

UC Riverside

UC Riverside Electronic Theses and Dissertations

Title

A Study of the MOS4-Associated Complex in RNA Metabolism in Arabidopsis

Permalink

<https://escholarship.org/uc/item/45h57490>

Author

Jia, Tianran

Publication Date

2017

Peer reviewed|Thesis/dissertation

UNIVERSITY OF CALIFORNIA
RIVERSIDE

A Study of the MOS4-Associated Complex in RNA Metabolism in *Arabidopsis*

A Dissertation submitted in partial satisfaction
of the requirements for the degree of

Doctor of Philosophy

in

Plant Biology

by

Tianran Jia

December 2017

Dissertation Committee:

Dr. Xuemei Chen, Chairperson

Dr. Fedor V. Karginov

Dr. Jaimie Van Norman

Copyright by
Tianran Jia
2017

The Dissertation of Tianran Jia is approved:

Committee Chairperson

University of California, Riverside

ACKNOWLEDGEMENTS

I would like to express my sincere appreciation to everyone who helped me during my PhD studies at the University of California, Riverside (UCR). This work would not have been possible without their generous help.

First, I would like to thank my PhD advisor, Dr. Xuemei Chen. She is one of the best scientists I have ever known, and she is a great role model for me. I have learned how to work efficiently, think strategically and collaborate with others effectively from her. Without her patient and insightful guidance, this work would not have been possible. What I have learned from her, not only about science, will greatly benefit my future life and career. I feel lucky as one of her students.

I would like to thank the other members of my dissertation committee: Drs. Fedor V. Karginov and Jaimie Van Norman for their insightful suggestions and critical technical support for my projects; and the members of my PhD qualifying committee: Drs. Julia Bailey-Serres, Patricia Springer, Zhenbiao Yang, Thomas Eulgem and Shouwei Ding for helping me build my background knowledge in plant sciences and molecular biology research. In addition, I would also like to express my genuine appreciation to Drs. Linda Walling, Eugene Nothnagel, Natasha Raikhel and all the other professors at UCR for their time and effort in teaching me during my studies here. The valuable knowledge and experience I have learned from them greatly benefit me now and in the future.

Furthermore, I would like to express my gratitude to all members whose time overlapped with mine in Dr. Xuemei Chen's lab, other labs' members, technicians and staffs at the UCR College of Natural and Agricultural Sciences. The research community at UCR is very impressive and helped me greatly during my graduate studies. I could not have imagined a better research community and fellow colleagues.

My special appreciation also goes to people who contributed to my PhD projects, including Drs. Chenjiang You and Brandon Le who performed and taught me RNA-seq analysis; Drs. Bailong Zhang and Liping Zeng who discovered and mapped the *mac7-1* mutant; Dr. Yong Zhang who contributed to the proteomic analysis; Li Liu and Dr. Wenrong He who performed the EMS mutagenesis; Dr. Shengjun Li who performed the BiFC analysis; Dr. Kaeli C.M. Johnson and ZhongShou Wu who performed the pathogen infection assays; Drs. Bin Yu, Xin Li and Yongli Qiao for sharing plant materials; Dr. Haiyan Zheng for helping with mass spectrometry analysis; Dr. Jianbo Song for sharing the small RNA-seq data; and Lorena Arroyo for helping me out with experiments and many other lab works. I will also need to thank Drs. Runlai Hang and Yongjian Qiu for their important contributions to Chapter 2 of my thesis.

Finally, I would like to thank my dearest parents and grandparents for their unconditional love and support through so many years. They are the rock of my life. I also need to thank my friends and especially my partner Mingyang Li, who brings love, friendship and memorable moments to my life.

In accordance with department guidelines, INTRODUCTION Chapter is part of a published manuscript in New Phytologist reprinted with permission [New Phytologist, 2017]. Chapter 1, in full, is a reprint of a published paper in The Plant Cell (DOI: <https://doi.org/10.1105/tpc.17.00370>). I acknowledge the permission to use the items from the above publishers.

AUTHOR CONTRIBUTIONS

INTRODUCTION

Tianran Jia and Xuemei Chen wrote the manuscript. Tianran Jia made Figure 1 and Table 1.

CHAPTER 1

Tianran Jia, Chenjiang You and Xuemei Chen wrote the manuscript. Tianran Jia organized all the figures and tables in Chapter 1. Tianran Jia, Bailong Zhang, Yong Zhang, Shengjun Li, Kaeli C.M. Johnson, Bin Yu, Xin Li and Xuemei Chen designed and/or performed the experiments. Tianran Jia, Chenjiang You and Liping Zeng analyzed the high-throughput sequencing data.

CHAPTER 2

Tianran Jia wrote the manuscript. Tianran Jia made Figure 2.1, Figure 2.3 and Table 2.1. Yongjian Qiu and Tianran Jia made Figure 2.2. Runlai Hang and Tianran Jia made Figure 2.4. Chenjiang You and Tianran Jia made Figure 2.5. Tianran Jia, Yongjian Qiu and Runlai Hang performed the experiments. Chenjiang You and Tianran Jia analyzed the RNA-seq data.

ABSTRACT OF THE DISSERTATION

A Study of the MOS4-Associated Complex in RNA Metabolism in *Arabidopsis*

by

Tianran Jia

Doctor of Philosophy, Graduate Program in Plant Biology
University of California, Riverside, December 2017
Dr. Xuemei Chen, Chairperson

In *Arabidopsis thaliana*, the MOS4-ASSOCIATED COMPLEX (MAC) is required for defense and development. The evolutionarily conserved putative RNA helicase MAC7 is a component of the *Arabidopsis* MAC, and the human MAC7 homolog, Aquarius, is implicated in precursor messenger RNA (pre-mRNA) splicing. In my thesis research, I studied the functions of MAC in RNA metabolism and other aspects of *Arabidopsis* development and defense. First, we show that the partial loss-of-function mutant *mac7-1* and two other MAC subunit mutants, *mac3a mac3b* and *pleiotropic regulatory locus 1 pleiotropic regulatory locus 2 (prl1 prl2)*, exhibit reduced microRNA (miRNA) levels, indicating that MAC promotes miRNA biogenesis. The *mac7-1* mutant shows reduced primary miRNA (pri-miRNA) levels without affecting miRNA gene (*MIR*) promoter activity or the half-life of pri-miRNA transcripts. As a nuclear protein, MAC7 is not concentrated in dicing bodies, but it affects the localization of HYPONASTIC LEAVES1 (HYL1), a key protein in pri-miRNA processing, to dicing bodies. Immunoprecipitation of HYL1 retrieved eleven known MAC subunits, including MAC7, indicating association

between HYL1 and MAC. We propose that MAC7 links *MIR* transcription to pri-miRNA processing. mRNA-seq analysis showed that down-regulated genes in MAC subunit mutants are mostly involved in plant defense and stimulus responses, confirming a role of MAC in biotic and abiotic stress responses. We also discovered global intron retention defects in mutants of three subunits of MAC, thus linking MAC function to mRNA splicing in *Arabidopsis*. Second, since previous studies showed that MAC has multiple functions in eukaryotes, including roles in transcription elongation, mRNA splicing, DNA repair, RNA export and the ubiquitin-proteasome pathway, we explored other potential functions of *Arabidopsis* MAC and the MAC7 subunit in particular beyond miRNA biogenesis and pre-mRNA splicing. We found that MAC7 has a potential role in photosynthetic pathways, as RNA-seq analysis showed that many photosynthesis-related genes are down-regulated in this mutant. In addition, MAC7 and PRL proteins are very likely to be involved in ribosomal RNA (rRNA) biogenesis, as many rRNA precursors accumulated in the *mac7-1* and *prl1 prl2* mutants.

