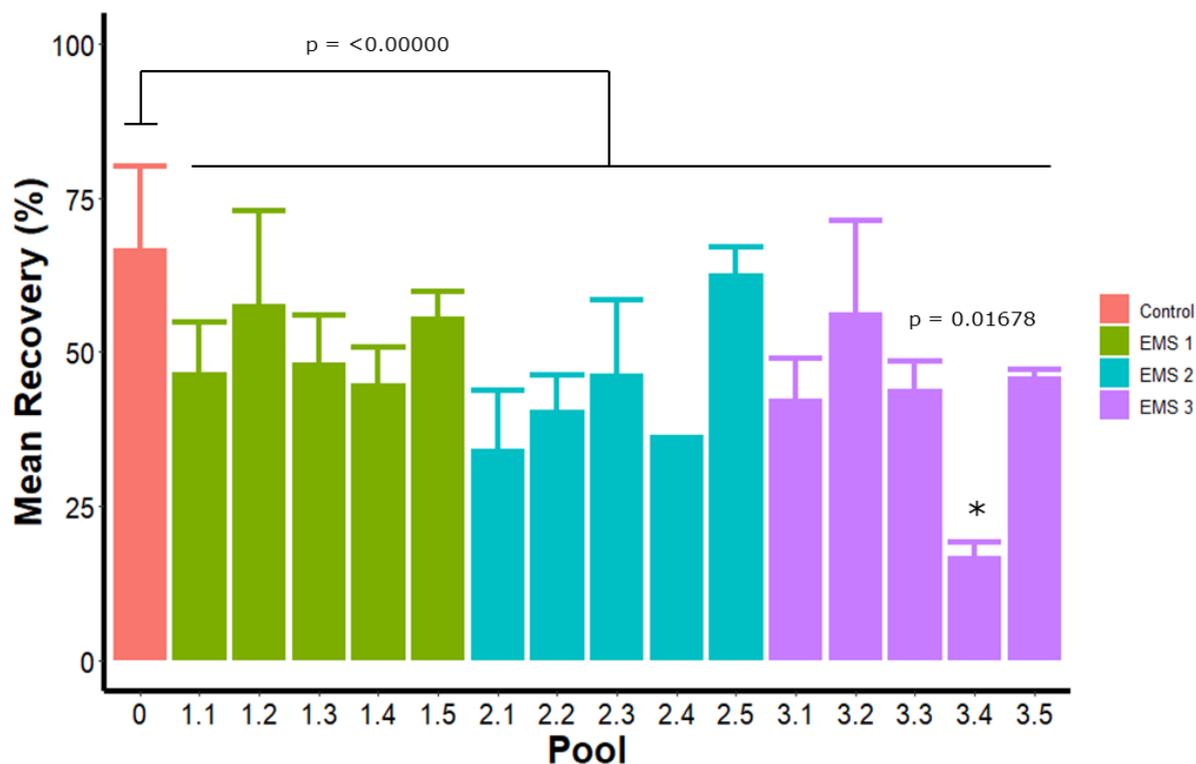
### Silencing of the transgene in the next generation

In order to verify if the altered recovery rate seen in the M2 putative mutants was truly caused by mutation, M3 seedlings taken from putative mutants were screened identically to the M2. In these experiments, instead of just one seedling, there were at approximately 48 for each genotype. However, in the M3 generation most seedlings no longer displayed any GFP, even when grown under MS conditions (i.e. in the absence of miR395).

Additionally, over the course of the screen it was identified that a number of individuals both in the EMS treated populations and the control populations did

not seem to recover GFP in the timeframe of screening. The number of plants which did not recover GFP in the screening timeframe was counted and a percentage of total recovery was calculated for each pool and treatment of EMS treated plants (Figure 7.2). The populations have been separated by the EMS treatment intensity (EMS 1/2/3), and then `sub-pools` (1.1/2/3 etc.) which are technical replicates of the corresponding treatment intensity.

None of the pools, including the non-EMS treated control had a 100% recovery rate. There was no significant difference within different sub-pools for each EMS treatment, with the exception of EMS 3, in which pool 3.4 differed from the other recovery rates. None of the mean recovery rates of each EMS treatment differed significantly from another. However, all three of the EMS treatments had a significantly lower rate of GFP recovery than the control when compared by ANOVA. One reason for this could be that the EMS treated populations are several generations further along than the control lines, and transcriptional silencing can get stronger over generations (Devanapally et al., 2020) (Lang-Mladek et al., 2010).



## Figure 7.2 – Percentage of control and EMS seedlings that recovered

**A:** Mean percentage of recovery of non-mutagenised and mutagenised seedlings. Pool refers to the EMS treatment intensity the seedlings were exposed to in earlier generations (the first number) and the seed-bulking group the lines were in (the second number). Asterixes signify statistical significance. Error bars represent one standard deviation.

### Phenotypic validation in the next generation

As many seeds were now available for each putative mutant background, I decided to directly assay miR395 levels in the roots of seedlings grown under screening conditions from extracted RNA. The quantity of RNA from the roots of these seedlings was a limiting factor, as for most seedlings it fell below what was typically required for a sRNA northern blot. I therefore decided to employ rt-qPCR using two-tailed hemiprobcs (Androvic et al., 2017) to quantify miR395 levels in seedlings, as this method required only approximately 50ng of RNA from each sample.

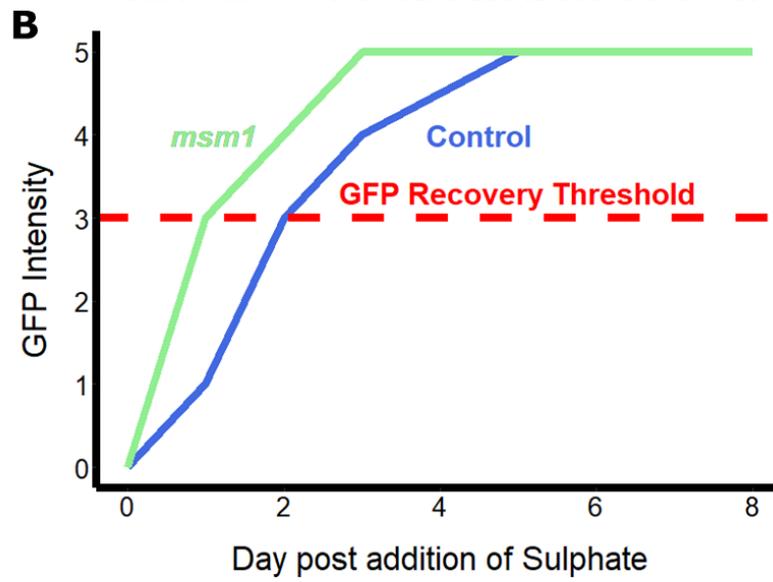
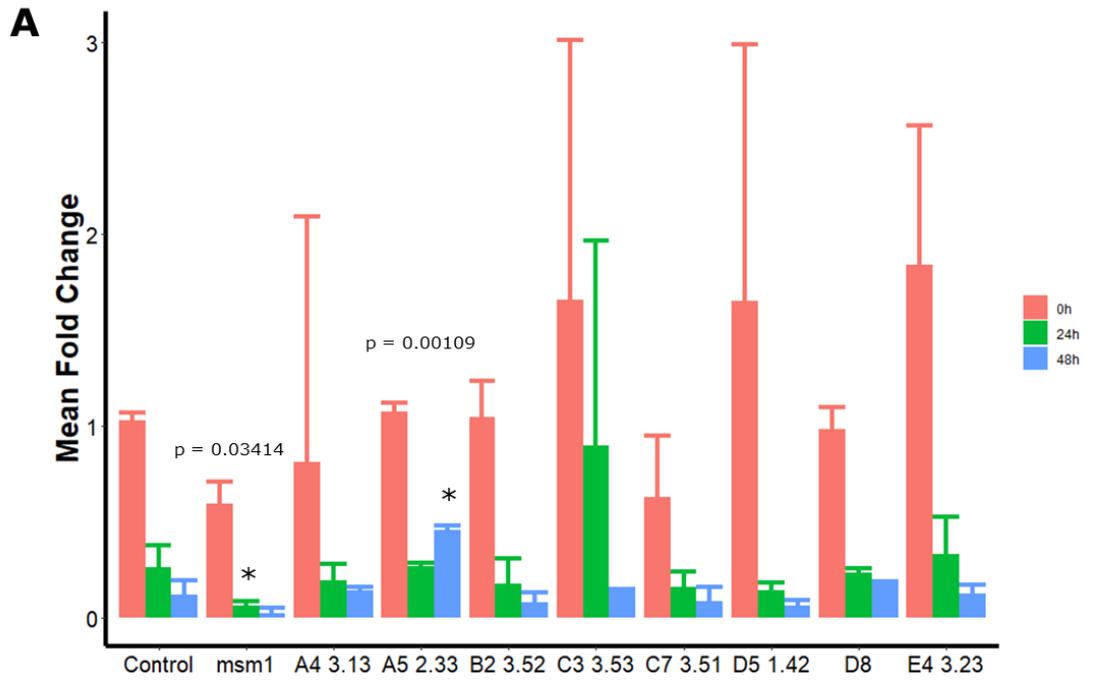
This method of quantifying mature miRNAs was recently pioneered by Androvic et al. (2017), and a number of studies have since been published which demonstrate the efficacy of this method (Androvic et al., 2017; Androvic et al. 2019; Damayanti et al., 2019; Wright et al., 2019).

A combination of early and late response mutants were tested in this way. All of them showed a downward trend consistent with the degradation of miR395 levels over time, inverse to the GFP recovery observed in these seedlings. These results aligned with the known biology of miR395 decay over time, giving confidence that the technique was working. One such mutant isolated in this way was named MicroRNA Stability Mutant 1 (*msm1*). This was retroactively named differently from the other mutants as it was confirmed later on to have a phenotype. The other mutants are so named based on a combination of the location on the 48-well plate used for screening, their EMS treatment batch and their seed bulk group.

When miR395 fold change was compared between control seedlings and mutant seedlings at the same time points, only 24h *msm1* and 48h A5 2.33 were significantly different (Figure 7.3A). For *msm1*, at 24h the levels of miR395 were

lower than the control seedlings at 24h, which corroborated the difference seen in the GFP – at 24h GFP had recovered in *msm1* but not in the control in the previous generation (Figure 7.3B).

In A5 2.33, the levels of miR395 were significantly higher than the control at 48h. However, this background was isolated as a putative early response mutant in the EMS screen, and therefore this result did not align with the observation in the previous generation. As such, A5 2.33 was excluded from further investigation.



### **Figure 7.3 - qPCR testing of the M3 generations of putative mutants identified through the screen**

**A:** qPCR data showing fold change of miR395 levels in non-EMS treated AtSUC2::*GFP*-miR395 seedlings and putative mutants isolated from the EMS screen at discrete time points after the relief of sulphate starvation. Three biological replicates were used for each background and four technical replicates for each. Asterixes denote statistical significance compared to the corresponding time point in the control. Error bars represent 1 standard deviation.

**B:** Measured GFP recovery in *msm1* and non-EMS treated AtSUC2::*GFP*-miR395 seedlings. The red line indicates the point at which they passed the defined `Clear GFP` threshold.

## Discussion

### Variance between EMS populations and control

It did appear as if the recovery window of EMS treated plants was shifted relative to the control. This can be seen in Figure 7.1D, where the percentage of recovery in EMS treated plants is lower than the control at days 2 and 5, but then higher at day 8. Seeds should have germinated at the same time as they were synchronised with a 2-day stratification at 4°C, so it is unlikely this is due to this reason.

The most likely reason is due to transcriptional silencing of the transgene. This type of silencing increases in strength from generation to generation (Lang-Mladek et al., 2010). The control seedlings used for these screening experiments were several generations younger than the EMS treated populations, as the EMS treated populations were taken from the same progenitor line, but then had to be taken two generations further after the EMS treatment in order to get to the M2 stage.

The decision was made to express recovery as a percentage of the total plants that did recover, as for all lines there were a number of plants which did not appear to recover GFP in the timeframe where GFP screening was possible. The

number of plants for which there was no recovery was considerably larger in the EMS treated plants than the control plants.

The window of time in which plants could be screened for GFP recovery was relatively short, limited to approximately day 16 of growth. This was because as the plants matured, the aerial tissues and lateral roots obscured the primary root. This restricted the amount of time in which a late response mutant could be identified. The apparent influence of transcriptional silencing of the GFP transgene further confounded the accurate identification of putative late response mutants. It was unclear whether GFP was not recovering due to a miRNA decay phenotype, or because of transcriptional silencing of the GFP. Therefore, the highest confidence class of mutant which this screen could identify were early response mutants.

### Types of mutants

An early response mutant recovers GFP earlier than the control. In the context of miR395, this could be the result of a loss of a protective factor, which would cause an increase in the rate of degradation of the miRNA. In this scenario, it would be expected that at all time points, including 0h, miR395 levels in this background would be lower than in the control. However, as the plants are screened for background fluorescence before the addition of MgSO<sub>4</sub>, a strong mutation of this type, such as *hen1* (Park et al., 2002), would not be identified by this screening protocol.

A second type of early response mutant which could be identified by this screening method would be a miRNA biogenesis mutant. As with the loss of a protective factor, strong mutant phenotypes such as *dcl1* (Schauer et al., 2002) would be excluded by the background screening.

A late response mutant recovers GFP later than the control, in principle because miR395 is degraded more slowly. This would most likely be the result of a loss of a degrader, such as an SDN (Ramachandran and Chen, 2008). This type of mutant would likely form the largest contribution to the field that this screen could be expected to produce, as there are still a number of unknown enzymatic activities that must be performing this function – for example an exonuclease that degrades uridylylated mature miRNAs. It is also theoretically possible that this type of phenotype could be caused by the loss of an inducer of miRNA

degradation, for example a mutation in AGO10 might cause a reduction in degradation of miR165/6 (Yu et al., 2017). Whilst this would be interesting, if it is part of a mechanism similar to the AGO10:miR165/6 degradation system it would likely be very miRNA specific, and not general.

### Improving the screen

There are some ways in which this screen could be improved as a tool to discover novel genes involved in miRNA turnover. However, by the standard set by many successful EMS mutagenesis screens, for which there are ample, well documented examples, this particular screen is lower throughput. This is because it requires repeated, manual monitoring of individual seedlings over a number of days. This contrasts with many other documented EMS screens where there is either a single screening point which does not rely on microscopy (Jia et al., 2017), or there is a lethal outcome for individuals which do not contain mutations in genes involved in a process of interest (Page and Grossniklaus, 2002); the latter meaning much higher numbers can be planted and managed by a single person. As such, any modifications to the screening procedure aimed to improve its capacity to identify novel mutations of interest must also take into consideration the addition of labour that such a modification would add.

One way the screen could be improved without adding too much extra labour would be as follows; instead of discounting EMS-treated plants which display fluorescence before the addition of MgSO<sub>4</sub>, these plants should still be tracked, but not have MgSO<sub>4</sub> added to them. If the fluorescence reduces over time then they can be excluded, and they are likely to be the result of asynchronous development. However, if these plants do not exhibit a reduction in fluorescence despite the absence of MgSO<sub>4</sub>, then they should be taken to the next generation, as they could contain a mutation which enhances the degradation of mature miRNA which could be novel, and could well be a component of miRNA decay pathways.

### Validating the phenotype by qPCR

Unfortunately, the two-tailed qPCR-based validation of the M3 generation of mutants displayed a large amount of variation. This could have potentially excluded some genuine mutants from being taken to the next stage of

investigation, as they would not differ from the control in a statistically significant way. This variability is likely compounded by the limited availability of qPCR primer design sites. This is because the two-tailed hemiprobes used to capture the miRNAs are small (roughly 70 – 80bp), and the forward qPCR primer must contain a majority of the miRNA sequence in order to discriminate between `captured` miRNAs and empty, non-hybridised probes (Androvic et al., 2017). However, as root tissue was used to best reflect the screening conditions, based on the quantity of RNA routinely extracted from these tissues this method of quantification still appears the most justifiable. It could potentially be improved by adding in more technical replicates and perhaps another biological replicate. The one mutant whose phenotype was validated by this approach was *msm1* (Figure 7.3A). The error bars for this were most consistent of all the mutants, and so confidence could be had in the result.

## Conclusion

Taken together, *msm1* was taken to the next stages of investigation following the analysis of the M3 generation. This was because it showed a reduction in miR395 accumulation in the qPCR data (Figure 7.3A), as well as demonstrating a clear morphological phenotype commonly associated with miRNA pathway mutants – this is shown and discussed in chapter 4.

This approach did prove capable of discriminating differences in rates of GFP recovery, which in some cases did seem indicative of differences in miR395 levels when directly assayed by qPCR (Figure 7.3A). However, as is also clear from the qPCR data, many of the mutants which were isolated from this screen appeared to be false positives indicating a high false discovery rate, particularly for late response mutants.

## Chapter 4: Characterisation of a miRNA phenotype in *msm1*

### Introduction

In the previous chapter, putative mutants identified through the EMS screen were tested in the M3 generation for miR395 accumulation. From these, *msm1* was selected for further study as it had both an apparent miR395 accumulation phenotype, and also a clear morphological phenotype (Figure 8.1A).

*Msm1* was isolated as an early response mutant. This meant that it recovered GFP earlier than the control, which suggested that it had a reduced miRNA accumulation. As described in the screening chapter, mutants were pre-screened before the addition of sulphate for GFP, and those that were fluorescent were discarded. Therefore, the miRNA accumulation phenotype of *msm1* is not expected to be dramatic, but still sufficient to alter the level of gene silencing.

The purpose of the EMS screen was to identify mutants in miRNA turnover pathways. However, it is not yet known if the reduced accumulation of miR395 in *msm1* is a result of increased degradation or reduced biogenesis of the mature miRNA. Additionally, all miRNA phenotyping of *msm1* has currently focussed around miR395. This miRNA is central to sulphate starvation. However, sulphate pathway mutants could also produce a similar phenotype.

In order to determine whether the difference in miR395 levels seen in *msm1* was due to a sulphate starvation mutant, it could be compared against another stress-responsive miRNA which accumulates in response to a different, unrelated stress. There exist other plant miRNAs which regulate response to nutrient deficiency stresses, such as miR399. This miRNA, similar to miR395, is upregulated in response to low levels of inorganic phosphate ( $P_i$ ), however in contrast to miR395 it exerts its regulatory effects in a canonical way. As miR399 levels increase, its target transcript, that of a ubiquitin-conjugating enzyme (UBC) involved in response to low  $P_i$  levels, decreases (Fujii et al., 2005). This miRNA represents a good initial candidate to test in *msm1*, as it exists in a

different regulatory network to miR395. It can therefore serve to exclude the possibility that the phenotype of *msm1* is sulphate starvation dependent.

In order to characterise *msm1* as a miRNA degradation mutant, it would be expected that mature miRNA would accumulate to a lower level, without an accompanying decrease in the level of pri-miRNA transcription. It would also be expected that such a mutation would affect the accumulation of all mature miRNAs in a similar way. This was because degradation should represent a key regulatory step that all mature miRNAs go through. Additionally, from previous studies performed into the SDN family of exonucleases, knockdown of SDN1/2/3 appeared to result in an increased accumulation of all miRNAs assayed (Ramachandran and Chen, 2008), suggesting a common degradation pathway.

It is also necessary to assay the levels of other miRNAs. Both miR395 and miR399 are stress responsive miRNAs and so, in principle, a mutant with an impaired stress response could also produce an apparent miRNA phenotype if only stress-responsive miRNA were analysed. Therefore, it is important that both stress responsive and non-stress responsive miRNAs are studied. One technique that would be highly appropriate for this would be sRNA sequencing. This technique involves the isolation and next generation sequencing of sRNAs (Lopez et al., 2015). From the data produced, all of the expressed sRNAs can be quantified and a corresponding miRNA profile of a sample can be constructed. This allows for conclusions to be drawn about the global populations of miRNAs, such as if there is a global decrease in total miRNA abundance in *msm1* relative to a control.

In addition to molecular phenotypes, based on studies of previously described miRNA accumulation mutants, there are also some morphological and physiological phenotypes which appear to be common. Most global miRNA accumulation mutants have reduced fertility – this is true of HYL1, HEN1, DCL1, HST, and AGO1 (Oliver et al., 2017). This can be characterised by truncated siliques, which are the seed containing organs of *Arabidopsis thaliana* (Oliver et al., 2017). A short root phenotype is also commonly seen in many miRNA accumulation mutants, such as *hen1* and *mac7* (Jia et al., 2017).

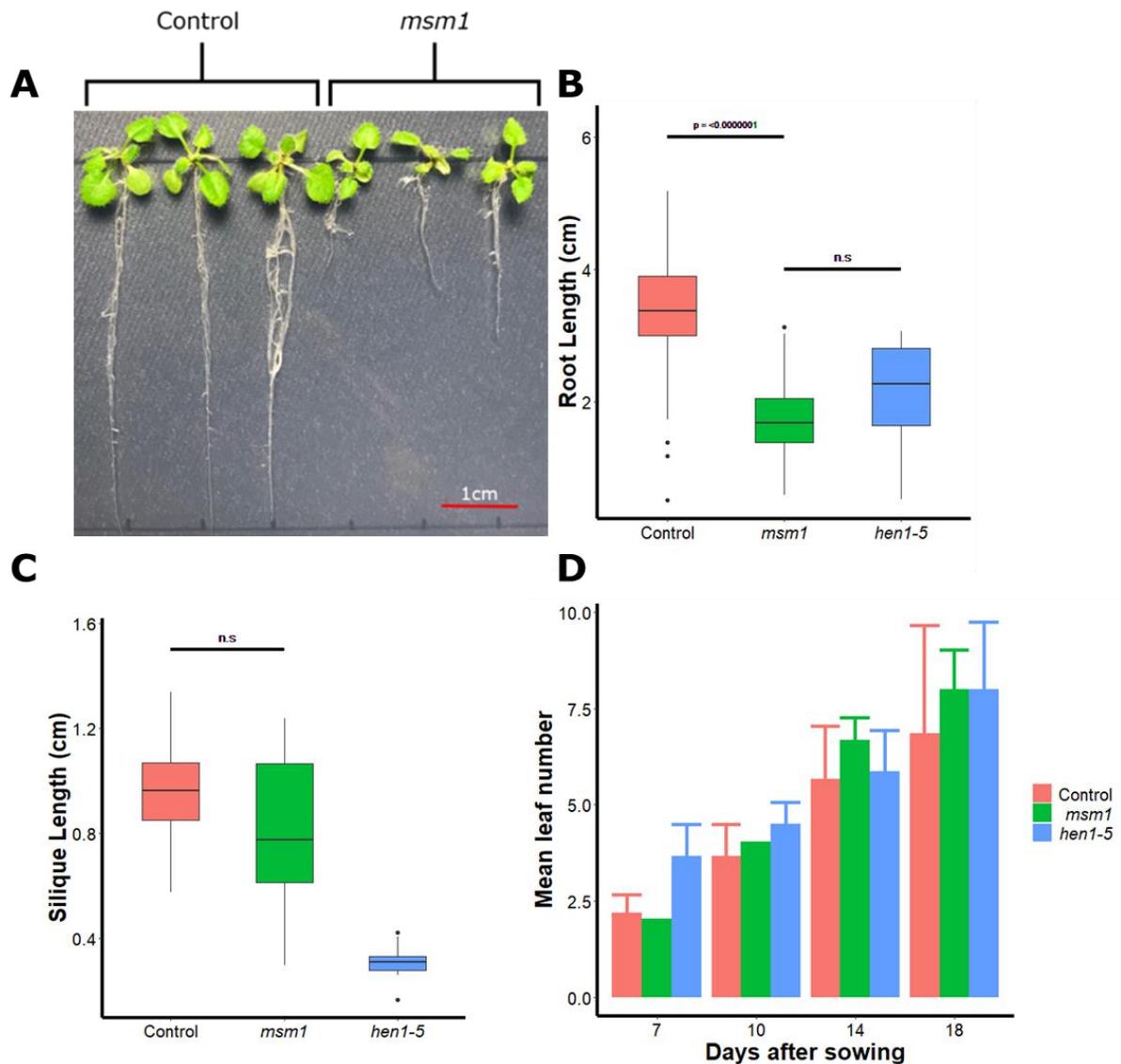
## Results

### *Msm1* morphology

The clearest morphological phenotype associated with *msm1* is the short root (Figure 8.1A). On average *msm1* roots are 2cm shorter than the corresponding control (Figure 8.1B) at 16 days. As *msm1* was an `early-response` mutant, it was expected to have reduced miRNA accumulation. Therefore, when assessing its morphological and physiological phenotype, it was compared to a known, global miRNA accumulation mutant, *hen1-5*. A significantly shorter root is also seen in *hen1-5*, when compared to the control. There was no significant difference between the mean root length of *hen1-5* and *msm1* seedlings at 16 days (Figure 8.1B).

Mature siliques were then measured from control, *msm1* and *hen1-5* plants. As expected, *hen1-5* had significantly shorter siliques than the control. This has been previously documented (Oliver et al., 2017). Interestingly, there was no significant difference in the mean silique lengths of *msm1* and the control (Figure 8.1C), suggesting *msm1* did not have reduced fertility.

Finally, as miRNAs are critical for growth and development (de Lima et al., 2012), I reasoned that a miRNA accumulation mutant may have an impaired growth rate. To this, the phyllochron of *msm1* was measured by counting the number of leaves on *msm1* (Boyes et al., 2001), control and *hen1-5* seedlings on pre-determined days. This was carried out in order to determine if the rate of growth of *msm1* was reduced relative to the control, which might be expected of a miRNA accumulation mutant. However, none of the backgrounds which were assessed demonstrated any significant difference in phyllochron measurements (Figure 8.1D).



**Figure 8.1 Quantification of morphological phenotypes displayed by *msm1***

**A:** Roots from pSUC2:GFP:395 and *msm1* seedlings at 16 days.

**B:** Boxplot showing root measurements taken from populations of pSUC2:GFP:395, *msm1* and *hen1-5* seedlings at 16 days. Black dots represent outliers.

**C:** Boxplot showing mature silique measurements taken from populations of pSUC2:GFP:395, *msm1* and *hen1-5* plants.

**D:** Bar chart showing mean leaf emergence in populations of pSUC2:GFP:395, *msm1* and *hen1-5* seedlings on the specified days. Error bars represent standard deviations of each population of  $n = 6$ .

### *Msm1* miRNA accumulation:

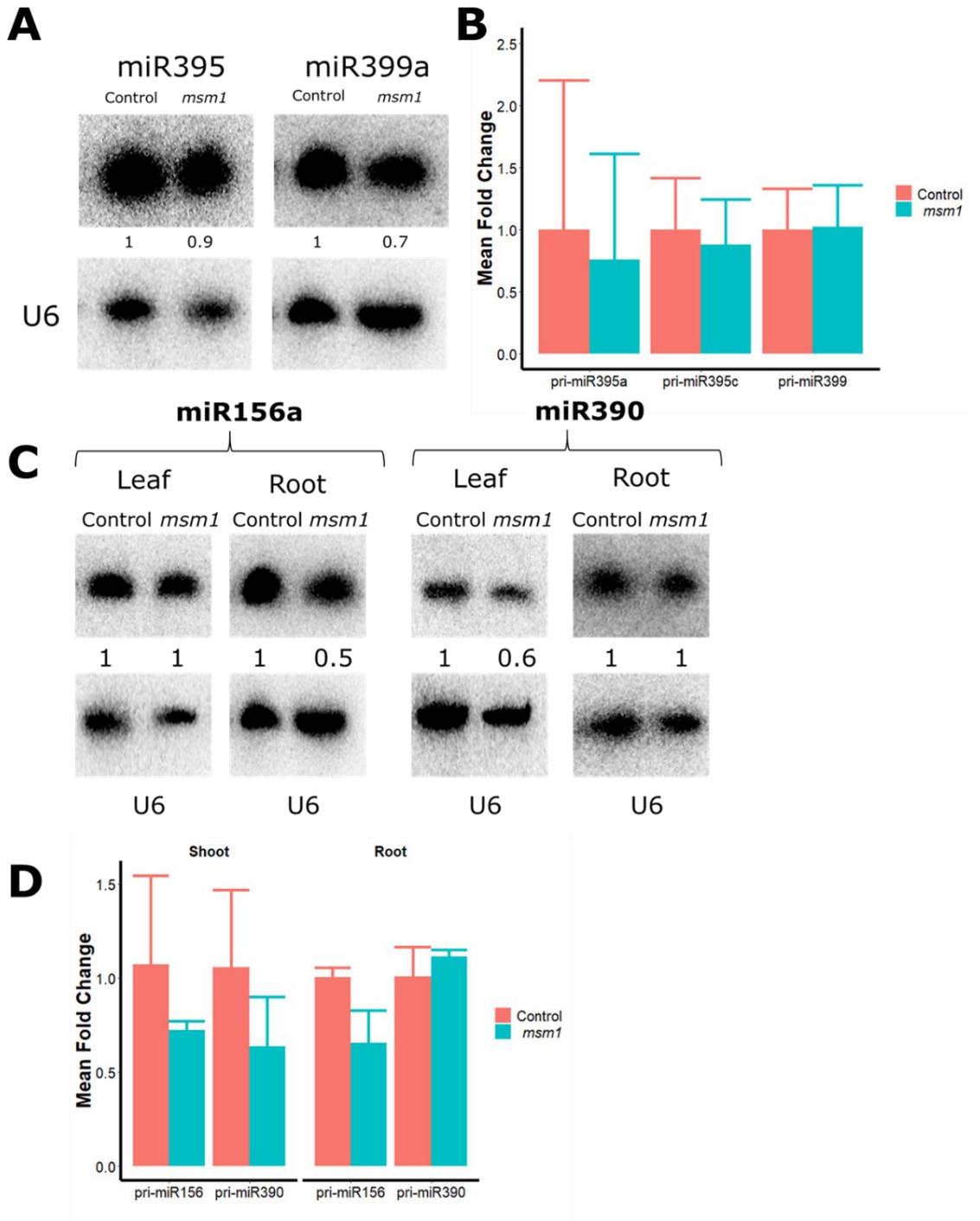
In order to quantify mature miRNA accumulation in the next generation of *msm1*, sRNA northern blots were performed. This was because the quantity of RNA was no longer a limiting factor, and this was deemed to be a higher confidence technique to assay mature miRNA in a quantitative manner. Initially, root tissue was used from plants grown on low sulphate media and low phosphate media (Figure 8.2A). The membranes were hybridised to probes for the corresponding nutrient stress responsive miRNA – miR395 for low sulphate (Kawashima et al., 2011) and mir399 for low phosphate (Fujii et al., 2005). The accumulation of miR395 in *msm1* as shown by Northern blot (Figure 8.2A) was reduced, although not as substantially so as had previously been shown by two-tailed RT-qPCR (Figure 7.3). However, this reduction was consistent over 3 replicates, and the hybridisation was performed using a custom locked nucleic acid (LNA) probe, which should enhance its accuracy (Válóczi et al., 2004). The reduction seen in accumulation of phosphate starvation responsive miR399 was even more dramatic than that seen in miR395 (Figure 8.2A). This was encouraging as it disentangled the apparent miRNA phenotype from the sulphate assimilation pathway.

A reduction in abundance of mature miRNA could arise not only from an increase in degradation, but also from a reduction in biogenesis. To determine whether or not these miRNAs had reduced abundance because they had reduced expression, I performed RT-qPCR on the pri-miRNA transcripts of selected loci. These were pri-miR395a & c as these were the most active miR395 loci (Kawashima et al., 2009), and pri-miR399a. None of these loci appeared to show any significant difference between the control and *msm1* (Figure 8.2B).

RNA from the P- experiment was used for Northern blots in Figure 8.2C as there was an abundance of this left over. This was not expected to confound the results as both genotypes received the same treatment, and so should still be comparable. At this stage, both root and shoot tissue were used, however the two were kept separate. This was because we were looking for a global effect on miRNA accumulation across the whole plant, however we also did not want

to dilute any potentially significant difference between the roots by combining with the shoot tissue, as this represented a greater biomass. The selected miRNAs were miR156 and miR390, as these are both well characterised Arabidopsis miRNAs known to be expressed at high levels in both root and shoot tissue (Allen et al., 2005; Gao et al., 2018). Interestingly, it appeared as if the accumulation of the same miRNA differed in the root and aerial tissue between the control and *msm1*. In the shoot tissue, miR390 accumulated less highly in *msm1* than in the control (Figure 8.2C), in a way that was conserved across three biological replicates. This was not true of the root tissue, however, in which the miRNA accumulation appeared to be the same (Figure 8.2C). The converse was true of miR156, which accumulated to the same level in *msm1* shoots as the control, but had roughly half the accumulation in *msm1* roots.

Again, these differences in accumulation could be attributable to an altered level of transcription. In order to rule this out, qPCR was performed on pri-miRNA transcripts of the assayed miRNAs in the corresponding tissues (Figure 8.2D). Whilst there was no significant difference between the levels of any of the pri-miRNA transcripts, it should be noted that the difference between pri-miR156 levels in *msm1* roots versus control roots came close to the significance cut off ( $p = 0.059$ ).



## **Figure 8.2 – Quantification of candidate miRNAs by northern blot and of precursors by qPCR**

**A:** Northern blot showing accumulation of miR395 and miR399 in *msm1* and pSUC2:GFP:395. Normalised to U6.

**B:** Bar chart showing qPCR data for pri-miR395a/b and pri-miR399a. Error bars represent standard deviation from 3 biological replicates.

**C:** Northern blot data showing miR156a and miR390 accumulation in *msm1* and pSUC2:GFP:395 shoot tissue and root tissue. Normalised to U6.

**D:** qPCR data from *msm1* and pSUC2:GFP:395 shoots and roots, showing relative fold change of pri-miR156a and pri-miR390. Error bars represent standard deviation from 3 biological replicates.

### sRNA sequencing

As there was preliminary evidence that some miRNAs had differential abundance in *msm1*, we decided to perform sRNA sequencing on *msm1*. For this experiment, seedlings were grown on normal MS media for 21 days. Root and shoot tissues were separated on harvest, and sRNA libraries were prepared for each separately, as there appeared to be a difference between roots and shoots in the accumulation of different miRNAs in *msm1* relative to the control. This would allow for the assessment of the levels of all miRNAs, rather than a select few.

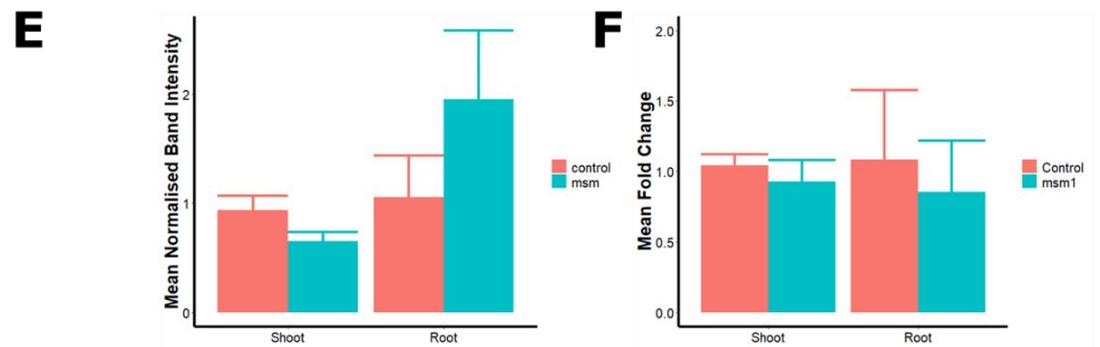
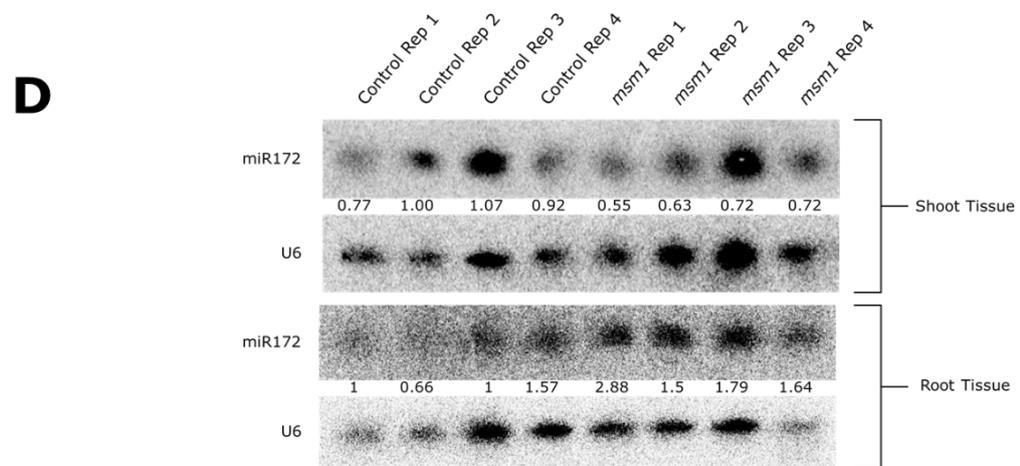
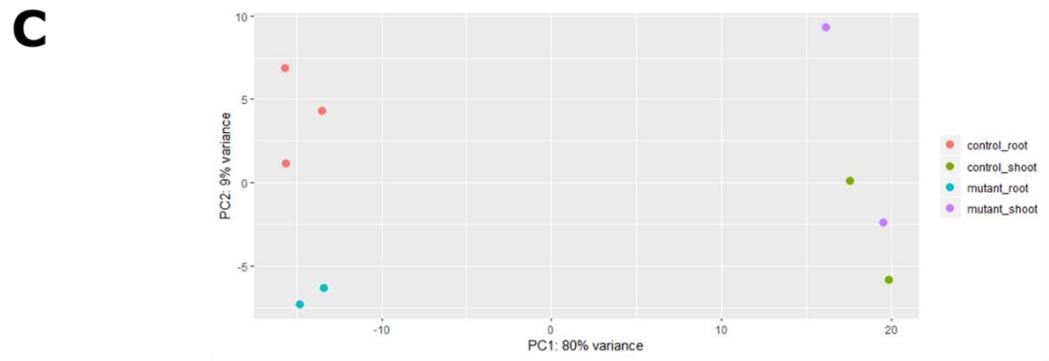
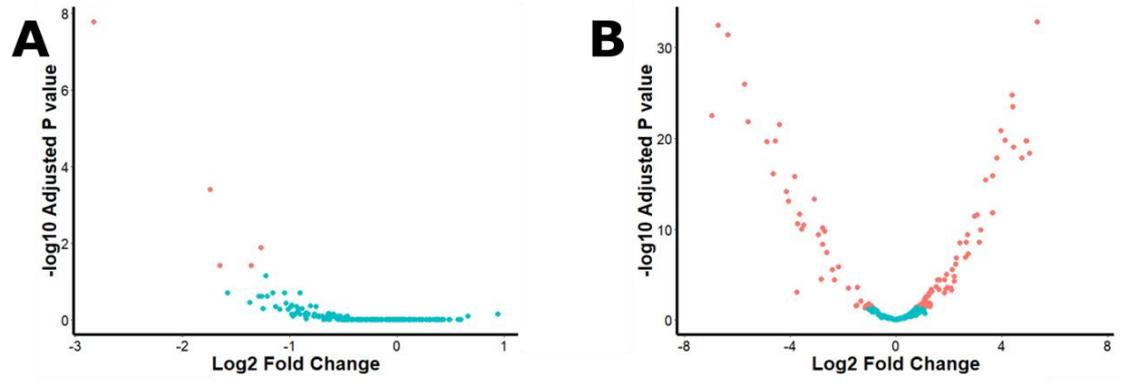
The miRNA profile of *msm1* roots was substantially different to the control root (Figure 8.3A). In *msm1* roots, 101 miRNAs had statistically significant levels of differential expression. Among these is miR156, which was previously demonstrated to be lower in *msm1* roots by sRNA Northern blot (Figure 8.2C). This strengthens the sRNA sequencing results. Interestingly, roughly a similar number of mature miRNAs have increased abundance as do decreased abundance in *msm1* roots, which did not fit with hypothesis that *msm1* had global reduction in mature miRNA levels (Figure 8.3A).