TABLE OF CONTENTS

INTRODUCTION	1
ABSTRACT	1
INTRODUCTION	3
FIGURES	16
TABLES	18
REFERENCES	26
CHAPTER 1	34
ABSTRACT	34
INTRODUCTION	36
RESULTS	40
DISCUSSION	53
METHODS	57
FIGURES	67
TABLES	92
REFERENCES	101
CHAPTER 2	108
ABSTRACT	108
INTRODUCTION	109
RESULTS	113
DISCUSSION	115
METHODS	118
FIGURES	121
TABLES	128
REFERENCES	131
Appendix A	135
INTRODUCTION	135
FIGURES	141
REFERENCES	143

LIST OF FIGURES

Figure 1. Illustrations of major steps in microRNA (miRNA) biogenesis.	17
Figure 1.1 The silencing suppressor mutant <i>mac7-1</i> exhibits pleiotropic phenotypes and compromised miRNA accumulation.....	68
Figure 1.2 The <i>mac7-1</i> mutant shows pleiotropic developmental phenotypes.....	69
Figure 1.3 Reduced miRNA accumulation in both seedlings and inflorescences in the <i>mac7-1</i> mutant.	71
Figure 1.4 The <i>mac7-1</i> mutation in the Col-0 background does not affect <i>CH42</i> expression at either the mRNA or protein level.....	72
Figure 1.5 A point mutation in <i>MAC7</i> is responsible for the morphological and molecular phenotypes in the <i>mac7-1</i> mutant.	73
Figure 1.6 The <i>MAC7</i> T-DNA insertion line <i>mac7-2</i> is a single-locus, recessive, embryo-lethal mutant.....	74
Figure 1.7 Amino acid sequence alignment of <i>Arabidopsis</i> <i>MAC7</i> and its orthologs in other species.	76
Figure 1.8 <i>MAC7</i> promotes pri-miRNA production.	77
Figure 1.9 <i>MAC7</i> does not affect the expression of the endogenous <i>SUC2</i> gene or key miRNA biogenesis factors.	79
Figure 1.10 <i>MAC7</i> is a nuclear protein associated with other MAC subunits.	80

Figure 1.11 MAC7 shows dispersed distribution in the nucleoplasm, while DCL1 and HYL1 concentrate in dicing bodies in the nucleus.	81
Figure 1.12 The MAC subunit genes <i>MAC3a</i> , <i>MAC3b</i> , <i>PRL1</i> and <i>PRL2</i> also promote miRNA biogenesis.	83
Figure 1.13 MAC7 affects HYL1 subcellular localization.	84
Figure 1.14 Down-regulated genes in the <i>mac</i> mutants are significantly related to stress responses, while up-regulated genes are involved in various biosynthetic and metabolic processes.	85
Figure 1.15 Growth of virulent <i>P.s.m. ES4326</i> at 0 and 3 days post-inoculation.	86
Figure 1.16 MAC subunits affect pre-mRNA splicing.	88
Figure 1.17 MAC affects pre-mRNA splicing and gene expression independently.	89
Figure 1.18 Intron-containing pri-miRNAs show similar splicing patterns in Col-0 and <i>mac7-1</i>	90
Figure 1.19 A model for MAC functions in miRNA biogenesis and pre-mRNA processing.	91
Figure 2.1 MAC7 has a potential role in photosynthetic pathways.	122
Figure 2.2 <i>mac7-1</i> is hyposensitive to red light compared to Col-0.	123
Figure 2.3 Global histone methylation levels were not altered in <i>mac7-1</i>	124
Figure 2.4 Ribosomal RNA precursors accumulated aberrantly in <i>mac7-1</i> and <i>prl1 prl2</i>	126

Figure 2.5 <i>mac7-1</i> did not cause global change in RNA export from the nucleus to the cytoplasm.	127
Figure A.1 Illustrations of miRNA cell-to-cell and systemic movement in plants.	142

LIST OF TABLES

Table 1. Proteins involved in miRNA biogenesis in <i>Arabidopsis</i>	18
Table 1.1 Phenotypic and genotypic segregation in the progeny of selfed heterozygous <i>mac7-2</i> plants. The <i>mac7-2</i> allele harbors a T-DNA insertion (SALK_129044).	92
Table 1.2 MAC7-associated proteins identified by immunoprecipitation followed by mass spectrometry.	93
Table 1.3 HYL1-associated MAC subunits identified by immunoprecipitation followed by mass spectrometry.	95
Table 1.4 Oligonucleotide sequences.	97
Table 2.1 Conserved MAC components in <i>Arabidopsis</i> , human and yeast complexes. .	128

INTRODUCTION

MicroRNA biogenesis in plants

ABSTRACT

MicroRNAs (miRNAs) are small non-coding RNAs, typically 20-24 nt, that regulate gene expression post-transcriptionally or transcriptionally through sequence complementarity. Since the identification of the first miRNA, *lin-4*, in the nematode *Caenorhabditis elegans* in 1993, thousands of miRNAs have been discovered in animals and plants. The prevalence of miRNAs throughout the animal and plant kingdoms indicates the critical biological functions of miRNAs. For example, miRNAs have essential functions in animal development, homeostasis and the progression of human diseases. In plants, miRNAs regulate development and responses to biotic and abiotic stresses by targeting transcription factors, stress response proteins and enzymes. The plant miRNA pathway has been extensively studied in the past decade. Recent studies have unveiled the complexity of the miRNA pathway, particularly miRNA biogenesis. Each step of miRNA biogenesis, from miRNA gene (*MIR*) transcription to the formation of the AGO-containing RNA-induced silencing complex (RISC), is tightly regulated by many factors. Previous studies have also implicated restricted subcellular locations where miRNA biogenesis takes place. In this introduction chapter, we discuss not only the players involved in these processes but also the subcellular sites where these processes are known or implicated to take place. We hope

to raise awareness that the cell biology of miRNAs holds the key to fully understanding these enigmatic molecules.

INTRODUCTION

MicroRNA (*MIR*) genes encoding microRNA (miRNAs) are transcribed by RNA polymerase II (Pol II) into primary miRNAs (pri-miRNAs). The stem-loop-containing pri-miRNAs are processed by the RNase III family enzyme DICER-LIKE1 (DCL1) into miRNA/miRNA* duplexes. These duplexes are 2'-*O*-methylated at the 3'-ends by the methyltransferase HUA ENHANCER1 (HEN1). One strand from the duplex is incorporated into ARGONAUTE1 (AGO1) to form an active RNA-induced silencing complex (RISC) (1) (Figure 1). The following section focuses on the complex regulation of miRNA biogenesis unveiled by recent studies (Table 1).

***MIR* transcription and transcriptional regulation**

Similar to protein-coding genes, most *MIR* genes contain the TATA-box motif and transcription factor binding motifs, such as those of auxin response factors (ARFs) and MYC2, in their promoters, indicating that *MIR* transcription is regulated by general and specific transcription factors (2, 3).

Mediator, a general transcriptional coactivator, helps recruit Pol II to *MIR* loci (4). Other factors promoting general *MIR* transcription include NEGATIVE ON TATA LESS2 (NOT2); the putative MYB domain-containing DNA-binding protein CELL DIVISION CYCLE 5 (CDC5); and the Elongator complex, which is thought to assist transcriptional elongation (5-7). NOT2, CDC5 and Elongator all interact with Pol II and the dicing

complex (the plant miRNA precursor processing complex), implying their functions in bridging Pol II transcription and pri-miRNA processing (5-7). Pol II activity in *MIR* transcription is probably subject to phospho-regulation. MiRNA levels are significantly reduced in mutants of the *CDKF;1* (*CYCLIN-DEPENDENT KINASE F;1*) and *CDKD* (*CYCLIN-DEPENDENT KINASE D*) genes. These mutants also have reduced phosphorylation marks at the Pol II C-terminal domain (CTD) (8, 9).

Factors specifically controlling the transcription of certain miRNAs within a miRNA family have also been characterized. For instance, POWERDRESS promotes the transcription of *MIR172a*, *b* and *c* by enhancing Pol II occupancy at their promoters, without affecting *MIR172d* or *e*. Under phosphate starvation, the MYB2 transcription factor binds to the promoter of *MIR399f* to promote its transcription (1, 10, 11).

MiRNA precursor processing

The dicing complex

Nascent pri-miRNAs are capped at the 5' end and polyadenylated at the 3' end, and intron-containing pri-miRNAs are spliced or alternatively spliced (2, 12-14). Pri-miRNAs are processed by the dicing complex, which contains DICER-LIKE1 (DCL1), HYPONASTIC LEAVES1 (HYL1) and SERRATE (SE) as core components, to yield mature miRNA/miRNA* duplexes (15-18) (Figure 1).

Of the four DCL RNase III family endonucleases in *Arabidopsis*, DCL1 is the predominant

miRNA precursor-processing enzyme (17, 18). DCL2, DCL3 and DCL4 produce various types of small interfering RNAs (siRNAs), including endogenous siRNAs as well as viral and transgene siRNAs (16, 19-22). A notable exception is that several young miRNAs, such as miR822 and miR839, are generated by DCL4 instead of DCL1 (23). DCL proteins appear to function as molecular rulers that measure and cleave small RNA duplexes at a specific length (24). DCL1 mainly processes pri-miRNAs in a base-to-loop manner in two steps. The first cut is 15-17 nt away from a bulge or unstructured region within the loop-distal stem. The resulting precursor-miRNA (pre-miRNA) is further cleaved by DCL1 to produce a 21-nt miRNA/miRNA* duplex (25-27). Alternative processing modes include loop-to-base processing (28).

In a five-member family of DOUBLE-STRANDED RNA-BINDING PROTEINS (DRBs), HYL1/DRB1 is a major miRNA biogenesis factor, and DRB2 affects the accumulation of a few miRNAs (29-31). HYL1 interacts with DCL1 to facilitate efficient and precise miRNA precursor processing (32-35), and HYL1 homo-dimerization is essential for this processing function (34, 36). HYL1 also affects the splicing of some pri-miRNAs and strand selection from miRNA/miRNA* duplexes in AGO1 loading (12, 35, 37).

Recent research has uncovered regulatory mechanisms impacting the activity, stability and nuclear localization of HYL1 in miRNA precursor processing (35, 38-41). C-TERMINAL DOMAIN PHOSPHATASE-LIKE (CPL) proteins dephosphorylate HYL1 to facilitate accurate miRNA precursor processing and strand selection during AGO loading (35). The

K homology (KH) domain protein REGULATOR OF CBF GENE EXPRESSION 3 (RCF3) promotes HYL1 dephosphorylation through interaction with CPL proteins (39). In addition, a functional PP4 (Protein Phosphatase 4) complex, composed of two catalytic phosphatases plus two regulatory subunits, stabilizes HYL1 by targeting it for dephosphorylation (42). This dephosphorylation is antagonized by the protein kinase MITOGEN-ACTIVATED PROTEIN KINASE 3 (MPK3) and members of the SNF1-related protein kinase 2 (SnRK2) subfamily (40, 43). Phospho-regulation affects not only HYL1 activity but also its protein stability; e.g., a *snrk2* mutation leads to reduced levels of HYL1 (43). HYL1 protein stability is regulated by the RING-finger E3 ligase CONSTITUTIVE PHOTOMORPHOGENIC 1 (COP1) through light signaling. Specifically, under light conditions, COP1 shifts to the cytoplasm and suppresses HYL1 cleavage by an unidentified protease, while in darkness, COP1 enters the nucleus and releases the protease that cleaves HYL1 (38). KETCH1 (KARYOPHERIN ENABLING THE TRANSPORT OF THE CYTOPLASMIC HYL1), a well-conserved importin-beta protein, transports HYL1 from the cytoplasm to the nucleus to form the dicing complex. *KETCH1* knockdown mutants show similar phenotypes as miRNA biogenesis mutants, including reduced miRNA levels and compromised pri-miRNA processing, indicating the importance of HYL1 nuclear localization in miRNA biogenesis (41).

In addition to DCL1 and HYL1, SE is also considered a core member of the miRNA processing complex in *Arabidopsis*. A mutation in *SE* results in reduced levels of mature miRNAs, increased levels of pri-miRNAs and defects in pri-miRNA splicing (44-47). As

a zinc-finger protein with RNA-binding activity, SE also functions outside of miRNA biogenesis. For example, SE interacts with U1 small nuclear ribonucleoprotein (snRNP) components (48). In contrast to DCL1 and HYL1, SE is distributed in a heterogeneous subnuclear pattern, reminiscent of nuclear speckles in which serine/arginine (SR) splicing factors are enriched (49). SE also affects the alternative splicing of some *Arabidopsis* mRNAs (44, 50).

Other proteins that influence miRNA precursor processing and/or *MIR* transcription

In the past decade, many proteins that influence miRNA precursor processing or *MIR* gene transcription have been identified (Figure 1). On the one hand, these proteins promote miRNA biogenesis in general, as most miRNAs accumulate to lower levels in mutants of these genes. On the other hand, the functions of these genes are not specific to miRNAs; in fact, many have functions in precursor mRNA (pre-mRNA) splicing. In terms of miRNA biogenesis, these proteins appear to act prior to or during pri-miRNA processing, as mutants in the corresponding genes have either higher or lower levels of pri-miRNAs (Table 1). Below, we categorize these proteins into two groups based on their effects on pri-miRNA accumulation and discuss their potential roles in miRNA biogenesis.

A large group of proteins appears to promote pri-miRNA processing, as loss/reduction-of-function mutants in the corresponding genes have reduced levels of mature miRNAs and increased levels of pri-miRNAs (Table 1). Proteins belonging to this group include CAP-BINDING PROTEIN 80 (CBP80) and CAP-BINDING PROTEIN 20 (CBP20) (44, 51);

STABILIZED 1 (STA1) (37); THO1/HPR1/EMU and THO2 (52, 53); SICKLE (54); TOUGH (55); PSR1-INTERACTING PROTEIN 1 (PINP1) (56); and MODIFIER OF SNC1, 2 (MOS2) (57). GLYCINE-RICH RNA-BINDING PROTEIN 7 (*AtGRP7*) may repress miRNA biogenesis, as its overexpression causes reduced levels of mature miRNAs and increased accumulation of pri-miRNAs (58). All these proteins have demonstrated or predicted ability to associate with or act on RNAs. CBP80 and CBP20 are subunits of the nuclear cap binding complex (CBC), which interacts with SE (44). The human and yeast homolog of STA1 is the U5 snRNP-associated protein Pre-mRNA-Processing Factor 6 (PRPF6), a confirmed splicing factor (37, 59). THO1/HPR1/EMU and THO2 are subunits of the THO/TREX (suppressor of the Transcriptional defects of *Hpr1* mutants by Overexpression/TRanscription-EXport) complex, a conserved multi-subunit complex involved in pre-mRNA co-transcriptional processing and mRNA export in yeast and animals (60). SICKLE is a plant-specific protein that interacts with many RNA-processing proteins (54, 61). TOUGH and MOS2 are RNA-binding proteins (55, 57). PINP1 is a putative RNA helicase (56). *AtGRP7* is a heterogeneous nuclear ribonucleoprotein (hnRNP)-like glycine-rich RNA-binding protein (58, 62). The molecular functions of these proteins in pri-miRNA processing are currently unknown.

A second group of proteins may act differently from the previously discussed group in miRNA biogenesis. Loss/reduction-of-function mutants in genes in this second group have reduced accumulation of both mature miRNAs and pri-miRNAs (Table 1). Proteins in this group include CDC5, NOT2, Elongator, PRL1 (PLEIOTROPIC REGULATORY LOCUS

1) and DDL (DAWDLE) (5-7, 63, 64). Although the reduced pri-miRNA accumulation in the mutants suggests a role of the proteins in *MIR* transcription or pri-miRNA stability, these proteins also seem to affect miRNA precursor processing. For example, mutants in *NOT2* show an increase in the number of dicing bodies, while those in *MOS2* and Elongator subunits mutants have a reduced number of dicing bodies. CDC5, NOT2, Elongator, PRL1 and DDL were all found to interact with DCL1 and may help recruit DCL1 to pri-miRNAs (5-7, 63, 64). Perhaps the most parsimonious hypothesis is that these proteins promote miRNA biogenesis by recruiting the dicing complex to nascent pri-miRNAs during transcription.