In *msm1* shoots, only 5 miRNAs had significant differential abundance (Figure 8.3B). This is perhaps not unexpected as the morphological phenotype of *msm1* appears to be restricted to the roots, however it again does not support the

hypothesis that *msm1* has global reduction in mature miRNA levels. MiR390 which had previously been shown to have increased abundance in *msm1* shoots but not roots (Figure 8.2C) was not significantly different in *msm1* shoot sRNA libraries. This may be because the shoot sRNA libraries were of lower quality than the root libraries. The root sRNA libraries cluster more tightly in the principle component analysis (Figure 8.3C) within replicates than do the shoot sRNA libraries, and there is also a greater separation between the genotype clusters for the root libraries than the shoot. The increased between-replicate variation in the shoot libraries reduces the statistical strength of testing, and therefore a number of significantly different miRNAs may have been missed in the sRNA library analysis. A full list of all the miRNAs tested along with their log2fold change and associated p value can be found in supplemental tables S2 and S3.

From this sRNA sequencing analysis, a number of mature miRNAs were determined to have differential abundance. In order to then validate these differences, a further set of Northern blots were performed on seedlings grown on MS media. The statistically strongest candidate miRNA that had differential abundance both in shoot and root was miR172. This was determined to be lower in *msm1* shoots than control shoots, and higher in *msm1* roots than control roots. The Northern blot data supported these predictions (Figure 8.3D), however the scale of the change was not as significant as reported by the sRNA sequencing results. The Northern blot is more likely to be accurate as sRNA sequencing has many steps at which bias can be introduced, such as PCR and adapter ligation steps (Baran-Gale et al., 2015).

The levels of pri-miR172a were then quantified in the same RNA samples used to prepare the libraries. This miR172 locus was selected because it represented the highest abundance locus in the sRNA sequencing data. As with the other pri-miRNA loci tested, no significant difference was identified between pri-miR172a accumulation in *msm1* and the control in any tissue tested (Figure 8.3D).



### **Figure 8.3 Analysis of differentially abundant miRNAs identified through sRNA sequencing**

**A:** Volcano plot showing log<sub>2</sub> fold change and  $-\log_{10}$  p value of miRNAs measured in *msm1* shoots with fold change relative to pSUC2:GFP:395 shoot data.

**B:** As in A but measurements taken from *msm1* roots and fold change relative to pSUC2:GFP:395 root data.

**C:** Principle component analysis output from DESeq2 analysis of *msm1* and control root and shoot sRNA libraries.

**D:** Northern blot showing levels of miR172 in *msm1* and control shoots and roots.

**E:** Bar chart showing average normalised band intensity taken from Northern blots shown in (D). Error bars represent standard deviation of four biological replicates.

**F:** Bar chart showing qPCR of pri-miR172 in Control and *msm1* shoot and root tissue. Error bars represent the standard deviation of three biological replicates.

#### Differential Truncation and Tailing analysis:

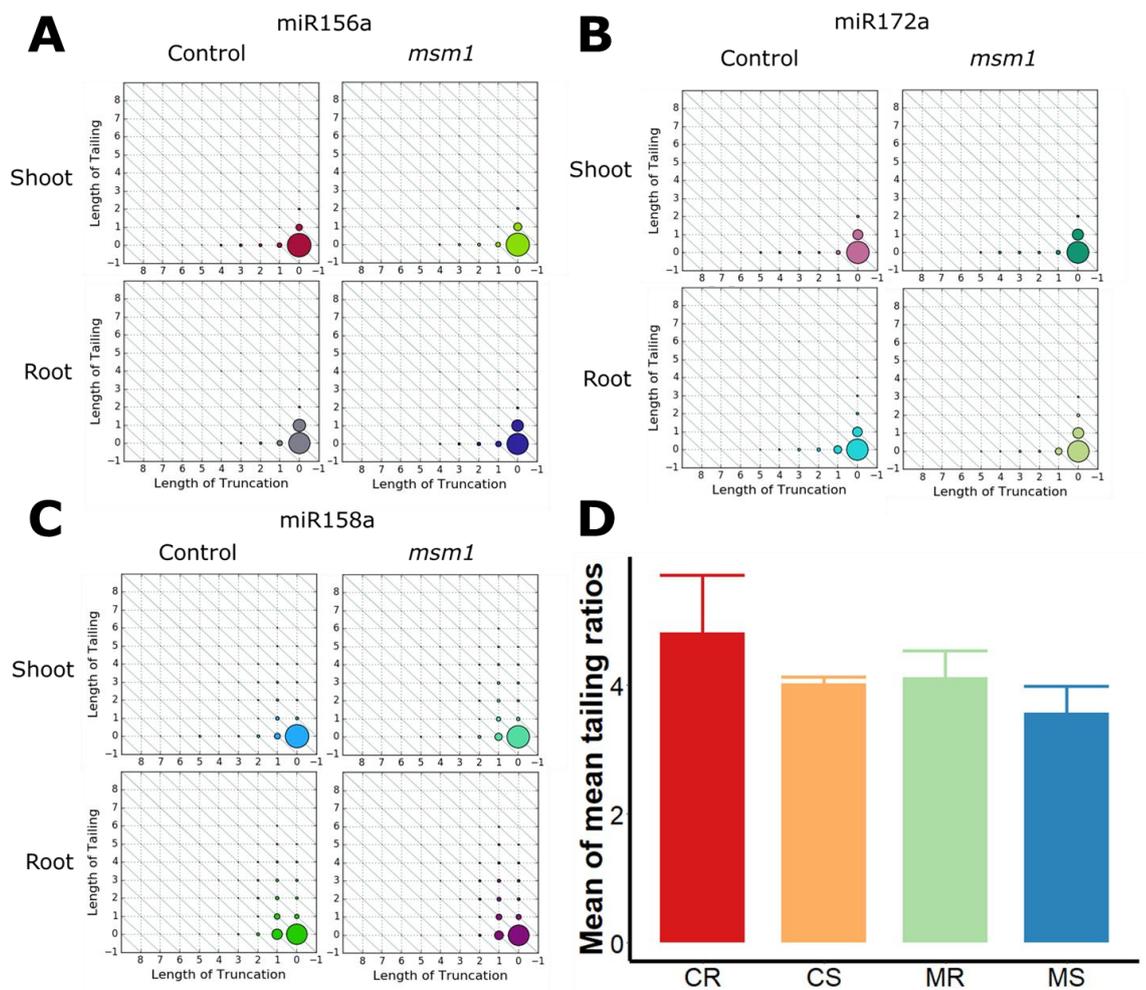
Degradation of mature miRNAs has been shown to leave behind hallmarks which are detectable by next generation sequencing (Zhai and Meyers, 2012). There are three clear types of these. One is truncated sRNAs which map to the annotated miRNA in the 5' region, but have missing nucleotides from the 3' end. The second are mature miRNAs which have been tailed by the non-templated addition of uridines, which map perfectly to the annotated miRNA but have additional nucleotides at the 3' end, and the third hallmark is a combination of both truncation and tailing (Zhai and Meyers, 2012). A tool, known as miTRATA, already exists for the detection and quantification of these degradation intermediates (Patel et al., 2016).

Control and *msm1* sRNA sequencing data was processed using miTRATA. The truncation and tailing levels were visualised for miR156 and miR172, as both of these miRNAs had been shown to be differentially abundant both in the sRNA

sequencing data and also in the Northern blot data (Figure 8.4A, B & C). MiR158 was also visualised as a positive control, as populations of this mature miRNA have previously been shown to be partially un-methylated, and therefore undergo relatively high levels of truncation and tailing in a wild-type background (Zhai et al., 2013).

In both the roots and shoots of *msm1*, miR156a and miR172a did not appear to show any clear difference in the pattern of truncation and tailing (Figure 8.4A & C). This was also true of miR158a, which, despite having more truncation and tailing still did not show a difference in pattern between the tested genotypes. The truncation and tailing ratios for each miRNA in each condition were then tested against the truncation and tailing ratios of the corresponding miRNA in the contrasting condition by Wilcoxon rank sum test. None of the tested miRNAs demonstrated a significantly different truncation and tailing ratio in one background relative to another (data not shown).

It was also feasible that *msm1* may have an altered global truncation and tailing rate, which could be missed by testing each miRNA individually. Therefore, the mean truncation and tailing ratio of all miRNAs in each library was calculated and tested against the mean truncation/tailing ratio in the contrary condition. However, no significant difference was found between the mean truncation and tailing ratios globally between any of the conditions (Figure 8.4D). Taken together this suggested that *msm1* did not have an altered level of miRNA degradation detectable by NGS.



**Figure 8.4 – Analysis of miRNA decay intermediates by miTRATA**

**A/B/C:** miTRATA plot depicting the levels of truncation and tailing of the respective miRNA in the corresponding genotype and tissue. Radius of the circle represents the number of reads supporting isoforms which matched the number of nucleotides removed or added. Colours are arbitrarily assigned by miTRATA software to differentiate different libraries.

**D:** Bar chart showing mean truncation and tailing ratio of all miRNAs in the corresponding condition. Error bars represent the standard deviation from two biological replicates.

## Discussion

The short root phenotype seen in *msm1* was initially promising, as there are a number of other miRNA accumulation mutants which demonstrate a similar phenotype. However, it should be noted that many known miRNA accumulation mutants do also exhibit reduced fertility (Oliver et al., 2017), which *msm1* did not (Figure 8.1C). This could suggest that *msm1* is either a previously undescribed gene involved in the miRNA pathway, or a previously undescribed allele of a known gene. Finally, the levels of miRNA reduction appear subtle in *msm1*, and so they may not be sufficient to impact fertility levels.

The reduction in miR399 levels in addition to the reduction in miR395 levels was encouraging, as these two miRNAs operate in different regulatory networks. There is some overlap between these networks – for example upregulation of miR399 by low phosphate treatment results in a reduction of miR395 levels (Hsieh et al., 2009), but in *msm1* both miRNA levels decrease in their respective induction treatments which is what would be expected of a miRNA accumulation mutant, rather than a phosphate or sulphate pathway mutant.

The levels of miRNA transcription and degradation were quantified, and seemingly no difference was found between them (Figure 8.2B, 3.2D, 3.3F & Figure 8.4). It is possible that there may be some reduction in the levels of some pri-miRNA transcripts, as in many cases there did appear to be a lower level in *msm1* that came close to statistical significance (Figure 8.2D). It is also worth noting that some of the error bars on the pri-miRNA qPCRs were quite large. Reducing these was challenging for a number of reasons. Firstly, the sites for primer design are limited, as they must flank the miRNA stem-loop sequence in order to capture pri-miRNA transcripts which haven't been processed. They still need to adhere to qPCR amplicon size design principles (Bustin and Huggett, 2017). Additionally, many pri-miRNA transcripts have not been fully mapped, so it is safest to keep primer sites as close to the stem-loop as possible as there is a higher likelihood that they will fall within the real pri-miRNA transcript. Secondly, many pri-miRNA transcripts are present only at very low abundances. Twice as much template was used for the pri-miRNA qPCRs to attempt to offset this, however the Ct values were still very high indicating low abundance. Finally, following completion of the qPCR, it was noted that miR399d is the most

active miR399 locus, not miR399a (Lin et al., 2008). Therefore, it would have been better to have quantified this pri-miRNA transcript rather than pri-miR399a.

It may have been more discerning to compare the sum of truncation and tailing in each genotype rather than the mean, as it would have been more sensitive to outliers which could have been of biological significance. However, the sRNA libraries were not all sequenced to the exact same depth due to error in the pooling. This meant that for some libraries, many more truncation and tailing intermediates could be detected purely because the sequencing depth was greater. Additionally, the pairwise testing of all the miRNAs individually between genotypes should have been sufficient to identify any values which differed significantly in any given condition.

If the pri-miRNA qPCR data is to be believed, then the mature miRNAs which display a difference in abundance, do so without a corresponding decrease in transcription or increase in degradation. Therefore, the underlying cause of this difference must exist between these two stages of the miRNA `lifecycle`. This would most likely be in a precursor processing stage such as export of the miRNA duplex from the nucleus by HASTY (Park et al., 2005). Whilst it is still possible that there could be an increased or reduced rate of pri-miRNA processing to pre-miRNA, it is less likely. This is because the pri-miRNA qPCR primers flanked the miRNA stem-loop. Therefore, when the stem-loop is liberated from the pri-miRNA transcript, the primer site containing regions will be separated from one another into two transcripts, which will be susceptible to degradation (Nagarajan et al., 2013). Therefore, an altered rate of processing should result in an altered accumulation of pri-miRNA transcript, which would be detected by the pri-miRNA qPCR.

An altered rate of pri-miRNA to pre-miRNA processing could still be tested by performing qPCR on pre-miRNA transcripts and assessing for differences between genotypes. This was not performed with this cDNA, however, for technical reasons. The cDNA libraries were prepared using an oligo-dT RT primer, in order to capture polyadenylated RNA pol II transcripts which included pri-miRNA transcripts (Lee et al., 2004). However, pre-miRNAs are liberated from the pri-miRNA transcript by DCL1 mediated endonucleolytic cleavage

(Kurihara and Watanabe, 2004), meaning the polyA tail is no longer present on these sequences. Therefore, to detect these RNA species, random hexamers or gene specific RT primers would have to have been used.

The miRNA profile of *msm1* roots was substantially different to the roots of control plants. Such a difference was not seen in the shoots. This was consistent with the morphological phenotype seen in *msm1*. Unexpectedly, some miRNAs appeared to increase in *msm1* as well as decrease. There is a precedent for factors involved in miRNA pathways affecting different miRNA species differently. For example, FIERY1, which promotes miRNA accumulation by reducing ribosome derived siRNA AGO1 competition, shows increases in some miRNAs and decreases of others on mutation (You et al., 2019). This is also seen in mutants of TOUGH, which enhances the cleavage efficiency of DCL1 and aids in the recruitment of pri-miRNAs (Ren et al., 2012). Therefore, the observation that the direction of change of miRNAs is not unilateral in *msm1* does not necessarily conflict with the hypothesis that *msm1* is involved with miRNA processing.

MiR399, miR395, miR156 and miR172 were all shown to have an altered accumulation in *msm1* roots by Northern blot. There is some functional relationship between these miRNAs. As previously stated, phosphate starvation induces miR399 and reduces miR395 (Hsieh et al., 2009). Additionally, miR156 levels have been shown to increase in shoot tissue in response to low phosphate, however not in root (Hsieh et al., 2009). MiR156 and miR172 have also been shown to be involved in some of the same pathways. MiR156 has been shown to act upstream of miR172, and repress SPL9, which is a positive regulator of miR172 expression. This regulatory axis controls the juvenile to adult transition, with miR156 often being considered the master regulator of the juvenile stage (Wu et al., 2009). Therefore, if miR156 levels were lower in *msm1* relative to the control, it would be expected based on the biology of this miRNA, that miR172 levels would be higher. This matches the data. In this scenario, where a knockdown of miR156 causes an increase in miR172, it would be expected that pri-miRNAs would also differ. This was not seen in my data, although as previously stated the accuracy of these qPCR data are limited, and miR156 pri-miRNA levels came close to being significantly reduced ( $p = 0.059$ ).

As there are a number of such regulatory circuits in which one miRNA indirectly regulates another, the overall miRNA profile becomes harder to interpret. This is because many of the differentially abundant miRNAs may be differentially expressed as a result of an increase or decrease in the abundance of another miRNA. Therefore, in the event of a mutation which alters the accumulation of a selection of miRNAs - for example in a specific transcription factor or splicing factor - one might expect to see a much larger number of miRNAs affected because of indirect interactions, similar to what is seen *msm1* roots.

It is difficult to assess any broad scale, functional enrichment of all the mature miRNAs that are different. This is because many miRNAs have a number of different targets, and a majority of these targets are transcription factors (Samad et al., 2017), which can have a variety of different functions depending on age, tissue and environment. For this reason, it is also probably not that useful to look for functional enrichment in predicted targets of miRNAs, as the identified target transcripts may not be expressed in the same tissues at the same time. In order to gain a better understanding of the types of genes which are differentially expressed in *msm1*, mRNA sequencing would be more appropriate.

The apparent tissue specificity of the *msm1* phenotype is interesting, as miRNAs are involved in most if not all physiological processes (Samad et al., 2017). Therefore, it would be expected that miRNA pathway components would have ubiquitous expression across all tissue types. However, based on the root morphology and the tissue sRNA sequencing profiles, this does not appear to be the case in *msm1*. This would not be unexpected of a transcription factor, however the breadth of pathways that would need to be regulated by a single transcription factor sufficient to cause the changes in the number of different miRNAs in different regulatory networks would be vast. A splicing factor may also be a potential candidate, as these can influence the levels of expression of many different genes and are increasingly being linked to the miRNA pathway (Wang et al., 2019a) (Jia et al., 2017). There is also evidence that some splicing factors exhibit tissue specific expression patterns, which could explain why the main phenotype is restricted to the root (Fang et al., 2004).

Another consideration that must be taken into account is that *msm1* could be homozygous for multiple phenotypic mutations. The likelihood of this depends on the density of mutations induced by the EMS treatment, which is not currently known. Therefore, it is possible that the root phenotype and the apparent miRNA phenotype could be caused by two separate mutations. However, it is not possible to determine the number of phenotypic mutations at the M3 stage. This would require backcrossing, screening for the phenotypic segregants, and then mapping the causative mutation in these individuals. This will be covered in greater depth in Chapter 6.

In conclusion, *msm1* appears to have differential abundance of some miRNAs, but not all. This is most dramatic in the roots, which is accompanied by a short root phenotype similar to that seen in many miRNA accumulation mutants. Roughly equal numbers of miRNAs have decreased and increased expression in *msm1* roots, seemingly without a change in the abundance of pri-miRNA transcripts or rate of degradation. It is not yet possible to verify at this stage if *msm1* contains a mutation in a miRNA pathway component, a transcription factor, a splicing factor, or a combination of all three.

## Chapter 5: Identifying alternative splicing phenotypes in *msm1*

### Introduction

In the previous chapter, *msm1* was found to have an altered miRNA profile that was most prominent in the roots. It was not possible to distinguish whether *msm1* contained a mutation in a miRNA pathway factor, transcription factor, splicing factor, or combination of all three. Based on the miRNA phenotype seen in *msm1*, as well as the short root phenotype, I perceived a splicing factor mutation to be most likely in *msm1*.

A number of *Arabidopsis* alternative splicing factor mutants have accompanying miRNA phenotypes (Jia et al., 2017; Wang et al., 2019a; Li et al., 2018). In these backgrounds, it is not uncommon to see miRNA accumulation both increasing and decreasing, similar to what was seen in the *msm1* root sRNA sequencing. There are a number of reasons why this may occur. Firstly, between 42 – 61% of genes in *Arabidopsis* undergo alternative splicing (Reddy et al., 2013). This means that alterations in the levels of alternative splicing have the potential to dysregulate a substantial number of biological pathways. Many of these pathways may involve miRNA regulation, either by promoting the expression of certain miRNAs or repressing the expression of certain miRNAs. Therefore, it would not be unexpected to see bidirectional changes in miRNA levels in this scenario.

Secondly, miRNAs themselves are subject to splicing. A number of plant miRNAs contain introns (Szarzynska et al., 2009), which must be removed in order to form the necessary structures required for DICER processing (Bielewicz et al., 2013). One example of particular interest in *Arabidopsis* is miR846 and miR842. Both of these miRNAs are contained within the same primary transcript, and alternative splicing determines which of the two mature miRNAs are expressed, in an abscisic acid (ABA) dependent mechanism (Jia and Rock, 2013).

As described in greater depth in Chapter 1, ABA is intimately linked with alternative splicing. Next generation sequencing studies in *Arabidopsis* have shown that treatment with external ABA comprehensively affects alternative splicing profiles (Zhu et al., 2017). ABA itself also relies on alternative splicing for its own function. For example, the splicing factor DRT111 is responsible for the splicing of many genes involved in ABA responses. When it is knocked out in *Arabidopsis*, plants display defects in ABA mediated stomatal closure and hypersensitivity to ABA during seed germination (Punzo et al., 2020). Therefore, if *msm1* is an alternative splicing mutant, it would be likely to have an ABA phenotype.

Detection of alternative splicing has become much more accessible with the advent of NGS. As described in Chapter 1 (Alternative Splicing), alternative splicing is the term used to describe the phenomenon when a decision is made as to which splice site is used for a transcript, where the number of splice isoforms is greater than one. Short read RNA sequencing is commonly used in both quantitative and qualitative studies into alternative splicing as it provides a high sequencing depth and good gene coverage (Bedre et al., 2019). Using the data produced by these methods, it is possible to map the transcripts back to the genome and quantify the differing permutations of exons and retained introns (Marquez et al., 2012). This can be achieved with a multitude of publicly available bioinformatic tools.

One such tool that can identify alternative splicing events is SplAdder (Kahles et al., 2016). This tool calculates a Percentage Spliced In (PSI) value for each individual alternative splicing event for which it can find evidence for the junction in the RNAseq data, by dividing the number of reads supporting the inclusion by the total number of reads for the transcript. The use of this PSI value therefore negates the effects of differential expression of genes, as each event is represented as proportion of the total pool of transcripts detected. In addition to reporting the number of significantly different alternative splicing events, SplAdder also provides the type of alternative splicing event from a pre-defined set. These types are intron retention, alternative 3' splice site usage,

alternative 5' splice site usage, exon skipping and multiple exon skipping, which are described in Chapter 1.

## Results:

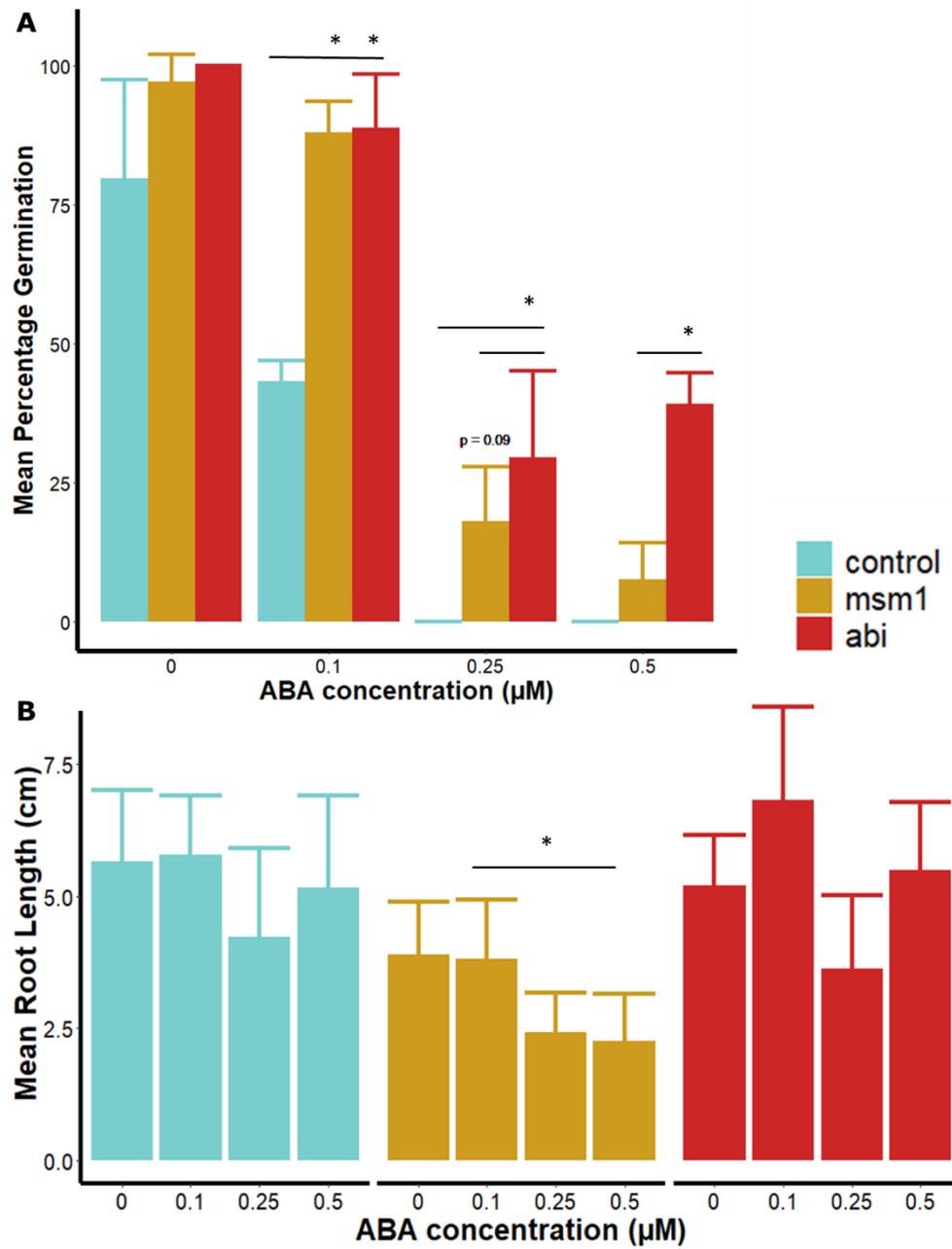
### ABA phenotype of *msm1*:

As many alternative splicing mutants exhibit altered ABA responses, I decided to test the ABA response of *msm1*. It has long been known that ABA inhibits seed germination (Schopfer et al., 1979). One simple method of assaying ABA sensitivity is through its germination rate on exposure to increasing concentrations of exogenous ABA.

Control, *msm1* and abscisic acid insensitive 4 (*abi-4*) seedlings were germinated on increasing concentrations of ABA. *Abi-4* was included as a positive control for the method, as it has previously been shown to have reduced sensitivity to ABA (Reeves et al., 2011). After 7 days of incubation in the growth chamber, the percentage germination was calculated for each background and each concentration of ABA. The germination rate of control seedlings decreased as the concentration of ABA increased, as expected. Similarly, as expected, the known mutant *abi-4* was resistant to increasing concentrations of ABA, showing no decrease at 0.1 $\mu$ M. At 0.25  $\mu$ M onwards, the germination rate dropped, but was still the highest of the three genotypes tested. Interestingly, *msm1* fell somewhere in between the two conditions. At 0.1 $\mu$ M it showed no decrease in the rate of germination. At 0.25 $\mu$ M, there was a decrease in germination, however the rate did not differ from that of *abi-4*. Finally, at 0.5 $\mu$ M, *msm1*'s germination rate had further decreased, such that it was lower than *abi-4*, but still higher than the control. Taken together this demonstrated that *msm1* has reduced sensitivity to ABA, which further supported the hypothesis that it was an alternative splicing mutant (Figure 9.1A).

Surprisingly, it was also found that in concentrations of ABA higher than 0.25 $\mu$ M *msm1* appeared to display a further reduction in root length (Figure 9.1B). This was not seen in either the control nor *abi-4*. However, it is plausible that these

seedlings germinated later than the seedlings grown in lower concentrations of ABA, which would explain the reduced root length.



**Figure 9.1 - Quantification of *msm1* germination and root length in response to increasing ABA concentrations**

**A:** Mean percentage germination of pSUC2:GFP:395, *msm1* and *abi-4* seedlings germinated on increasing concentrations of ABA. Error bars represent one standard deviation. Black lines and asterisks represent statistical comparisons and significance. N = ~50

**B:** Mean root length of pSUC2:GFP:395, *msm1* and *abi-4* seedlings measured at 14 days. Error bars, lines and asterisk used as before in (A).

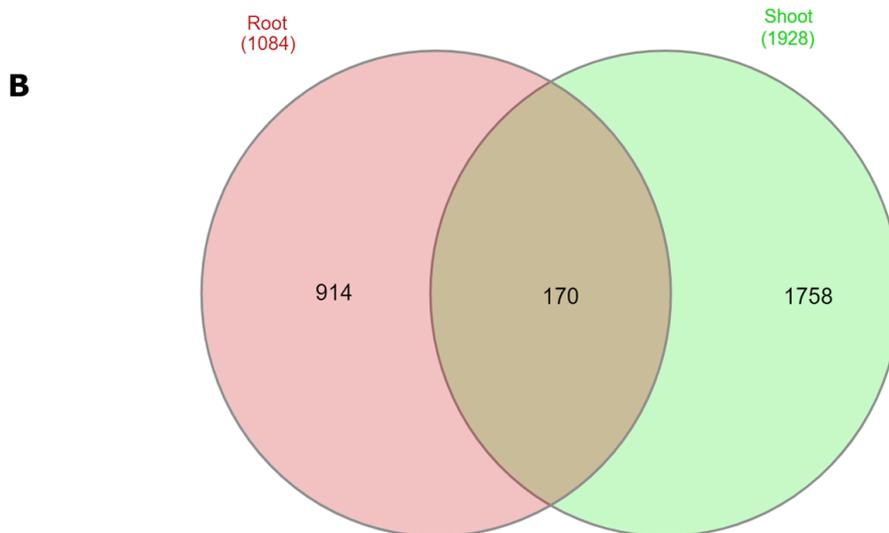
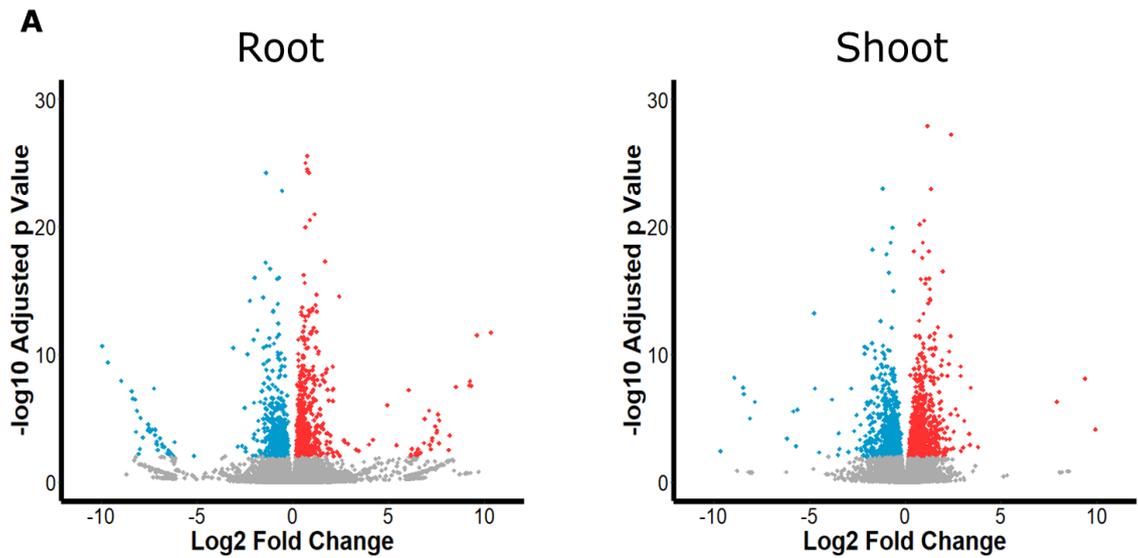
Differential expression analysis:

In order to identify and quantify any differences in the splicing profiles of *msm1* and the corresponding control, mRNA sequencing was performed separately on *msm1* roots and shoots, and control seedling roots and shoots. The same RNA used to prepare the sRNA libraries in chapter 4 was used for the mRNA sequencing. Contrasting the miRNA profile of *msm1*, both the root and the shoot showed significantly altered mRNA expression profiles (Figure 9.2A) relative to the control.

Further contrasting the miRNA profile seen in *msm1*, the greatest number of differentially expressed transcripts was in *msm1* shoots, in which 1928 transcripts passed the significance threshold ( $p = 0.01$ ). In roots, 1084 transcripts passed this threshold, although it is noteworthy that in *msm1* roots the scale of change was greater. These numbers include splice variants of genes.

Of these differentially expressed genes in *msm1* roots and shoots, 170 transcripts are commonly differentially expressed in both tissues (Figure 9.2B). A majority of the differentially expressed transcripts which did not overlap between root and shoot were present in the mRNA sequencing of the contrary tissue. 84% of the non-overlapping differentially expressed transcripts in the root tissue were present in the shoot tissue sequencing, and 90% of the non-overlapping differentially expressed transcripts in the shoot tissue were present in the root tissue sequencing. This indicated that there were tissue specific differences in the expression patterns of many genes. It also suggested that the

170 transcripts that were commonly differentially expressed were so for reasons other than coincidence of expression. A full list of significantly differentially expressed transcripts in *msm1* can be found in supplementary tables S4 and S5.



## Figure 9.2 – Comparison of mRNA differential expression profiles between *msm1* roots and shoots

**A:** Volcano plots showing log<sub>2</sub> fold change and associated p value of transcripts in *msm1* roots (left) and shoots (right). Blue points indicate a negative log<sub>2</sub> fold change and p ≤ 0.01, red points indicate a positive log<sub>2</sub> fold change and p ≤ 0.01, and grey points indicate p > 0.01.

**B:** Venn diagram showing the overlap between significantly differentially expressed transcripts in root (red) and shoot (green).

### GO Enrichment of differentially expressed genes:

Transcripts which were differentially expressed were analysed for `GO Biological Process` term enrichment (Ashburner et al., 2000) using ShinyGO v0.61 (Ge et al., 2020). A significance cut-off of 0.05 FDR was selected, and the top 30 most significant results were taken. A `Gene Ratio` was then calculated for each of the significantly enriched terms by taking the number of genes in the list matching the term and dividing by the total number of genes in *Arabidopsis* annotated with the respective term. Therefore, the Gene Ratio represents the proportion of the total number of genes with a GO term annotation that are present in the differentially expressed genes. For all GO figures the y axis is ordered in descending significance, so that the terms with the highest y axis values are the most significant.

Initially, the complete lists of differentially expressed genes for *msm1* shoots and *msm1* roots were analysed using ShinyGO (Figures 4.3A (top) and 4.3B (top)). This did not take into account whether the genes were upregulated or downregulated, and therefore represent a profile of all the biological processes which are altered in *msm1*, without directional information. Interestingly, in both *msm1* roots and shoots the most significant biological process terms that were enriched pertained to sulphur compound biosynthesis and metabolism. Many of the other terms which were significantly enriched in *msm1* shoots were related to abiotic stress responses, for example response to water deprivation, response to light stimulus, response to osmotic stress and response to cold (Figures 4.3A and 4.3B). Response to ABA was also one of the most significantly

enriched terms in *msm1* shoots, which agrees with the ABA response phenotype seen in *msm1* (Figure 9.1). Whilst there were also some abiotic stress related terms enriched in *msm1* roots, unexpectedly many of the terms concerned DNA and RNA synthesis, for example ribonucleoside metabolic process, purine nucleoside metabolic process and nucleobase-containing small molecule metabolic process (Figure 9.3B).

Transcripts were then separated based on whether they had a significantly increased log<sub>2</sub> fold change, or a significantly decreased log<sub>2</sub> fold change. These new lists were analysed with ShinyGO using the same settings as before. In both *msm1* shoots and *msm1* roots the sulphur compound metabolic and biosynthetic processes were most significantly enriched in genes which were upregulated (Figures 4.3A (middle) and 4.3B (middle)), and not present in the top 30 most significant terms for the downregulated genes (Figures 4.3A (bottom) and 4.3B (bottom)), suggesting a unilateral upregulation of these genes.

In *msm1* shoots, the biological process terms pertaining to abiotic stress response appear enriched in both upregulated genes and downregulated genes, suggesting a potential global dysregulation of abiotic stress response. It should be noted however, that more of these terms appeared in the downregulated gene list (Figure 9.3B (bottom)).

In *msm1* roots, the DNA and RNA metabolism terms previously commented on appear only in the upregulated gene list (Figure 9.3B (middle)), as well as the stress responsive terms. In the downregulated gene list, the most enriched terms relate to formation of cell walls, for example plant-type secondary wall biogenesis, hemicellulose metabolic process and glucuronoxylans metabolic process (Carpita, 2011). This is perhaps unsurprising given the morphological phenotype of the *msm1* root.

Finally, ShinyGO was used to analyse the 170 transcripts which were commonly differentially expressed between root and shoot. Sulphur compound biosynthesis and metabolism, and response to various abiotic stresses were

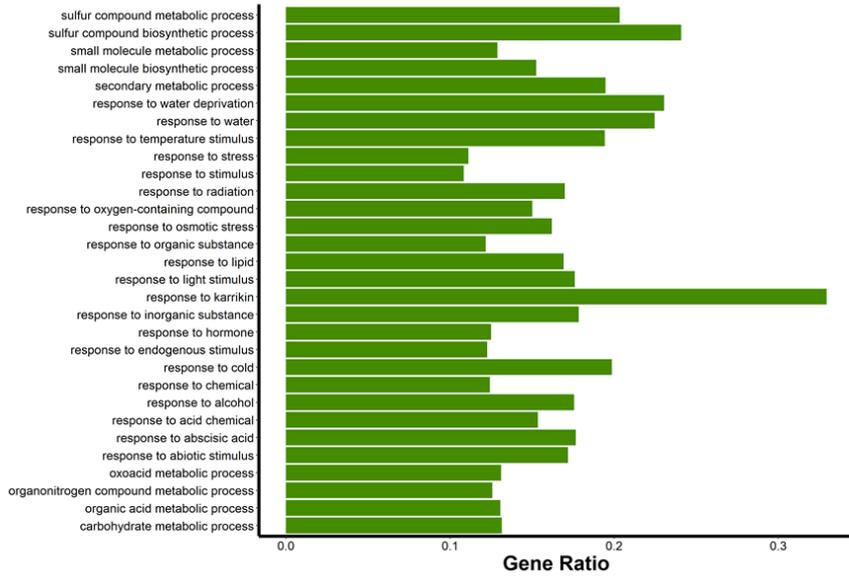
again significantly enriched (Figure 9.3C). This further reinforced that the underlying causative mutation in *msm1* was involved in stress response.

For all of the gene lists analysed using ShinyGO, no significantly enriched transcription factor binding motifs were found.