A prominent feature of these proteins, regardless of their effects on pri-miRNAs, is their demonstrated or predicted roles in splicing. Splicing defects in both pri-miRNAs and pre-mRNAs were detected in *abh1/cbp80*, *cbp20*, *sic-1* and *stal-1* mutants (44, 51). *AtGRP7* overexpression results in changes in pri-miRNA splicing (58, 62). *MOS2* is required for proper splicing of *SNCI* (*SUPPRESSOR OF NPRI-1*, *CONSTITUTIVE 1*), which encodes a Toll Interleukin 1 Receptor Nucleotide Binding Leucine-Rich Repeat (TIR-NB-LRR) class of protein involved in plant defense responses (65, 66). Alternatively spliced SR gene transcripts were detected in *tho1* and *tho2* mutants (52, 53). PRL1 belongs to the NineTeen Complex (NTC) or MOS4-associated complex (MAC), a protein complex with nineteen conserved members in yeast, human and plants. MAC is involved in spliceosome assembly and pre-mRNA splicing in all eukaryotic model organisms (67-69). *Arabidopsis* TOUGH colocalizes or interacts with splicing factor SR proteins, indicating a potential role in

general pre-mRNA splicing (55, 70). Although there is no direct evidence demonstrating PINP1 involvement in pre-mRNA processing in *Arabidopsis*, its yeast homolog, Prp16, is a confirmed splicing factor (71).

The large number of proteins acting in both RNA splicing and miRNA biogenesis begs the question of whether or how these two nuclear RNA processing events are related. Since some pri-miRNAs harbor introns (12), splicing may be an essential step in miRNA biogenesis. However, pri-miRNAs without introns are also affected in mutants in some of the aforementioned genes. For example, the levels of pri-miR159a, which contains no introns, were altered in the mutants of *CBC*, *STA1*, *PRL1*, *MOS2*, *THO2* and *SICKLE* (12, 37, 44, 52, 54, 57, 64, 65). At least for pri-miRNAs without introns, the aforementioned proteins cannot act in miRNA biogenesis through their functions in RNA splicing. Another formal possibility is that these proteins only act in splicing; in loss/reduction-of-function mutants of these genes, the accumulation of unspliced introns sequesters the dicing complex and thus inhibits miRNA biogenesis. It was found that a mutation in the intron lariat debranching gene *DBR1* (*LARIAT DEBRANCHING ENZYME 1*) results in the over-accumulation of intronic RNAs, which compete with pri-miRNAs for the dicing complex (72). However, many of the aforementioned proteins interact with the dicing complex or pri-miRNAs, which implies a direct role in miRNA biogenesis. Perhaps many of the proteins act broadly in nuclear RNA metabolism, with RNA splicing and miRNA biogenesis being two independent processes in which they participate.

MiRNA stabilization and RISC formation

The miRNA/miRNA* duplex is stabilized through 3'-terminal 2'-*O*-methylation by HEN1 (Figure 1). HEN1 was first discovered in *Arabidopsis* as a methyltransferase that specifically methylates small RNAs (73, 74). HEN1 homologs with similar functions were later discovered in other plants, animals and fungi (75-77). The crystal structure of an *Arabidopsis* HEN1-small RNA complex suggests that the small RNA duplex is bound by the HEN1 double-stranded RNA-binding domain (dsRBD), with one terminus at the methyltransferase (MTase) active site and methylated in a Mg^{2+} -dependent manner (78). A recent study suggested that HEN1 might interact with DCL1 and HYL1 based on yeast two-hybrid and *in vitro* pull-down assays (79). However, further *in vivo* analysis is needed to confirm this interaction.

During AGO loading, one strand of the small RNA duplex is selected as the guide strand while the passenger strand is removed (Figure 1). The current model for *Arabidopsis* RISC loading is as follows. 1) AGO1 and a dimer of HEAT SHOCK PROTEIN 90 (HSP90) form a complex. 2) The binding of adenosine triphosphate (ATP) to HSP90 causes a conformational change of AGO1 that allows the small RNA duplex to be incorporated into the AGO1-HSP90 protein complex. 3) ATP hydrolysis induces AGO1 dissociation from HSP90. 4) The AGO1 conformational change caused by HSP90 dissociation removes the passenger strand and results in a mature RISC (80).

Two AGO1-interacting importin-beta family proteins, ENHANCED MIRNA ACTIVITY 1 (EMA1) and TRANSPORTIN 1 (TRN1), negatively and positively regulate miRNA loading into AGO1, respectively. As importin-beta family proteins, the most intuitive hypothesis would be that they mediate AGO1 or miRNA nuclear-cytoplasmic shuttling. However, the nuclear-cytoplasmic partitioning of AGO1 and miRNAs is unchanged in these mutants (81, 82). Nevertheless, the fact that importin-beta family proteins affect the loading of miRNAs into AGO1 suggests that RISC formation occurs at specific subcellular locations.

The selection of miRNA guide strands is not random. In *Arabidopsis*, guide strand selection is known to be affected by miRNA precursor processing factors, the nature of the 5' end nucleotide and the structure of the small RNA duplex. HYL1 and the HYL1 phosphatase CPL1 facilitate guide strand selection, with *hyl1* and *cpl1* mutants exhibiting accumulated miRNA* strands (35, 83). The nature of the 5' nucleotides directs AGO loading. Most miRNA guide strands start with a 5'-terminal uridine and are incorporated into AGO1. In contrast, few miRNA* strands have a 5'-terminal uridine. MiRNA* strands with 5'-terminal adenosine are largely associated with AGO2, while those with 5'-terminal cytosine are associated with AGO5 (84). The loading of miRNAs into AGO proteins is also affected by the bulges in miRNA/miRNA* duplex structures. AGO2 favors miRNA duplexes without central mismatches, while AGO1 prefers duplexes with central mismatches, and the preference of AGO10 for miR165/6 relies on the internal base mismatches of the miR166 precursor (85, 86).

The cell biology of miRNA biogenesis: dicing bodies, nuclear export and RISC loading

The nucleus is the site of pri-miRNA processing in *Arabidopsis* (87). Live-cell imaging revealed that DCL1 and HYL1 colocalize in round and membrane-less nuclear bodies referred to as dicing bodies, which range in number from zero to four in each cell (49, 88, 89). DCL1 and HYL1 also exhibit diffuse patterns in the nucleoplasm but are excluded from nucleoli (49, 72). *In vivo* tracking of a pri-miRNA showed its colocalization with dicing bodies, indicating the role of dicing bodies in pri-miRNA processing (49). Dicing bodies resemble Cajal bodies in shape, size and number. However, colocalization analysis demonstrated that dicing bodies and Cajal bodies are distinct structures (49, 89) (Figure 1).

Different mechanisms have been proposed for the formation of membrane-less nuclear bodies such as dicing bodies, including stochastic assembly, ordered assembly and seeded assembly (90). Low complexity sequences, which are enriched in many RNA- and DNA-binding proteins, contribute to the formation of higher-order RNA- and protein-containing structures (91, 92). *Arabidopsis* DCL1 contains two dsRBDs. The second dsRBD is truncated in the *dcl1-9* mutant, which exhibits severe miRNA biogenesis defects (17, 93); moreover, the truncated DCL1-9 protein fails to localize to dicing bodies (49). Similarly, the N-terminal dsRBDs of HYL1 are essential for HYL1 localization to dicing bodies (94).

Many other miRNA biogenesis factors, such as SE, RCF3 and THO2, largely form splicing speckles and partially colocalize with dicing bodies (39, 49, 52). NOT2, MOS2 and PINP1 show diffuse nucleoplasmic patterns and also partially colocalize with dicing bodies (5, 56, 57). The subcellular localization patterns of the above factors indicate their roles in both miRNA and mRNA biogenesis. Dicing body formation is affected by several miRNA biogenesis factors. Mutants of *MOS2* and Elongator subunits have a reduced number of dicing bodies, while those of *PINP1*, *NOT2* and *DBR1* have more dicing bodies than wild-type; thus, opposite effects are observed although all of the above mutants have compromised miRNA levels (5, 7, 56, 57, 72).

The export of miRNAs from the nucleus to the cytoplasm is fundamental for miRNA activity (1, 95-97). Exportin 5, a RanGTP-dependent dsRNA-binding protein, mediates the nuclear export of pre-miRNAs in animals (96, 98, 99). In *Arabidopsis*, miRNA/miRNA* duplexes are probably excised from pre-miRNAs in the nucleus (as DCL1 acts in the nucleus) and are thought to be transported to the cytoplasm by HASTY (87, 97) (Figure 1). In the *hasty* mutant, the nuclear-cytoplasmic partitioning of miRNAs is not altered (97). Therefore, the functions of HASTY in miRNA nuclear export in *Arabidopsis* still require further investigation. As described above, THO/TREX complex components are required for miRNA biogenesis (52, 53). Since THO/TREX mediates transcription-coupled mRNA export through interactions with the nuclear pore complex, it is also possible that THO/TREX plays a role in miRNA export (95).

It is still unclear whether miRNAs are exported to the cytoplasm prior to RISC loading or if RISC loading precedes export to the cytoplasm. However, the involvement of cytoplasmic HSP90 in RISC loading is one line of evidence in favor of RISC loading in the cytoplasm (84, 100).

FIGURES

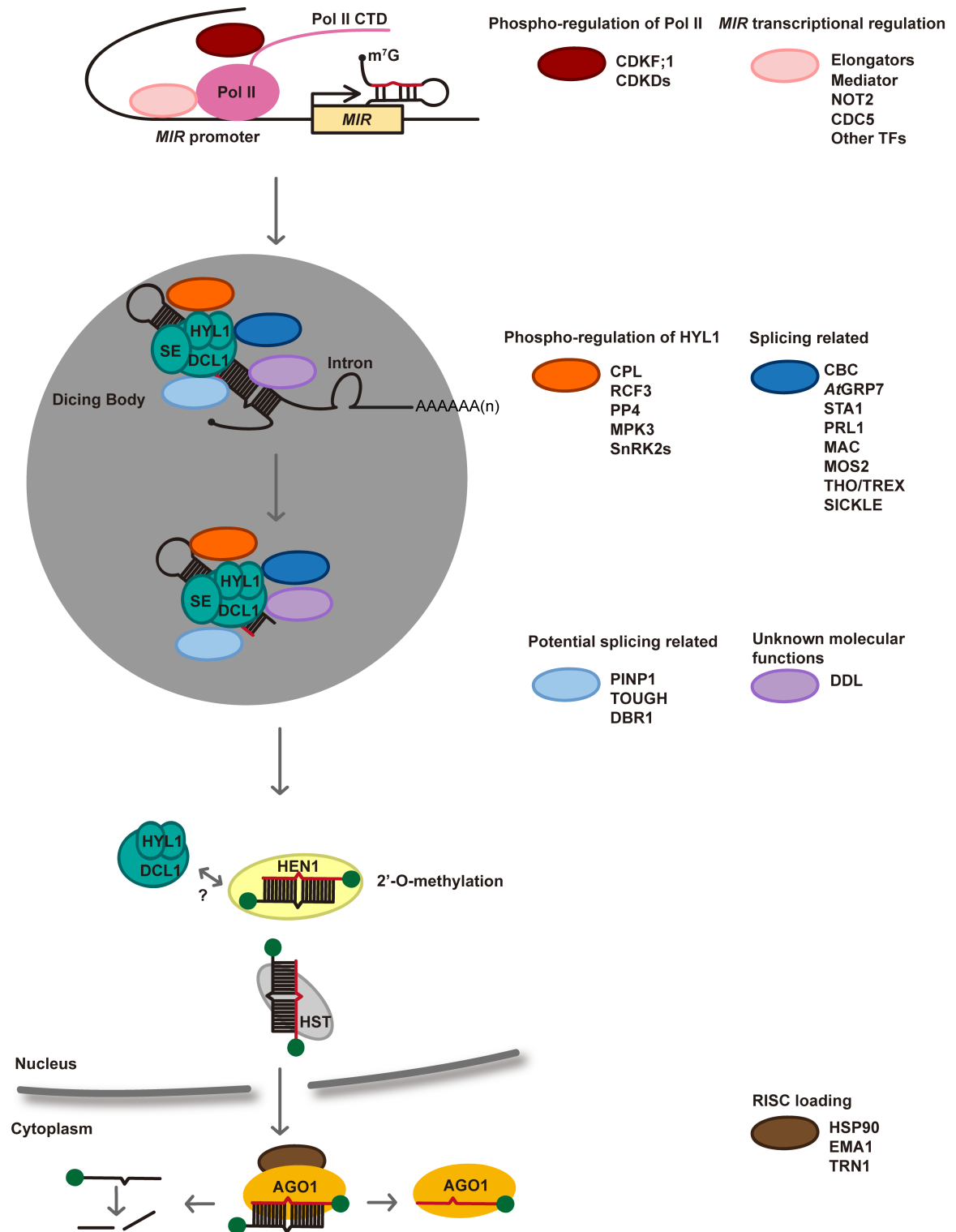


Figure 1. Illustrations of major steps in microRNA (miRNA) biogenesis.

RNA polymerase II (Pol II)-mediated miRNA gene (*MIR*) transcription is regulated by multiple transcription factors (TFs). Pol II activity itself is also subjected to phospho-regulation at its C-terminal domain (CTD). MiRNA precursors are processed at the dicing bodies by the dicing complex, which is mainly composed of DICER-LIKE 1 (DCL1), HYPONASTIC LEAVES 1 (HYL1) and SERRATE (SE). Many other protein factors contribute to miRNA precursor processing through phospho-regulation, RNA splicing and other unknown molecular mechanisms. It remains unclear whether the dicing complex interacts with HUA ENHANCER 1 (HEN1) (question mark) and contributes to miRNA/miRNA* duplex export and RNA-induced silencing complex (RISC) assembly. During RISC loading, one strand of the small RNA duplex is selected as the guide strand (red) and incorporated into ARGONAUTE 1 (AGO1) to form a functional RISC, while the other strand (the passenger strand) is removed and degraded. Proteins are color-coded according to their known molecular functions in phospho-regulation of Pol II (red), *MIR* transcription (pink), phospho-regulation of HYL1 (orange), splicing/RNA binding (dark blue), potential splicing/RNA binding (light blue) and RISC assembly (brown). The core dicing complex components are colored green, and the purple component indicates a protein with unknown molecular functions. m⁷G: 7-methylguanylate cap at the 5' end of primary miRNAs; CDKF;1: CYCLIN-DEPENDENT KINASE F;1; CDKDs: CYCLIN-DEPENDENT KINASE D; NOT2: NEGATIVE ON TATA LESS 2; CDC5: CELL DIVISION CYCLE 5; CPL: C-TERMINAL DOMAIN PHOSPHATASE-LIKE; RCF3: REGULATOR OF CBF GENE EXPRESSION 3; PP4: Protein Phosphatase 4 complex; MPK3: MITOGEN-ACTIVATED PROTEIN KINASE 3; SnRK2s: SNF1-related protein kinase 2 subfamily; CBC: cap binding complex; *At*GRP7: GLYCINE-RICH RNA-BINDING PROTEIN 7; STA1: STABILIZED 1; PRL1: PROTEIN PLEIOTROPIC REGULATORY LOCUS 1; MAC: MOS4-associated complex; MOS2: MODIFIER OF SNC1, 2; THO/TREX: suppressor of the Transcription defects of *Hpr1* mutants by Overexpression/TRanscription-EXport complex; PINP1: PSR1-INTERACTING PROTEIN 1; DBR1: LARIAT DEBRANCHING ENZYME 1; DDL: DAWDLE; HST: HASTY; HSP90: HEAT SHOCK PROTEIN 90; EMA1: ENHANCED MIRNA ACTIVITY 1; TRN1: TRANSPORTIN 1.