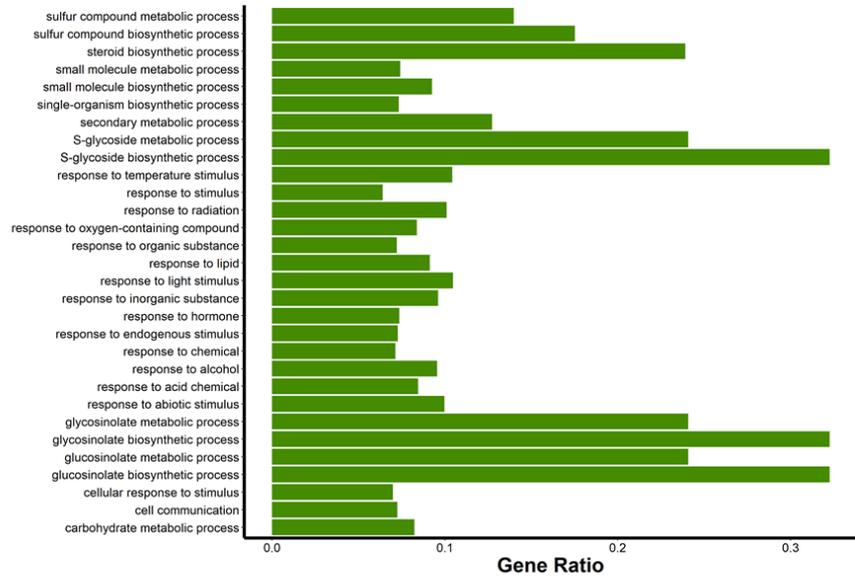
A

# Shoot

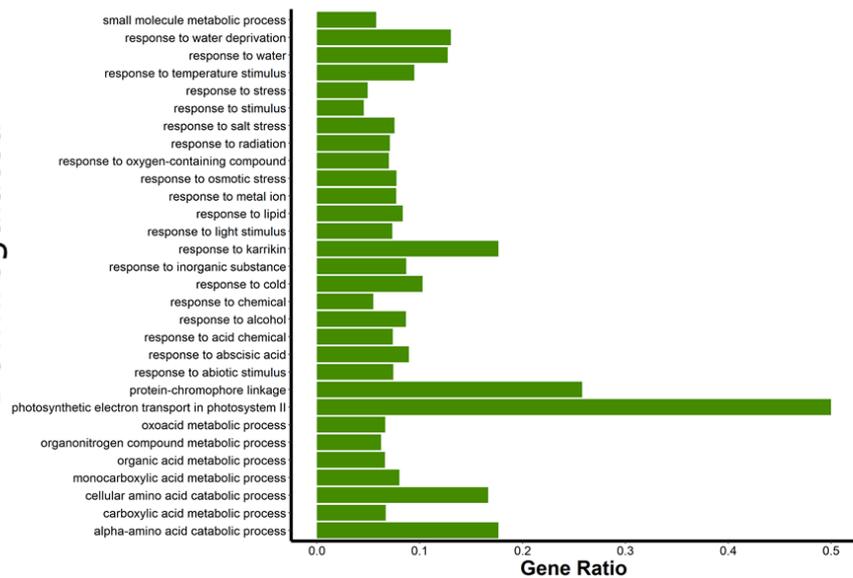
Differentially expressed

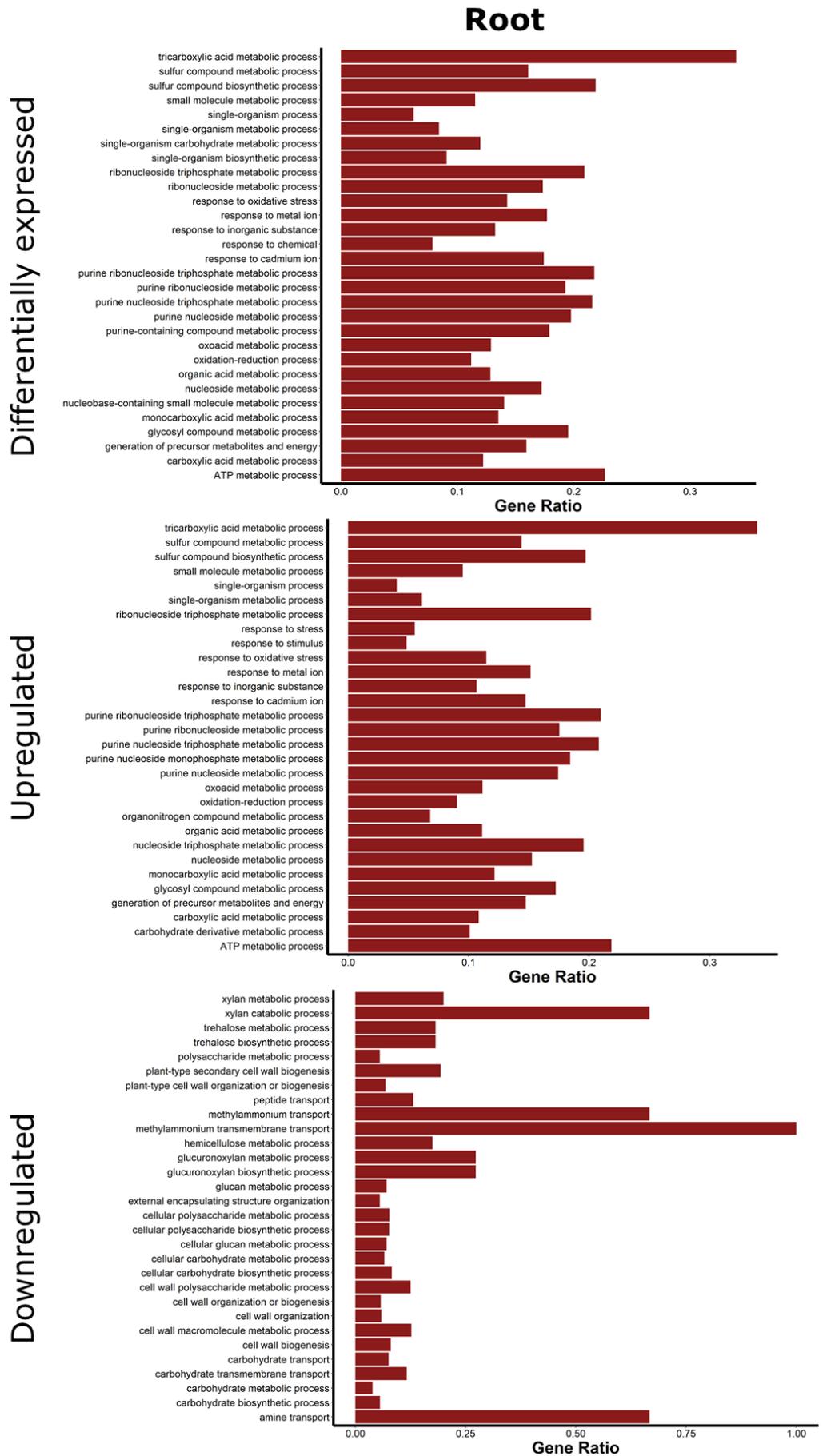


Upregulated

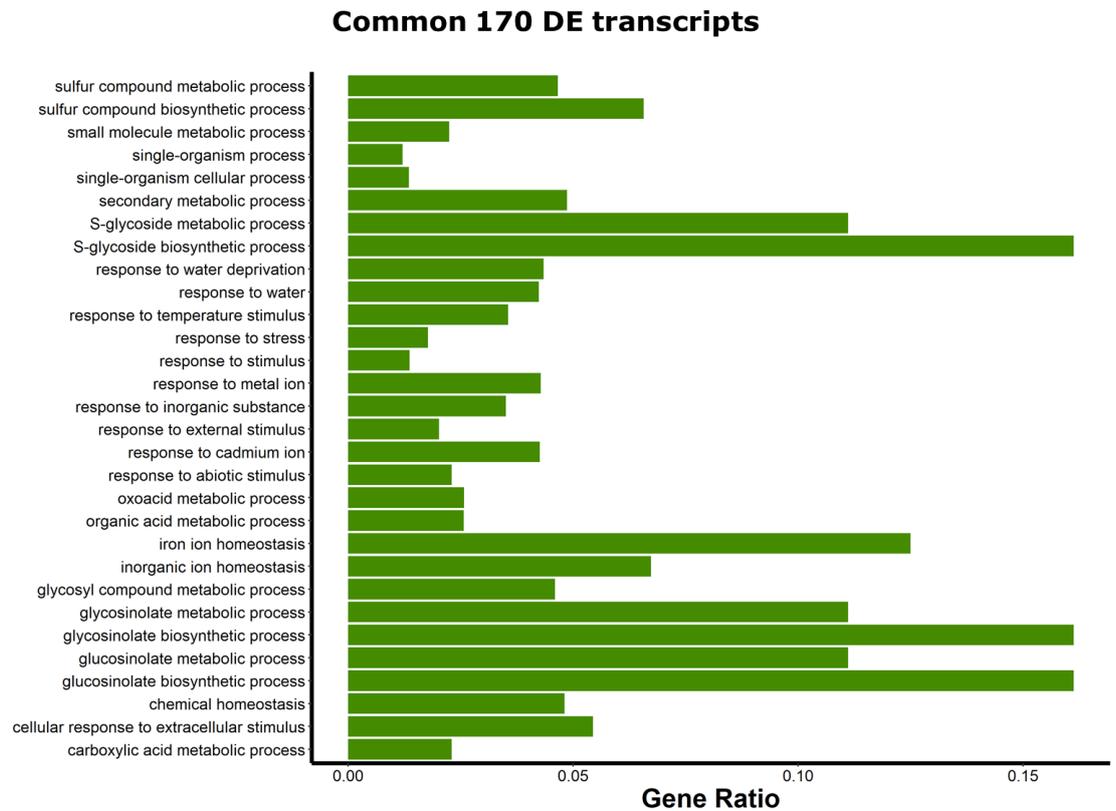


Downregulated



**B**

C



**Figure 9.3 GO profiles of differentially expressed genes in *msm1* roots and shoots**

**A:** GO term enrichment analysis of differentially expressed *msm1* shoots in all differentially expressed transcripts (top), transcripts with a positive log<sub>2</sub> fold change (middle) and transcripts with a negative log<sub>2</sub> fold change relative to the control (bottom).

**B:** Same as (A) but with *msm1* root mRNA sequencing

**C:** GO term enrichment analysis of all commonly differentially expressed transcripts between *msm1* shoot and root.

## SplAdder analysis of *msm1*

These PSI values of splicing events reported by SplAdder were then tested on an event by event basis between a number of genotypes.

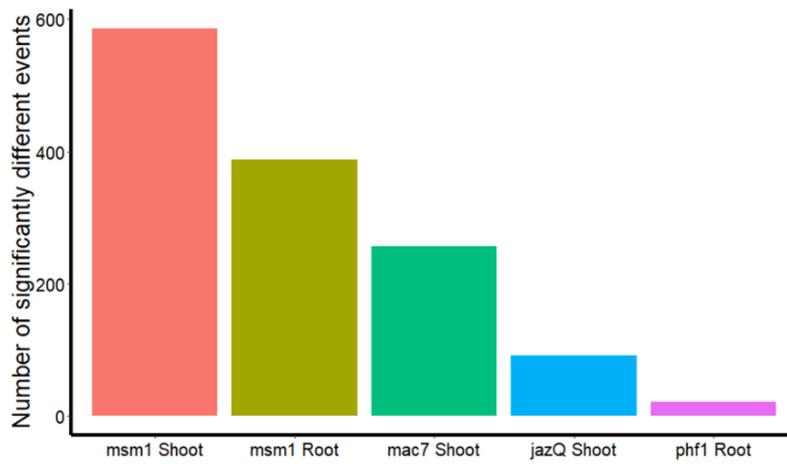
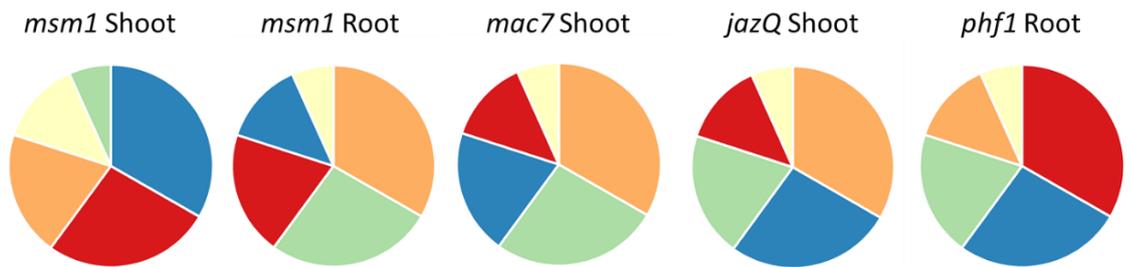
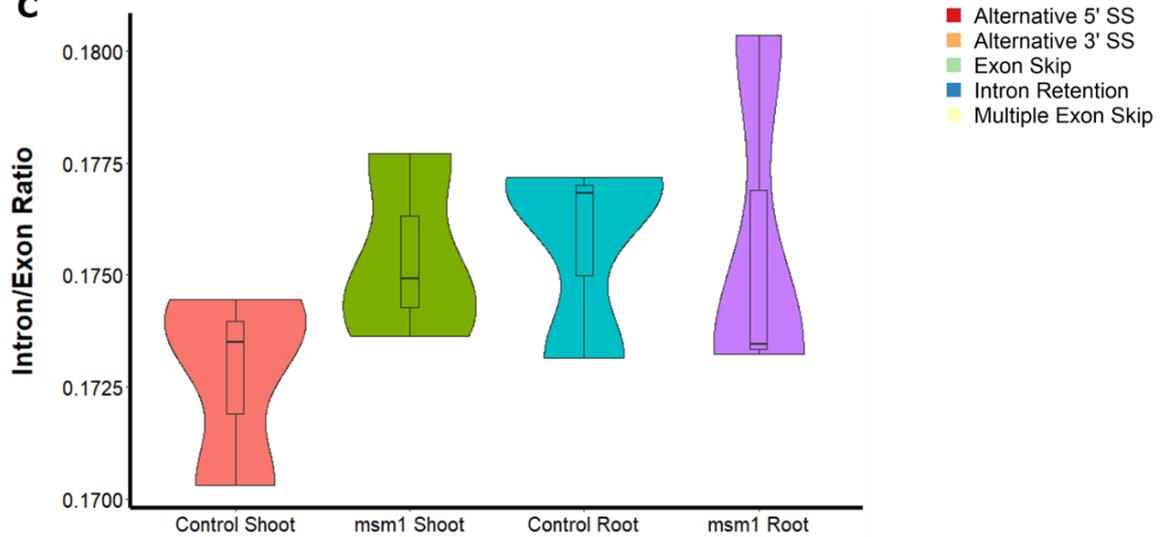
Firstly, the number of significantly different alternative splicing events between the control shoot and root and *msm1* shoot and root were calculated. In addition, as a control for the method, mRNA sequencing from a known alternative splicing mutant with a miRNA phenotype (*mac7-1*) was run alongside *msm1* samples, as well as mRNA sequencing from two mutants not considered to have an alternative splicing phenotype (*jazQ*, *phf1*). These datasets were chosen as they were all sequenced at an equivalent depth of 20 million reads, and were publicly available for download.

Both *msm1* roots and shoots appeared to have a larger number of significantly different alternative splicing events than any of the other genotypes tested (Figure 9.4A). This number was highest in *msm1* shoots with 586 events which were significantly different from the corresponding control shoots. The second highest was in *msm1* roots, at 387. Both of these numbers were higher than the published alternative splicing mutant *mac7-1*, although it is important to note that this is just the number of unique events, without taking into account the number of reads supporting them. There were still many events in the non-alternative splicing phenotype mutants *jazQ* and *phf1*. This is perhaps to be expected as alternative splice variants, particularly *phf1* as abiotic stress response and alternative splicing are strongly linked (Calixto et al., 2018; English et al., 2010).

Profiles were constructed of the types of significantly different alternative splicing events in each of the genotypes tested relative to their corresponding controls. These profiles represented the total sum of the supporting counts of every significantly different alternative splicing event junction of a given type. They therefore show the most common type of differential alternative splicing (Figure 9.4B) on a gene by gene basis.

In *msm1* shoots, which had the highest number of significantly different alternative splicing events, intron retention events were by far the most numerous event type that differed from the control. Following this were alternative 5' and alternative 3' splice site usage events. In *msm1* roots, interestingly intron retention events represented only the fourth most abundant significantly different event type. The most common events were alternative 3' splice site usage, followed by exon skipping and then alternative 5' splice site usage (Figure 9.4B).

The positive control, *mac7-1*, which has been verified as an intron retention mutant (Jia et al., 2017) surprisingly showed the lowest number of significantly different intron retention events amongst all the tested genotypes. However, the intron retention phenotype of *mac7-1* was demonstrated at a global level, which could be missed by subsetting based on gene by gene significant differences. For example, many genes could have a slightly higher intron retention level that is not sufficient to pass pairwise statistical tests, but that cumulatively result in a globally significant increase in intron retention. In order to check whether this was also the case in *msm1*, the number of reads mapping to introns were divided by the number of reads mapping to exons for *msm1* shoots and roots and their corresponding controls (Figure 9.4C). There was no clear significant difference between any of these values.

**A****B****C**

**Figure 9.4 Quantification of alternative splicing events in mRNA sequencing from *msm1* and controls**

**A:** The number of significantly different events in the corresponding genotype identified by SplAdder. Colours represent different genotypes.

**B:** Type and weighting of each alternative splicing event in the corresponding genotype. The colour represents the type of alternative splicing event.

**C:** Violin plots describing the intron/exon ratios in pSUC2:GFP:395 and *msm1* shoots and roots. Colour represents different condition.

Coverage Plots and counts supporting alternative splicing events

Coverage plots were produced for genes with a significantly different level of alternative splicing in *msm1* in order to visualise examples. On manual inspection of the most significant differential intron retention events in *msm1*, none stood out as dramatic. This was because the total gene count numbers were high for many of the most different genes, and the number of supporting reads for the differential alternative splicing events were proportionally extremely low. For example, the gene with the largest intron retention difference between *msm1* shoots and the control had a total count of over 2 million reads across three libraries, but only 325 counts supporting the intron retention event. The coverage plots also looked identical (data not shown). Many of the alternative splice site usage events were supported by much higher counts relative to the gene expression level in both *msm1* shoots and roots. Of these, AT1G67090 was significantly different in the shoot tissue and AT2G21330 was significantly different in the root tissue. Both of these fell in the category of alternative 3' splice site usage. In both cases, in *msm1* there appeared to be an extension in 3' region of the transcript, despite similar read numbers in both conditions. This was consistent with the SplAdder results, which defines alternative 3' splice site usage by inclusion of a new end-terminal region (Figures 4.5A and B).

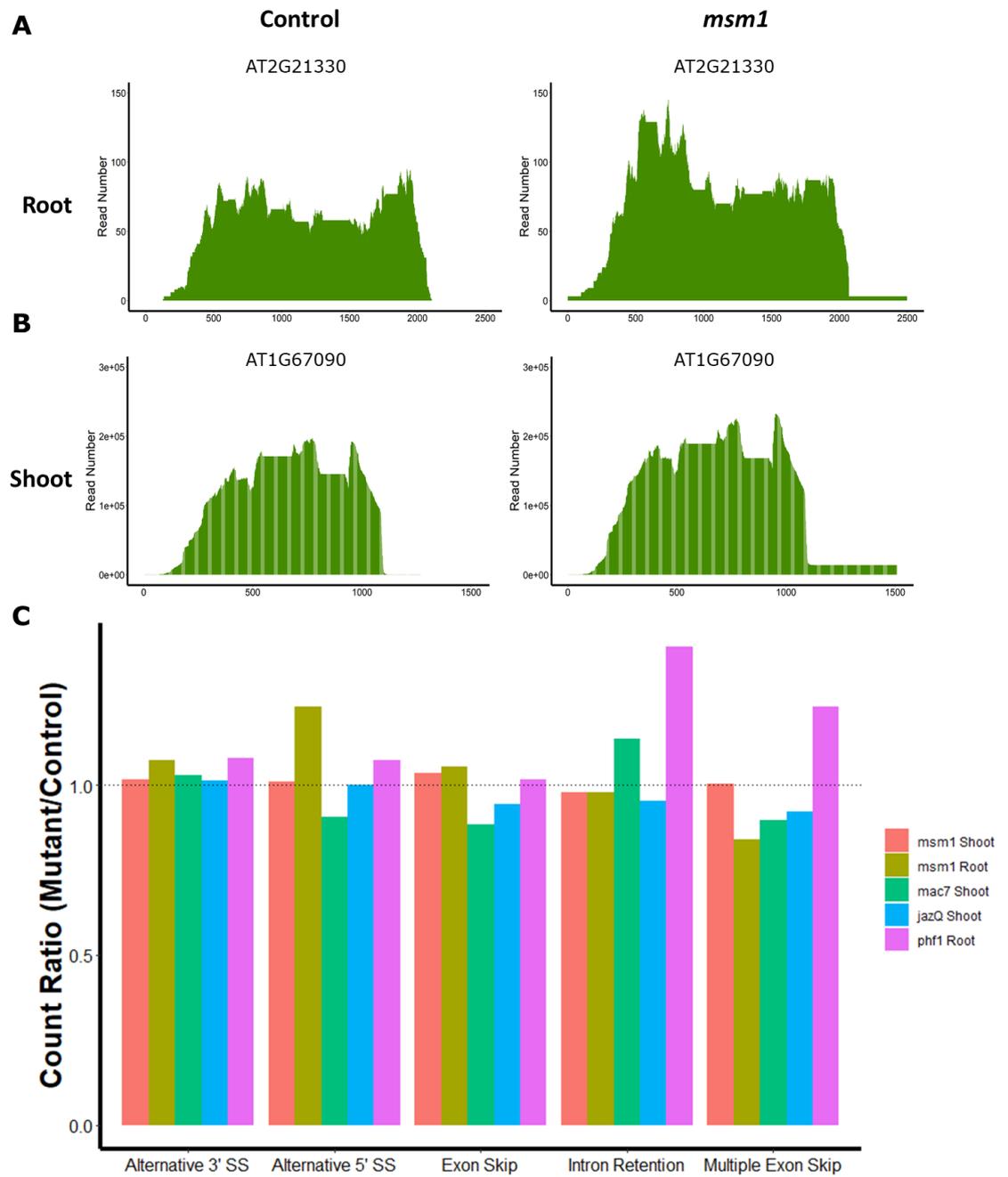
Finally, for each mutant, a count ratio was calculated for every type of alternative splicing event detected by SplAdder. This was calculated by taking the mean count of each event in the mutant, and dividing by the mean count of events in the control, regardless of whether or not they were significantly different. A ratio of 1 indicates no difference between the mutant and the control globally, less than one indicates a decrease in the number of supporting counts in the mutant, and greater than one indicates an increase in the number of counts in the mutant (Figure 9.5C).

Despite having a higher number of significantly different alternative splicing events, on a global level *msm1* shoots did not differ substantially from the control in any of the measured categories of alternative splicing. Intron retention was the most impacted category, consistent with the significantly different profiles in Figure 9.4B. The Count Ratios in *msm1* shoots were below 1, indicating that the levels of intron retention in *msm1* were lower than the control. This was mirrored in *msm1* roots as well, however *msm1* roots differed more substantially from the control. Despite alternative 3' splice site usage being the most common significantly different alternative splicing even in *msm1* roots, the global levels of alternative 5' splice site usage were considerably higher than any other category (Figure 9.5C). In total, *msm1* roots appeared to have the most altered total splicing profile of the two *msm1* conditions.

Aligning with the published literature, *mac7-1* showed a global increase in intron retention events (Figure 9.5C). Whilst *phf1* appeared to show considerably higher levels of alternative splicing relative to its control profile, this was to be expected on account of the phosphate starvation stress phenotype seen in this mutant (González et al., 2005), and the strong link between alternative splicing and abiotic stress response (Shang et al., 2017).

Interestingly, *phf1* demonstrated the greatest change in count ratio for intron retention and multiple exon skipping. This was unexpected as it had the lowest number of significantly different alternative splicing events (Figure 9.4A). This may be a result of the abiotic stress phenotype present in this background, as

intron retention has been shown to be highly stress sensitive (Filichkin et al., 2018).



**Figure 9.5 – Coverage plots of example alternatively spliced genes in *msm1* and count ratios of events**

**A:** Coverage plot of AT2G21330 in pSUC2:GFP:395 and *msm1* shoots. X axis represents distance along the gene in nucleotides.

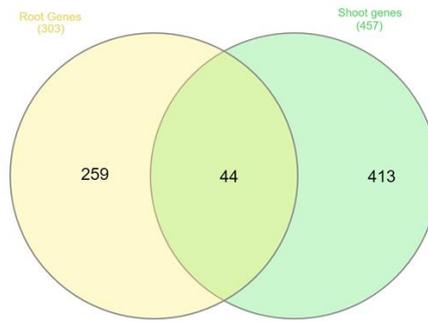
**B:** Coverage of AT1G67090 in pSUC2:GFP:395 and *msm1* roots. X axis as in (A)

**C:** Total count ratios (mutant / control) for each alternative splicing event type in the corresponding conditions. Dotted line represents a ratio of 1 which indicates no difference.

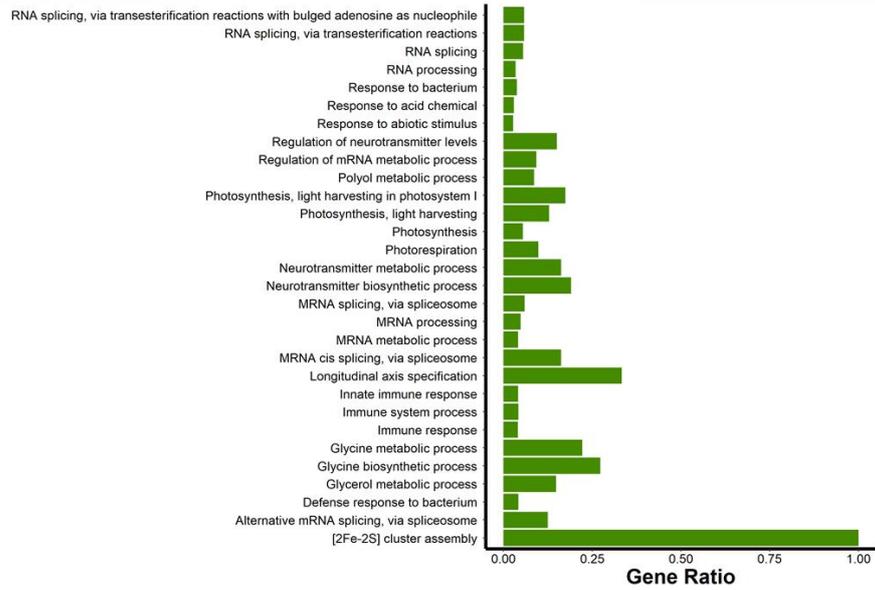
Gene ontology of alternatively spliced genes

ShinyGO was run as before on the genes, which were differentially alternative spliced in root, shoot and the 44 genes commonly alternatively spliced between both root and shoot (Figure 9.6A). Among the terms common to alternatively spliced genes in *msm1* roots and shoots were photosynthesis, RNA splicing, mRNA processing and response to abiotic stimulus. In *msm1* roots, the GO terms enriched in alternatively spliced genes pertained to developmental processes, circadian rhythm and various metabolic pathways such as hexose glucose metabolism (Figure 9.6C). In the shoots, in addition to abiotic stress response terms there were also a number of biotic stress related terms such as response to bacterium, innate immune response and immune system process (Figure 9.6B). GO enrichment analysis of the 44 commonly alternative spliced genes had only 6 terms significantly enriched. These were stomatal movement, response to CO<sub>2</sub>, response to bacterium, response to anaesthetic, regulation of stomatal movement and carbon utilisation (data not shown).

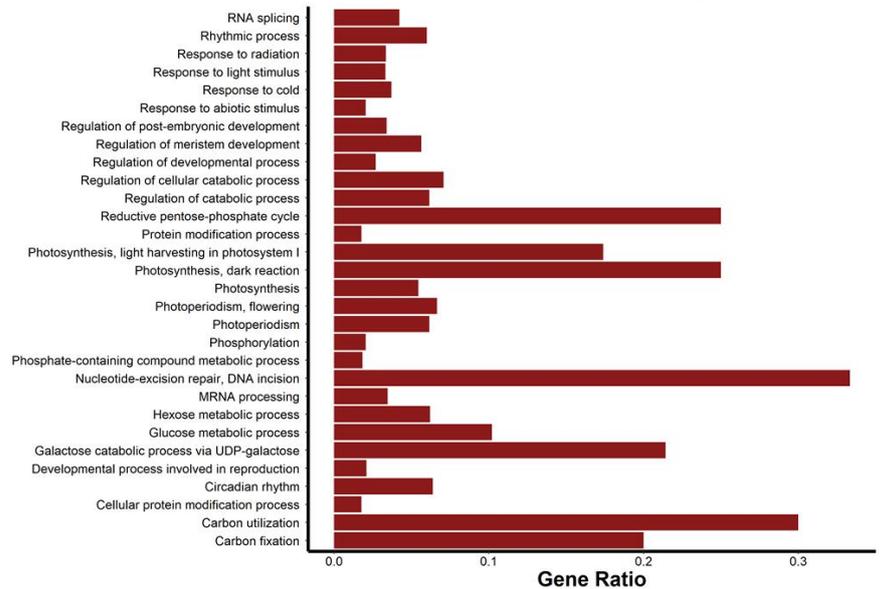
**A**



**B**



**C**



### **Figure 9.6 – GO terms enriched in differentially alternatively spliced genes in *msm1***

**A:** Overlap of significantly different alternative splicing events between *msm1* roots and *msm1* shoots.

**B:** GO term enrichment in genes with significantly different levels of alternative splicing in *msm1* shoots.

**C:** As in (B) but in *msm1* roots.

### Discussion:

#### *Msm1* has an altered transcriptome profile

The difference between the three genotypes tested for ABA sensitivity were clearly visible. Whilst *msm1* demonstrated an ABA sensitivity phenotype, it was not as strong as the phenotype seen in *abi-4* (Figure 9.1A). The reduced germination inhibition by ABA in *msm1* suggested that this genotype had reduced sensitivity to ABA. However, the further reduction in root length of *msm1* under higher concentrations of ABA indicate an increased sensitivity to ABA, as it has long been known that external application of ABA inhibits root growth (Sharp and LeNoble, 2002). Therefore, it is more accurate to say that *msm1* has an altered ABA sensitivity phenotype, rather than an increased or decreased sensitivity. This is not difficult to reconcile as ABA signalling is complex and integrates many different signalling pathways, and so the same mutation could affect these different pathways in myriad ways. It is often difficult to untangle ABA phenotypes from alternative splicing phenotypes, as they typically co-occur and are both capable of regulating each other (Z. Wang et al., 2015) (Cruz et al., 2014). However, as the change in ABA phenotype was not clearly unilateral, it could indicate that it was the result of altered regulation at various points, which I viewed as suggestive of an alternative splicing mutation.

The differential expression profiles of *msm1* were surprising, as they did not mirror the sRNA sequencing data. Despite the roots of *msm1* being the only tissue with a clearly altered expression profile in the sRNA sequencing data

(Chapter 4 Figure 8.3A), both roots and shoots had substantially different expression profiles in the mRNA sequencing data. Additionally, a greater number of transcripts appeared to be differentially expressed in the shoots of *msm1* relative to the roots, despite no clear accompanying morphological phenotype. The 170 commonly differentially expressed genes in *msm1* roots and shoots may represent a core set of genes which are directly impacted by the causative mutation(s) in *msm1*. This mutation may indeed be contained within this population of transcripts. The remaining transcripts which are not commonly differentially expressed between roots and shoots might therefore represent indirectly effected gene networks which were impacted by the perturbations of the core set of transcripts. If this is the case, then the GO biological term enrichment analysis conducted on the core 170 commonly differentially expressed transcripts would represent the most useful of the GO analyses, as it would consist of the processes that the gene with the causative mutation in *msm1* is directly involved in.

Amongst the most statistically significant, commonly enriched terms in *msm1* roots and shoots were terms pertaining to sulphur compound biosynthetic processes (Figure 9.3C). These terms only appeared in the top 30 most enriched terms for genes which were upregulated in *msm1* relative to the control (Figure 9.3A and B (top)). This is particularly surprising, as *msm1* was expected not to have a sulphur pathway phenotype based on previous experiments showing phosphate pathways were equally affected by the causative mutation (Chapter 4 Figure 8.2A). However, it should be noted that the `sulphur compound biosynthetic process` GO term does not encompass sulphate assimilation or reduction, for which there are distinct terms (Ashburner et al., 2000). Therefore, an upregulation of sulphur compound biosynthetic processes does not belie a sulphur starvation phenotype. Indeed, sulphur containing compounds have been linked to a wide variety of different physiological processes in plants such as biotic and abiotic stress response (Nwachukwu et al., 2012). This is particularly salient in this context as many other terms which were significantly enriched in both *msm1* shoots and roots pertained to abiotic stress response. Therefore, it is entirely plausible that the enrichment of sulphur compound biosynthetic processes was a result of increased expression in stress responsive pathways.

Many of the processes with enriched terms in the GO analysis could also support an alternative splicing phenotype in *msm1*. For example, in the core 170 genes commonly differentially expressed between *msm1* roots and shoots response to water deprivation and response to temperature stimulus were both significantly enriched, despite the plants being grown under normal conditions. Both temperature sensing and drought response have been shown to be regulated by alternative splicing, as described in Chapter 1. Mutations in splicing factors have also been shown to result in their dysregulation (Liu et al., 2013, Posé et al., 2013).

Despite these similarities, there were some notable differences in the biological processes, which were enriched in *msm1* shoots and roots. This was particularly striking in the terms enriched in genes which were downregulated in *msm1* roots (Figure 9.3B (bottom)), which largely pertained to cell wall synthesis. There was also an enrichment of terms relating to nucleotide synthesis in genes, which were upregulated in the roots, not seen in *msm1* shoots. Interestingly, neither of these sets of terms were present in the GO analysis conducted on the commonly differentially expressed 170 genes (Figure 9.3C). This could suggest that the root phenotype and the global enrichment of genes involved in stress response are attributable to different causative mutations.

### Splicing profiles in *msm1* roots and shoots

There was some disparity in the type of differential alternative splicing event in *msm1* shoots and roots. In roots, intron retention made up only approximately 1/6<sup>th</sup> of the differential alternative splicing events. However, in the shoots it was closer to 1/3<sup>rd</sup> (Figure 9.4B). In both *msm1* roots and shoots alternative 3' and 5' splice site usage were in the top 3 most frequent differential splicing events, and made up approximately half of the events. As discussed in the results, the number of reads supporting these alternative splice site events proportionally to the total number of reads mapping to the gene were much higher than in the intron retention events. The intron retention events were more numerous, but on average had a much lower supporting read count. Therefore, the alternative splice site usage differences are much more likely to be biologically significant

than the intron retention differences. The number of significantly different alternative splice site usages were roughly the same between both shoot and root.

Only 4 out of the 170 commonly differentially expressed genes exhibited differential alternative splicing. This could be indicative of a number of things. Firstly, the genes, which undergo differential expression in *msm1* may do so as a consequence of alternative splicing differences upstream. As the majority of alternative splicing products that are differentially expressed in *msm1* are the result of isoform switching, it is possible that this is caused by isoform competition similar to that seen in FLM, described in Chapter 1 (Posé et al., 2013). Secondly, the genes, which are differentially expressed could be the result of an alternative function of the mutant gene in *msm1*. Many genes involved in alternative splicing have been shown to interact with other protein complexes and exert biological activity independent of their capacity as a splicing component (Reddy et al., 2013). Thirdly, the differentially expressed genes may be the result on one mutant gene, and the altered splicing profile may be the result of another. This is probably the least likely scenario, as the GO profile showed an enrichment of many terms with a strong relation to alternative splicing.

As previously discussed in Chapter 1, one common mechanism of negative regulation by alternative splicing involves the introduction of premature stop codons. This results in nonsense mediated decay and therefore a reduction in the levels of transcript (Chaudhary et al., 2019). This is most common in intron retention events. However, most of the genes, which displayed a significant level of differential alternative splicing did not appear to be differentially expressed. Therefore, it is unlikely that this mechanism of alternative splicing regulation is perturbed in *msm1*.

The total splicing profiles shown in Figure 9.5C also represent an important metric by which alternative splicing levels can be assessed. As many of the read numbers supporting individual alternative splicing events can be low, identifying statistical significance is challenging. This could result in significantly different

alternative splicing events being missed as a result of low read depth. One way in which this can be circumvented is by taking the cumulative total of all the reads supporting each event for libraries with an equivalent read depth and comparing them, as was done in Figure 9.5C.

In *msm1* shoots, the total alternative splicing profile did not differ substantially from the control at a global level. This suggested the possibility that the transcripts, which underwent differential alternative splicing in *msm1* shoots did so as a result of contextual regulation, rather than due to a general alternative splicing phenotype. Interestingly, the total alternative splicing profile of *msm1* roots was considerably more dramatic than the shoots relative to the control. This disparity reflects the morphological phenotype of *msm1*. The roots of *msm1* are more likely to be under stress conditions than the shoot, as the extent of the difference from the control is greatest (Figure 9.2A). As stress and alternative splicing are linked, this is perhaps to be expected.

### Functional consequences of alternative splicing in *msm1*

The two genes in the coverage plots demonstrate typical alternative 3' splice site usage in *msm1* (Figure 9.5A and B). AT1G67090, which was alternatively spliced in the shoots, encodes a member of the Rubisco small subunit multigene family, which are necessary for photosynthesis (Rutschow et al., 2008). In the control condition AT1G67090.2 appears to be the main isoform of the gene expressed. This isoform is 1267bp long. In *msm1*, there appears to be a large increase in the number of copies of AT1G67090.1, which is a longer, 1510bp isoform. It is not clear what the functional significance is of this. In roots, AT2G21330 was used as an example. This gene encodes fructose-bisphosphate aldolase 1, and is involved in gluconeogenesis and chloroplastic glycolysis (Mininno et al., 2012). There are three known alternative splicing isoforms of this gene, and as in AT1G67090 a longer isoform which is not present in the control appears to be upregulated in *msm1*. The functional significance of this alternative splicing in plants has not yet been elucidated, however in yeast, alternative splicing of this transcript can result in a phenomenon termed spliceosome mediated decay (Volanakis et al., 2013). Interestingly, neither of

these genes were found to be significantly differentially expressed in *msm1*, despite having significantly different alternative splicing.

Both of these genes had implications for photosynthesis. Interestingly, GO terms relating to photosynthesis were commonly enriched in the alternatively spliced genes in *msm1* roots and *msm1* shoots (Figure 9.6B). There is a body of research which demonstrates a link between alternative splicing and response to light, which is chloroplast dependent and has implications for photosynthesis (Godoy Herz et al., 2019). Interestingly, the study by Godoy Herz et al. (2019) showed that in the dark, the rate of RNA pol II elongation was slower, and resulted in longer splice isoforms. This was consistent with what was seen with many of the alternative 3' splice site usage events seen in *msm1*, and could potentially suggest a shared mechanism.

Abiotic stress response terms were also commonly significantly enriched. This further reinforces the conclusion drawn from the differential expression analysis, that *msm1* has a stress response phenotype. As discussed in Chapter 1, alternative splicing is highly important in early plant development (Szakonyi and Duque, 2018). In *msm1* roots, many of the genes, which were alternatively spliced had a significant enrichment for terms related to the regulation of early development (Figure 9.6B). This, coupled with the morphological phenotype of *msm1* roots, could suggest that roots are at an earlier stage in development than their corresponding controls. In the 44 commonly alternatively spliced genes, both regulation of stomatal movement and stomatal movement terms were significantly enriched. This process is known to be regulated by ABA (Wang, 2014), and can therefore likely be explained by the ABA phenotype seen in *msm1* (Figure 9.1). Response to anaesthetic was another significantly enriched term in these genes. Plants have been documented to produce endogenous anaesthetic compounds to deal with stress (Yokawa et al., 2019), and so this fits in with the general stress response phenotype of *msm1*.

## Limitations

Establishing statistically significant differences in individual alternative splicing events proved challenging in this study. This was because the number of reads supporting each individual splicing event was relatively low, averaging at about 20 counts per event junction. One of the limiting factors of how useful an mRNA sequencing dataset is for a splicing analysis is the read depth. As the read depth increases, the number of reads supporting each individual splicing event would increase, therefore increasing the robustness of statistical comparison. Additionally, the number of splicing events detected would also increase, as many splicing events are likely missed due to insufficient reads to support them. This also has implications for situations where there is differential expression between two backgrounds. Whilst the use of PSI values should normalise to the level of expression, if a transcript has significant upregulation or downregulation, then at the same sequencing depth one would expect to detect a greater variation of alternative splicing events. This further reduces the confidence when testing low abundance alternative splicing events in transcripts which exhibit differential expression.

Unfortunately, there is bias present in the significantly different splicing profiles (Figure 9.4B). This was because count data from *msm1* was used to represent the proportion of the pie chart for significantly different events. Therefore, events, which increase in frequency will occupy a larger percentage of the chart than events, which decrease in frequency, and so the total splice profile is more reliable than the significantly different splice profile in representing differences between genotypes. The alternative was to use the number of significantly different events to populate the pie chart. However, when this was done, events which had many significantly different frequencies but low supporting count data dominated the chart – in *msm1* shoots this was multiple exon skipping. I felt that this was not the appropriate way of displaying the data as it inflated the significance of high frequency, low count results and also contained no information about the direction of change.

One further step that could have been taken would have been wet lab verification of the significantly different alternative splicing events. For genes such as AT1G67090, where the control expresses only the short isoform and *msm1* expresses both short and long isoforms (Figure 9.5A/B), primers could be designed to discriminate between the two splice isoforms. End point PCR could be used on cDNA libraries to discriminate between presence and absence of splice isoforms, and qPCR could be used to assay quantitative differences.

## Conclusions

Taken together, the data presented in this chapter demonstrated that *msm1* had a mild alternative splicing phenotype. This phenotype most commonly presented as an isoform switching phenotype, with a majority of differential alternative splicing events consisting of alternative 3' and 5' splice site usage. A number of splicing factors have been described which are involved in 3' splice site recognition. One example discussed in the introduction is the splicing factor DRT111, which also had an ABA phenotype (Punzo et al., 2020). However, this mutant displayed hypersensitivity to ABA during seed germination, and *msm1* shows reduced sensitivity (Figure 9.1A), and so it is unlikely to be a mutant in the same gene. In addition to the alternative splicing phenotype, *msm1* had a substantially altered transcriptome profile when compared to control seedlings, indicating a large number of biological processes were altered by the causative mutation(s). The next step is to identify the causative mutation in *msm1*, in order to determine if it is a novel mutant, or a mutant in a gene already described.

# Chapter 6: Identification of the causative SNP(s) in *msm1*

## Introduction

In the previous chapter, *msm1* was found to have a substantially different transcriptome profile to the corresponding control, as well as potentially a weak alternative splicing phenotype. It was unknown what the causative mutation(s) was that underpinned the observed phenotype.

As discussed in Chapter 3, *msm1* is the product of an EMS screen. EMS typically induces SNPs, which take the form of a base pair switch from GC to AT and AT to GC (Sega, 1984). This can result in an alteration of the amino acid encoded within the codon. As a consequence, this can lead to a missense mutation or premature stop codon, which can result in a loss or gain of function of a gene. Less commonly, EMS treatment can cause base pair insertions or deletions. The effects of these types of mutations, whilst less frequent, are likely to be more dramatic. This is because an insertion of base pairs of any number that is not a multiple of three will result in a frameshift mutation, thus disrupting the transcript to varying degrees, based on the position ("Frameshift Mutation - MeSH - NCBI," n.d.).

There are a number of ways in which an unknown mutation can be identified in *Arabidopsis*. Firstly, is map based cloning. This method typically involves an outcross to a different ecotype, isolation of phenotype segregating offspring and then chromosome walking to identify the causative locus (Jander et al., 2002). Whilst this technique has advantages, it requires a considerable amount of time to perform, which was in limited supply at this point in my PhD.

More modern techniques to accomplish this involve the use of NGS. This is typically done at the DNA level. As there is a high-quality *Arabidopsis* reference genome available, it is possible to sequence mutants and identify regions of their DNA that do not align with the reference. As NGS is also quantitative, this allows for conclusions to be drawn about zygosity of the mutation as well. This is usually performed on phenotypic segregants in the F2 stage following a backcross.

There are a range of bioinformatics tools available which can perform this type of mutation mapping. One example of such a tool is the Simple Mapping Pipeline (SIMPLE) tool (Wachsman et al., 2017). In addition to identifying non-reference reads in the mutant sequencing pool, it also identifies those which have either a heterozygous or homozygous ratio. In doing this, it prevents hundreds of results being returned detailing every single SNP present, and returns a much smaller list of potential candidates which have a mendelian ratio. This is useful as it reduces the benchwork component required to validate the causative mutation(s).

As these tools only take fastq files as input, they can be used both on sequenced DNA but also RNAseq data. This also has the added benefit of only testing the genes which are currently expressed in the corresponding tissue. This could help to exclude some candidates from further study as they may not be expressed in the plant at that point in development or under those conditions, and therefore is not likely to be responsible for the phenotype seen under normal conditions.

## Results

### Identifying putative causative mutations

The SIMPLE bioinformatic tool was run using the mRNA sequencing data from *msm1* and the corresponding control. The *msm1* seedlings were at the M3 generation in this data. This output a list of candidate mutations, which were predicted to result in a functional consequence, with the corresponding reads of reference sequence nucleotide and altered nucleotide for both the control and mutant conditions. The SNPs also had a frequency consistent with a heterozygous or homozygous trait. There were 20 total candidates, which are shown in supplemental table S7.

This table was further subset manually, to list only mutations that were homozygous mutant in *msm1* and homozygous WT in the control (Table 2). Four out of five of these potential causative mutations resulted in a missense variant and the other resulted in a premature stop codon. A literature search was then carried out into each of the candidate mutations to identify whether or not any of them best explained the phenotype seen in *msm1*.

A second table was compiled to show the Biological process GO terms ascribed to each of the genes in Table 2, taken from TAIR (Table 3). Based on the observed phenotypes in *msm1*, CPL3 and OPS were taken forward as the two most likely candidates.

**Table 2 – Homozygous mutations detected in *msm1***

Curated output from the SIMPLE tool. If the Name is the same as the Gene ID then no name has currently been ascribed. Numeric columns indicate the number of counts supporting the reference or alternate sequence.

Gene ID	Name	Mutant Reference	Mutant Alternative	Control Reference	Control Alternative	Consequence
AT1G74310	CLPB1	0	135	158	0	missense variant
AT2G33540	CPL3	0	45	53	0	missense variant
AT3G09070	OPS	0	43	134	0	stop gained
AT4G33990	EMB2758	0	15	17	0	missense variant
AT5G49960	AT5G49960	0	35	28	0	missense variant

**Table 3 – GO terms ascribed to homozygous genes**

Biological process GO terms taken from TAIR for each of the genes in Table 2.

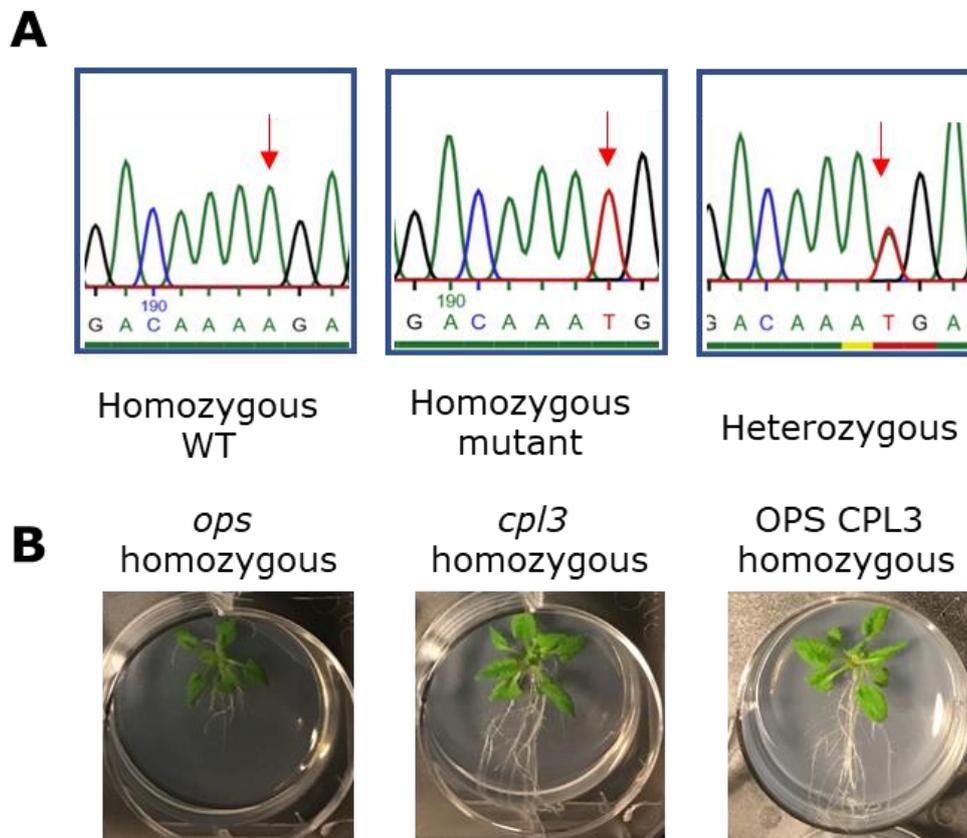
Gene ID	Gene	GO Biological Process terms
AT1G74310	CLPB1	cellular response to hypoxia, positive regulation of translation, protein metabolic process, protein unfolding, response to heat
AT2G33540	CPL3	dephosphorylation of RNA polymerase II C-terminal domain, negative regulation of RNA interference, negative regulation of abscisic acid-activated signaling pathway, response to salt stress
AT3G09070	OPS	cotyledon vascular tissue pattern formation, phloem development, phloem transport, regulation of root development, root system development
AT4G33990	EMB2758	RNA modification, embryo development ending in seed dormancy
AT5G49960	AT5G49960	None

## Backcross genotyping

*Msm1* M2 plants, and therefore M3 plants were homozygous for all five of the mutations shown in Table 2. In order to identify which mutations were responsible for which phenotypes, these various alleles needed to be segregated from one another, so that they could be investigated in a reductionist manner. *Msm1* M3 plants were backcrossed to pSUC2:GFP:395 plants and taken to the F2 stage. These were germinated on normal MS media and inspected for segregation. Out of 71 seedlings which germinated, 12 exhibited the short root phenotype. This equated to 16%, which was lower than expected for a homozygous recessive trait, but could have been an artifact of a low N number.

From the literature search conducted on the output from SIMPLE, OPS and CPL3 were considered to be the most likely causative mutant genes. This was because the GO terms ascribed to these genes most closely matched the observed phenotype in *msm1*. Backcross seedlings, which segregated for the short root phenotype were genotyped for the described mutations in OPS and CPL3 by PCR and sanger sequencing. The traces were used to determine if each individual was homozygous mutant, homozygous wild-type or heterozygous for the SNP, as shown in Figure 10.1A.

The short root phenotype co-segregated with *ops*, which confirmed that this mutation was the causative mutation or was closely linked to the short root phenotype. This aligned with the study by Truernit *et al.* (2012) which demonstrated that *ops* had significantly shorter roots than wild-type seedlings. There was no pattern of co-segregation of *cp13* and short roots, indicating that *cp13* was not related to the observed root phenotype (Figure 10.1B).



**Figure 10.1 – Genotyping *ops* and *cpl3* in *msm1* backcrosses**

**A:** Example of sanger sequencing trace for homozygous WT, homozygous mutant and heterozygous individuals. The bar underneath the letters represents the confidence of each call in the trace. Green is high confidence, yellow intermediate and red low. Red arrow indicates the SNP position.

**B:** Morphology of plants homozygous for *ops*, *cpl3* and WT OPS CPL3.

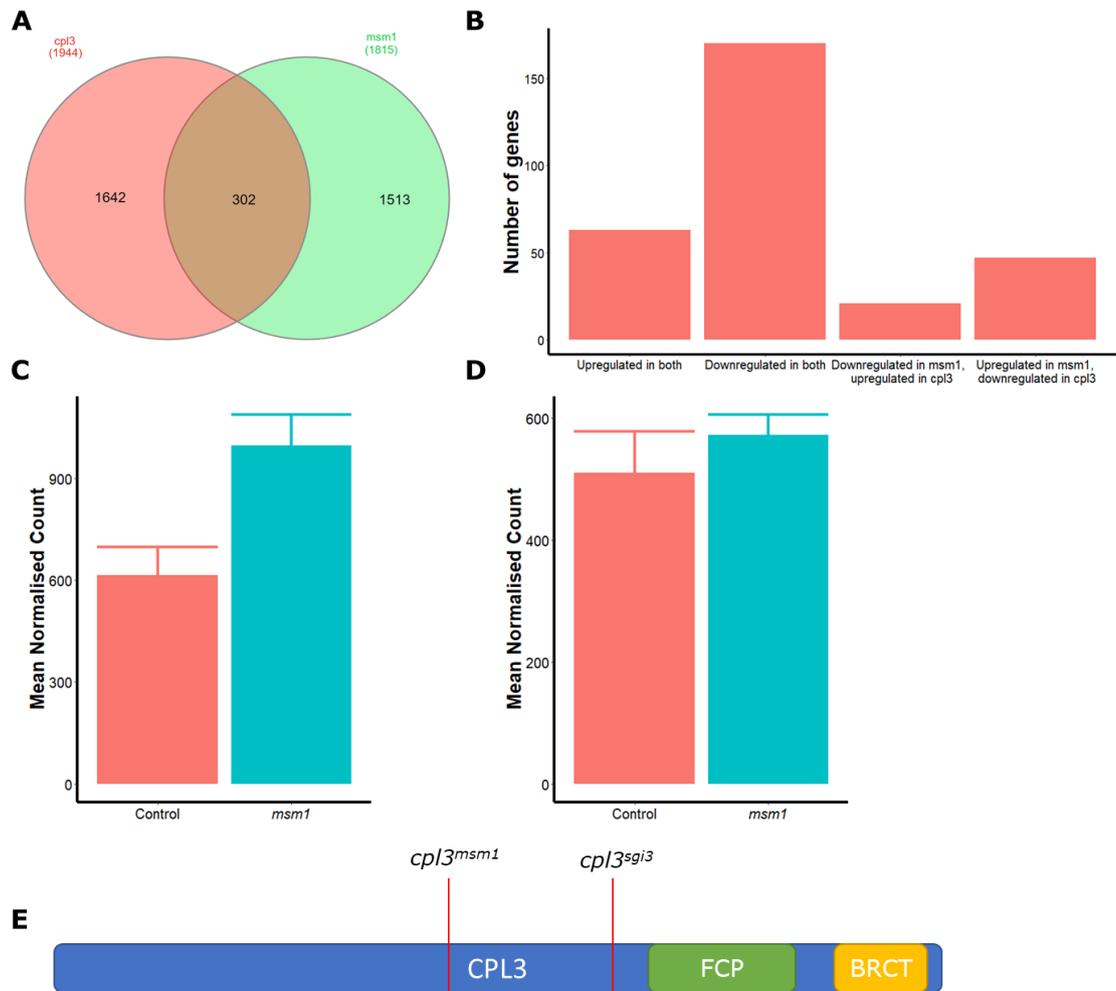
## Evidence for *cp13* mutation

Many of the GO biological process terms as well as published literature on CPL3 and CPL family members reflected aspects of the phenotype seen in *msm1*. As discussed in greater depth in Chapter 1, CPL3 is a regulator of response to ABA, salt stress, cold stress and hyperosmolarity (Koiwa et al., 2002). It is also a regulator of plant immune responsive genes (Li et al., 2014). All of these processes appear in the GO terms significantly enriched in the differentially expressed genes in *msm1*. There was also phenotypic evidence to suggest that *msm1* exhibited an altered ABA phenotype, shown in Chapter 5 Figure 9.1. Therefore, I suspected that *msm1* also contained a loss of function mutation of CPL3.

CPL3 was recently published on by Li et al. (2019), and in this study mRNA sequencing was performed on aerial tissue of *cp13* and differential expression was calculated. Differentially expressed genes were taken from the supplemental data of this study and were compared against differentially expressed genes in *msm1* shoot tissue. A similar number of genes were differentially expressed in both backgrounds. However, only 302 genes were commonly differentially expressed between *msm1* and *cp13* (Figure 10.2A). The number of differentially expressed genes in *msm1* shoots in 5.2A is lower than in Figure 9.2A, as splice isoforms have been removed.

The direction of change for most of these transcripts was also the same (Figure 10.2B). Among these transcripts is RD29, which has been previously shown to be hyperinduced on knockout of CPL3 (Koiwa et al., 2002). This also occurs in *msm1* (Figure 10.2C).

CPL3 levels in the RNAseq were slightly higher in *msm1* than in the control (Figure 10.2D). This suggested a possible compensatory response for reduced function. The SNP in *cp13<sup>msm1</sup>* was located at nucleotide 2555 in the CDS, falling within exon 6 (Figure 10.2E). This is upstream of the causative SNP from the study carried out on *cp13*.



**Figure 10.2 – *msm1* and *cpl3<sup>sgi3</sup>* have common differentially expressed genes including *cpl3* induced RD29**

**A:** Overlap of differentially expressed genes in published *cpl3<sup>sgi3</sup>* mutant and *msm1*.

**B:** Direction of change of commonly differentially expressed genes between *cpl3<sup>sgi3</sup>* and *msm1*.

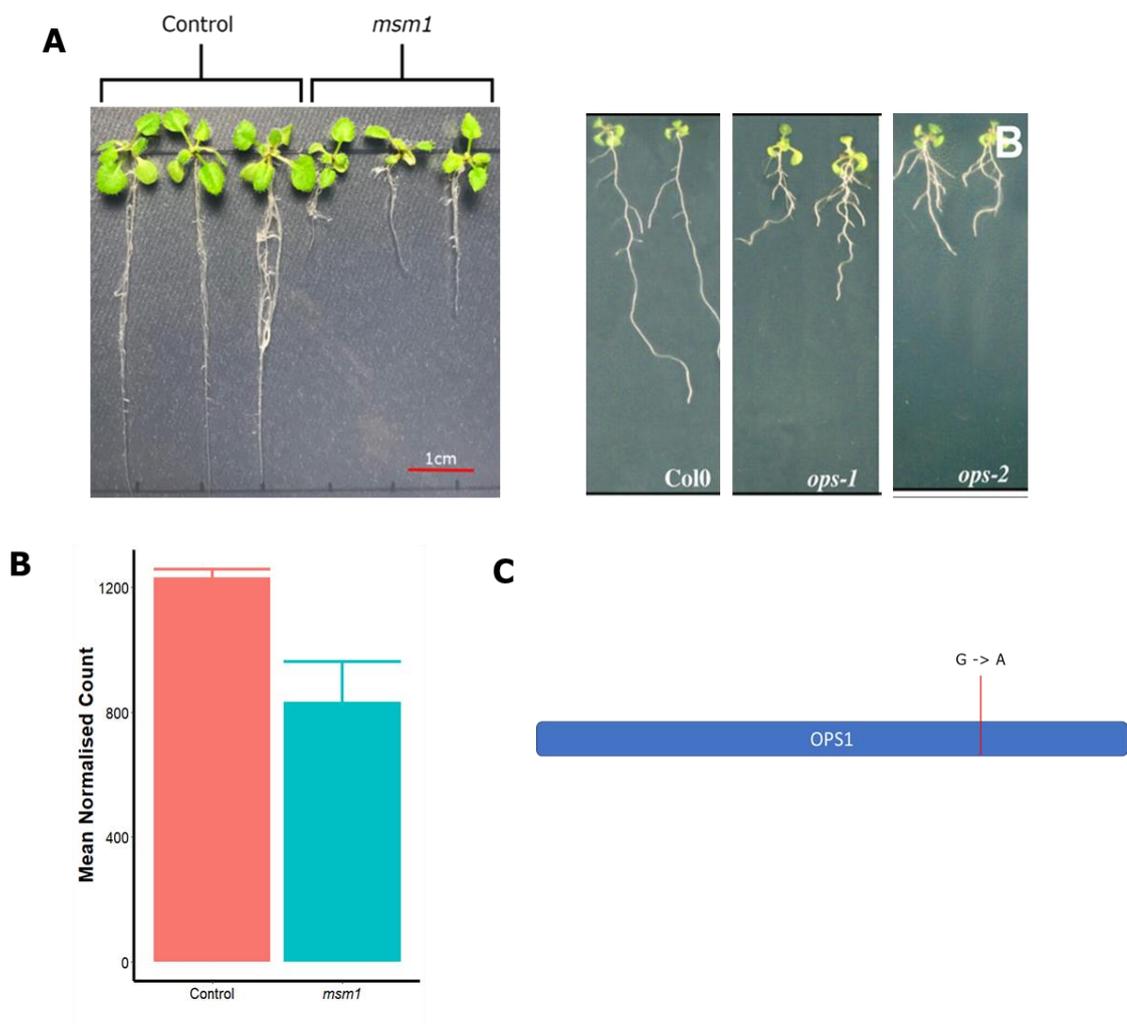
**C:** Mean normalised count of RD29 in pSUC2:GFP:395 and *msm1* seedlings taken from RNAseq data.

**D:** Mean normalised count of CPL3 in control and *msm1* seedlings taken from RNAseq data.

**E:** Graphical representation of CPL3 coding sequence. FCP and BRCT functional domains are delineated with green and yellow segments respectively. Red line with accompanying annotation shows the position of the SNP in the corresponding mutant background.

## Evidence of OPS mutation

The roots of *msm1* resemble those of published OPS mutants (Figure 10.3A) (Truernit et al., 2012). The roots were not only shorter than the control, but they also exhibited a greater degree of branching. In addition, the GO biological process term enrichment of *msm1* roots in Chapter 5 Figure 9.3 reflect the published literature on this. The *ops* in *msm1* has a premature stop codon, which typically results in degradation of the transcript by NMD (Hug et al., 2016). OPS levels were lower in *msm1* than in the control line (Figure 10.3B), suggesting that NMD may be working to reduce the levels. Finally, Figure 10.3C demonstrates the location along the transcript of the SNP in *msm1*.



**Figure 10.3 – Comparison of OPS phenotype and transcript levels in *msm1* compared to the control.**

**A:** OPS mutants taken from Truernit et al., 2012 and *msm1*

**B:** OPS levels in pSUC2:GFP:395 and *msm1* seedlings in mRNA sequencing data.

**C:** Gene model of OPS. Light blue represents UTRs, dark blue represents the single exon, and the red line and accompanying text represents the SNP present in *msm1*.

## Discussion

### CPL3 and OPS are the most likely causative mutations

The CPL3 SNP in *msm1* is a T to A change in exon 6, which results in an amino acid change from lysine to methionine. Lysine contains a positive charge in one of its side groups, and often plays a key role in protein structure and enzymatic activity (Betts and Russell, 2003). This SNP does not fall directly within either the FCP or the BRCT domains previously described as essential for function (Bang et al., 2006), however it could impact the folding in such a way that these domains are compromised. In the recent study by Li et al. ), *cp/3* was also isolated through an EMS screen. In this study, the causative SNP was also upstream of the FCP and BRCT domains (Li et al., 2019).

This slight increase in *msm1* may be a compensatory response for the reduction in function in *msm1*. The presence of a stress response and ABA phenotype consistent with CPL3 knockout, coupled with the observation that the difference in levels of CPL3 between *msm1* and the control is not substantial, implies that the missense mutation in *msm1 cp/3* likely has altered function.

The differential expression overlaps between *cp/3* and *msm1* was lower than might be expected. This could be for two reasons. Firstly, the CPL3 mutation in

*msm1* is in a different location to the mutation in *cp13*. This could result in a weaker phenotype in *msm1*, or compromise some functionalities of CPL3 but not others, as CPL3 is involved in a number of different processes. Secondly, in the study by Li et al. (2019), a number of genes which are dysregulated as a result of *cp13* are altered because of an RDR6 dependent interaction. The background used for *msm1* is *rdr6-15*, so genes affected by this interaction in *cp13* might not be affected in *msm1*, further ameliorating the phenotype.

CLPB1 was an interesting candidate that was identified by the SIMPLE tool. One of its associated GO terms – response to heat – loosely aligned with the GO terms enriched in the differentially expressed genes of *msm1*. However, this mutant was excluded, as all described mutants in this gene exhibited no phenotype under normal conditions. They only exhibited a phenotype under heat stress conditions.

AT5G49960 had no ascribed GO terms on TAIR. However, it appeared to encode a chloroplastic ion channel. It seemed unlikely that a mutation in this type of gene would be sufficient to cause the phenotypes observed, and so it was excluded from further analysis, until such a time that the other, more promising candidates had been excluded.

EMB2758 also appeared to be homozygous for a SNP in *msm1*. The read numbers supporting this were the lowest of any of the other candidate mutations (Table 2), suggesting that this transcript was likely expressed at a much lower level. This gene is part of a much larger family of Embryo-Defective (EMB) genes, generally defined by an abnormal embryo phenotype (Meinke, 2020). The GO term ascribed to this gene pertaining to embryo development is `embryo development ending in seed dormancy`. However, the seed dormancy phenotype shown in Chapter 5 Figure 9.1A is clearly ABA dependent, as the untreated *msm1* seedlings have the same germination rate as controls. As no embryo defective phenotype was observed in *msm1* under normal conditions, this mutant was also excluded from further analysis as it did not seem to have an altered function.

It is interesting that *ops* was pulled out from this genetic screen, particularly as an early response mutant. Many of the dysregulated genes in *msm1* are likely the result of *ops*, as this is most likely the most dramatic mutation in *msm1*.

SUC2 levels are lower in *msm1*, which may be the result of *ops*. The SUC2 promoter was used to express the GFP transgene, and so it would be expected that this would therefore translate to a reduction in the intensity of GFP fluorescence. This was not the case for *msm1*, as it was isolated as an early response mutant, meaning it recovered GFP faster than control plants. It is possible that the reduction in miR395 levels was sufficient to counterbalance this, however the reduction in miR395 in the next generation seemed mild (Figure 8.2A). It is also possible that the reduction in SUC2 levels is a result of the transcriptional silencing discussed in Chapter 3. As the SUC2 promoter was used to drive the expression of the GFP transgene, which did exhibit transcriptional silencing in the M3 generation, it is possible the silencing spread to the promoter. As this is a small RNA guided phenomenon (Kim and Rossi, 2009), it would not be possible for the plant to discriminate between the transgene SUC2 promoter and the endogenous SUC2 promoter.

OPS was shown by Truernit et al. (2012) to be a regulator of entry into phloem differentiation in *Arabidopsis*. *msm1* is most likely to contain an OPS loss of function, due to severity of the mutation and similarity to published loss of function. The roots of *msm1* are therefore more likely to be at an earlier stage of differentiation than their control counterpart. This likely explains the disparity between the miRNA profiles of *msm1* roots and control roots, shown in Chapter 4 (Figure 8.3A & B), as different miRNAs are expressed at different stages of development in *Arabidopsis* (de Lima et al., 2012).

## Limitations

The SIMPLE tool was designed for use with genomic DNA sequencing. Whilst it can in principle be used for mRNA sequencing as was performed here, there are some limitations to doing this. For example, only mutations in transcribed sequences will be picked up. Mutations which prevent the expression of a gene in *msm1* would therefore not be identified through this method. This could be caused by mutations in promoter regions which could affect the level of transcriptional initiation of a gene. Therefore, to be fully certain of a causative mutation, the highest confidence approach should be to subsequently complement the mutant with a wild-type copy of the gene and see if it rescues

the phenotype. Unfortunately, this was not performed in this PhD due to time limitations.

## Conclusion

In *msm1*, there appear to be consequential mutations in both OPS and CPL3. Whilst the OPS mutation has a clear morphological phenotype, CPL3 appears to have a subtler, molecular phenotype. Therefore, it is likely that *msm1* is a loss of function of OPS and a reduction in function of CPL3. As both of these mutants have previously been described, the next stages would be to try and identify novel aspects of the phenotypes caused by the mutant alleles generated here. Additionally, segregants with each individual mutation need to be taken to the next generation, so that the SNPs can be studied in a reductionist manner.

## Chapter 7: Investigation of *cp13* through *msm1*

### Introduction:

In Chapter 6, *msm1* was shown to have a molecular phenotype likely caused by a SNP in CPL3. In addition to gene expression changes, CPL3 has recently been shown to be involved in S-PTGS and the production of RNA-quality control siRNAs (rqc-siRNAs) (Li et al., 2019). In the study by Li et al. (2019), mutations to *cp13* resulted in the accumulation of a large number of siRNAs, which appeared to derive from endogenous transcripts. Additionally, it resulted in an increase in transgene silencing. The authors of this study also demonstrated that this phenomenon was RDR6 dependent, however no sequencing experiments were performed on their *rdr6 cp13* line. Therefore, if *msm1* contains a *cp13* allele, then it represents an opportunity to examine the full consequences of *rdr6 cp13* on the transcriptome and on the sRNAome, as *msm1* is in an *rdr6-15* background. This would be novel and the two sequencing datasets have already been generated in earlier steps of this investigation.

In order to verify this sRNA phenotype, sRNA sequencing data would need to be mapped to the Arabidopsis transcriptome. The RDR6-dependent 21mer rqc-siRNA population described by Li et al is not expected to be present, and therefore *msm1* would be expected to have a sRNA profile more similar to WT in this respect. However, this exercise may also reveal populations of RDR6-independent sRNAs, which accumulate as a result of *cp13* that have not yet been reported on. In the study by Li et al. (2019), RDR6 was implicated in the biogenesis of these rqc-siRNAs by qPCR performed on transcripts which were found to exhibit a canonical sRNA:target relationship in *cp13*, which was abrogated in *rdr6 cp13*. However, this provides no information on other transcripts which may produce sRNAs and be regulated by them in *rdr6 cp13* relative to *cp13*. In the sRNA and mRNA sequencing datasets generated from *msm1*, these populations may be searched for and investigated.

In order for siRNAs to be produced from an RNA, that RNA must be double-stranded. This requirement is necessitated by the substrate-preferences of the DCLs. There are a number of routes through which double-strandedness can be achieved. Firstly, and perhaps most commonly, is through the activity of RDRs,

such as RDR6 and RDR2. These polymerases bind to a transcript, usually with the help of another protein such as SGS3. They then use the transcript as a template and produce a second, complementary strand (Curaba and Chen, 2008). In this scenario, sRNAs are generally produced from the whole length of the transcript. A second mechanism through which a double-stranded region of RNA can be produced is by the secondary structure, which is a result of folding of the primary sequence (Wu and Tinoco, 1998). In this situation, foldback structures can be formed by regions of base pair complementarity within a transcript; this is how miRNAs are produced (Wang et al., 2019b). In this case, the sRNAs are not likely to map to the entire transcript. Rather, they will map to discrete locations where the double stranded region is formed, as is seen when miRNAs are mapped back to their pri-miRNA transcripts.

In addition to mapping distribution, the likeliness of sRNAs which are attributable to secondary structure can be inferred from the Minimum Free Energy (MFE) of the transcript they derive from. This is ultimately a measure of how likely a structure is to form, with a lower value meaning the more likely. This measure takes into account the number of nucleotides, the composition of nucleotides and the arrangement of nucleotides (Clote et al., 2005). In the case of *msm1*, it would be expected that there would be more sRNAs, which formed as a result of secondary structure folding than RDR activity, as RDR6 is compromised. This would translate to a lower overall average MFE in transcript derived sRNAs. A variety of bioinformatic tools are readily available to study the MFE energy of transcripts. One such bioinformatic tools is RNAfold, in the ViennaRNA package (Lorenz et al., 2011).

## Results

### *msm1* has populations of transcripts which produce a significantly different number of siRNAs

Firstly, sRNA sequencing data produced from *msm1* and the corresponding control in Chapter 4 were used to identify any genes which may have a significantly different number of sRNAs mapping to them in *msm1*. This analysis was carried out in a conceptually identical way to the analysis performed by Li et al. (2019) The sRNA sequencing data from the study carried out by Li et al.

(2019) were also analysed in this manner. In the study performed by Li et al. (2019), an EMS screen was performed and a number of mutants were isolated. These were originally named *Suppressor of Growth Inhibition 1 to 7 (sgi)*, however when the causative mutants were mapped they were renamed such that the mutated gene was given a superscript annotation referring to the original *sgi* name. Therefore, to reference the *cp13* allele identified by Li et al. (2019), the name *cp13<sup>sgi3</sup>* will be used.

Root sRNA data from *msm1* was compared against *cp13<sup>sgi3</sup>*. The tissues were different, which applies some restrictions to direct comparisons. However, this decision was made because the PCA showed better clustering of *msm1* root sRNA data (Chapter 4 Figure 8.3C) which made statistical analyses easier and more reliable. It was not possible to draw conclusions with *msm1* shoot data due to the quality and spread of it. 1128 transcripts in *msm1* had a significantly different number of sRNAs mapping to them, compared to 580 in *cp13<sup>sgi3</sup>* (Figure 11.11A). Of these, only 20 overlapped.

The low degree of overlap of transcripts with a significantly different number of sRNAs mapping to them between the two conditions could have been due to tissue specific expression patterns. For example, the 20 commonly significantly different transcripts may be the only 20 which are commonly expressed between *cp13<sup>sgi3</sup>* shoots and *msm1* roots. Therefore, to test this, a list of expressed transcripts was generated from *msm1* root mRNA sequencing data which had a count greater than 20, a figure selected after manual examination of the count data. This was then checked for overlap with the transcripts which had a significantly different number of sRNAs mapping to them in *cp13<sup>sgi3</sup>* (Figure 11.1B). 439 transcripts were commonly expressed in both *msm1* roots and *cp13<sup>sgi3</sup>* shoots. Therefore, the 20 overlapping transcripts with a significantly different number of sRNAs mapping to them in *cp13<sup>sgi3</sup>* and *msm1* roots (Figure 11.1A) were not attributable only to coincidence of expression.

The transcripts in Figure 11.1A have a significantly different number of sRNAs mapping to them than their corresponding controls. However, this takes into account a heterogeneous pool of sRNAs ranging from 20 to 25 nucleotides in length. As discussed in Chapter 1, the difference in length in these various sub-populations of sRNA belie different origins and modes of action. Therefore, to

study the data in greater resolution, the sRNAs were segregated by size and re-mapped to the Arabidopsis transcriptome.

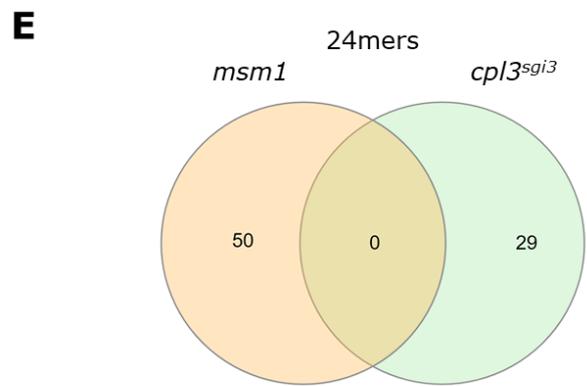
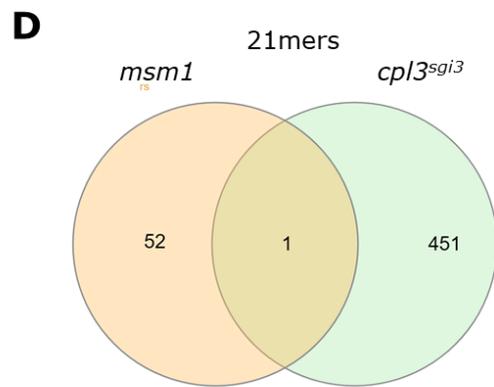
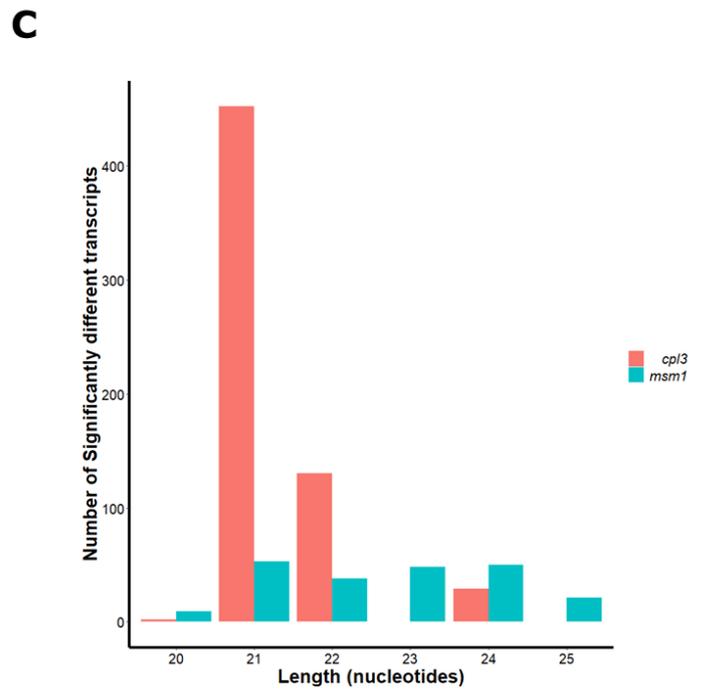
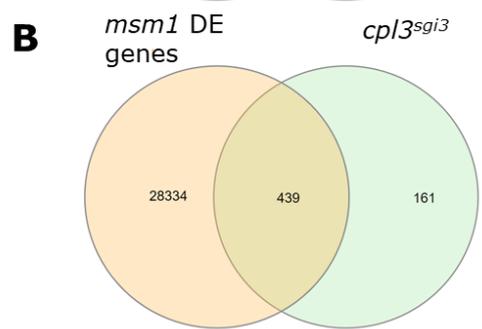
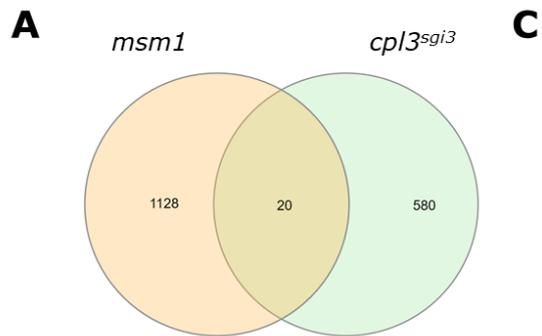
The number of significantly different events were quantified and plotted (Figure 11.1C) for each size segregated pool of sRNAs. The results for *cp13<sup>sg13</sup>* matched what was published by Li *et al.* (2019), with 21nt species dominating as the most numerous significantly different sRNA class. There were also notable peaks at 22nt and 24nt as was also reported (Li *et al.*, 2019). Interestingly, in *msm1* there appeared to be a similar number of transcripts with a significantly different number of sRNAs mapping to them in each size cluster between 21 and 24 nucleotides. This pattern could be consistent with non-DICER degradation products; this possibility is addressed in the next section. The number of transcripts with a significantly different number of 24nt sRNAs mapping to them was roughly the same for *msm1* and *cp13<sup>sg13</sup>*. The number of transcripts producing a significantly higher number of 21mers in *msm1* was roughly 10 times lower than in *cp13<sup>sg13</sup>*, which further supports the conclusions of Li *et al.* (2019) that the production of 21mers in *cp13* is RDR6 dependent (Figure 11.1C). For almost all cases, the number of sRNAs seemingly originating from endogenous transcripts were greater in the mutant background than the control, with only one or two transcripts producing fewer in each size category.