TABLES

Table 1. Proteins involved in miRNA biogenesis in *Arabidopsis*.

* Protein complexes are highlighted in boldface.

Proteins/Complex*	Human or Yeast Homolog	Key Domain(s)/Motif(s)	Functions in MiRNA Biogenesis	Pri-miRNA Levels in the Mutant	MiRNA Levels in the Mutant	Proved Interactions	(Predicted) Subcellular Localizations	References
Pol II	Pol II	N/A	Producing pri-miRNAs	Reduced	Reduced	CDC5, PRL1, NOT2a & NOT2b, Mediator	Nucleoplasm	(2, 4-6, 64, 101, 102)
CDKF;1	No homologs, plant specific	Kinase	Phosphorylation of serine in Pol II CTD, and regulating Pol II activity	Reduced	Reduced	Pol II CTD	Nucleoplasm	(8)
CDKDs	TFIIH-associated Kin28/CDK7 class	Kinase	Phosphorylation of serine in Pol II CTD, and regulating Pol II activity	Reduced	Reduced	Pol II CTD	Nucleoplasm	(8)
Mediator	Mediator	N/A	Transcriptional coactivator; Promoting	Reduced	Reduced	Pol II	Nucleoplasm	(4, 102)

			Pol II-mediated miRNA transcription					
Elongator	Elongator	N/A	Promoting transcription elongation and processing of pri-miRNAs	Reduced	Reduced	DCL1, SE, HYL1	Nucleoplasm	(7)
NOT2a & NOT2b	CNOT2	Gly-rich, Gln-rich	Transcription factors of <i>MIR</i>	Reduced	Reduced	DCL1, SE, CBP80, CBP20, Pol II	Nucleoplasm	(5)
CDC5	CDC5L	<i>MYB</i> domain	Transcription factor; Binding to <i>MIR</i> promoters and promoting transcription	Reduced	Reduced	DCL1, SE, Pol II	Nucleoplasm	(6)
DCL1	DICER	DExH/H-box-RNA helicase, DUF283 (dsRBD-like fold), PAZ, 2 × RNaseIII, 2 × dsRBD	Pri-miRNA cleavage and excision of miRNA/miRNA* duplex	Increased	Reduced	SE, HYL1, TGH, DDL1, HEN1, PRL1, CDC5, NOT2,	Nucleoplasm/Dicing bodies	(5-7, 32, 55, 63, 64, 79, 103)

						Elongator		
HYL1/DRB1	TRBP	2 × dsRBD	Accuracy of pri-miRNA cleavage by DCL1	Increased	Reduced	DCL1, SE, HYL1, TGH, HEN1, PRL1	Nucleoplasm/Dicing bodies	(32, 36, 46, 55, 64, 79)
RCF3	-	5 × KH domain	Interacting with CPL to regulate HYL1 phosphorylation	No change	Reduced	SE, HYL1, CPL1	Nucleoplasm	(39)
CPL1	-	FCP1, 2 × dsRBD	HYL1 dephosphorylation	Reduced	Reduced	SE, HYL1	Nucleoplasm	(35)
PP4	PP4	N/A	HYL1 dephosphorylation	Unknown	Reduced	HYL1	Nucleoplasm	(42)
MPK3	MAPK14, HOG1	Kinase	HYL1 phosphorylation	Unknown	Increased	HYL1	Nucleoplasm	(40)
SnRK2 Kinases	SNF1 protein kinases, AMP-activated protein kinase	Kinases	SE, HYL1 phosphorylation	Increased	Reduced	SE, HYL1	Nucleoplasm	(43)

COP1	RFWD2	RING-type zinc finger, 7 × WD repeats	Regulation of HYL1 stability	Unknown	Reduced	Unknown	Both nucleus and cytoplasm	(38)
KETCH1	Importin β	HEAT repeats	Imports HYL1 to nucleus for miRNA biogenesis	Increased	Reduced	HYL1	Both nucleus and cytoplasm	(41)
SE	ARS2	C2H2—zinc finger	Accuracy of pri-miRNA cleavage by DCL1 and pri-miRNA splicing regulation	Increased	Reduced	DCL1, HYL1, SE, CBP80, CBP20, TGH, CPL1, RACK1, PRL1, CDC5; NOT2, Elongator	Nucleoplasm	(5-7, 35, 46, 50, 55, 64, 103, 104)
CBC	CBC	N/A	Cap binding and pri-miRNA processing regulation	Increased	Reduced	SE, NOT2	Nucleoplasm	(5, 44)

<i>AtGRP7</i>	-	Gly-rich, RRM	Pri-miRNA processing regulation	Increased in GRP7-ox plants	Unclear	Unknown	Nucleoplasm	(58)
STA1	PRPF6	Ubiquitin-like, 3 × TPR repeats, 15 × HAT repeats	Pri-miRNA processing regulation and pri-miRNA intron splicing	Increased	Reduced	Unknown	Nucleoplasm	(37)
PRL1	PLRG1	7 × WD repeat	Pri-miRNA biogenesis and processing	Reduced	Reduced	DCL1, SE, HYL1, MAC	Nucleoplasm	(64)
MAC (including PRL1, CDC5, MAC3, MAC7 and other subunits)	MAC	N/A	Pri-miRNA biogenesis and processing	Reduced	Reduced	DCL1, SE, HYL1, Pol II	Nucleoplasm	(6, 64, 105)
MOS2	GRKOW	KOW, D111/G-patch	Pri-miRNA binding and processing	Increased	Reduced	No interaction with known miRNA biogenesis factors	Nucleoplasm	(57)

THO/TREX	THO/TR EX	N/A	MiRNA precursor processing	Increas ed	Reduc ed	No interacti on with known miRNA biogenes is factors	Nucleoplasm	(52, 53)
SICKLE	-	Proline-rich	Pri-miRNA processing regulation, pri-miRNA splicing	Increas ed	Reduc ed	Unclear	Co-localization with HYL1 in distinct nuclear bodies	(54)
PINP1	Prp16	(DEAH)-box RNA helicase	Pri-miRNA processing	Increas ed	Reduc ed	Unknow n	Nucleoplasm	(56)
TOUGH	-	G-patch, SWAP, Arg/Ser rich	Pri-miRNA and pre- miRNA binding, and processing regulation	Increas ed	Reduc ed	DCL1, SE, HYL1		(55)
DBR1	DBR1	Metallophosphoes terase domain, Lariat recognition loop	Pri-miRNA processing	Increas ed	Reduc ed	Unknow n	Nucleoplasm, partially colocalized with HYL1	(72)
DDL	SNIP1	FHA, Arg-rich	Pri-miRNA stabilization	Increas ed	Reduc ed	DCL1	Nucleoplasm	(63)
RACK1	GNB2L1	7 × WD repeat	Pri-miRNA processing	Variou s	Reduc ed	SE	Nucleoplasm	(104)

				changes				
HEN1	HENMT1	dsRBD, HTH La-type RNA binding	2'-O-methylation of 3' terminal ribose	Unclear	Reduced	DCL1, HYL1	Unclear	(74, 75, 79)
HST	Exportin-5 (XPO5)	ARM-like	Export miRNA/miRNA* from the nucleus to the cytoplasm	Unclear	Reduced	Unknown	Both nucleus and cytoplasm	(97)
HSP90	HSP family	-	Assisting miRNA loading into AGO1	Unclear	Unclear	AGO1	Cytoplasm	(80)
EMA1	Importin β	HEAT repeats	Modulating miRNA activity by preventing the loading of miRNAs into AGO1 complexes	Unclear	Increased	Unknown	Unknown	(82)
TRN1	Importin β	HEAT repeats	Positively regulating miRNA activity by promoting the association of	Reduced	No change	AGO1	Unknown	(81)

			miRNAs with AGO1					
AGO1	AGOs	PAZ, MID, PIWI domains	Binding to miRNAs to form functional RISC	Unclear	Reduced	TRN1, HSP90	Both nucleus and cytoplasm, mainly cytoplasm	(80, 81)