21mers and 24mers were of particular interest owing to the large body of literature demonstrating their capacity for silencing. The transcripts producing a significantly different number of 21mers in *msm1* and *cp13<sup>sg13</sup>* were compared for overlap, in order to determine if the same transcripts were producing a different number in both (Figure 11.1D). This did not appear to be the case, as only 1 transcript was common to both. The 24mer producing transcripts were similarly compared, and there were no overlaps between the two populations (Figure 11.1E).

Consistent with the findings of Li *et al.* (2019) in *cp13<sup>sg13</sup>*, the transcript derived sRNAs in *msm1* did not typically exhibit a canonical sRNA:target relationship with their corresponding transcripts. In *msm1*, in the transcripts that produced more 21mers, only 4 of them displayed differential expression. In all four cases however the level of expression of the transcript was higher (Figure 11.1F), which is not what would be expected if the 21mers were enacting silencing. In the transcripts which produced more 24mers, there were also 4 which displayed

differential expression. Again, the levels of expression are higher (Figure 11.1G).

As described by Li et al. (2019), 56 transcripts in *cp13<sup>sg13</sup>* displayed a canonical relationship with their progenitor transcripts. This was not depicted here as it was well described in Li et al. (2019). The 24mer producing transcripts were not specifically analysed for sRNA:target relationships, and so these data are shown here. Of the transcripts which produce more 24mers, 5 are differentially expressed and two in a canonical fashion. 24mers are still associated with repression of transcript levels, however they operate via transcriptional silencing (Blevins et al., n.d.). It should be observed that there were no transcripts in *cp13<sup>sg13</sup>* which exclusively produced more 24mers, and so these data also apply to the 21mers (Figure 11.1H).





**Figure 11.1 – Investigating transcript derived sRNAs in *msm1* and *cp13<sup>sgl3</sup>* and their relationship with their transcripts.**

**A:** Venn diagram representing the overlap between transcripts with a significantly different number of sRNAs mapping in *msm1* and *cp13<sup>sgl3</sup>*.

**B:** Venn diagram representing the overlap between transcripts which are expressed in *msm1* and transcripts with a significantly different number of sRNAs mapping in *cp13<sup>sgl3</sup>*.

**C:** Number of transcripts with a significantly different number of the corresponding size class of sRNA.

**D:** Overlap between transcripts with a significantly different number of 21mer sRNAs mapping to them between *msm1* and *cp13<sup>sgl3</sup>*.

**E:** As in **D** but for transcripts with a significantly different number of 24mer sRNAs mapping to them.

**F:** Relationship between log<sub>2</sub> fold change of 21mer sRNAs and mRNA in *msm1*. Each annotated segment represents a transcript which had a significantly different number of 21mer sRNAs mapping to it in *msm1* relative to the control. The inner ring represents the log<sub>2</sub> fold change in the sRNA sequencing and the outer ring represents the log<sub>2</sub> fold change in the mRNA sequencing, both relative to the control. Red indicates positive change.

**G:** As in **F** but with transcripts producing a significantly different number of 24mer sRNAs mapping to the corresponding transcripts. The background is still *msm1*.

Many siRNAs produced from endogenous transcripts in *msm1* are unlikely to be degradation debris

As *msm1* accumulated significantly different numbers of all size clusters of sRNA, rather than just 21, 22 and 24 as in *cp13<sup>sgl3</sup>*, it was possible that these samples displayed a greater amount of degradation. This would result in an apparent increase of all size clusters of sRNA, however they would be the product of degradation rather than of controlled sRNA biogenesis. Therefore, in this case it would be expected that the same transcripts would be producing a significantly different number of multiple size clusters. A six-way Venn diagram

was constructed to assess the degree of overlap between transcripts producing significantly different sRNAs of each size cluster.

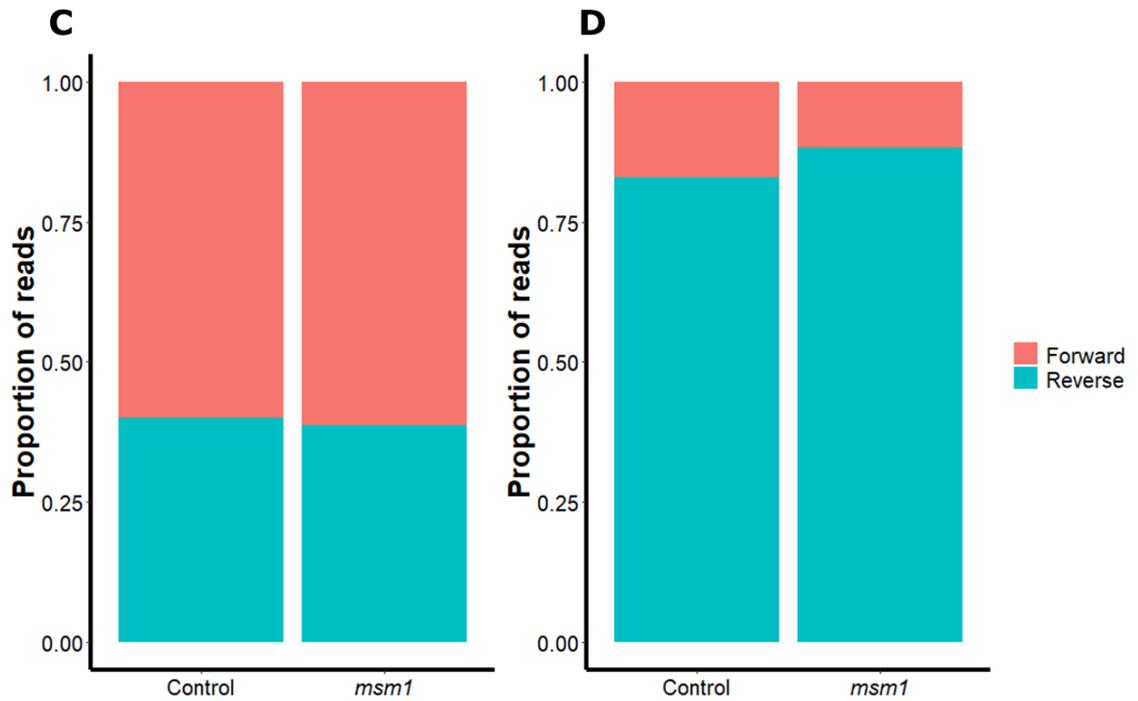
For 21mers and 24mers, which represent the most abundant species of sRNAs, there was the least overlap, with roughly half of the transcripts only producing a significantly different number of just those size clusters in *msm1* (Figure 11.2A). The 23mers, which were significantly differentially expressed for 48 different transcripts and displayed amongst the highest average count difference, a majority of these transcripts overlapped with other size producing transcripts (Figure 11.2A). This implied that these may potentially be the product of an increase in degradation of the original transcript, however it should be observed that there is no significant change in the levels of the mRNA transcripts. This data did contrast with the data produced by Li et al. (2019), as in their study there was a high degree of overlap between the transcripts producing 21mers, 22mers and 24mers (Figure 11.2B).

In both *cp13<sup>sgi3</sup>* and *msm1* there are transcripts which produce more sRNAs of size 21 and 24 nucleotides than their corresponding controls. However, in both of the backgrounds this was only a small sub-population of transcripts expressed, and there was virtually no overlap in the transcripts producing more sRNAs between the two backgrounds (Figure 11.1A). It was shown in both this study (Figures 6.1F, G & 1H) and by Li et al. (2019), that this is not simply due to an increase in expression of the transcripts, as many of the transcripts producing more sRNAs did not show a change in expression from the control. Additionally, in mRNA sequencing data generated from *msm1* and *cp13<sup>sgi3</sup>*, many transcripts which do accumulate at higher levels in the mutant do not appear to generate more sRNAs relative to the control. It is therefore not yet clear what determines which transcripts produce more sRNAs in *cp13<sup>sgi3</sup>* and *msm1* than in the control.

Finally, one clear indicator of whether or not sRNAs derive from RDR products is whether or not they map to both the sense strand of the transcript and also the antisense strand (Martínez de Alba et al., 2015). Conversely, random degradation products would be expected to map only to the sense strand, as there would not exist an antisense strand. sRNAs from transcripts which produced a significantly different number of 21mers (Figure 11.2C) and 24mers (Figure 11.2D) were analysed for their strand distribution. For each of these

categories the data were pooled. The sRNAs mapping to transcripts which produce a significantly different number of only 21mers demonstrated a roughly even split between sRNAs mapping to the sense strand and sRNAs mapping to the antisense strand. There were approximately 10% more reads mapping to the sense strand (Figure 11.2C), which could be indicative of some level of degradation. There was also no clear difference in the ratio between *msm1* and the control background. Interestingly, there appeared to be a strong negative strand bias for sRNAs deriving from transcripts which produced a significantly different number of 24mers (Figure 11.2D). This was even more pronounced in *msm1*. Despite this, between 10 and 25% of the reads mapped to the antisense strand of the transcripts. For both the 21mer and 24mer producing transcripts individual genes were manually tested and inspected to ensure that each gene produced a combination of both sense and anti-sense mappers.





**Figure 11.2 – Venn diagrams of transcripts producing single size clusters and the mapping of these sRNAs**

**A:** 6-way venn diagram constructed of transcript lists which have a significantly different number of 20mer, 21mer, 22mer, 23mer, 24mer and 25mers mapping. Red arrows indicate transcripts which have either only a significantly different number of 21mers or 24mers mapping in *msm1*.

**B:** 4-way venn diagram as in A, constructed using transcript lists from *cp13<sup>sgi3</sup>*. 25mers were excluded as there were no significantly different 25mer producing transcripts in *cp13<sup>sgi3</sup>*.

**C:** Strand distribution of sRNAs deriving from transcripts producing a significantly different number of 21mer sRNAs.

**D:** Strand distribution of sRNAs deriving from transcripts producing a significantly different number of 24mer sRNAs.

Transcripts producing a greater number of sRNAs in *msm1* are stress related

In order to determine if the transcripts which produce more sRNAs in *cp13<sup>sgi3</sup>* and *msm1* were random or not, they were first subjected to a GO analysis using

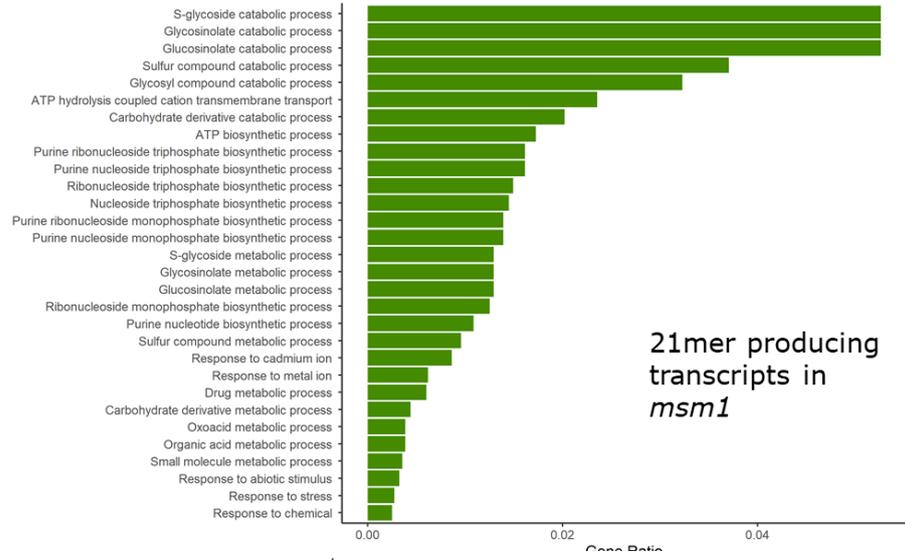
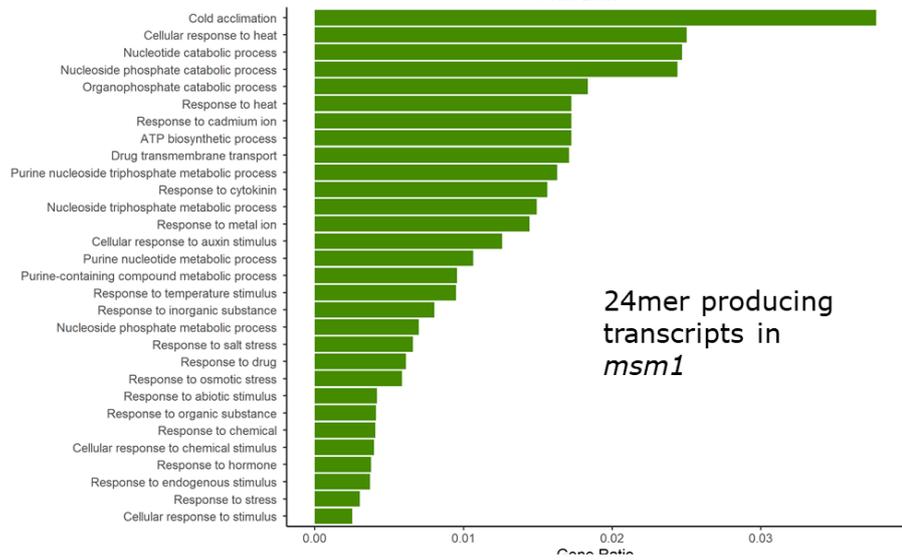
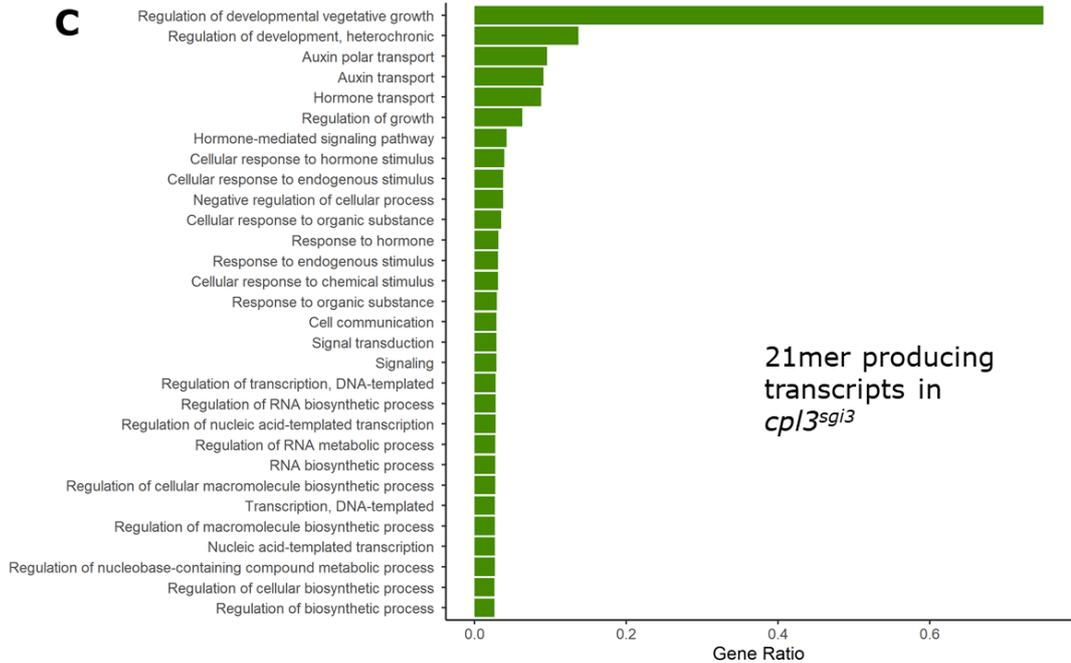
ShinyGO, as in Chapter 5. The reasoning behind this is that transcripts which are related to each other in function may also be structurally related, as gene duplication is a common phenomenon which facilitates functional diversification (Ha et al., 2009) – CPL3 is a good example of this. For this analysis, only the transcripts which produced more sRNAs than the corresponding control were considered, as it was deemed that a separate set of criteria would likely determine if more or fewer sRNAs are produced. However, as in the majority of cases, more sRNAs were transcript-derived in the mutants relative to the control, so this only translated to a small number of exclusions.

Transcripts which produced more 21mers and more 24mers in *msm1* had six overlapping significantly enriched GO terms. These can be simplified to two distinct categories; response to abiotic stress and ATP biosynthesis. These two broad, simplified categories can also be applied to the GO term enrichment of the 21mer producing transcripts and the 24mer producing transcripts separately (Figures 6.3A & B). Interestingly, the specific types of stress responsive gene which produce more 21mers were not the same as those that produced 24mers. For example, the 24mer producing transcripts have significant term enrichment for response to osmotic stress, response to salt and response to hormone - none of which appear in the 21mer producing transcripts (Figure 11.3B). In the 21mer producing transcripts, there is a significant enrichment of terms pertaining to sulphur compound catabolic process (Figure 11.3A). As discussed in Chapter 5, this is likely reflective of the stress response phenotype (Nwachukwu et al., 2012).

Many of the GO terms which were enriched in transcripts which produced more 21mers and 24mers in *msm1* were also present in the GO analysis performed in Chapter 5 on differentially expressed genes (Figure 9.3). However, as demonstrated in Figures 6.1F & G, many of these sRNA producing transcripts are not differentially expressed in *msm1*. Therefore, despite this overlap in GO terms between differentially expressed transcripts and sRNA producing transcripts, the transcripts themselves are different.

GO enrichment was also computed for the transcripts which produced more 21mers and 24mers in *cp13<sup>sg13</sup>*, in order to determine if they displayed the same pattern as *msm1*. There was no significant GO term enrichment for transcripts which produced a greater number of 24mers in *cp13<sup>sg13</sup>*. The 21mers, however,

did show significant enrichment of a number of GO terms (Figure 11.3C). The GO terms enriched in the 21mer producing transcripts of *cp13<sup>sg13</sup>* were more diverse than *msm1*. In addition to some stress responsive terms, the GO enrichment of 21mer producing transcripts in *cp13<sup>sg13</sup>* pertained to many developmental events, such as `Regulation of developmental vegetative growth` (Figure 11.3C).

**A****B****C**

**Figure 11.3 – GO term enrichment in transcripts producing more rqc-siRNAs in *msm1* or *cpl3<sup>sgi3</sup>***

**A:** Gene ontology enrichment of transcripts producing a significantly different number of 21mers in *msm1*.

**B:** As in **A** but for transcripts producing a significantly different number of 24mers in *msm1*.

**C:** Gene ontology enrichment of transcripts producing a significantly different number of 21mers in *cpl3<sup>sgi3</sup>*.

sRNA distribution across progenitor transcripts is even implicating an RDR activity

In the study on CPL3 performed by Li et al. (2019), they reported that the sRNAs produced in the absence of CPL3 which were derived from endogenous transcripts did not show any mapping bias for any region of the transcripts. They mapped evenly across the entire length of the transcript. This analysis considered all size clusters of sRNA between 18 – 25nt in a single analysis, and did not involve subcategorization into each size class. The sRNAs produced in *msm1* were analysed in the same way, but were then also subset into size 21nt and 24nt separately and re-analysed.

Transcripts which produced both more 21mers and 24mers in *msm1* similar to those seen in *cpl3<sup>sgi3</sup>* were observed, despite RDR6 not being functional in *msm1*. These transcripts produced a significantly different number of 21mers and 24mers in the mutant predominantly in exonic regions (Figure 11.4A). AT3G06435 was used as an example here, as it produced a significantly different number of 21mers and 24mers in *msm1*, and had high mapping count numbers. It also appeared to be representative of the larger trend seen from manual inspection of the top 15 transcripts which produced the greatest number of sRNAs in *msm1*. On manual inspection, there were some cases in which this did not appear to be the case, in which a majority of the reads appeared to map to the 3' end (data not shown). However, these examples had comparatively low count data relative to the transcripts with even spread distribution, and were therefore not considered to be biologically significant.

AT3G06435 had the largest distribution of 21mers close the 3' end, however when the scale is taken into account it had a substantial number of reads spread across the rest of the transcript (Figure 11.4A). This pattern was also seen in the control but with lower read numbers. The 24mers had a more even distribution, with the peak occurring roughly mid-way through the first exon (Figure 11.4B). The peak counts for the 24mers were lower than the 21mers, however a similar difference between the mutant and control were seen. Finally, all sRNAs between 20-25nt were examined for AT3G06435 (Figure 11.4C). This also demonstrated an even spread of sRNAs mapping across the whole of the exonic regions, indicating that this transcript was producing a large number of sRNAs.

Coverage plots were produced for genes which produced a significantly different number of 21mers exclusively and 24mers exclusively in *msm1*. These did not demonstrate notably high numbers of sRNAs compared with some of the other examples used, for example AT3G06435. However, on manual inspection, the distribution of sRNAs across the 24mer producing transcripts was largely uniform, with only a small number demonstrating a clustered distribution close to the 3' end. It is possible that these products are the result of mRNA degradation. This data is not presented as the count numbers were too low. Overall this even spread of sRNAs, coupled with the earlier observation that sRNAs mapped to both forward and reverse strands (Figures 6.2C & D), implicates an RDR activity.

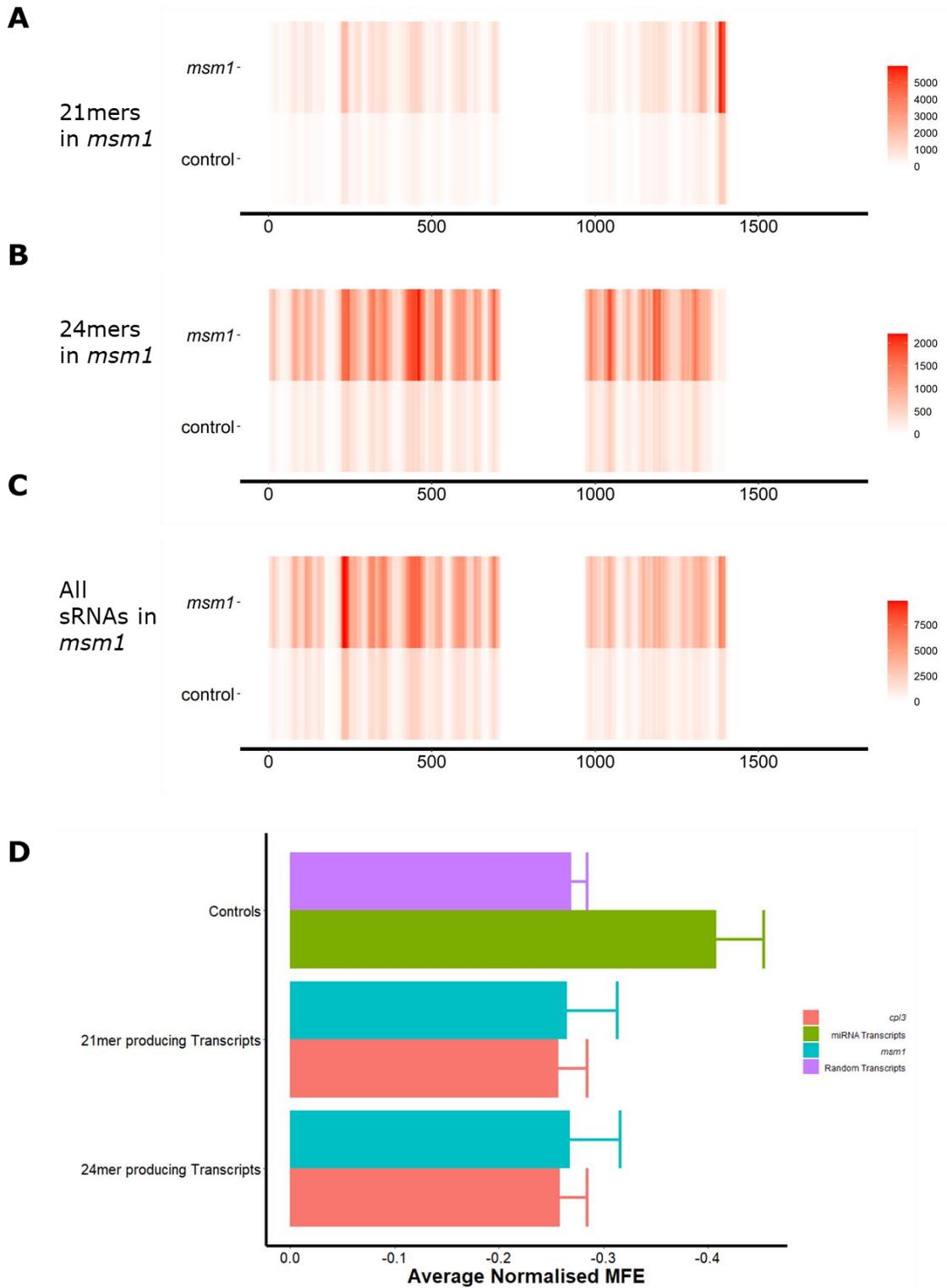
I then looked for differences in the average MFE of transcripts producing more 21mers and 24mers separately in *cp13<sup>sgl3</sup>* and *msm1*. Transcripts in *msm1* which produced more 21mers exclusively, transcripts which produced more 24mers exclusively and transcripts producing more 21mers and 24mers in *cp13<sup>sgl3</sup>* were folded using RNAfold in the ViennaRNA package (Lorenz et al., 2011). The whole transcript was used as the distribution of sRNAs was even across the whole length. The MFE was then taken from these folded structures for each gene, and then the average was calculated for all genes in each condition (Figure 11.4D). The MFE was also calculated for a comparable number of random transcripts which were expressed in the roots of *msm1*. This was done to represent a standard of MFE expected of transcripts which did not have any association with each other. As a final control, 18 *Arabidopsis* miRNA transcripts were also taken

from TAIR10, and folded alongside the other transcript populations. These transcripts have been experimentally shown to produce the necessary secondary hairpin structures required to generate mature miRNAs, and therefore represent the MFE of structured, sRNA producing transcripts.

After the MFE was calculated for each transcript, the `normalised MFE` was then calculated for each by dividing the MFE by the transcript length in nucleotides, in order to account for differences in length. This was an important step as longer transcripts will have lower MFEs, as they are naturally more likely to adopt more secondary structures. Finally, the average normalised MFE for each population of transcripts was plotted (Figure 11.4D).

Statistical significance was tested using pairwise T tests between each condition. The miRNA producing transcripts were significantly different from every other population of transcripts. Transcripts which produced more 21mers and more 24mers in *cp/3<sup>sgl3</sup>* differed significantly from the random selection of transcripts. In these cases, the MFE in the *cp/3<sup>sgl3</sup>* transcript populations was higher, suggesting that they were less likely to form secondary structures than any of the other populations. This supports RDR6 operating on these transcripts to produce sRNAs. This was not so for any of the *msm1* populations (Table 4). It was also not possible to determine any significant difference between any of the 21mer producing populations and the 24mer producing populations (Table 4).

**AT3G06435.1**



**Figure 11.4 – Coverage plots of sRNAs deriving from endogenous transcripts in *msm1* and their MFE**

**A:** Coverage plot heatmap of 21mers mapping to example gene AT3G06435.1 produced from *msm1* sRNA sequencing data.

**B:** As in **A** but with 24mers

**C:** As in **A** and **B** but with all size sRNAs.

**D:** Average Normalised MFE in corresponding categories of transcript. Colours represent the background / type of control, error bars represent 1 standard deviation.

**Table 4 – Statistical analyses of MFE data**

Pairwise results of two sample T tests performed between corresponding groups from Figure 9D.

<b>Group 1</b>	<b>Group 2</b>	<b>p value</b>
21_cpl3	random transcripts	4.62E-05
21_cpl3	pri-miRNA transcript	2.21E-10
21_msm1	pri-miRNA transcript	1.31E-11
24_msm1	pri-miRNA transcript	1.18E-11
random transcripts	pri-miRNA transcript	4.22E-10
24_cpl3	random transcripts	0.004209
24_cpl3	pri-miRNA transcript	3.20E-11

## *msm1* has a significantly increased level of TE silencing

Finally, as there were a number of genes which produced a significantly greater number of 24mers in *msm1*, I decided to check whether or not the levels of transposable elements (TEs) were different. The activity of TEs are routinely suppressed in *Arabidopsis* both by post-transcriptional and transcriptional mechanisms. Transcriptional silencing through the activity of RDR2 and DCL3 (Blevins et al., n.d.) are among the main mediators of this repression; as previously described this pathway produces 24mers. Both mRNA sequencing data from *msm1* and sRNA sequencing data from *msm1* were quantitatively mapped to the TAIR10 TE database taken from TAIR ("TAIR - Download - TAIR10 transposable elements," n.d.).

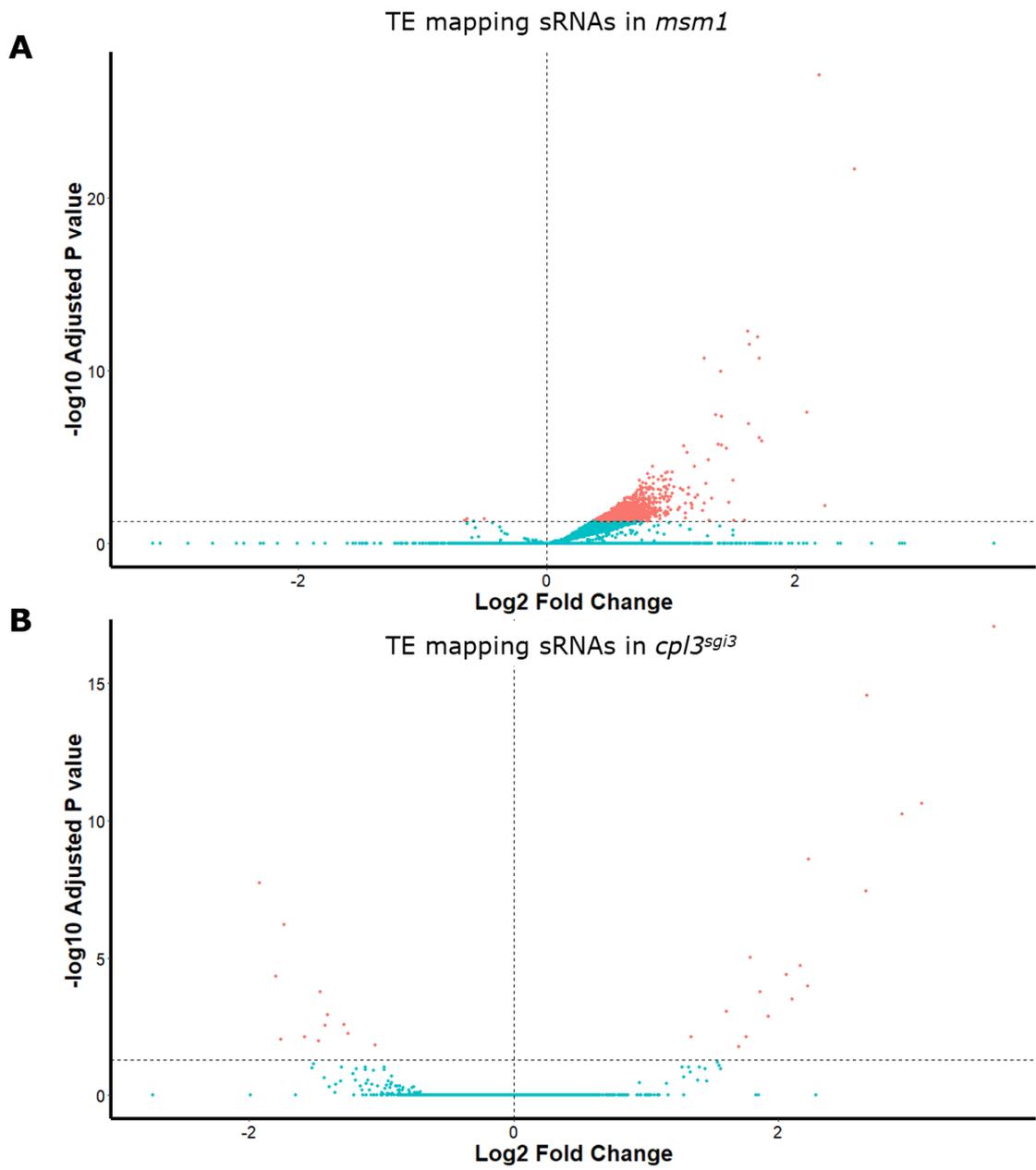
In the mRNA sequencing, only one known TE exhibited a differential abundance in *msm1*. This was AT5TE47100, and was only present in *msm1*. The counts averaged in *msm1* at 312, and 0 in the control background. It is possible that this was a false positive as it was the only one.

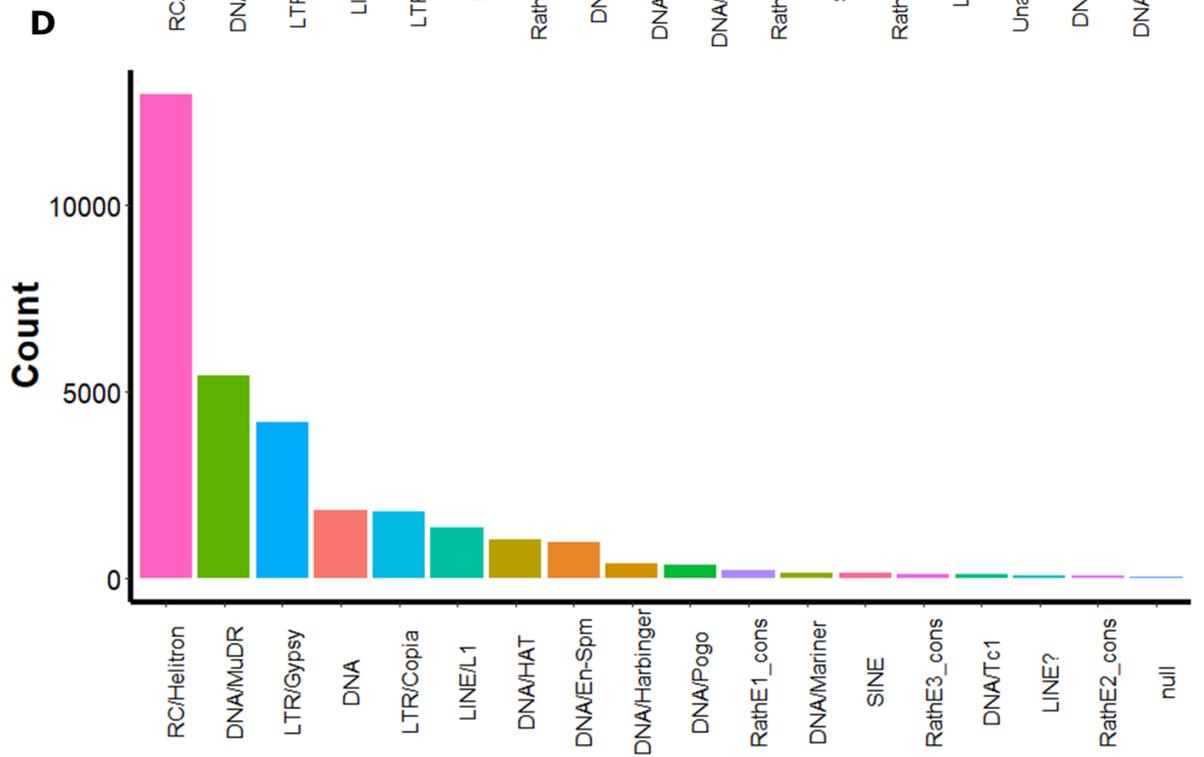
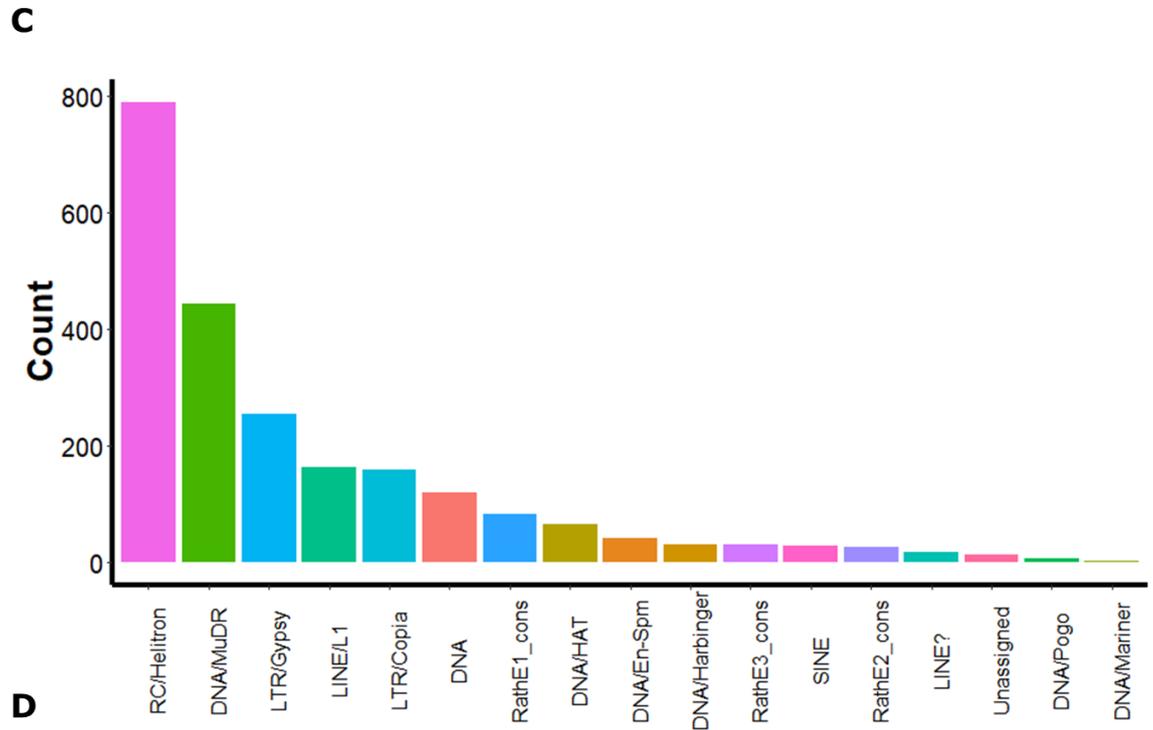
Despite only one transposon having a significantly different expression level in the mRNA sequencing data taken from *msm1*, the sRNA sequencing data demonstrated a much more dramatic difference. In *msm1*, over 2000 TEs have a significantly different number of 24mers mapping to them. Of these, only 3 had a lower number of sRNAs mapping in *msm1* than in the control; for the rest there was a significantly higher number of sRNAs mapping in *msm1* than the control. These count differences were considerable, with *msm1* displaying thousands more counts for some TEs than the control (Figure 11.5A).

The same analysis was performed on the mRNA and sRNA sequencing data published by Li et al. (2019) (Figure 11.5B). In their mRNA sequencing data, 7 TEs displayed a lower level of expression in *cp13<sup>sgl3</sup>*, and 102 TEs displayed a higher level of expression in *cp13<sup>sgl3</sup>*. Contrasting the >2000 TEs which displayed a significantly different number of sRNAs mapping to them in *msm1*, only 29 TEs displayed such a phenomenon in *cp13<sup>sgl3</sup>*. Of these, 11 had fewer sRNAs mapping to them in *cp13<sup>sgl3</sup>*, and the rest had more.

Finally, a count was performed for each TE superfamily whose members had a significantly different number of sRNAs mapping to them in *msm1*. RC/Helitron

represented the superfamily with the largest representation in the data; the count was almost double that of the next most represented family. DNA/MuDR and LTR/Gypsy both had the next highest representation, after which the frequency tails off for the rest of the superfamilies (Figure 11.5C). In order to determine if there was a significance to the superfamilies which were represented, or rather if they simply reflected the superfamilies by number, the number of TEs belonging to each superfamily were plotted. These mirrored the data seen in Figure 11.5C, with RC/Helitron being the most abundant, followed by DNA/MuDR and LTR/Gypsy (Figure 11.5D). This suggested that the number of siRNAs mapping to TEs in *msm1* was proportional to the number of TEs, and there did not appear to be a specific enrichment which could not be accounted for by the number of TEs in the superfamily. A list of the counts for each of the TE for which there was a significantly different number of mapping sRNAs in *msm1* can be found in supplementary table S6.





## Figure 11.5 – Mapping sRNAs to TEs in *msm1* and *cpl3<sup>sgi3</sup>*

**A:** Volcano plot showing number of sRNAs mapping to TEs in *msm1* versus pSU2:GFP:395. Red dots indicate significance value of or below  $p=0.05$ . Vertical dashed line shows log<sub>2</sub> fold change of 0 for ease of interpreting increases and decreases. Horizontal dashed line demonstrates the significance threshold of  $p=0.05$ .

**B:** As in A) but the data used is the sRNA mapping to TEs in *cpl3<sup>sgi3</sup>* versus the control.

**C:** Frequency of each type of TE with a significantly different number of sRNAs mapping in *msm1*.

**D:** Number of known TE in each superfamily taken from TAIR.

## Discussion

### Limitations

As described, the Salmon quantitative mapping tool was used to map sRNAs to endogenous transcripts (Patro et al., 2017). This was used to identify the transcripts from which the sRNAs were derived. As they are transcript derived siRNAs, they are not expected to contain mismatches. Owing to this, it is expected that the siRNAs which derive from the antisense strand produced by RDR activity would negatively regulate the transcripts from which they derive, as they contain a perfect match and can therefore guide the appropriate AGO protein, forming a type of cis-regulation (Hutvagner and Simard, 2008). However, it is known that sRNA targeting rules can still permit binding and regulation by sRNA in the event of a number of mismatches, which is dependent on where these mismatches occur between the sRNA and target site (Kim et al., 2016). Therefore, despite there being little evidence that these sRNAs were

affecting the level of expression of the transcripts they seemingly derived from (Figures 6.1F, G, H), it is possible that they could work in trans and negatively regulate the expression of related genes. This could potentially explain why the GO enrichment terms of genes producing more sRNAs in *msm1* overlap with the GO terms of differentially expressed genes in *msm1* despite the transcripts being different.

Li et al. (2019) demonstrate that the rqc-siRNA populations which accumulate at higher levels in *cp13<sup>sg13</sup>* do so in an RDR6-dependent mechanism. However, *msm1* contains mutations in both *cp13* but also in *rdr6*. Therefore, for genes which have a significantly different number of sRNAs mapping to them in *msm1* roots but not *cp13<sup>sg13</sup>*, these sRNAs may be RDR6 independent. If this were the case, then it would be expected that the transcripts which produced a significantly different number of 21mers in *msm1* might have a lower MFE, as they would rely on folding for the induction of double-stranded secondary structures. This was not so, as the average normalised MFE of 21mer producing transcripts in *msm1* did not differ significantly from either random transcripts or 21mer producing transcripts in *cp13<sup>sg13</sup>*. This would also serve as an explanation as to why the overlap between transcripts which produce more or fewer 21mers in *msm1* and *cp13<sup>sg13</sup>* is so low, as those in *cp13<sup>sg13</sup>* would be RDR6 independent and those in *msm1* would be RDR6 dependent.

However, a limitation to this approach of normalising the MFE of a transcript based on length is that the transcript itself could contain discrete regions which fold to produce secondary structures, which are diluted out by the rest of the transcript which do not exhibit such a property. This may also introduce some bias, as the transcripts could have a lower MFE because they are longer, and that could be the cause for them producing more sRNAs. It may not be a confounding factor but rather the defining factor. This has been discussed at length in (Trotta, 2014), and mathematically more complex methods were suggested to circumvent this. However, this is unlikely cause an issue in this case, as the coverage plots of the sRNAs when mapped to the transcripts they derive from generally show an even coverage across the whole transcript (Figure 11.4). Therefore, it is not likely that they are the product of discrete folding sites within the transcript, and are more likely to derive from RDR products.

## 24mers in *msm1* and their consequence

The transcripts which produce a significantly different number of 24mers in either background potentially represent a different route of sRNA biogenesis. This can be assumed for a number of reasons. Firstly, biogenesis of 24mer sRNA species usually relies on RDR2, rather than RDR6. This has been clearly demonstrated using knockout studies and sRNA sequencing (Blevins et al., n.d.). RDR2 represents the RNA polymerase activity, and these substrates are then typically processed by DCL3, which is what determines the length of the sRNA. This is probably also the source of the 24mers seen in *cp13<sup>sgl3</sup>*, and indeed there is a precedent for RDR2 being involved in S-PTGS of transgenes, which was the primary focus of the study by Li et al. (2019) (Jauvion et al., 2012). There was evidence for an RDR activity in the strand distribution analysis (Figure 11.2C & D), albeit there was a strong bias for the antisense strand. This distribution is reminiscent of the strand distribution seen in sRNA sequencing when studying mature miRNAs, and exists because the guide strand is selected by AGO1 and the passenger strand is degraded (Okamura et al., 2009). It is possible a similar explanation can be applied in this context too, however the selecting AGO would more likely be AGO4 than AGO1 (Wang and Axtell, 2017), and the consequence would be that these loaded, active siRNAs target only the sense strand of the gene in the DNA (Lewsey et al., 2016).

Secondly, in *msm1*, which does not have functional RDR6, the number of transcripts producing a significantly different number of 24nt sRNAs is equivalent to that seen in *cp13<sup>sgl3</sup>* (Figure 11.11C). However, RDR2 has been shown to be a nuclear localised protein (Pontes et al., 2006). This does not seem compatible with the model proposed by Li et al. (2019), in which P bodies become saturated with aberrant transcripts, and the overflow is processed by the spatially-linked, cytoplasmic siRNA body (Li et al., 2019), as the nuclear bodies which contain RDR2 are isolated from this body. This could be explained however, as RDR2 acting to maintain DNA methylation induced by RDR6-dependent 21mer induced methylation, for which there is a precedent (Jauvion et al., 2012). This does not explain why there is an increase in 24mers in *msm1*. Indeed, the same study by Jauvion et al. (2012) demonstrated that RDR2 can partially antagonise the production of RDR6-dependent siRNA in S-PTGS,

evidenced by the observation that S-PTGS is more efficient in *rdr2* than WT. It might be expected then that in the absence of RDR6, RDR2 mediated silencing might be more efficient, and therefore a greater number of 24mers would be produced. This might explain why the number of transcripts with a significantly different number of 24mers mapping was greater in the *rdr6-15* backgrounds than *msm1* was in. However, this does not explain the differences between *msm1* and its corresponding *rdr6-15* control.

When the 24mer producing transcripts in *msm1* and *cp13<sup>sg13</sup>* are compared, there is only one overlapping transcript which produces a differential number of 24mers. This is not unexpected given the low degree of overlap between all the sRNA producing transcripts in *msm1* and *cp13<sup>sg13</sup>*. That the number of 24mer producing transcripts did not decrease in *msm1* in the way that the 21mer producing transcripts did, relative to *cp13<sup>sg13</sup>*, reinforces the deduction that these sRNAs are produced as a result of an RDR activity that is not RDR6.

As in the study by Li et al. (2019), a majority of these sRNAs did not appear to affect the levels of the transcripts from which they were derived. A number of reasons were suggested for this, such as the sRNAs not reaching a level sufficient to enact PTGS. For the genes that did exhibit a canonical relationship in *cp13*, this was shown to be more pronounced under stress conditions. This observation fits well with previous literature on *cp13*, which provided evidence to suggest that CPL3 is typically involved in reprogramming transcription in response to environmental cues (Bang et al., 2006). Additionally, *cp13* does not exhibit developmental defects such as those seen in other *cp1* knockouts such as *cp14*, further suggesting it has adopted a more specialised and contextual function. In *msm1*, there did not appear to be any transcripts which displayed a canonical sRNA:target relationship. This is perhaps to be expected, as transcripts which displayed a canonical relationship in *cp13<sup>sg13</sup>* with sRNAs derived from them, did not in *cp13<sup>sg13</sup> rdr6* (Li et al., 2019). This is likely because the 21mers produced in *cp13<sup>sg13</sup>*, which are most likely AGO1 bound, would likely have a much faster and more direct effect on the transcriptome as they work post-transcriptionally (Okamura et al., 2004). This contrasts the main mode of action of 24mers, which facilitate transcriptional silencing (Blevins et al., n.d.).

It was interesting that a small proportion of genes had a larger number of sRNAs mapping to them in the control than the mutant, for both *msm1* and *cp13<sup>sg13</sup>*.

Additionally, even for genes with a significantly higher number of sRNAs mapping to them in the mutant in both *msm1* and *cp13<sup>sg13</sup>*, the control also had sRNAs mapping to these genes. These counts were not inconsiderable for some of them, with some demonstrating counts of up to 1000. This suggests that these rqc-siRNAs can accumulate under normal conditions too. This was also clear in the coverage plots (Figure 11.4), which demonstrated that for a number of genes sRNAs accumulate both in the mutant and control, with the levels being higher in the mutant, consistent with what was shown by Li et al. (2019). Interestingly, the distribution of sRNAs across the transcripts appeared to be the same in the mutant as in the control. This could also be the result of aberrant transcripts entering into decay pathways, as in *cp13<sup>sg13</sup>*; *cp13* could just exacerbate the number.

### Purpose of the sRNAs?

That the GO analysis of transcripts in *msm1* which produced more 21mers and 24mers reflected the GO analysis performed in Chapter 5, without the same genes being analysed is particularly interesting. There are a number of potential explanations for this. Firstly, it could be the result of *cp13* transcription factor activity, which determines the types of genes that are *cp13* responsive. In this scenario, however, it is not clear as to why some of the genes are differentially expressed and some produce more sRNAs. Secondly, it could be the downstream result of the pathways perturbed by the causative mutations in *cp13*. Thirdly, this could link in with the potential alternative splicing phenotype displayed by *msm1* and described in Chapter 5. Some stress responsive genes, such as HsfA2, are constitutively expressed, but alternatively spliced in such a way that when they are not required, they contain a premature stop codon (Liu et al., 2013). This results in NMD of the transcript which could leave footprints behind in the form of small, RNA fragments. If the splicing profile of *msm1* is altered then some of these transcripts may undergo this nonsense mediated decay in inappropriate ways. This could also explain why in *msm1* many of these transcripts produce sRNAs under control conditions as well, and there are simply more of them in *msm1*.

As described in Chapter 5, a majority of the GO terms enriched in *msm1* differentially expressed transcripts, and now in transcripts which produce a

significantly different number of transcripts, pertain to stress response. The role of CPL3 and stress response has already been described (Bang et al., 2006), and so this is perhaps unsurprising. Strangely, ATP biosynthesis GO terms were also significantly enriched in both populations of transcripts in *msm1*. Exactly why this is, is not clear, however, it is likely to be due to the mutation in CPL3 as more sRNAs derive from these transcripts as well as them being differentially expressed.

### TE silencing in *msm1*

The continuous production of siRNAs which direct the silencing of TEs represents an area of debate in the literature. However, one of the leading theories for why this happens is to dynamically define the borders of the TEs. This is supported by the observation that the 24mers involved in RdDM appear to preferentially target the edges of the TEs when mapped back (Zemach et al., 2013). The purpose of this border-defining is likely to prevent the expression of TEs situated near active genes, so that the open chromatin state does not result in TEs being expressed along with the gene. This would undoubtedly have fitness implications.

The significantly increased numbers of 24mers mapping to TE in *msm1* were not immediately clear. Li et al demonstrated in their study that in *cp13<sup>sgl3</sup>*, the greatest number of sRNAs that accumulated differentially in the mutant background were those derived from transgenes (Li et al., 2019). The exact criteria through which these genes are distinguished from endogenous genes are not known, however these may also discriminate TEs from coding genes. This represented the largest difference in *msm1* from the control in the sRNA data; the scale of change was on par with the scale of change seen in *cp13<sup>sgl3</sup>*.

It did not appear that there was any specific enrichment of any of the TE superfamilies in *msm1*. The number of TEs in each superfamily which had a significantly different number of sRNAs mapping in *msm1* mirrored the number of TEs present in each superfamily (Figure 11.5C & D), indicating that there was no clear enrichment of any superfamily. The enrichment of the specific TE families also reflected this (data not shown). This suggests that the underlying mechanism which is perturbed in *msm1* is likely more general, and not specific to certain TE superfamilies.

The disparity between sRNAs mapping to TEs between *cp13<sup>sgl3</sup>* and *msm1* is interesting. In *cp13<sup>sgl3</sup>* the number of such cases where there is a significantly different number of sRNAs mapping to a TE is substantially lower than in *msm1*. A higher proportion of these also decrease in the mutant, than do in *msm1*; there is not the clear unilateral increase in *cp13<sup>sgl3</sup>* as in *msm1* (Figure 11.5A). This could indicate competition between the RDR6 dependent rqc-siRNA pathway and the pathway resulting in an increase in TE-mapping siRNAs seen in *msm1*. It could demonstrate that whilst accumulation of rqc-siRNAs in coding genes is RDR6 dependent, the increase in TE mapping sRNAs is not. In which case, in the absence of functional RDR6 this level of regulation may have been elevated, which would explain why it is so much more pronounced in *msm1*.

Competition could also be artificially introduced in the sequencing process, as all libraries were sequenced to the same depth, however *cp13<sup>sgl3</sup>* had an additional population of RDR6-dependent, 21mer rqc-siRNAs that *cp13<sup>msm1</sup>* did not have. These could have diluted out some of the TE derived siRNAs which accumulate at a higher level in *cp13<sup>msm1</sup>*. Alternatively, the TE-mapping siRNAs which increase in abundance in *msm1* may be a result of *cp13<sup>msm1</sup>*, as the SNP in this *cp13* allele is in a different place than the SNP in *cp13<sup>sgl3</sup>*. Finally, it could potentially be a result of one of the other mutant alleles in *msm1*. This is not considered to be as likely as the previous two scenarios, as none of the putative mutant alleles identified in Chapter 6 have ascribed functions related in any way to this phenomenon (Table 3).

The production of siRNAs which mediate the silencing of TEs in plants relies on the activity of RNA pol IV (Law et al., 2013), and is therefore independent of RNA pol II activity which is compromised by mutation to CPL3. This likely means that the enhanced silencing of TEs is unconnected to the production of aberrant transcripts described by Li et al. (2019) in *cp13<sup>sgl3</sup>*. The transcription of active TEs does, however, rely on RNA pol II. Nonetheless, a majority of these TEs were not readily detectable in the mRNA sequencing data of either *msm1* and *cp13<sup>sgl3</sup>* – although it is noteworthy that more were detectable in *cp13<sup>sgl3</sup>*. As there was only one TE in *msm1* that had a differential expression level in the mRNA sequencing when compared to the control, it can be concluded that in *msm1* there is an increase in TE silencing without a concomitant change in expression of the TEs which were detectable. In times when these transposons do become

active, such as in the meiocytes, endosperm or nurse cells of gametophytes, this may result in a functional consequence (Martínez and Slotkin, 2012).

One potential mechanism is that *cp13* acts as a negative regulator of RNA pol IV. This could be either by direct interaction with RNA pol IV or through an indirect interaction with one of its many associated proteins such as SHH1 or CLSY1 (Zhou and Law, 2015). In *msm1*, the nature of the SNP in CPL3 may result in a loss of this functionality, resulting in an increase in the rate of RNA pol IV transcription, which is required for the creation of 24mers, which target TEs. This would account for the increase in 24nt TE mapping siRNAs observed. This model does not take into account or necessarily require RDR6; the differentiating factor between *cp13<sup>sgl3</sup>* and *msm1* would be the location and therefore consequence of the SNP.

A second potential mechanism is that in *cp13*, there are two conflicting consequences. The RDR6 dependent accumulation of rqc-siRNAs competes with the enhanced production of 24mer TE mapping siRNAs. The differentiating factor here would be the presence or absence of functional RDR6. However, this is not mutually exclusive with the previously proposed mechanism, nor do the two combined represent the only potential explanations.

## Conclusion

Taken together, it seems unlikely that the genes which produce a greater number of sRNAs in *cp13<sup>sgl3</sup>* and *msm1* are random. Rather, in *cp13<sup>sgl3</sup>* genes related to development produce more sRNAs, and in *msm1* genes related to stress response produce more sRNAs. These appear to be the result of RDR activities as the distribution of these sRNAs appears to span the whole transcript, and there is no distinct difference between the normalised average MFE.

More dramatically, in *msm1*, there appears to be a significant increase in the levels of TE silencing, not seen in *cp13<sup>sgl3</sup>*. This was inferred by there being a larger number of sRNAs mapping to over 2000 TE in *msm1* than the control. This is likely to be a general mechanism that does not appear to affect any specific superfamily of transposons in a preferential way.

## Chapter 8: Discussion and conclusions

### miRNA and splicing phenotype in *msm1*

When all data is considered together, it seems unlikely that *msm1* has a general miRNA phenotype. The differences seen in chapter 4 (Figure 8.3 B), which appear to be substantial only in the root, are likely due to the *ops* mutation. As OPS regulates phloem differentiation, it is likely that the difference in miRNA profile between control seedlings and *msm1* is the result of comparing root tissue at different stages of development.

There is also a clear ABA phenotype displayed in *msm1* (Figures 4.1 and 4.3). This phenotype has been documented in other CPL mutants (Koiwa et al., 2002), suggesting that this element of the phenotype seen in *msm1* is likely attributable to *cp13<sup>msm1</sup>*. As previously discussed, there is also a strong link between ABA and alternative splicing (Zhu et al., 2017). Therefore, it is not clear whether *msm1* has a true alternative splicing phenotype, or whether this is an indirect effect of the altered ABA signalling. Further work would need to be performed in order to disentangle these two elements.

### Understanding CPL3 through *msm1*

The analysis of the transcriptome and sRNA profile of *msm1* has yielded novel insights into the likely function of CPL3 in an *rdr6* background. This does, however, rely on complementation with WT CPL3 followed by phenotyping to be certain.

Firstly, in a *cp13* background, in the absence of RDR6, rqc-siRNAs which are 21nt long are lost, but there is no clear reduction in the number of 24nt rqc-siRNAs. This follows on from the work performed by Li et al. (2019), in which they showed a number of genes which appeared to be regulated by rqc-siRNA in *cp13* returned to WT levels in *cp13 rdr6*. This also suggests that the 24nt sRNAs are likely not RDR6 dependent, however this may also not be the case as the genes producing 24nt sRNAs in *msm1* are largely different to the genes producing more in *cp13*.

Secondly, in *cp13 rdr6* there appears to be a large number of sRNAs which map to TEs. This was not commented on by Li et al, however on analysis of their data it did not appear to occur at a comparable level. There is not sufficient evidence to say why this is, however several hypothesis can be suggested.

The first hypothesis as to why these TE mapping sRNAs accumulate in *msm1* and not in published *cp13* data is that the position and nature of the SNP is different. In the published *cp13* mutant a glutamine residue was converted to a premature stop codon (Li et al., 2019), however in *msm1* a lysine residue is converted to a non-charged methionine, and the location of this was further upstream than the SNP in the published *cp13* mutant. This could be tested by complementing the public *cp13* mutant background with *cp13<sup>msm1</sup>*, followed by sRNA sequencing and mapping to TE.

The second hypothesis is that this production of TE mapping sRNAs in *msm1* only occurs because the background is *rdr6*. It has been shown both in this work and the published study that many of the rqc-siRNAs that accumulate in *cp13* are RDR6 dependent (Li et al., 2019). Therefore, in the absence of RDR6 a lower level of CPL3 regulation may be elevated and detectable in the sRNA sequencing. This could be tested in a number of ways. Firstly, the published study on CPL3 did generate an *rdr6 cp13* mutant, but they didn't perform sRNA sequencing on it. sRNA libraries could be generated from this line and sequenced, and if this too has an increase in TE mapping sRNAs then it would support the hypothesis.

The third hypothesis is that the TE mapping sRNAs produced in *msm1* are CPL3 independent, and are the result of one of the other mutations present in *msm1*. This seems unlikely based on the GO of the genes identified in the RNA sequencing in Chapter 6, especially considering that CPL3 is ascribed as a negative regulator of RNA interference. However, it is possible that these genes may have other functions which have not yet been identified, and therefore would not have been annotated as such. It is also possible that as the mutations were identified based on the RNA sequencing, some mutations which would not be represented at the transcriptome level could be responsible for the TE mapping sRNA phenotype. This could be tested by complementing *msm1* with wild type CPL3 and then repeating the sRNA sequencing, to see if the TE mapping sRNAs are lost.

If these TE mapping sRNAs are indeed the result of *msm1<sup>cp/3</sup>*, then the next rational question to ask is why and how they are produced. There are several different hypotheses as to why this could be. Firstly, sRNAs which repress TEs rely on RNA pol IV (Law et al., 2013). Therefore, the increase in these sRNAs in *msm1* could indicate that CPL3 acts as a negative regulator of RNA pol IV in a wild type background. This is not conceptually difficult to imagine as CPL3 usually interacts with RNA pol II subunits (Koiwa et al., 2002).

A second reason why there are so many more TE mapping sRNAs is that there could be an increased level of TE activation in *msm1*, which would necessitate an increase in RdDM to bring the levels back under control. This also has precedent, as a number of studies have linked transposon activation with biotic and abiotic stresses (Ito, 2013), and as shown in Chapter 5, Figure 9.2, *msm1* has an upregulation of stress responsive genes even under normal conditions. However, when transposon levels were quantified in the RNA sequencing of *msm1* and its corresponding control, only one TE displayed differential expression that was statistically significant. This evidence is not conclusive however, despite there not being a clear difference between the TE expression in control vs *msm1* seedlings, this could be because the levels of TE are brought back down by the increase in sRNAs.

### Main limitations of this study

The main and clear limitation of this study is that it was not possible to identify the causative mutations for the sRNA phenotypes seen in *msm1* in the timeframe of this PhD. Whilst evidence has been provided to suggest that *cp/3* is the most likely candidate, complementation with a wild-type copy followed by sRNA sequencing was not performed. Additionally, without this data there could also be mutations causing the sRNA phenotype which are not represented in the RNA sequencing data, such as SNPs in promoters or intergenic regions.

Additionally, *msm1* is a genetically complex background owing to the phenotypic *ops* mutation and potentially others. Therefore, once the causative mutation of the TE sRNA phenotype of *msm1* has been identified by complementation, the sRNA sequencing should be repeated in a simplified background, such as a SALK line or segregant from the backcrosses generated in Chapter 6 which are homozygous for the causative mutation but none of the other phenotypic ones.

## Future work

As discussed, the next logical steps to this work would be to systematically complement the *msm1* background with wild-type copies of the genes which have been identified as mutated in Chapter 6. Each of these complemented backgrounds would then be screened for the TE sRNA mapping phenotype, either by sRNA sequencing or, more cheaply, sRNA northern blot using radiolabelled TE probes (Alwine et al., 1977). Any background which loses the sRNAs would indicate that the complemented gene is the causative mutation for the TE mapping sRNAs – this would most likely be CPL3. If none of them lost this phenotype, then it would be appropriate to genome sequence *msm1* and look for other mutations which may not have been present in the RNA sequencing.

Another useful experiment to perform on *msm1* would be bisulfite sequencing. This method of sequencing involves treating DNA with bisulfite, which converts methylated cytosines to uracil residues (Li and Tollefsbol, 2011). Next-generation sequencing of DNA libraries prepared this way can then be analysed for differential levels of methylation by looking at the proportion of non-mapping uracils to mapping cytosines at a given position. If *msm1* has an increased number of 24mer sRNAs, then if they are functional it would be expected that they would have a higher proportion of genome methylation at targeted sites than their corresponding controls.

## Wider Relevance of this thesis

In this thesis, a screening method has been devised for use with the AtSUC2::GFP-miR395 line which can be used to identify potentially novel genes involved in miRNA turnover. As has been shown in this thesis, this screen also has the capacity to identify a wider range of PTGS mutants. Further screening of this EMS collection by subsequent individuals, following the protocol developed in this thesis and its' suggested corrections, could identify many more mutants in miRNA decay and PTGS.

Understanding the myriad mechanisms by which TEs are activated is of biological importance. Firstly, the activity of TEs in a genome can shape the evolutionary trajectory of a species. TEs are the single most variable component of plant genomes, and even closely related plant species can be host to wildly different TE populations (Lisch, 2013). Selection pressures can be applied to these TEs, potentially resulting in epigenetic and genetic variation in gene function. Therefore, understanding the mechanisms of TE regulation can enhance our understanding of evolution and speciation at a molecular level.

Secondly, there is an increasing body of evidence that TE activation can occur in response to various abiotic and biotic stresses. Mutations in these TEs often results in phenotypes, particularly when challenged with abiotic stresses (Joly-Lopez et al., 2017). Many of these are temperature stresses (Ito, 2013), and understanding these is of increasing importance in the wake of a changing climate in the 21<sup>st</sup> century (Difflenbaugh and Scherer, 2011).

## Chapter 9: References

- Allen, E., Xie, Z., Gustafson, A.M., Carrington, J.C., 2005. microRNA-Directed Phasing during Trans-Acting siRNA Biogenesis in Plants. *Cell* 121, 207–221. <https://doi.org/10.1016/j.cell.2005.04.004>
- Alwine, J.C., Kemp, D.J., Stark, G.R., 1977. Method for detection of specific RNAs in agarose gels by transfer to diazobenzyloxymethyl-paper and hybridization with DNA probes. *Proc. Natl. Acad. Sci. U. S. A.* 74, 5350–5354. <https://doi.org/10.1073/pnas.74.12.5350>
- Androvic, P., Romanyuk, N., Urdzikova-Machova, L., Rohlova, E., Kubista, M., Valihrach, L., 2019. Two-tailed RT-qPCR panel for quality control of circulating microRNA studies. *Sci. Rep.* 9. <https://doi.org/10.1038/s41598-019-40513-w>
- Androvic, P., Valihrach, L., Elling, J., Sjoback, R., Kubista, M., 2017. Two-tailed RT-qPCR: a novel method for highly accurate miRNA quantification. *Nucleic Acids Res.* 45, e144. <https://doi.org/10.1093/nar/gkx588>
- Archambault, J., Chambers, R.S., Kobor, M.S., Ho, Y., Cartier, M., Bolotin, D., Andrews, B., Kane, C.M., Greenblatt, J., 1997. An essential component of a C-terminal domain phosphatase that interacts with transcription factor IIF in *Saccharomyces cerevisiae*. *Proc. Natl. Acad. Sci. U. S. A.* 94, 14300–14305.
- Ashburner, M., Ball, C.A., Blake, J.A., Botstein, D., Butler, H., Cherry, J.M., Davis, A.P., Dolinski, K., Dwight, S.S., Eppig, J.T., Harris, M.A., Hill, D.P., Issel-Tarver, L., Kasarskis, A., Lewis, S., Matese, J.C., Richardson, J.E., Ringwald, M., Rubin, G.M., Sherlock, G., 2000. Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. *Nat. Genet.* 25, 25–29. <https://doi.org/10.1038/75556>
- Bang, W., Kim, S., Ueda, A., Vikram, M., Yun, D., Bressan, R.A., Hasegawa, P.M., Bahk, J., Koiwa, H., 2006a. Arabidopsis Carboxyl-Terminal Domain Phosphatase-Like Isoforms Share Common Catalytic and Interaction Domains But Have Distinct in Planta Functions. *Plant Physiol.* 142, 586–594. <https://doi.org/10.1104/pp.106.084939>
- Bang, W., Kim, S., Ueda, A., Vikram, M., Yun, D., Bressan, R.A., Hasegawa, P.M., Bahk, J., Koiwa, H., 2006b. Arabidopsis Carboxyl-Terminal Domain Phosphatase-Like Isoforms Share Common Catalytic and Interaction Domains But Have Distinct in Planta Functions. *Plant Physiol.* 142, 586–594. <https://doi.org/10.1104/pp.106.084939>
- Bang, W., Kim, S., Ueda, A., Vikram, M., Yun, D., Bressan, R.A., Hasegawa, P.M., Bahk, J., Koiwa, H., 2006c. Arabidopsis Carboxyl-Terminal Domain Phosphatase-Like Isoforms Share Common Catalytic and Interaction Domains But Have Distinct in Planta Functions. *Plant Physiol.* 142, 586–594. <https://doi.org/10.1104/pp.106.084939>
- Baran-Gale, J., Kurtz, C.L., Erdos, M.R., Sison, C., Young, A., Fannin, E.E., Chines, P.S., Sethupathy, P., 2015. Addressing Bias in Small RNA Library Preparation for Sequencing: A New Protocol Recovers MicroRNAs that Evade Capture by Current Methods. *Front. Genet.* 6. <https://doi.org/10.3389/fgene.2015.00352>
- Bartel, D.P., 2009. MicroRNA Target Recognition and Regulatory Functions. *Cell* 136, 215–233. <https://doi.org/10.1016/j.cell.2009.01.002>
- Bartel, D.P., 2004. MicroRNAs: Genomics, Biogenesis, Mechanism, and Function. *Cell* 116, 281–297. [https://doi.org/10.1016/S0092-8674\(04\)00045-5](https://doi.org/10.1016/S0092-8674(04)00045-5)
- Baumberger, N., Baulcombe, D.C., 2005. Arabidopsis ARGONAUTE1 is an RNA Slicer that selectively recruits microRNAs and short interfering RNAs. *Proc. Natl. Acad. Sci.* 102, 11928–11933. <https://doi.org/10.1073/pnas.0505461102>
- Bedre, R., Irigoyen, S., Petrillo, E., Mandadi, K.K., 2019. New Era in Plant Alternative Splicing Analysis Enabled by Advances in High-Throughput Sequencing (HTS) Technologies. *Front. Plant Sci.* 10. <https://doi.org/10.3389/fpls.2019.00740>

- Betts, M.J., Russell, R.B., 2003. Amino Acid Properties and Consequences of Substitutions, in: *Bioinformatics for Geneticists*. John Wiley & Sons, Ltd, pp. 289–316. <https://doi.org/10.1002/0470867302.ch14>
- Bielewicz, D., Kalak, M., Kalyna, M., Windels, D., Barta, A., Vazquez, F., Szweykowska-Kulinska, Z., Jarmolowski, A., 2013. Introns of plant pri-miRNAs enhance miRNA biogenesis. *EMBO Rep.* 14, 622–628. <https://doi.org/10.1038/embor.2013.62>
- Blevins, T., Podicheti, R., Mishra, V., Marasco, M., Wang, J., Rusch, D., Tang, H., Pikaard, C.S., n.d. Identification of Pol IV and RDR2-dependent precursors of 24 nt siRNAs guiding de novo DNA methylation in Arabidopsis. *eLife* 4. <https://doi.org/10.7554/eLife.09591>
- Bohmert, K., Camus, I., Bellini, C., Bouchez, D., Caboche, M., Benning, C., 1998. AGO1 defines a novel locus of Arabidopsis controlling leaf development. *EMBO J.* 17, 170–180. <https://doi.org/10.1093/emboj/17.1.170>
- Bollman, K.M., Aukerman, M.J., Park, M.-Y., Hunter, C., Berardini, T.Z., Poethig, R.S., 2003. HASTY, the Arabidopsis ortholog of exportin 5/MSN5, regulates phase change and morphogenesis. *Development* 130, 1493–1504. <https://doi.org/10.1242/dev.00362>
- Bouché, N., Laressergues, D., Gascioli, V., Vaucheret, H., 2006. An antagonistic function for Arabidopsis DCL2 in development and a new function for DCL4 in generating viral siRNAs. *EMBO J.* 25, 3347–3356. <https://doi.org/10.1038/sj.emboj.7601217>
- Boyes, D.C., Zayed, A.M., Ascenzi, R., McCaskill, A.J., Hoffman, N.E., Davis, K.R., Görlach, J., 2001. Growth stage-based phenotypic analysis of Arabidopsis: a model for high throughput functional genomics in plants. *Plant Cell* 13, 1499–1510. <https://doi.org/10.1105/tpc.010011>
- Braun, J.E., Huntzinger, E., Izaurrealde, E., 2013. The role of GW182 proteins in miRNA-mediated gene silencing. *Adv. Exp. Med. Biol.* 768, 147–163. [https://doi.org/10.1007/978-1-4614-5107-5\\_9](https://doi.org/10.1007/978-1-4614-5107-5_9)
- Bustin, S., Huggett, J., 2017. qPCR primer design revisited. *Biomol. Detect. Quantif.* 14, 19–28. <https://doi.org/10.1016/j.bdq.2017.11.001>
- Calixto, C.P.G., Guo, W., James, A.B., Tzioutziou, N.A., Entizne, J.C., Panter, P.E., Knight, H., Nimmo, H.G., Zhang, R., Brown, J.W.S., 2018a. Rapid and Dynamic Alternative Splicing Impacts the Arabidopsis Cold Response Transcriptome[CC-BY]. *Plant Cell* 30, 1424–1444. <https://doi.org/10.1105/tpc.18.00177>
- Calixto, C.P.G., Guo, W., James, A.B., Tzioutziou, N.A., Entizne, J.C., Panter, P.E., Knight, H., Nimmo, H.G., Zhang, R., Brown, J.W.S., 2018b. Rapid and Dynamic Alternative Splicing Impacts the Arabidopsis Cold Response Transcriptome[CC-BY]. *Plant Cell* 30, 1424–1444. <https://doi.org/10.1105/tpc.18.00177>
- Carpita, N.C., 2011. Update on Mechanisms of Plant Cell Wall Biosynthesis: How Plants Make Cellulose and Other (1→4)-β-d-Glycans. *Plant Physiol.* 155, 171–184. <https://doi.org/10.1104/pp.110.163360>
- Carthew, R.W., Sontheimer, E.J., 2009. Origins and Mechanisms of miRNAs and siRNAs. *Cell* 136, 642–655. <https://doi.org/10.1016/j.cell.2009.01.035>
- Chan, P.P., Lowe, T.M., 2009. GtRNADB: a database of transfer RNA genes detected in genomic sequence. *Nucleic Acids Res.* 37, D93–D97. <https://doi.org/10.1093/nar/gkn787>
- Chasin, L.A., 2007. Searching for splicing motifs. *Adv. Exp. Med. Biol.* 623, 85–106. [https://doi.org/10.1007/978-0-387-77374-2\\_6](https://doi.org/10.1007/978-0-387-77374-2_6)
- Chaudhary, S., Khokhar, W., Jabre, I., Reddy, A.S.N., Byrne, L.J., Wilson, C.M., Syed, N.H., 2019. Alternative Splicing and Protein Diversity: Plants Versus Animals. *Front. Plant Sci.* 10. <https://doi.org/10.3389/fpls.2019.00708>
- Chekanova, J.A., Gregory, B.D., Reverdatto, S.V., Chen, H., Kumar, R., Hooker, T., Yazaki, J., Li, P., Skiba, N., Peng, Q., Alonso, J., Brukhin, V., Grossniklaus, U., Ecker, J.R., Belostotsky, D.A., 2007. Genome-Wide High-Resolution Mapping of Exosome Substrates Reveals Hidden Features in the Arabidopsis Transcriptome. *Cell* 131, 1340–1353. <https://doi.org/10.1016/j.cell.2007.10.056>

- Chen, X., 2005. MicroRNA biogenesis and function in plants. *FEBS Lett.* 579, 5923–5931. <https://doi.org/10.1016/j.febslet.2005.07.071>
- Chong, G.L., Foo, M.H., Lin, W.-D., Wong, M.M., Verslues, P.E., 2019. Highly ABA-Induced 1 (HAI1)-Interacting protein HIN1 and drought acclimation-enhanced splicing efficiency at intron retention sites. *Proc. Natl. Acad. Sci.* 116, 22376–22385. <https://doi.org/10.1073/pnas.1906244116>
- CLOTE, P., FERRÉ, F., KRANAKIS, E., KRIZANC, D., 2005. Structural RNA has lower folding energy than random RNA of the same dinucleotide frequency. *RNA* 11, 578–591. <https://doi.org/10.1261/rna.7220505>
- Cruz, T.M.D., Carvalho, R.F., Richardson, D.N., Duque, P., 2014a. Abscisic Acid (ABA) Regulation of Arabidopsis SR Protein Gene Expression. *Int. J. Mol. Sci.* 15, 17541–17564. <https://doi.org/10.3390/ijms151017541>
- Cruz, T.M.D., Carvalho, R.F., Richardson, D.N., Duque, P., 2014b. Abscisic Acid (ABA) Regulation of Arabidopsis SR Protein Gene Expression. *Int. J. Mol. Sci.* 15, 17541–17564. <https://doi.org/10.3390/ijms151017541>
- Cui, P., Chen, T., Qin, T., Ding, F., Wang, Z., Chen, H., Xiong, L., 2016. The RNA Polymerase II C-Terminal Domain Phosphatase-Like Protein FIERY2/CPL1 Interacts with eIF4AIII and Is Essential for Nonsense-Mediated mRNA Decay in Arabidopsis. *Plant Cell* 28, 770–785. <https://doi.org/10.1105/tpc.15.00771>
- Curaba, J., Chen, X., 2008. Biochemical Activities of Arabidopsis RNA-dependent RNA Polymerase 6. *J. Biol. Chem.* 283, 3059–3066. <https://doi.org/10.1074/jbc.M708983200>
- Dalmay, T., 2013. Mechanism of miRNA-mediated repression of mRNA translation. *Essays Biochem.* 54, 29–38. <https://doi.org/10.1042/bse0540029>
- Dalmay, T., Hamilton, A., Rudd, S., Angell, S., Baulcombe, D.C., 2000. An RNA-dependent RNA polymerase gene in Arabidopsis is required for posttranscriptional gene silencing mediated by a transgene but not by a virus. *Cell* 101, 543–553. [https://doi.org/10.1016/s0092-8674\(00\)80864-8](https://doi.org/10.1016/s0092-8674(00)80864-8)
- Damayanti, F., Lombardo, F., Masuda, J., Shinozaki, Y., Ichino, T., Hoshikawa, K., Okabe, Y., Wang, N., Fukuda, N., Ariizumi, T., Ezura, H., 2019. Functional Disruption of the Tomato Putative Ortholog of HAWAIIAN SKIRT Results in Facultative Parthenocarpy, Reduced Fertility and Leaf Morphological Defects. *Front. Plant Sci.* 10. <https://doi.org/10.3389/fpls.2019.01234>
- de Felippes, F.F., Marchais, A., Sarazin, A., Oberlin, S., Voinnet, O., 2017. A single miR390 targeting event is sufficient for triggering TAS3-tasiRNA biogenesis in Arabidopsis. *Nucleic Acids Res.* 45, 5539–5554. <https://doi.org/10.1093/nar/gkx119>
- de Lima, J.C., Loss-Morais, G., Margis, R., 2012. MicroRNAs play critical roles during plant development and in response to abiotic stresses. *Genet. Mol. Biol.* 35, 1069–1077.
- Devanapally, S., Raman, P., Allgood, S., Etefa, F., Diop, M., Chey, M., Lin, Y., Cho, Y.E., Yin, R., Jose, A.M., 2020. Recovery from transgenerational RNA silencing is driven by gene-specific homeostasis. *bioRxiv* 148700. <https://doi.org/10.1101/148700>
- Diaz-Pendon, J.A., Li, F., Li, W.-X., Ding, S.-W., 2007. Suppression of Antiviral Silencing by Cucumber Mosaic Virus 2b Protein in Arabidopsis Is Associated with Drastically Reduced Accumulation of Three Classes of Viral Small Interfering RNAs. *Plant Cell* 19, 2053–2063. <https://doi.org/10.1105/tpc.106.047449>
- Diffenbaugh, N.S., Scherer, M., 2011. Observational and model evidence of global emergence of permanent, unprecedented heat in the 20th and 21st centuries. *Clim. Change* 107, 615–624. <https://doi.org/10.1007/s10584-011-0112-y>
- Ding, F., Cui, P., Wang, Z., Zhang, S., Ali, S., Xiong, L., 2014. Genome-wide analysis of alternative splicing of pre-mRNA under salt stress in Arabidopsis. *BMC Genomics* 15, 431. <https://doi.org/10.1186/1471-2164-15-431>
- Doench, J.G., Petersen, C.P., Sharp, P.A., 2003. siRNAs can function as miRNAs. *Genes Dev.* 17, 438–442. <https://doi.org/10.1101/gad.1064703>