REFERENCES

1. K. Rogers, X. Chen, Biogenesis, turnover, and mode of action of plant microRNAs. *Plant Cell* **25**, 2383-2399 (2013).
2. Z. Xie *et al.*, Expression of Arabidopsis MIRNA genes. *Plant Physiol* **138**, 2145-2154 (2005).
3. M. Megraw *et al.*, MicroRNA promoter element discovery in Arabidopsis. *Rna* **12**, 1612-1619 (2006).
4. Y. J. Kim *et al.*, The role of Mediator in small and long noncoding RNA production in Arabidopsis thaliana. *Embo J* **30**, 814-822 (2011).
5. L. Wang *et al.*, NOT2 proteins promote polymerase II-dependent transcription and interact with multiple MicroRNA biogenesis factors in Arabidopsis. *Plant Cell* **25**, 715-727 (2013).
6. S. Zhang, M. Xie, G. Ren, B. Yu, CDC5, a DNA binding protein, positively regulates posttranscriptional processing and/or transcription of primary microRNA transcripts. *Proc Natl Acad Sci U S A* **110**, 17588-17593 (2013).
7. X. Fang, Y. Cui, Y. Li, Y. Qi, Transcription and processing of primary microRNAs are coupled by Elongator complex in Arabidopsis. *Nat Plants* **1**, 15075 (2015).
8. M. Hajheidari, S. Farrona, B. Huettel, Z. Koncz, C. Koncz, CDKF;1 and CDKD protein kinases regulate phosphorylation of serine residues in the C-terminal domain of Arabidopsis RNA polymerase II. *Plant Cell* **24**, 1626-1642 (2012).
9. M. Hajheidari, C. Koncz, D. Eick, Emerging roles for RNA polymerase II CTD in Arabidopsis. *Trends Plant Sci* **18**, 633-643 (2013).
10. D. Baek *et al.*, Regulation of miR399f transcription by AtMYB2 affects phosphate starvation responses in Arabidopsis. *Plant Physiol* **161**, 362-373 (2013).
11. R. E. Yumul *et al.*, POWERDRESS and diversified expression of the MIR172 gene family bolster the floral stem cell network. *PLoS Genet* **9**, e1003218 (2013).
12. B. Szarzynska *et al.*, Gene structures and processing of Arabidopsis thaliana HYL1-dependent pri-miRNAs. *Nucleic Acids Res* **37**, 3083-3093 (2009).
13. A. Zielezinski *et al.*, mirEX 2.0 - an integrated environment for expression profiling of plant microRNAs. *BMC Plant Biol* **15**, 144 (2015).
14. A. Stepień *et al.*, Posttranscriptional coordination of splicing and miRNA biogenesis in plants. *Wiley Interdiscip Rev RNA*, (2016).

15. Y. Kurihara, Y. Watanabe, Arabidopsis micro-RNA biogenesis through Dicer-like 1 protein functions. *Proc Natl Acad Sci U S A* **101**, 12753-12758 (2004).
16. A. Fukudome, T. Fukuhara, Plant dicer-like proteins: double-stranded RNA-cleaving enzymes for small RNA biogenesis. *J Plant Res* **130**, 33-44 (2017).
17. W. Park, J. Li, R. Song, J. Messing, X. Chen, CARPEL FACTORY, a Dicer homolog, and HEN1, a novel protein, act in microRNA metabolism in Arabidopsis thaliana. *Curr Biol* **12**, 1484-1495 (2002).
18. B. J. Reinhart, E. G. Weinstein, M. W. Rhoades, B. Bartel, D. P. Bartel, MicroRNAs in plants. *Genes Dev* **16**, 1616-1626 (2002).
19. N. Bouche, D. Laressergues, V. Gascioli, H. Vaucheret, An antagonistic function for Arabidopsis DCL2 in development and a new function for DCL4 in generating viral siRNAs. *Embo J* **25**, 3347-3356 (2006).
20. V. Gascioli, A. C. Mallory, D. P. Bartel, H. Vaucheret, Partially redundant functions of Arabidopsis DICER-like enzymes and a role for DCL4 in producing trans-acting siRNAs. *Curr Biol* **15**, 1494-1500 (2005).
21. S. Mlotshwa *et al.*, DICER-LIKE2 plays a primary role in transitive silencing of transgenes in Arabidopsis. *PLoS One* **3**, e1755 (2008).
22. Z. Xie, E. Allen, A. Wilken, J. C. Carrington, 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 (2005).
23. R. Rajagopalan, H. Vaucheret, J. Trejo, D. P. Bartel, A diverse and evolutionarily fluid set of microRNAs in Arabidopsis thaliana. *Genes Dev* **20**, 3407-3425 (2006).
24. I. J. Macrae *et al.*, Structural basis for double-stranded RNA processing by Dicer. *Science* **311**, 195-198 (2006).
25. L. Song, M. J. Axtell, N. V. Fedoroff, RNA secondary structural determinants of miRNA precursor processing in Arabidopsis. *Curr Biol* **20**, 37-41 (2010).
26. C. Liu, M. J. Axtell, N. V. Fedoroff, The helicase and RNaseIIIa domains of Arabidopsis Dicer-Like1 modulate catalytic parameters during microRNA biogenesis. *Plant Physiol* **159**, 748-758 (2012).
27. H. Zhu *et al.*, Bidirectional processing of pri-miRNAs with branched terminal loops by Arabidopsis Dicer-like1. *Nat Struct Mol Biol* **20**, 1106-1115 (2013).

28. N. G. Bologna, J. L. Mateos, E. G. Bresso, J. F. Palatnik, A loop-to-base processing mechanism underlies the biogenesis of plant microRNAs miR319 and miR159. *Embo J* **28**, 3646-3656 (2009).
29. S. J. Curtin *et al.*, The roles of plant dsRNA-binding proteins in RNAi-like pathways. *FEBS Lett* **582**, 2753-2760 (2008).
30. A. L. Eamens, K. W. Kim, S. J. Curtin, P. M. Waterhouse, DRB2 is required for microRNA biogenesis in *Arabidopsis thaliana*. *PLoS One* **7**, e35933 (2012).
31. A. Hiraguri *et al.*, Specific interactions between Dicer-like proteins and HYL1/DRB-family dsRNA-binding proteins in *Arabidopsis thaliana*. *Plant Mol Biol* **57**, 173-188 (2005).
32. Y. Kurihara, Y. Takashi, Y. Watanabe, The interaction between DCL1 and HYL1 is important for efficient and precise processing of pri-miRNA in plant microRNA biogenesis. *Rna* **12**, 206-212 (2006).
33. Z. Dong, M. H. Han, N. Fedoroff, The RNA-binding proteins HYL1 and SE promote accurate in vitro processing of pri-miRNA by DCL1. *Proc Natl Acad Sci U S A* **105**, 9970-9975 (2008).
34. X. Yang *et al.*, Homodimerization of HYL1 ensures the correct selection of cleavage sites in primary miRNA. *Nucleic Acids Res* **42**, 12224-12236 (2014).
35. P. A. Manavella *et al.*, Fast-forward genetics identifies plant CPL phosphatases as regulators of miRNA processing factor HYL1. *Cell* **151**, 859-870 (2012).
36. S. W. Yang *et al.*, Structure of *Arabidopsis* HYPONASTIC LEAVES1 and its molecular implications for miRNA processing. *Structure* **18**, 594-605 (2010).
37. S. Ben Chaabane *et al.*, STA1, an *Arabidopsis* pre-mRNA processing factor 6 homolog, is a new player involved in miRNA biogenesis. *Nucleic Acids Res* **41**, 1984-1997 (2013).
38. S. K. Cho, S. Ben Chaabane, P. Shah, C. P. Poulsen, S. W. Yang, COP1 E3 ligase protects HYL1 to retain microRNA biogenesis. *Nat Commun* **5**, 5867 (2014).
39. P. Karlsson *et al.*, KH domain protein RCF3 is a tissue-biased regulator of the plant miRNA biogenesis cofactor HYL1. *Proc Natl Acad Sci U S A* **112**, 14096-14101 (2015).
40. B. Raghuram, A. H. Sheikh, Y. Rustagi, A. K. Sinha, MicroRNA biogenesis factor DRB1 is a phosphorylation target of mitogen activated protein kinase MPK3 in both rice and *Arabidopsis*. *FEBS J* **282**, 521-536 (2015).