- Duque, P., 2011. A role for SR proteins in plant stress responses. *Plant Signal. Behav.* 6, 49–54. <https://doi.org/10.4161/psb.6.1.14063>
- Dvinge, H., 2018. Regulation of alternative mRNA splicing: old players and new perspectives. *FEBS Lett.* 592, 2987–3006. <https://doi.org/10.1002/1873-3468.13119>
- English, A.C., Patel, K.S., Loraine, A.E., 2010. Prevalence of alternative splicing choices in *Arabidopsis thaliana*. *BMC Plant Biol.* 10, 102. <https://doi.org/10.1186/1471-2229-10-102>
- Enright, A.J., John, B., Gaul, U., Tuschl, T., Sander, C., Marks, D.S., 2003. MicroRNA targets in *Drosophila*. *Genome Biol.* 5, R1. <https://doi.org/10.1186/gb-2003-5-1-r1>
- Fagegaltier, D., Bougé, A.-L., Berry, B., Poisot, É., Sismeiro, O., Coppée, J.-Y., Théodore, L., Voinnet, O., Antoniewski, C., 2009. The endogenous siRNA pathway is involved in heterochromatin formation in *Drosophila*. *Proc. Natl. Acad. Sci.* 106, 21258–21263. <https://doi.org/10.1073/pnas.0809208105>
- Fahlgren, N., Montgomery, T.A., Howell, M.D., Allen, E., Dvorak, S.K., Alexander, A.L., Carrington, J.C., 2006. Regulation of AUXIN RESPONSE FACTOR3 by TAS3 ta-siRNA affects developmental timing and patterning in *Arabidopsis*. *Curr. Biol.* CB 16, 939–944. <https://doi.org/10.1016/j.cub.2006.03.065>
- Fang, Y., Hearn, S., Spector, D.L., 2004. Tissue-specific Expression and Dynamic Organization of SR Splicing Factors in *Arabidopsis*. *Mol. Biol. Cell* 15, 2664–2673. <https://doi.org/10.1091/mbc.E04-02-0100>
- Filichkin, S.A., Hamilton, M., Dharmawardhana, P.D., Singh, S.K., Sullivan, C., Ben-Hur, A., Reddy, A.S.N., Jaiswal, P., 2018. Abiotic Stresses Modulate Landscape of Poplar Transcriptome via Alternative Splicing, Differential Intron Retention, and Isoform Ratio Switching. *Front. Plant Sci.* 9. <https://doi.org/10.3389/fpls.2018.00005>
- Finnegan, E.J., Margis, R., Waterhouse, P.M., 2003. Posttranscriptional Gene Silencing Is Not Compromised in the *Arabidopsis* CARPEL FACTORY (DICER-LIKE1) Mutant, a Homolog of Dicer-1 from *Drosophila*. *Curr. Biol.* 13, 236–240. [https://doi.org/10.1016/S0960-9822\(03\)00010-1](https://doi.org/10.1016/S0960-9822(03)00010-1)
- Folkes, L., Moxon, S., Woolfenden, H.C., Stocks, M.B., Szittyá, G., Dalmay, T., Moulton, V., 2012. PAREsnip: a tool for rapid genome-wide discovery of small RNA/target interactions evidenced through degradome sequencing. *Nucleic Acids Res.* 40, e103. <https://doi.org/10.1093/nar/gks277>
- Frameshift Mutation - MeSH - NCBI [WWW Document], n.d. URL <https://www.ncbi.nlm.nih.gov/mesh?Db=mesh&Cmd=DetailsSearch&Term=%22Frameshift+Mutation%22%5BMeSH+Terms%5D> (accessed 5.5.20).
- Franco-Zorrilla, J.M., Valli, A., Todesco, M., Mateos, I., Puga, M.I., Rubio-Somoza, I., Leyva, A., Weigel, D., García, J.A., Paz-Ares, J., 2007. Target mimicry provides a new mechanism for regulation of microRNA activity. *Nat. Genet.* 39, 1033–1037. <https://doi.org/10.1038/ng2079>
- Fuchs Wightman, F., Giono, L.E., Fededa, J.P., de la Mata, M., 2018. Target RNAs Strike Back on MicroRNAs. *Front. Genet.* 9. <https://doi.org/10.3389/fgene.2018.00435>
- Fujii, H., Chiou, T.-J., Lin, S.-I., Aung, K., Zhu, J.-K., 2005a. A miRNA Involved in Phosphate-Starvation Response in *Arabidopsis*. *Curr. Biol.* 15, 2038–2043. <https://doi.org/10.1016/j.cub.2005.10.016>
- Fujii, H., Chiou, T.-J., Lin, S.-I., Aung, K., Zhu, J.-K., 2005b. A miRNA involved in phosphate-starvation response in *Arabidopsis*. *Curr. Biol.* CB 15, 2038–2043. <https://doi.org/10.1016/j.cub.2005.10.016>
- Gantier, M.P., McCoy, C.E., Rusinova, I., Saulep, D., Wang, D., Xu, D., Irving, A.T., Behlke, M.A., Hertzog, P.J., Mackay, F., Williams, B.R.G., 2011. Analysis of microRNA turnover in mammalian cells following Dicer1 ablation. *Nucleic Acids Res.* 39, 5692–5703. <https://doi.org/10.1093/nar/gkr148>

- Gao, R., Wang, Y., Gruber, M.Y., Hannoufa, A., 2018. miR156/SPL10 Modulates Lateral Root Development, Branching and Leaf Morphology in Arabidopsis by Silencing AGAMOUS-LIKE 79. *Front. Plant Sci.* 8. <https://doi.org/10.3389/fpls.2017.02226>
- Garcia, D., Garcia, S., Voinnet, O., 2014. Nonsense-mediated decay serves as a general viral restriction mechanism in plants. *Cell Host Microbe* 16, 391–402. <https://doi.org/10.1016/j.chom.2014.08.001>
- Ge, S.X., Jung, D., Yao, R., 2020a. ShinyGO: a graphical gene-set enrichment tool for animals and plants. *Bioinformatics* 36, 2628–2629. <https://doi.org/10.1093/bioinformatics/btz931>
- Ge, S.X., Jung, D., Yao, R., 2020b. ShinyGO: a graphical gene-set enrichment tool for animals and plants. *Bioinformatics* 36, 2628–2629. <https://doi.org/10.1093/bioinformatics/btz931>
- Ghasemi, A., Zahediasl, S., 2012. Normality Tests for Statistical Analysis: A Guide for Non-Statisticians. *Int. J. Endocrinol. Metab.* 10, 486–489. <https://doi.org/10.5812/ijem.3505>
- Ghini, F., Rubolino, C., Climent, M., Simeone, I., Marzi, M.J., Nicassio, F., 2018. Endogenous transcripts control miRNA levels and activity in mammalian cells by target-directed miRNA degradation. *Nat. Commun.* 9, 3119. <https://doi.org/10.1038/s41467-018-05182-9>
- Godoy Herz, M.A., Kubaczka, M.G., Brzyżek, G., Servi, L., Krzyszton, M., Simpson, C., Brown, J., Swiezewski, S., Petrillo, E., Kornblihtt, A.R., 2019. Light Regulates Plant Alternative Splicing through the Control of Transcriptional Elongation. *Mol. Cell* 73, 1066–1074.e3. <https://doi.org/10.1016/j.molcel.2018.12.005>
- González, E., Solano, R., Rubio, V., Leyva, A., Paz-Ares, J., 2005. PHOSPHATE TRANSPORTER TRAFFIC FACILITATOR1 Is a Plant-Specific SEC12-Related Protein That Enables the Endoplasmic Reticulum Exit of a High-Affinity Phosphate Transporter in Arabidopsis. *Plant Cell* 17, 3500–3512. <https://doi.org/10.1105/tpc.105.036640>
- Greene, E.A., Codomo, C.A., Taylor, N.E., Henikoff, J.G., Till, B.J., Reynolds, S.H., Enns, L.C., Burtner, C., Johnson, J.E., Odden, A.R., Comai, L., Henikoff, S., 2003. Spectrum of chemically induced mutations from a large-scale reverse-genetic screen in Arabidopsis. *Genetics* 164, 731–740.
- Griffiths-Jones, S., Grocock, R.J., van Dongen, S., Bateman, A., Enright, A.J., 2006. miRBase: microRNA sequences, targets and gene nomenclature. *Nucleic Acids Res.* 34, D140–D144. <https://doi.org/10.1093/nar/gkj112>
- Guo, Y., Liu, J., Eifenbein, S.J., Ma, Y., Zhong, M., Qiu, C., Ding, Y., Lu, J., 2015a. Characterization of the mammalian miRNA turnover landscape. *Nucleic Acids Res.* 43, 2326–2341. <https://doi.org/10.1093/nar/gkv057>
- Guo, Y., Liu, J., Eifenbein, S.J., Ma, Y., Zhong, M., Qiu, C., Ding, Y., Lu, J., 2015b. Characterization of the mammalian miRNA turnover landscape. *Nucleic Acids Res.* 43, 2326–2341. <https://doi.org/10.1093/nar/gkv057>
- Ha, M., Kim, E.-D., Chen, Z.J., 2009. Duplicate genes increase expression diversity in closely related species and allopolyploids. *Proc. Natl. Acad. Sci.* 106, 2295–2300. <https://doi.org/10.1073/pnas.0807350106>
- Hamilton, A.J., Baulcombe, D.C., 1999. A species of small antisense RNA in posttranscriptional gene silencing in plants. *Science* 286, 950–952. <https://doi.org/10.1126/science.286.5441.950>
- Heberle, H., Meirelles, G.V., da Silva, F.R., Telles, G.P., Minghim, R., 2015. InteractiVenn: a web-based tool for the analysis of sets through Venn diagrams. *BMC Bioinformatics* 16, 169. <https://doi.org/10.1186/s12859-015-0611-3>
- Howe, K.L., Contreras-Moreira, B., De Silva, N., Maslen, G., Akanni, W., Allen, J., Alvarez-Jarreta, J., Barba, M., Bolser, D.M., Cambell, L., Carbajo, M., Chakiachvili, M., Christensen, M., Cummins, C., Cuzick, A., Davis, P., Fexova, S., Gall, A., George, N., Gil, L., Gupta, P., Hammond-Kosack, K.E., Haskell, E., Hunt, S.E., Jaiswal, P., Janacek, S.H., Kersey, P.J., Langridge, N., Maheswari, U., Maurel, T., McDowall, M.D., Moore, B., Muffato, M., Naamati, G., Naithani, S., Olson, A., Papatheodorou, I., Patricio, M., Paulini, M., Pedro, H., Perry, E., Preece, J., Rosello, M., Russell, M., Sitnik, V., Staines, D.M., Stein, J., Tello-Ruiz,

- M.K., Trevanion, S.J., Urban, M., Wei, S., Ware, D., Williams, G., Yates, A.D., Flicek, P., 2020. Ensembl Genomes 2020—enabling non-vertebrate genomic research. *Nucleic Acids Res.* 48, D689–D695. <https://doi.org/10.1093/nar/gkz890>
- Hsieh, L.-C., Lin, S.-I., Shih, A.C.-C., Chen, J.-W., Lin, W.-Y., Tseng, C.-Y., Li, W.-H., Chiou, T.-J., 2009. Uncovering Small RNA-Mediated Responses to Phosphate Deficiency in Arabidopsis by Deep Sequencing. *Plant Physiol.* 151, 2120–2132. <https://doi.org/10.1104/pp.109.147280>
- Huang, Y., Ji, L., Huang, Q., Vassilyev, D.G., Chen, X., Ma, J.-B., 2009. Structural insights into mechanisms of the small RNA methyltransferase HEN1. *Nature* 461, 823–827. <https://doi.org/10.1038/nature08433>
- Hug, N., Longman, D., Cáceres, J.F., 2016. Mechanism and regulation of the nonsense-mediated decay pathway. *Nucleic Acids Res.* 44, 1483–1495. <https://doi.org/10.1093/nar/gkw010>
- Hunter, C., Willmann, M.R., Wu, G., Yoshikawa, M., de la Luz Gutiérrez-Nava, M., Poethig, S.R., 2006. Trans-acting siRNA-mediated repression of ETTIN and ARF4 regulates heteroblasty in Arabidopsis. *Dev. Camb. Engl.* 133, 2973–2981. <https://doi.org/10.1242/dev.02491>
- Hutvagner, G., Simard, M.J., 2008. Argonaute proteins: key players in RNA silencing. *Nat. Rev. Mol. Cell Biol.* 9, 22–32. <https://doi.org/10.1038/nrm2321>
- Ibrahim, F., Rymarquis, L.A., Kim, E.-J., Becker, J., Balassa, E., Green, P.J., Cerutti, H., 2010. Uridylation of mature miRNAs and siRNAs by the MUT68 nucleotidyltransferase promotes their degradation in Chlamydomonas. *Proc. Natl. Acad. Sci.* 107, 3906–3911. <https://doi.org/10.1073/pnas.0912632107>
- Ito, H., 2013. Small RNAs and regulation of transposons in plants. *Genes Genet. Syst.* 88, 3–7. <https://doi.org/10.1266/ggs.88.3>
- Jacobsen, S.E., Running, M.P., Meyerowitz, E.M., 1999. Disruption of an RNA helicase/RNase III gene in Arabidopsis causes unregulated cell division in floral meristems. *Development* 126, 5231–5243.
- Jander, G., Norris, S.R., Rounsley, S.D., Bush, D.F., Levin, I.M., Last, R.L., 2002. Arabidopsis Map-Based Cloning in the Post-Genome Era. *Plant Physiol.* 129, 440–450.
- Jauvion, V., Rivard, M., Bouteiller, N., Elmayan, T., Vaucheret, H., 2012. RDR2 Partially Antagonizes the Production of RDR6-Dependent siRNA in Sense Transgene-Mediated PTGS. *PLOS ONE* 7, e29785. <https://doi.org/10.1371/journal.pone.0029785>
- Ji, L., Chen, X., 2012. Regulation of small RNA stability: methylation and beyond. *Cell Res.* 22, 624–636. <https://doi.org/10.1038/cr.2012.36>
- Ji, L., Liu, X., Yan, J., Wang, W., Yumul, R.E., Kim, Y.J., Dinh, T.T., Liu, J., Cui, X., Zheng, B., Agarwal, M., Liu, C., Cao, X., Tang, G., Chen, X., 2011. ARGONAUTE10 and ARGONAUTE1 Regulate the Termination of Floral Stem Cells through Two MicroRNAs in Arabidopsis. *PLOS Genet.* 7, e1001358. <https://doi.org/10.1371/journal.pgen.1001358>
- Jia, F., Rock, C.D., 2013a. MIR846 and MIR842 comprise a cistronic MIRNA pair that is regulated by abscisic acid by alternative splicing in roots of Arabidopsis. *Plant Mol. Biol.* 81, 447–460. <https://doi.org/10.1007/s11103-013-0015-6>
- Jia, F., Rock, C.D., 2013b. MIR846 and MIR842 comprise a cistronic MIRNA pair that is regulated by abscisic acid by alternative splicing in roots of Arabidopsis. *Plant Mol. Biol.* 81, 447–460. <https://doi.org/10.1007/s11103-013-0015-6>
- Jia, T., Zhang, B., You, C., Zhang, Y., Zeng, L., Li, S., Johnson, K.C.M., Yu, B., Li, X., Chen, X., 2017a. The Arabidopsis MOS4-Associated Complex Promotes MicroRNA Biogenesis and Precursor Messenger RNA Splicing. *Plant Cell* 29, 2626–2643. <https://doi.org/10.1105/tpc.17.00370>
- Jia, T., Zhang, B., You, C., Zhang, Y., Zeng, L., Li, S., Johnson, K.C.M., Yu, B., Li, X., Chen, X., 2017b. The Arabidopsis MOS4-Associated Complex Promotes MicroRNA Biogenesis and Precursor Messenger RNA Splicing. *Plant Cell* 29, 2626–2643. <https://doi.org/10.1105/tpc.17.00370>
- Joly-Lopez, Z., Forczek, E., Vello, E., Hoen, D.R., Tomita, A., Bureau, T.E., 2017. Abiotic Stress Phenotypes Are Associated with Conserved Genes Derived from Transposable Elements. *Front. Plant Sci.* 8. <https://doi.org/10.3389/fpls.2017.02027>

- Jones-Rhoades, M.W., Bartel, D.P., 2004. Computational identification of plant microRNAs and their targets, including a stress-induced miRNA. *Mol. Cell* 14, 787–799. <https://doi.org/10.1016/j.molcel.2004.05.027>
- Kahles, A., Ong, C.S., Zhong, Y., Ratsch, G., 2016a. SplAdder: identification, quantification and testing of alternative splicing events from RNA-Seq data. *Bioinformatics* 32, 1840–1847. <https://doi.org/10.1093/bioinformatics/btw076>
- Kahles, A., Ong, C.S., Zhong, Y., Ratsch, G., 2016b. SplAdder: identification, quantification and testing of alternative splicing events from RNA-Seq data. *Bioinformatics* 32, 1840–1847. <https://doi.org/10.1093/bioinformatics/btw076>
- Kalyna, M., Simpson, C.G., Syed, N.H., Lewandowska, D., Marquez, Y., Kusenda, B., Marshall, J., Fuller, J., Cardle, L., McNicol, J., Dinh, H.Q., Barta, A., Brown, J.W.S., 2012. Alternative splicing and nonsense-mediated decay modulate expression of important regulatory genes in Arabidopsis. *Nucleic Acids Res.* 40, 2454–2469. <https://doi.org/10.1093/nar/gkr932>
- Kashima, I., Yamashita, A., Izumi, N., Kataoka, N., Morishita, R., Hoshino, S., Ohno, M., Dreyfuss, G., Ohno, S., 2006. Binding of a novel SMG-1–Upf1–eRF1–eRF3 complex (SURF) to the exon junction complex triggers Upf1 phosphorylation and nonsense-mediated mRNA decay. *Genes Dev.* 20, 355–367. <https://doi.org/10.1101/gad.1389006>
- Kasschau, K.D., Fahlgren, N., Chapman, E.J., Sullivan, C.M., Cumbie, J.S., Givan, S.A., Carrington, J.C., 2007. Genome-wide profiling and analysis of Arabidopsis siRNAs. *PLoS Biol.* 5, e57. <https://doi.org/10.1371/journal.pbio.0050057>
- Kawashima, C.G., Matthewman, C.A., Huang, S., Lee, B.-R., Yoshimoto, N., Koprivova, A., Rubio-Somoza, I., Todesco, M., Rathjen, T., Saito, K., Takahashi, H., Dalmay, T., Kopriva, S., 2011a. Interplay of SLIM1 and miR395 in the regulation of sulfate assimilation in Arabidopsis. *Plant J. Cell Mol. Biol.* 66, 863–876. <https://doi.org/10.1111/j.1365-313X.2011.04547.x>
- Kawashima, C.G., Matthewman, C.A., Huang, S., Lee, B.-R., Yoshimoto, N., Koprivova, A., Rubio-Somoza, I., Todesco, M., Rathjen, T., Saito, K., Takahashi, H., Dalmay, T., Kopriva, S., 2011b. Interplay of SLIM1 and miR395 in the regulation of sulfate assimilation in Arabidopsis. *Plant J.* 66, 863–876. <https://doi.org/10.1111/j.1365-313X.2011.04547.x>
- Kawashima, C.G., Yoshimoto, N., Maruyama-Nakashita, A., Tsuchiya, Y.N., Saito, K., Takahashi, H., Dalmay, T., 2009a. Sulphur starvation induces the expression of microRNA-395 and one of its target genes but in different cell types. *Plant J. Cell Mol. Biol.* 57, 313–321. <https://doi.org/10.1111/j.1365-313X.2008.03690.x>
- Kawashima, C.G., Yoshimoto, N., Maruyama-Nakashita, A., Tsuchiya, Y.N., Saito, K., Takahashi, H., Dalmay, T., 2009b. Sulphur starvation induces the expression of microRNA-395 and one of its target genes but in different cell types. *Plant J. Cell Mol. Biol.* 57, 313–321. <https://doi.org/10.1111/j.1365-313X.2008.03690.x>
- Keller, M., Hu, Y., Mesihovic, A., Fragkostefanakis, S., Schleiff, E., Simm, S., 2017. Alternative splicing in tomato pollen in response to heat stress. *DNA Res. Int. J. Rapid Publ. Rep. Genes Genomes* 24, 205–217. <https://doi.org/10.1093/dnares/dsw051>
- Kim, D., Langmead, B., Salzberg, S.L., 2015. HISAT: a fast spliced aligner with low memory requirements. *Nat. Methods* 12, 357–360. <https://doi.org/10.1038/nmeth.3317>
- Kim, D., Sung, Y.M., Park, Jinman, Kim, Sukjun, Kim, J., Park, Junhee, Ha, H., Bae, J.Y., Kim, SoHui, Baek, D., 2016. General rules for functional microRNA targeting. *Nat. Genet.* 48, 1517–1526. <https://doi.org/10.1038/ng.3694>
- Kim, D.H., Rossi, J.J., 2009. Transcriptional gene silencing using small RNAs. *Methods Mol. Biol. Clifton NJ* 555, 119–125. [https://doi.org/10.1007/978-1-60327-295-7\\_9](https://doi.org/10.1007/978-1-60327-295-7_9)
- Kim, Y., Schumaker, K.S., Zhu, J.-K., 2006. EMS mutagenesis of Arabidopsis. *Methods Mol. Biol. Clifton NJ* 323, 101–103. <https://doi.org/10.1385/1-59745-003-0:101>
- Kloc, A., Martienssen, R., 2008. RNAi, heterochromatin and the cell cycle. *Trends Genet. TIG* 24, 511–517. <https://doi.org/10.1016/j.tig.2008.08.002>

- Koiwa, H., Barb, A.W., Xiong, L., Li, F., McCully, M.G., Lee, B., Sokolchik, I., Zhu, J., Gong, Z., Reddy, M., Sharkhuu, A., Manabe, Y., Yokoi, S., Zhu, J.-K., Bressan, R.A., Hasegawa, P.M., 2002a. C-terminal domain phosphatase-like family members (AtCPLs) differentially regulate *Arabidopsis thaliana* abiotic stress signaling, growth, and development. *Proc. Natl. Acad. Sci. U. S. A.* 99, 10893–10898. <https://doi.org/10.1073/pnas.112276199>
- Koiwa, H., Barb, A.W., Xiong, L., Li, F., McCully, M.G., Lee, B., Sokolchik, I., Zhu, J., Gong, Z., Reddy, M., Sharkhuu, A., Manabe, Y., Yokoi, S., Zhu, J.-K., Bressan, R.A., Hasegawa, P.M., 2002b. C-terminal domain phosphatase-like family members (AtCPLs) differentially regulate *Arabidopsis thaliana* abiotic stress signaling, growth, and development. *Proc. Natl. Acad. Sci.* 99, 10893–10898. <https://doi.org/10.1073/pnas.112276199>
- Kopriva, S., 2006. Regulation of Sulfate Assimilation in *Arabidopsis* and Beyond. *Ann. Bot.* 97, 479–495. <https://doi.org/10.1093/aob/mcl006>
- Krol, J., Busskamp, V., Markiewicz, I., Stadler, M.B., Ribi, S., Richter, J., Duebel, J., Bicker, S., Fehling, H.J., Schübeler, D., Oertner, T.G., Schrott, G., Bibel, M., Roska, B., Filipowicz, W., 2010. Characterizing light-regulated retinal microRNAs reveals rapid turnover as a common property of neuronal microRNAs. *Cell* 141, 618–631. <https://doi.org/10.1016/j.cell.2010.03.039>
- Krueger, F., 2020. FelixKrueger/TrimGalore.
- Ku, H.-Y., Lin, H., 2014. PIWI proteins and their interactors in piRNA biogenesis, germline development and gene expression. *Natl. Sci. Rev.* 1, 205–218. <https://doi.org/10.1093/nsr/nwu014>
- Kurihara, Y., Watanabe, Y., 2004a. *Arabidopsis* micro-RNA biogenesis through Dicer-like 1 protein functions. *Proc. Natl. Acad. Sci. U. S. A.* 101, 12753–12758. <https://doi.org/10.1073/pnas.0403115101>
- Kurihara, Y., Watanabe, Y., 2004b. *Arabidopsis* micro-RNA biogenesis through Dicer-like 1 protein functions. *Proc. Natl. Acad. Sci.* 101, 12753–12758. <https://doi.org/10.1073/pnas.0403115101>
- Lagos-Quintana, M., Rauhut, R., Lendeckel, W., Tuschl, T., 2001. Identification of Novel Genes Coding for Small Expressed RNAs. *Science* 294, 853–858. <https://doi.org/10.1126/science.1064921>
- Laloum, T., Martín, G., Duque, P., 2018. Alternative Splicing Control of Abiotic Stress Responses. *Trends Plant Sci.* 23, 140–150. <https://doi.org/10.1016/j.tplants.2017.09.019>
- Lang-Mladek, C., Popova, O., Kiok, K., Berlinger, M., Rakic, B., Aufsatz, W., Jonak, C., Hauser, M.-T., Luschnig, C., 2010. Transgenerational Inheritance and Resetting of Stress-Induced Loss of Epigenetic Gene Silencing in *Arabidopsis*. *Mol. Plant* 3, 594–602. <https://doi.org/10.1093/mp/ssq014>
- Lau, N.C., Lim, L.P., Weinstein, E.G., Bartel, D.P., 2001. An abundant class of tiny RNAs with probable regulatory roles in *Caenorhabditis elegans*. *Science* 294, 858–862. <https://doi.org/10.1126/science.1065062>
- Law, J.A., Du, J., Hale, C.J., Feng, S., Krajewski, K., Palanca, A.M.S., Strahl, B.D., Patel, D.J., Jacobsen, S.E., 2013. Polymerase IV occupancy at RNA-directed DNA methylation sites requires SHH1. *Nature* 498, 385–389. <https://doi.org/10.1038/nature12178>
- Lee, R.C., Ambros, V., 2001. An extensive class of small RNAs in *Caenorhabditis elegans*. *Science* 294, 862–864. <https://doi.org/10.1126/science.1065329>
- Lee, R.C., Feinbaum, R.L., Ambros, V., 1993. The *C. elegans* heterochronic gene *lin-4* encodes small RNAs with antisense complementarity to *lin-14*. *Cell* 75, 843–854. [https://doi.org/10.1016/0092-8674\(93\)90529-Y](https://doi.org/10.1016/0092-8674(93)90529-Y)
- Lee, Y., Ahn, C., Han, J., Choi, H., Kim, J., Yim, J., Lee, J., Provost, P., Rådmark, O., Kim, S., Kim, V.N., 2003. The nuclear RNase III Drosha initiates microRNA processing. *Nature* 425, 415–419. <https://doi.org/10.1038/nature01957>

- Lee, Y., Kim, M., Han, J., Yeom, K.-H., Lee, S., Baek, S.H., Kim, V.N., 2004a. MicroRNA genes are transcribed by RNA polymerase II. *EMBO J.* 23, 4051–4060. <https://doi.org/10.1038/sj.emboj.7600385>
- Lee, Y., Kim, M., Han, J., Yeom, K.-H., Lee, S., Baek, S.H., Kim, V.N., 2004b. MicroRNA genes are transcribed by RNA polymerase II. *EMBO J.* 23, 4051–4060. <https://doi.org/10.1038/sj.emboj.7600385>
- Lee, Y., Kim, M., Han, J., Yeom, K.-H., Lee, S., Baek, S.H., Kim, V.N., 2004c. MicroRNA genes are transcribed by RNA polymerase II. *EMBO J.* 23, 4051–4060. <https://doi.org/10.1038/sj.emboj.7600385>
- Lewis, B.P., Shih, I., Jones-Rhoades, M.W., Bartel, D.P., Burge, C.B., 2003. Prediction of Mammalian MicroRNA Targets. *Cell* 115, 787–798. [https://doi.org/10.1016/S0092-8674\(03\)01018-3](https://doi.org/10.1016/S0092-8674(03)01018-3)
- Lewsey, M.G., Hardcastle, T.J., Melnyk, C.W., Molnar, A., Valli, A., Urich, M.A., Nery, J.R., Baulcombe, D.C., Ecker, J.R., 2016. Mobile small RNAs regulate genome-wide DNA methylation. *Proc. Natl. Acad. Sci.* 113, E801–E810. <https://doi.org/10.1073/pnas.1515072113>
- Li, F., Cheng, C., Cui, F., de Oliveira, M.V.V., Yu, X., Meng, X., Intorne, A.C., Babilonia, K., Li, M., Li, B., Chen, S., Ma, X., Xiao, S., Zeng, Y., Fei, Z., Metz, R., Johnson, C.D., Koiwa, H., Sun, W., Li, Z., de Souza Filho, G.A., Shan, L., He, P., 2014. Modulation of RNA polymerase II phosphorylation downstream of pathogen perception orchestrates plant immunity. *Cell Host Microbe* 16, 748–758. <https://doi.org/10.1016/j.chom.2014.10.018>
- Li, J., Yang, Z., Yu, B., Liu, J., Chen, X., 2005. Methylation protects miRNAs and siRNAs from a 3'-end uridylation activity in Arabidopsis. *Curr. Biol.* 15, 1501–1507. <https://doi.org/10.1016/j.cub.2005.07.029>
- Li, S., Xu, R., Li, A., Liu, K., Gu, L., Li, M., Zhang, H., Zhang, Y., Zhuang, S., Wang, Q., Gao, G., Li, N., Zhang, C., Li, Y., Yu, B., 2018. SMA1, a homolog of the splicing factor Prp28, has a multifaceted role in miRNA biogenesis in Arabidopsis. *Nucleic Acids Res.* 46, 9148–9159. <https://doi.org/10.1093/nar/gky591>
- Li, T., Natran, A., Chen, Y., Vercruyse, J., Wang, K., Gonzalez, N., Dubois, M., Inzé, D., 2019a. A genetics screen highlights emerging roles for CPL3, RST1 and URT1 in RNA metabolism and silencing. *Nat. Plants* 5, 539–550. <https://doi.org/10.1038/s41477-019-0419-7>
- Li, T., Natran, A., Chen, Y., Vercruyse, J., Wang, K., Gonzalez, N., Dubois, M., Inzé, D., 2019b. A genetics screen highlights emerging roles for CPL3, RST1 and URT1 in RNA metabolism and silencing. *Nat. Plants* 5, 539–550. <https://doi.org/10.1038/s41477-019-0419-7>
- Li, T., Natran, A., Chen, Y., Vercruyse, J., Wang, K., Gonzalez, N., Dubois, M., Inzé, D., 2019c. A genetics screen highlights emerging roles for CPL3, RST1 and URT1 in RNA metabolism and silencing. *Nat. Plants* 5, 539–550. <https://doi.org/10.1038/s41477-019-0419-7>
- Li, Y., Tollefsbol, T.O., 2011. DNA methylation detection: Bisulfite genomic sequencing analysis. *Methods Mol. Biol.* Clifton NJ 791, 11–21. [https://doi.org/10.1007/978-1-61779-316-5\\_2](https://doi.org/10.1007/978-1-61779-316-5_2)
- Lin, S.-I., Chiang, S.-F., Lin, W.-Y., Chen, J.-W., Tseng, C.-Y., Wu, P.-C., Chiou, T.-J., 2008. Regulatory Network of MicroRNA399 and PHO2 by Systemic Signaling. *Plant Physiol.* 147, 732–746. <https://doi.org/10.1104/pp.108.116269>
- Lisch, D., 2013. How important are transposons for plant evolution? *Nat. Rev. Genet.* 14, 49–61. <https://doi.org/10.1038/nrg3374>
- Liu, Jinjie, Sun, N., Liu, M., Liu, Jiancheng, Du, B., Wang, X., Qi, X., 2013. An Autoregulatory Loop Controlling Arabidopsis HsfA2 Expression: Role of Heat Shock-Induced Alternative Splicing. *Plant Physiol.* 162, 512–521. <https://doi.org/10.1104/pp.112.205864>
- Liu, Q., Wang, F., Axtell, M.J., 2014. Analysis of Complementarity Requirements for Plant MicroRNA Targeting Using a *Nicotiana benthamiana* Quantitative Transient Assay[W][OPEN]. *Plant Cell* 26, 741–753. <https://doi.org/10.1105/tpc.113.120972>

- Llave, C., Xie, Z., Kasschau, K.D., Carrington, J.C., 2002. Cleavage of Scarecrow-like mRNA Targets Directed by a Class of Arabidopsis miRNA. *Science* 297, 2053–2056. <https://doi.org/10.1126/science.1076311>
- Lopez, J.P., Diallo, A., Cruceanu, C., Fiori, L.M., Laboissiere, S., Guillet, I., Fontaine, J., Ragoussis, J., Benes, V., Turecki, G., Ernst, C., 2015. Biomarker discovery: quantification of microRNAs and other small non-coding RNAs using next generation sequencing. *BMC Med. Genomics* 8, 35. <https://doi.org/10.1186/s12920-015-0109-x>
- Lorenz, R., Bernhart, S.H., Höner zu Siederdisen, C., Tafer, H., Flamm, C., Stadler, P.F., Hofacker, I.L., 2011. ViennaRNA Package 2.0. *Algorithms Mol. Biol.* 6, 26. <https://doi.org/10.1186/1748-7188-6-26>
- Love, M.I., Huber, W., Anders, S., 2014. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* 15. <https://doi.org/10.1186/s13059-014-0550-8>
- Lu, S., Sun, Y.-H., Chiang, V.L., 2009. Adenylation of plant miRNAs. *Nucleic Acids Res.* 37, 1878–1885. <https://doi.org/10.1093/nar/gkp031>
- Lund, E., Güttinger, S., Calado, A., Dahlberg, J.E., Kutay, U., 2004. Nuclear Export of MicroRNA Precursors. *Science* 303, 95–98. <https://doi.org/10.1126/science.1090599>
- Mallory, A.C., Bartel, D.P., Bartel, B., 2005. MicroRNA-Directed Regulation of Arabidopsis AUXIN RESPONSE FACTOR17 Is Essential for Proper Development and Modulates Expression of Early Auxin Response Genes. *Plant Cell* 17, 1360–1375. <https://doi.org/10.1105/tpc.105.031716>
- Manavella, P.A., Hagmann, J., Ott, F., Laubinger, S., Franz, M., Macek, B., Weigel, D., 2012. Fast-Forward Genetics Identifies Plant CPL Phosphatases as Regulators of miRNA Processing Factor HYL1. *Cell* 151, 859–870. <https://doi.org/10.1016/j.cell.2012.09.039>
- Marquez, Y., Brown, J.W.S., Simpson, C., Barta, A., Kalyna, M., 2012. Transcriptome survey reveals increased complexity of the alternative splicing landscape in Arabidopsis. *Genome Res.* 22, 1184–1195. <https://doi.org/10.1101/gr.134106.111>
- Martínez, G., Slotkin, R.K., 2012. Developmental relaxation of transposable element silencing in plants: functional or byproduct? *Curr. Opin. Plant Biol.* 15, 496–502. <https://doi.org/10.1016/j.pbi.2012.09.001>
- Martínez de Alba, A.E., Moreno, A.B., Gabriel, M., Mallory, A.C., Christ, A., Bounon, R., Balzergue, S., Aubourg, S., Gautheret, D., Crespi, M.D., Vaucheret, H., Maizel, A., 2015a. In plants, decapping prevents RDR6-dependent production of small interfering RNAs from endogenous mRNAs. *Nucleic Acids Res.* 43, 2902–2913. <https://doi.org/10.1093/nar/gkv119>
- Martínez de Alba, A.E., Moreno, A.B., Gabriel, M., Mallory, A.C., Christ, A., Bounon, R., Balzergue, S., Aubourg, S., Gautheret, D., Crespi, M.D., Vaucheret, H., Maizel, A., 2015b. In plants, decapping prevents RDR6-dependent production of small interfering RNAs from endogenous mRNAs. *Nucleic Acids Res.* 43, 2902–2913. <https://doi.org/10.1093/nar/gkv119>
- Maruyama-Nakashita, A., Nakamura, Y., Tohge, T., Saito, K., Takahashi, H., 2006. Arabidopsis SLIM1 Is a Central Transcriptional Regulator of Plant Sulfur Response and Metabolism. *Plant Cell* 18, 3235–3251. <https://doi.org/10.1105/tpc.106.046458>
- Megraw, M., Baev, V., Rusinov, V., Jensen, S.T., Kalantidis, K., Hatzigeorgiou, A.G., 2006. MicroRNA promoter element discovery in Arabidopsis. *RNA* 12, 1612–1619. <https://doi.org/10.1261/rna.130506>
- Meinke, D.W., 2020. Genome-wide identification of EMBRYO-DEFECTIVE (EMB) genes required for growth and development in Arabidopsis. *New Phytol.* 226, 306–325. <https://doi.org/10.1111/nph.16071>
- Mi, S., Cai, T., Hu, Y., Chen, Y., Hodges, E., Ni, F., Wu, L., Li, S., Zhou, H., Long, C., Chen, S., Hannon, G.J., Qi, Y., 2008. Sorting of Small RNAs into Arabidopsis Argonaute Complexes Is Directed by the 5' Terminal Nucleotide. *Cell* 133, 116–127. <https://doi.org/10.1016/j.cell.2008.02.034>

- Millar, A.A., Waterhouse, P.M., 2005. Plant and animal microRNAs: similarities and differences. *Funct. Integr. Genomics* 5, 129–135. <https://doi.org/10.1007/s10142-005-0145-2>
- Mininno, M., Brugière, S., Pautre, V., Gilgen, A., Ma, S., Ferro, M., Tardif, M., Alban, C., Ravel, S., 2012. Characterization of Chloroplastic Fructose 1,6-Bisphosphate Aldolases as Lysine-methylated Proteins in Plants. *J. Biol. Chem.* 287, 21034–21044. <https://doi.org/10.1074/jbc.M112.359976>
- Morel, J.-B., Godon, C., Mourrain, P., Béclin, C., Boutet, S., Feuerbach, F., Proux, F., Vaucheret, H., 2002. Fertile hypomorphic ARGONAUTE (ago1) mutants impaired in post-transcriptional gene silencing and virus resistance. *Plant Cell* 14, 629–639. <https://doi.org/10.1105/tpc.010358>
- Moreno, A.B., Martínez de Alba, A.E., Bardou, F., Crespi, M.D., Vaucheret, H., Maizel, A., Mallory, A.C., 2013. Cytoplasmic and nuclear quality control and turnover of single-stranded RNA modulate post-transcriptional gene silencing in plants. *Nucleic Acids Res.* 41, 4699–4708. <https://doi.org/10.1093/nar/gkt152>
- Nagarajan, V.K., Jones, C.I., Newbury, S.F., Green, P.J., 2013. XRN 5'→3' exoribonucleases: Structure, mechanisms and functions. *Biochim. Biophys. Acta* 1829, 590–603. <https://doi.org/10.1016/j.bbagr.2013.03.005>
- Neph, S., Kuehn, M.S., Reynolds, A.P., Haugen, E., Thurman, R.E., Johnson, A.K., Rynes, E., Maurano, M.T., Vierstra, J., Thomas, S., Sandstrom, R., Humbert, R., Stamatoyannopoulos, J.A., 2012. BEDOPS: high-performance genomic feature operations. *Bioinformatics* 28, 1919–1920. <https://doi.org/10.1093/bioinformatics/bts277>
- Ner-Gaon, H., Halachmi, R., Savaldi-Goldstein, S., Rubin, E., Ophir, R., Fluhr, R., 2004. Intron retention is a major phenomenon in alternative splicing in Arabidopsis. *Plant J.* 39, 877–885. <https://doi.org/10.1111/j.1365-3113.2004.02172.x>
- Nikiforova, V., Freitag, J., Kempa, S., Adamik, M., Hesse, H., Hoefgen, R., 2003. Transcriptome analysis of sulfur depletion in Arabidopsis thaliana: interlacing of biosynthetic pathways provides response specificity. *Plant J.* 33, 633–650. <https://doi.org/10.1046/j.1365-3113.2003.01657.x>
- Nwachukwu, I.D., Slusarenko, A.J., Gruhlke, M.C.H., 2012. Sulfur and sulfur compounds in plant defence. *Nat. Prod. Commun.* 7, 395–400.
- Okamura, K., Ishizuka, A., Siomi, H., Siomi, M.C., 2004. Distinct roles for Argonaute proteins in small RNA-directed RNA cleavage pathways. *Genes Dev.* 18, 1655–1666. <https://doi.org/10.1101/gad.1210204>
- Okamura, K., Liu, N., Lai, E.C., 2009. Distinct mechanisms for microRNA strand selection by Drosophila Argonautes. *Mol. Cell* 36, 431–444. <https://doi.org/10.1016/j.molcel.2009.09.027>
- Oliver, C., Pradillo, M., Jover-Gil, S., Cuñado, N., Ponce, M.R., Santos, J.L., 2017. Loss of function of Arabidopsis microRNA-machinery genes impairs fertility, and has effects on homologous recombination and meiotic chromatin dynamics. *Sci. Rep.* 7, 1–14. <https://doi.org/10.1038/s41598-017-07702-x>
- Onodera, Y., Haag, J.R., Ream, T., Costa Nunes, P., Pontes, O., Pikaard, C.S., 2005. Plant nuclear RNA polymerase IV mediates siRNA and DNA methylation-dependent heterochromatin formation. *Cell* 120, 613–622. <https://doi.org/10.1016/j.cell.2005.02.007>
- Orphanides, G., Reinberg, D., 2002. A Unified Theory of Gene Expression. *Cell* 108, 439–451. [https://doi.org/10.1016/S0092-8674\(02\)00655-4](https://doi.org/10.1016/S0092-8674(02)00655-4)
- Page, D.R., Grossniklaus, U., 2002. The art and design of genetic screens: Arabidopsis thaliana. *Nat. Rev. Genet.* 3, 124–136. <https://doi.org/10.1038/nrg730>
- Papp, I., Mette, M.F., Aufsatz, W., Daxinger, L., Schauer, S.E., Ray, A., Winden, J. van der, Matzke, M., Matzke, A.J.M., 2003. Evidence for Nuclear Processing of Plant Micro RNA and Short Interfering RNA Precursors. *Plant Physiol.* 132, 1382–1390. <https://doi.org/10.1104/pp.103.021980>

- Parent, J.-S., Bouteiller, N., Elmayan, T., Vaucheret, H., 2015. Respective contributions of Arabidopsis DCL2 and DCL4 to RNA silencing. *Plant J.* 81, 223–232. <https://doi.org/10.1111/tpj.12720>
- Parizotto, E.A., Dunoyer, P., Rahm, N., Himber, C., Voinnet, O., 2004. In vivo investigation of the transcription, processing, endonucleolytic activity, and functional relevance of the spatial distribution of a plant miRNA. *Genes Dev.* 18, 2237–2242. <https://doi.org/10.1101/gad.307804>
- Park, W., Li, J., Song, R., Messing, J., Chen, X., 2002. CARPEL FACTORY, a Dicer Homolog, and HEN1, a Novel Protein, Act in microRNA Metabolism in Arabidopsis thaliana. *Curr. Biol.* 12, 1484–1495. [https://doi.org/10.1016/S0960-9822\(02\)01017-5](https://doi.org/10.1016/S0960-9822(02)01017-5)
- Patel, P., Ramachandruni, S.D., Kakrana, A., Nakano, M., Meyers, B.C., 2016a. miTRATA: a web-based tool for microRNA Truncation and Tailing Analysis. *Bioinformatics* 32, 450–452. <https://doi.org/10.1093/bioinformatics/btv583>
- Patel, P., Ramachandruni, S.D., Kakrana, A., Nakano, M., Meyers, B.C., 2016b. miTRATA: a web-based tool for microRNA Truncation and Tailing Analysis. *Bioinformatics* 32, 450–452. <https://doi.org/10.1093/bioinformatics/btv583>
- Patro, R., Duggal, G., Love, M.I., Irizarry, R.A., Kingsford, C., 2017a. Salmon: fast and bias-aware quantification of transcript expression using dual-phase inference. *Nat. Methods* 14, 417–419. <https://doi.org/10.1038/nmeth.4197>
- Patro, R., Duggal, G., Love, M.I., Irizarry, R.A., Kingsford, C., 2017b. Salmon: fast and bias-aware quantification of transcript expression using dual-phase inference. *Nat. Methods* 14, 417–419. <https://doi.org/10.1038/nmeth.4197>
- Peragine, A., Yoshikawa, M., Wu, G., Albrecht, H.L., Poethig, R.S., 2004. SGS3 and SGS2/SDE1/RDR6 are required for juvenile development and the production of trans-acting siRNAs in Arabidopsis. *Genes Dev.* 18, 2368–2379. <https://doi.org/10.1101/gad.1231804>
- Pikaard, C.S., 2006. Cell biology of the Arabidopsis nuclear siRNA pathway for RNA-directed chromatin modification. *Cold Spring Harb. Symp. Quant. Biol.* 71, 473–480. <https://doi.org/10.1101/sqb.2006.71.046>
- Pontes, O., Li, C.F., Nunes, P.C., Haag, J., Ream, T., Vitins, A., Jacobsen, S.E., Pikaard, C.S., 2006. The Arabidopsis Chromatin-Modifying Nuclear siRNA Pathway Involves a Nucleolar RNA Processing Center. *Cell* 126, 79–92. <https://doi.org/10.1016/j.cell.2006.05.031>
- Posé, D., Verhage, L., Ott, F., Yant, L., Mathieu, J., Angenent, G.C., Immink, R.G.H., Schmid, M., 2013. Temperature-dependent regulation of flowering by antagonistic FLM variants. *Nature* 503, 414–417. <https://doi.org/10.1038/nature12633>
- Pratt, A.J., MacRae, I.J., 2009. The RNA-induced Silencing Complex: A Versatile Gene-silencing Machine. *J. Biol. Chem.* 284, 17897–17901. <https://doi.org/10.1074/jbc.R900012200>
- Punzo, P., Ruggiero, A., Possenti, M., Perrella, G., Nurcato, R., Costa, A., Morelli, G., Grillo, S., Batelli, G., 2020. DRT111/SFPS splicing factor controls ABA sensitivity during seed development and germination. *Plant Physiol.* <https://doi.org/10.1104/pp.20.00037>
- Qi, Y., He, X., Wang, X.-J., Kohany, O., Jurka, J., Hannon, G.J., 2006. Distinct catalytic and non-catalytic roles of ARGONAUTE4 in RNA-directed DNA methylation. *Nature* 443, 1008–1012. <https://doi.org/10.1038/nature05198>
- Quast, C., Pruesse, E., Yilmaz, P., Gerken, J., Schweer, T., Yarza, P., Peplies, J., Glöckner, F.O., 2013. The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. *Nucleic Acids Res.* 41, D590–D596. <https://doi.org/10.1093/nar/gks1219>
- Quinlan, A.R., Hall, I.M., 2010. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* 26, 841–842. <https://doi.org/10.1093/bioinformatics/btq033>
- R Core Team, 2018. R: A Language and Environment for Statistical Computing.
- Ramachandran, V., Chen, X., 2008a. Degradation of microRNAs by a family of exoribonucleases in Arabidopsis. *Science* 321, 1490–1492. <https://doi.org/10.1126/science.1163728>

- Ramachandran, V., Chen, X., 2008b. Degradation of microRNAs by a Family of Exoribonucleases in Arabidopsis. *Science* 321, 1490–1492. <https://doi.org/10.1126/science.1163728>
- Reddy, A.S.N., Marquez, Y., Kalyna, M., Barta, A., 2013. Complexity of the Alternative Splicing Landscape in Plants. *Plant Cell* 25, 3657–3683. <https://doi.org/10.1105/tpc.113.117523>
- Rédei, G.P. (Ed.), 2008. Murashige & Skoog Medium (MS1), in: *Encyclopedia of Genetics, Genomics, Proteomics and Informatics*. Springer Netherlands, Dordrecht, pp. 1290–1290. [https://doi.org/10.1007/978-1-4020-6754-9\\_10971](https://doi.org/10.1007/978-1-4020-6754-9_10971)
- Reeves, W.M., Lynch, T.J., Mobin, R., Finkelstein, R.R., 2011. Direct targets of the transcription factors ABA-Insensitive(ABI)4 and ABI5 reveal synergistic action by ABI4 and several bZIP ABA response factors. *Plant Mol. Biol.* 75, 347–363. <https://doi.org/10.1007/s11103-011-9733-9>
- Reinhart, B.J., Weinstein, E.G., Rhoades, M.W., Bartel, B., Bartel, D.P., 2002. MicroRNAs in plants. *Genes Dev.* 16, 1616–1626. <https://doi.org/10.1101/gad.1004402>
- Ren, G., Chen, X., Yu, B., 2012a. Uridylation of miRNAs by hen1 suppressor1 in Arabidopsis. *Curr. Biol.* CB 22, 695–700. <https://doi.org/10.1016/j.cub.2012.02.052>
- Ren, G., Xie, M., Dou, Y., Zhang, S., Zhang, C., Yu, B., 2012b. Regulation of miRNA abundance by RNA binding protein TOUGH in Arabidopsis. *Proc. Natl. Acad. Sci. U. S. A.* 109, 12817–12821. <https://doi.org/10.1073/pnas.1204915109>
- Rissland, O.S., Hong, S.-J., Bartel, D.P., 2011. MicroRNA destabilization enables dynamic regulation of the miR-16 family in response to cell-cycle changes. *Mol. Cell* 43, 993–1004. <https://doi.org/10.1016/j.molcel.2011.08.021>
- Rüegger, S., Großhans, H., 2012. MicroRNA turnover: when, how, and why. *Trends Biochem. Sci.* 37, 436–446. <https://doi.org/10.1016/j.tibs.2012.07.002>
- Rutschow, H., Ytterberg, A.J., Friso, G., Nilsson, R., van Wijk, K.J., 2008. Quantitative proteomics of a chloroplast SRP54 sorting mutant and its genetic interactions with CLPC1 in Arabidopsis. *Plant Physiol.* 148, 156–175. <https://doi.org/10.1104/pp.108.124545>
- Samad, A.F.A., Sajad, M., Nazaruddin, N., Fauzi, I.A., Murad, A.M.A., Zainal, Z., Ismail, I., 2017. MicroRNA and Transcription Factor: Key Players in Plant Regulatory Network. *Front. Plant Sci.* 8. <https://doi.org/10.3389/fpls.2017.00565>
- Schauer, S.E., Jacobsen, S.E., Meinke, D.W., Ray, A., 2002. DICER-LIKE1: blind men and elephants in Arabidopsis development. *Trends Plant Sci.* 7, 487–491. [https://doi.org/10.1016/S1360-1385\(02\)02355-5](https://doi.org/10.1016/S1360-1385(02)02355-5)
- Schindelin, J., Arganda-Carreras, I., Frise, E., Kaynig, V., Longair, M., Pietzsch, T., Preibisch, S., Rueden, C., Saalfeld, S., Schmid, B., Tinevez, J.-Y., White, D.J., Hartenstein, V., Eliceiri, K., Tomancak, P., Cardona, A., 2012. Fiji: an open-source platform for biological-image analysis. *Nat. Methods* 9, 676–682. <https://doi.org/10.1038/nmeth.2019>
- Schopfer, P., Bajracharya, D., Plachy, C., 1979. Control of Seed Germination by Abscisic Acid. *Plant Physiol.* 64, 822–827.
- Sega, G.A., 1984. A review of the genetic effects of ethyl methanesulfonate. *Mutat. Res. Genet. Toxicol.* 134, 113–142. [https://doi.org/10.1016/0165-1110\(84\)90007-1](https://doi.org/10.1016/0165-1110(84)90007-1)
- Shang, X., Cao, Y., Ma, L., 2017. Alternative Splicing in Plant Genes: A Means of Regulating the Environmental Fitness of Plants. *Int. J. Mol. Sci.* 18. <https://doi.org/10.3390/ijms18020432>
- Sharp, R.E., LeNoble, M.E., 2002. ABA, ethylene and the control of shoot and root growth under water stress. *J. Exp. Bot.* 53, 33–37. <https://doi.org/10.1093/jexbot/53.366.33>
- Sheu-Gruttadauria, J., Pawlica, P., Klum, S.M., Wang, S., Yario, T.A., Schirle Oakdale, N.T., Steitz, J.A., MacRae, I.J., 2019. Structural Basis for Target-Directed MicroRNA Degradation. *Mol. Cell* 75, 1243–1255.e7. <https://doi.org/10.1016/j.molcel.2019.06.019>
- Sibley, C.R., Blazquez, L., Ule, J., 2016. Lessons from non-canonical splicing. *Nat. Rev. Genet.* 17, 407–421. <https://doi.org/10.1038/nrg.2016.46>
- Sun, Y.-H., Lu, S., Shi, R., Chiang, V.L., 2011. Computational prediction of plant miRNA targets. *Methods Mol. Biol.* Clifton NJ 744, 175–186. [https://doi.org/10.1007/978-1-61779-123-9\\_12](https://doi.org/10.1007/978-1-61779-123-9_12)

- Sunkar, R., Li, Y.-F., Jagadeeswaran, G., 2012. Functions of microRNAs in plant stress responses. *Trends Plant Sci.* 17, 196–203. <https://doi.org/10.1016/j.tplants.2012.01.010>
- Sunkar, R., Zhu, J.-K., 2004. Novel and Stress-Regulated MicroRNAs and Other Small RNAs from *Arabidopsis*. *Plant Cell* 16, 2001–2019. <https://doi.org/10.1105/tpc.104.022830>
- Szakonyi, D., Duque, P., 2018. Alternative Splicing as a Regulator of Early Plant Development. *Front. Plant Sci.* 9. <https://doi.org/10.3389/fpls.2018.01174>
- Szarzynska, B., Sobkowiak, L., Pant, B.D., Balazadeh, S., Scheible, W.-R., Mueller-Roeber, B., Jarmolowski, A., Szweykowska-Kulinska, Z., 2009. Gene structures and processing of *Arabidopsis thaliana* HYL1-dependent pri-miRNAs. *Nucleic Acids Res.* 37, 3083–3093. <https://doi.org/10.1093/nar/gkp189>
- TAIR - Download - TAIR10 transposable elements [WWW Document], n.d. URL [https://www.arabidopsis.org/download/index-auto.jsp?dir=%2Fdownload\\_files%2FGenes%2FTAIR10\\_genome\\_release%2FTAIR10\\_transposable\\_elements](https://www.arabidopsis.org/download/index-auto.jsp?dir=%2Fdownload_files%2FGenes%2FTAIR10_genome_release%2FTAIR10_transposable_elements) (accessed 8.18.20a).
- TAIR - Download - TAIR10 transposable elements [WWW Document], n.d. URL [https://www.arabidopsis.org/download/index-auto.jsp?dir=%2Fdownload\\_files%2FGenes%2FTAIR10\\_genome\\_release%2FTAIR10\\_transposable\\_elements](https://www.arabidopsis.org/download/index-auto.jsp?dir=%2Fdownload_files%2FGenes%2FTAIR10_genome_release%2FTAIR10_transposable_elements) (accessed 7.3.20b).
- TAIR10 - Genome - Assembly - NCBI [WWW Document], n.d. URL [https://www.ncbi.nlm.nih.gov/assembly/GCF\\_000001735.3/](https://www.ncbi.nlm.nih.gov/assembly/GCF_000001735.3/) (accessed 8.18.20).
- Tamaru, H., 2010. Confining euchromatin/heterochromatin territory: jumonji crosses the line. *Genes Dev.* 24, 1465–1478. <https://doi.org/10.1101/gad.1941010>
- Taochy, C., Gursansky, N.R., Cao, J., Fletcher, S.J., Dressel, U., Mitter, N., Tucker, M.R., Koltunow, A.M.G., Bowman, J.L., Vaucheret, H., Carroll, B.J., 2017. A Genetic Screen for Impaired Systemic RNAi Highlights the Crucial Role of DICER-LIKE 21[OPEN]. *Plant Physiol.* 175, 1424–1437. <https://doi.org/10.1104/pp.17.01181>
- Thomson, T., Lin, H., 2009. The Biogenesis and Function PIWI Proteins and piRNAs: Progress and Prospect. *Annu. Rev. Cell Dev. Biol.* 25, 355–376. <https://doi.org/10.1146/annurev.cellbio.24.110707.175327>
- Trotta, E., 2014. On the Normalization of the Minimum Free Energy of RNAs by Sequence Length. *PLoS ONE* 9. <https://doi.org/10.1371/journal.pone.0113380>
- Truernit, E., Bauby, H., Belcram, K., Barthélémy, J., Palauqui, J.-C., 2012. OCTOPUS, a polarly localised membrane-associated protein, regulates phloem differentiation entry in *Arabidopsis thaliana*. *Dev. Camb. Engl.* 139, 1306–1315. <https://doi.org/10.1242/dev.072629>
- Tu, B., Liu, L., Xu, C., Zhai, J., Li, Shengben, Lopez, M.A., Zhao, Y., Yu, Y., Ramachandran, V., Ren, G., Yu, B., Li, Shigui, Meyers, B.C., Mo, B., Chen, X., 2015. Distinct and Cooperative Activities of HESO1 and URT1 Nucleotidyl Transferases in MicroRNA Turnover in *Arabidopsis*. *PLoS Genet.* 11. <https://doi.org/10.1371/journal.pgen.1005119>
- Tuteja, N., 2007. Abscisic Acid and Abiotic Stress Signaling. *Plant Signal. Behav.* 2, 135–138.
- Underwood, C.J., Henderson, I.R., Martienssen, R.A., 2017. Genetic and epigenetic variation of transposable elements in *Arabidopsis*. *Curr. Opin. Plant Biol.* 36, 135–141. <https://doi.org/10.1016/j.pbi.2017.03.002>
- Válóczi, A., Hornyik, C., Varga, N., Burgyán, J., Kauppinen, S., Havelda, Z., 2004. Sensitive and specific detection of microRNAs by northern blot analysis using LNA-modified oligonucleotide probes. *Nucleic Acids Res.* 32, e175. <https://doi.org/10.1093/nar/gnh171>
- Vaucheret, H., 2008. Plant ARGONAUTES. *Trends Plant Sci.* 13, 350–358. <https://doi.org/10.1016/j.tplants.2008.04.007>
- Vaucheret, H., Mallory, A.C., Bartel, D.P., 2006. AGO1 homeostasis entails coexpression of MIR168 and AGO1 and preferential stabilization of miR168 by AGO1. *Mol. Cell* 22, 129–136. <https://doi.org/10.1016/j.molcel.2006.03.011>

- Vaucheret, H., Vazquez, F., Crété, P., 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 Dev.* 18, 1187–1197. <https://doi.org/10.1101/gad.1201404>
- Vazquez, F., Hohn, T., 2013. Biogenesis and Biological Activity of Secondary siRNAs in Plants. *Scientifica* 2013, 783253. <https://doi.org/10.1155/2013/783253>
- Vazquez, F., Vaucheret, H., Rajagopalan, R., Lepers, C., Gascioli, V., Mallory, A.C., Hilbert, J.-L., Bartel, D.P., Crété, P., 2004. Endogenous trans-acting siRNAs regulate the accumulation of Arabidopsis mRNAs. *Mol. Cell* 16, 69–79. <https://doi.org/10.1016/j.molcel.2004.09.028>
- Volanakis, A., Passoni, M., Hector, R.D., Shah, S., Kilchert, C., Granneman, S., Vasiljeva, L., 2013. Spliceosome-mediated decay (SMD) regulates expression of nonintronic genes in budding yeast. *Genes Dev.* 27, 2025–2038. <https://doi.org/10.1101/gad.221960.113>
- Wachsman, G., Modliszewski, J.L., Valdes, M., Benfey, P.N., 2017a. A SIMPLE Pipeline for Mapping Point Mutations1[OPEN]. *Plant Physiol.* 174, 1307–1313. <https://doi.org/10.1104/pp.17.00415>
- Wachsman, G., Modliszewski, J.L., Valdes, M., Benfey, P.N., 2017b. A SIMPLE Pipeline for Mapping Point Mutations. *Plant Physiol.* 174, 1307–1313. <https://doi.org/10.1104/pp.17.00415>
- Wang, F., Axtell, M.J., 2017. AGO4 is specifically required for heterochromatic siRNA accumulation at Pol V-dependent loci in Arabidopsis thaliana. *Plant J.* 90, 37–47. <https://doi.org/10.1111/tpj.13463>
- Wang, J., Chen, S., Jiang, N., Li, N., Wang, X., Li, Z., Li, X., Liu, H., Li, L., Yang, Y., Ni, T., Yu, C., Ma, J., Zheng, B., Ren, G., 2019a. Spliceosome disassembly factors ILP1 and NTR1 promote miRNA biogenesis in Arabidopsis thaliana. *Nucleic Acids Res.* 47, 7886–7900. <https://doi.org/10.1093/nar/gkz526>
- Wang, J., Chen, S., Jiang, N., Li, N., Wang, X., Li, Z., Li, X., Liu, H., Li, L., Yang, Y., Ni, T., Yu, C., Ma, J., Zheng, B., Ren, G., 2019b. Spliceosome disassembly factors ILP1 and NTR1 promote miRNA biogenesis in Arabidopsis thaliana. *Nucleic Acids Res.* 47, 7886–7900. <https://doi.org/10.1093/nar/gkz526>
- Wang, J., Mei, J., Ren, G., 2019c. Plant microRNAs: Biogenesis, Homeostasis, and Degradation. *Front. Plant Sci.* 10. <https://doi.org/10.3389/fpls.2019.00360>
- Wang, X., Wang, Y., Dou, Y., Chen, L., Wang, J., Jiang, N., Guo, C., Yao, Q., Wang, C., Liu, L., Yu, B., Zheng, B., Chekanova, J.A., Ma, J., Ren, G., 2018. Degradation of unmethylated miRNA/miRNA\*s by a DEDDy-type 3' to 5' exoribonuclease Atrimmer 2 in Arabidopsis. *Proc. Natl. Acad. Sci.* 115, E6659–E6667. <https://doi.org/10.1073/pnas.1721917115>
- Wang, X., Zhang, S., Dou, Y., Zhang, C., Chen, X., Yu, B., Ren, G., 2015. Synergistic and Independent Actions of Multiple Terminal Nucleotidyl Transferases in the 3' Tailing of Small RNAs in Arabidopsis. *PLOS Genet* 11, e1005091. <https://doi.org/10.1371/journal.pgen.1005091>
- Wang, Y., Liu, J., Huang, B., Xu, Y.-M., Li, J., HUANG, L.-F., LIN, J., ZHANG, J., MIN, Q.-H., YANG, W.-M., WANG, X.-Z., 2015. Mechanism of alternative splicing and its regulation. *Biomed. Rep.* 3, 152–158. <https://doi.org/10.3892/br.2014.407>
- Wang, Y.-F., 2014. ABA Regulation of Stomatal Movement, in: Zhang, D.-P. (Ed.), *Abscisic Acid: Metabolism, Transport and Signaling*. Springer Netherlands, Dordrecht, pp. 287–313. [https://doi.org/10.1007/978-94-017-9424-4\\_15](https://doi.org/10.1007/978-94-017-9424-4_15)
- Wang, Z., Ji, H., Yuan, B., Wang, S., Su, C., Yao, B., Zhao, H., Li, X., 2015. ABA signalling is fine-tuned by antagonistic HAB1 variants. *Nat. Commun.* 6, 1–12. <https://doi.org/10.1038/ncomms9138>
- Wickham, H., 2009. *ggplot2: Elegant Graphics for Data Analysis*, Use R! Springer-Verlag, New York. <https://doi.org/10.1007/978-0-387-98141-3>
- Wierzbicki, A.T., Ream, T.S., Haag, J.R., Pikaard, C.S., 2009. RNA polymerase V transcription guides ARGONAUTE4 to chromatin. *Nat. Genet.* 41, 630–634. <https://doi.org/10.1038/ng.365>
- Winter, J., Diederichs, S., 2011. Argonaute proteins regulate microRNA stability: Increased microRNA abundance by Argonaute proteins is due to microRNA stabilization. *RNA Biol.* 8, 1149–1157. <https://doi.org/10.4161/rna.8.6.17665>

- Wright, C., Rajpurohit, A., Burke, E.E., Williams, C., Collado-Torres, L., Kimos, M., Brandon, N.J., Cross, A.J., Jaffe, A.E., Weinberger, D.R., Shin, J.H., 2019. Comprehensive assessment of multiple biases in small RNA sequencing reveals significant differences in the performance of widely used methods. *BMC Genomics* 20. <https://doi.org/10.1186/s12864-019-5870-3>
- Wu, G., Park, M.Y., Conway, S.R., Wang, J.-W., Weigel, D., Poethig, R.S., 2009. The Sequential Action of miR156 and miR172 Regulates Developmental Timing in Arabidopsis. *Cell* 138, 750–759. <https://doi.org/10.1016/j.cell.2009.06.031>
- Wu, G., Poethig, R.S., 2006. Temporal regulation of shoot development in Arabidopsis thaliana by miR156 and its target SPL3. *Dev. Camb. Engl.* 133, 3539–3547. <https://doi.org/10.1242/dev.02521>
- Wu, M., Tinoco, I., 1998. RNA folding causes secondary structure rearrangement. *Proc. Natl. Acad. Sci. U. S. A.* 95, 11555–11560. <https://doi.org/10.1073/pnas.95.20.11555>
- Xie, Z., Allen, E., Fahlgren, N., Calamar, A., Givan, S.A., Carrington, J.C., 2005a. Expression of Arabidopsis MIRNA Genes. *Plant Physiol.* 138, 2145–2154. <https://doi.org/10.1104/pp.105.062943>
- Xie, Z., Allen, E., Wilken, A., Carrington, J.C., 2005b. DICER-LIKE 4 functions in trans-acting small interfering RNA biogenesis and vegetative phase change in Arabidopsis thaliana. *Proc. Natl. Acad. Sci. U. S. A.* 102, 12984–12989. <https://doi.org/10.1073/pnas.0506426102>
- Xie, Z., Johansen, L.K., Gustafson, A.M., Kasschau, K.D., Lellis, A.D., Zilberman, D., Jacobsen, S.E., Carrington, J.C., 2004. Genetic and functional diversification of small RNA pathways in plants. *PLoS Biol.* 2, E104. <https://doi.org/10.1371/journal.pbio.0020104>
- Xie, Z., Kasschau, K.D., Carrington, J.C., 2003. Negative Feedback Regulation of Dicer-Like1 in Arabidopsis by microRNA-Guided mRNA Degradation. *Curr. Biol.* 13, 784–789. [https://doi.org/10.1016/S0960-9822\(03\)00281-1](https://doi.org/10.1016/S0960-9822(03)00281-1)
- Xu, J., Yang, J.-Y., Niu, Q.-W., Chua, N.-H., 2006. Arabidopsis DCP2, DCP1, and VARICOSE Form a Decapping Complex Required for Postembryonic Development. *Plant Cell* 18, 3386–3398. <https://doi.org/10.1105/tpc.106.047605>
- Xu, P., Billmeier, M., Mohorianu, I.-I., Green, D., Fraser, W., Dalmay, T., 2015. An improved protocol for small RNA library construction using High Definition adapters. *Methods Gener. Seq.* 2. <https://doi.org/10.1515/mngs-2015-0001>
- Yan, K., Liu, P., Wu, C.-A., Yang, G.-D., Xu, R., Guo, Q.-H., Huang, J.-G., Zheng, C.-C., 2012. Stress-induced alternative splicing provides a mechanism for the regulation of microRNA processing in Arabidopsis thaliana. *Mol. Cell* 48, 521–531. <https://doi.org/10.1016/j.molcel.2012.08.032>
- Yang, A., Shao, T.-J., Bofill-De Ros, X., Lian, C., Villanueva, P., Dai, L., Gu, S., 2020. AGO-bound mature miRNAs are oligouridylylated by TUTs and subsequently degraded by DIS3L2. *Nat. Commun.* 11, 2765. <https://doi.org/10.1038/s41467-020-16533-w>
- Yang, G.D., Yan, K., Wu, B.J., Wang, Y.H., Gao, Y.X., Zheng, C.C., 2012. Genomewide analysis of intronic microRNAs in rice and Arabidopsis. *J. Genet.* 91, 313–324. <https://doi.org/10.1007/s12041-012-0199-6>
- Yang, X., Zhang, H., Li, L., 2012. Alternative mRNA processing increases the complexity of microRNA-based gene regulation in Arabidopsis. *Plant J. Cell Mol. Biol.* 70, 421–431. <https://doi.org/10.1111/j.1365-313X.2011.04882.x>
- Ye, J., Coulouris, G., Zaretskaya, I., Cutcutache, I., Rozen, S., Madden, T.L., 2012. Primer-BLAST: A tool to design target-specific primers for polymerase chain reaction. *BMC Bioinformatics* 13, 134. <https://doi.org/10.1186/1471-2105-13-134>
- Ye, R., Chen, Z., Lian, B., Rowley, M.J., Xia, N., Chai, J., Li, Y., He, X.-J., Wierzbicki, A.T., Qi, Y., 2016. A Dicer-Independent Route for Biogenesis of siRNAs that Direct DNA Methylation in Arabidopsis. *Mol. Cell* 61, 222–235. <https://doi.org/10.1016/j.molcel.2015.11.015>
- Yokawa, K., Kagenishi, T., Baluška, F., 2019. Anesthetics, Anesthesia, and Plants. *Trends Plant Sci.* 24, 12–14. <https://doi.org/10.1016/j.tplants.2018.10.006>

- Yoshikawa, M., Peragine, A., Park, M.Y., Poethig, R.S., 2005. A pathway for the biogenesis of trans-acting siRNAs in Arabidopsis. *Genes Dev.* 19, 2164–2175. <https://doi.org/10.1101/gad.1352605>
- You, C., He, W., Hang, R., Zhang, C., Cao, X., Guo, H., Chen, X., Cui, J., Mo, B., 2019a. FIERY1 promotes microRNA accumulation by suppressing rRNA-derived small interfering RNAs in Arabidopsis. *Nat. Commun.* 10, 4424. <https://doi.org/10.1038/s41467-019-12379-z>
- You, C., He, W., Hang, R., Zhang, C., Cao, X., Guo, H., Chen, X., Cui, J., Mo, B., 2019b. FIERY1 promotes microRNA accumulation by suppressing rRNA-derived small interfering RNAs in Arabidopsis. *Nat. Commun.* 10, 1–15. <https://doi.org/10.1038/s41467-019-12379-z>
- Yu, B., Bi, L., Zhai, J., Agarwal, M., Li, S., Wu, Q., Ding, S.-W., Meyers, B.C., Vaucheret, H., Chen, X., 2010. siRNAs compete with miRNAs for methylation by HEN1 in Arabidopsis. *Nucleic Acids Res.* 38, 5844–5850. <https://doi.org/10.1093/nar/gkq348>
- Yu, B., Yang, Z., Li, J., Minakhina, S., Yang, M., Padgett, R.W., Steward, R., Chen, X., 2005. Methylation as a Crucial Step in Plant microRNA Biogenesis. *Science* 307, 932–935. <https://doi.org/10.1126/science.1107130>
- Yu, Y., Ji, L., Le, B.H., Zhai, J., Chen, J., Luscher, E., Gao, L., Liu, C., Cao, X., Mo, B., Ma, J., Meyers, B.C., Chen, X., 2017a. ARGONAUTE10 promotes the degradation of miR165/6 through the SDN1 and SDN2 exonucleases in Arabidopsis. *PLoS Biol.* 15, e2001272. <https://doi.org/10.1371/journal.pbio.2001272>
- Yu, Y., Ji, L., Le, B.H., Zhai, J., Chen, J., Luscher, E., Gao, L., Liu, C., Cao, X., Mo, B., Ma, J., Meyers, B.C., Chen, X., 2017b. ARGONAUTE10 promotes the degradation of miR165/6 through the SDN1 and SDN2 exonucleases in Arabidopsis. *PLOS Biol.* 15, e2001272. <https://doi.org/10.1371/journal.pbio.2001272>
- Zemach, A., Kim, M.Y., Hsieh, P.-H., Coleman-Derr, D., Eshed-Williams, L., Thao, K., Harmer, S.L., Zilberman, D., 2013. The nucleosome remodeler DDM1 allows DNA methyltransferases to access H1-containing heterochromatin. *Cell* 153, 193–205. <https://doi.org/10.1016/j.cell.2013.02.033>
- Zeng, Y., Yi, R., Cullen, B.R., 2003. MicroRNAs and small interfering RNAs can inhibit mRNA expression by similar mechanisms. *Proc. Natl. Acad. Sci.* 100, 9779–9784. <https://doi.org/10.1073/pnas.1630797100>
- Zhai, J., Meyers, B.C., 2012. Deep Sequencing from hen1 Mutants to Identify Small RNA 3' Modifications. *Cold Spring Harb. Symp. Quant. Biol.* 77, 213–219. <https://doi.org/10.1101/sqb.2013.77.014779>
- Zhai, J., Zhao, Y., Simon, S.A., Huang, S., Petsch, K., Arikait, S., Pillay, M., Ji, L., Xie, M., Cao, X., Yu, B., Timmermans, M., Yang, B., Chen, X., Meyers, B.C., 2013a. Plant MicroRNAs Display Differential 3' Truncation and Tailing Modifications That Are ARGONAUTE1 Dependent and Conserved Across Species. *Plant Cell* 25, 2417–2428. <https://doi.org/10.1105/tpc.113.114603>
- Zhai, J., Zhao, Y., Simon, S.A., Huang, S., Petsch, K., Arikait, S., Pillay, M., Ji, L., Xie, M., Cao, X., Yu, B., Timmermans, M., Yang, B., Chen, X., Meyers, B.C., 2013b. Plant MicroRNAs Display Differential 3' Truncation and Tailing Modifications That Are ARGONAUTE1 Dependent and Conserved Across Species[W]. *Plant Cell* 25, 2417–2428. <https://doi.org/10.1105/tpc.113.114603>
- Zhang, H., Kolb, F.A., Jaskiewicz, L., Westhof, E., Filipowicz, W., 2004. Single Processing Center Models for Human Dicer and Bacterial RNase III. *Cell* 118, 57–68. <https://doi.org/10.1016/j.cell.2004.06.017>
- Zhang, W., Murphy, C., Sieburth, L.E., 2010. Conserved RNaseIII domain protein functions in cytoplasmic mRNA decay and suppresses Arabidopsis decapping mutant phenotypes. *Proc. Natl. Acad. Sci.* 107, 15981–15985. <https://doi.org/10.1073/pnas.1007060107>
- Zhao, Y., Yu, Y., Zhai, J., Ramachandran, V., Dinh, T.T., Meyers, B.C., Mo, B., Chen, X., 2012a. The Arabidopsis nucleotidyl transferase HESO1 uridylates unmethylated small RNAs to trigger their degradation. *Curr. Biol. CB* 22, 689–694. <https://doi.org/10.1016/j.cub.2012.02.051>

- Zhao, Y., Yu, Y., Zhai, J., Ramachandran, V., Dinh, T.T., Meyers, B.C., Mo, B., Chen, X., 2012b. HESO1, a nucleotidyl transferase in Arabidopsis, uridylylates unmethylated miRNAs and siRNAs to trigger their degradation. *Curr. Biol.* CB 22, 689–694. <https://doi.org/10.1016/j.cub.2012.02.051>
- Zhou, M., Law, J.A., 2015a. RNA Pol IV and V in Gene Silencing: Rebel Polymerases Evolving Away From Pol II's Rules. *Curr. Opin. Plant Biol.* 27, 154–164. <https://doi.org/10.1016/j.pbi.2015.07.005>
- Zhou, M., Law, J.A., 2015b. RNA Pol IV and V in Gene Silencing: Rebel Polymerases Evolving Away From Pol II's Rules. *Curr. Opin. Plant Biol.* 27, 154–164. <https://doi.org/10.1016/j.pbi.2015.07.005>
- Zhu, F.-Y., Chen, M.-X., Ye, N.-H., Shi, L., Ma, K.-L., Yang, J.-F., Cao, Y.-Y., Zhang, Y., Yoshida, T., Fernie, A.R., Fan, G.-Y., Wen, B., Zhou, R., Liu, T.-Y., Fan, T., Gao, B., Zhang, D., Hao, G.-F., Xiao, S., Liu, Y.-G., Zhang, J., 2017. Proteogenomic analysis reveals alternative splicing and translation as part of the abscisic acid response in Arabidopsis seedlings. *Plant J.* 91, 518–533. <https://doi.org/10.1111/tpj.13571>
- Zhu, H., Hu, F., Wang, R., Zhou, X., Sze, S.-H., Liou, L.W., Barefoot, A., Dickman, M., Zhang, X., 2011. Arabidopsis Argonaute10 Specifically Sequesters miR166/165 to Regulate Shoot Apical Meristem Development. *Cell* 145, 242–256. <https://doi.org/10.1016/j.cell.2011.03.024>