41. Z. Zhang *et al.*, KETCH1 imports HYL1 to nucleus for miRNA biogenesis in Arabidopsis. *Proc Natl Acad Sci U S A* **114**, 4011-4016 (2017).
42. C. Su *et al.*, The Protein Phosphatase 4 and SMEK1 Complex Dephosphorylates HYL1 to Promote miRNA Biogenesis by Antagonizing the MAPK Cascade in Arabidopsis. *Dev Cell* **41**, 527-539 e525 (2017).
43. J. Yan *et al.*, The SnRK2 kinases modulate miRNA accumulation in Arabidopsis. *PLoS Genet* **13**, e1006753 (2017).
44. S. Laubinger *et al.*, Dual roles of the nuclear cap-binding complex and SERRATE in pre-mRNA splicing and microRNA processing in Arabidopsis thaliana. *Proc Natl Acad Sci U S A* **105**, 8795-8800 (2008).
45. S. P. Grigg, C. Canales, A. Hay, M. Tsiantis, SERRATE coordinates shoot meristem function and leaf axial patterning in Arabidopsis. *Nature* **437**, 1022-1026 (2005).
46. L. Yang, Z. Liu, F. Lu, A. Dong, H. Huang, SERRATE is a novel nuclear regulator in primary microRNA processing in Arabidopsis. *Plant J* **47**, 841-850 (2006).
47. D. Lobbes, G. Rallapalli, D. D. Schmidt, C. Martin, J. Clarke, SERRATE: a new player on the plant microRNA scene. *EMBO Rep* **7**, 1052-1058 (2006).
48. K. Knop *et al.*, Active 5' splice sites regulate the biogenesis efficiency of Arabidopsis microRNAs derived from intron-containing genes. *Nucleic Acids Res*, (2016).
49. Y. Fang, D. L. Spector, Identification of nuclear dicing bodies containing proteins for microRNA biogenesis in living Arabidopsis plants. *Curr Biol* **17**, 818-823 (2007).
50. K. D. Raczynska *et al.*, The SERRATE protein is involved in alternative splicing in Arabidopsis thaliana. *Nucleic Acids Res* **42**, 1224-1244 (2014).
51. T. Gonatopoulos-Pournatzis, V. H. Cowling, Cap-binding complex (CBC). *Biochem J* **457**, 231-242 (2014).
52. A. G. Francisco-Mangilet *et al.*, THO2, a core member of the THO/TREX complex, is required for microRNA production in Arabidopsis. *Plant J* **82**, 1018-1029 (2015).
53. C. Furumizu, H. Tsukaya, Y. Komeda, Characterization of EMU, the Arabidopsis homolog of the yeast THO complex member HPR1. *Rna* **16**, 1809-1817 (2010).

54. X. Zhan *et al.*, Arabidopsis proline-rich protein important for development and abiotic stress tolerance is involved in microRNA biogenesis. *Proc Natl Acad Sci U S A* **109**, 18198-18203 (2012).
55. G. Ren *et al.*, Regulation of miRNA abundance by RNA binding protein TOUGH in Arabidopsis. *Proc Natl Acad Sci U S A* **109**, 12817-12821 (2012).
56. Y. Qiao, J. Shi, Y. Zhai, Y. Hou, W. Ma, Phytophthora effector targets a novel component of small RNA pathway in plants to promote infection. *Proc Natl Acad Sci U S A* **112**, 5850-5855 (2015).
57. X. Wu *et al.*, A role for the RNA-binding protein MOS2 in microRNA maturation in Arabidopsis. *Cell Res* **23**, 645-657 (2013).
58. T. Koster *et al.*, Regulation of pri-miRNA processing by the hnRNP-like protein AtGRP7 in Arabidopsis. *Nucleic Acids Res* **42**, 9925-9936 (2014).
59. C. L. Will, R. Luhrmann, Spliceosome structure and function. *Cold Spring Harb Perspect Biol* **3**, (2011).
60. C. G. Heath, N. Viphakone, S. A. Wilson, The role of TREX in gene expression and disease. *Biochem J* **473**, 2911-2935 (2016).
61. C. M. Marshall, V. Tartaglio, M. Duarte, F. G. Harmon, The Arabidopsis sickle Mutant Exhibits Altered Circadian Clock Responses to Cool Temperatures and Temperature-Dependent Alternative Splicing. *Plant Cell* **28**, 2560-2575 (2016).
62. C. Streitner *et al.*, An hnRNP-like RNA-binding protein affects alternative splicing by in vivo interaction with transcripts in Arabidopsis thaliana. *Nucleic Acids Res* **40**, 11240-11255 (2012).
63. B. Yu *et al.*, The FHA domain proteins DAWDLE in Arabidopsis and SNIP1 in humans act in small RNA biogenesis. *Proc Natl Acad Sci U S A* **105**, 10073-10078 (2008).
64. S. Zhang, Y. Liu, B. Yu, PRL1, an RNA-binding protein, positively regulates the accumulation of miRNAs and siRNAs in Arabidopsis. *PLoS Genet* **10**, e1004841 (2014).
65. C. Copeland, S. Xu, Y. Qi, X. Li, MOS2 has redundant function with its homolog MOS2H and is required for proper splicing of SNC1. *Plant Signal Behav* **8**, (2013).
66. Y. Zhang, Y. T. Cheng, D. Bi, K. Palma, X. Li, MOS2, a protein containing G-patch and KOW motifs, is essential for innate immunity in Arabidopsis thaliana. *Curr Biol* **15**, 1936-1942 (2005).

67. C. Koncz, F. Dejong, N. Villacorta, D. Szakonyi, Z. Koncz, The spliceosome-activating complex: molecular mechanisms underlying the function of a pleiotropic regulator. *Front Plant Sci* **3**, 9 (2012).
68. J. Monaghan *et al.*, Two Prp19-like U-box proteins in the MOS4-associated complex play redundant roles in plant innate immunity. *PLoS Pathog* **5**, e1000526 (2009).
69. K. C. M. Johnson, O. X. Dong, X. Li, The evolutionarily conserved MOS4-associated complex. *Central European Journal of Biology* **6**, 776-784 (2011).
70. L. I. Calderon-Villalobos *et al.*, The evolutionarily conserved TOUGH protein is required for proper development of *Arabidopsis thaliana*. *Plant Cell* **17**, 2473-2485 (2005).
71. Y. Wang, J. D. Wagner, C. Guthrie, The DEAH-box splicing factor Prp16 unwinds RNA duplexes in vitro. *Curr Biol* **8**, 441-451 (1998).
72. Z. Li *et al.*, Intron Lariat RNA Inhibits MicroRNA Biogenesis by Sequestering the Dicing Complex in *Arabidopsis*. *PLoS Genet* **12**, e1006422 (2016).
73. B. Yu *et al.*, Methylation as a crucial step in plant microRNA biogenesis. *Science* **307**, 932-935 (2005).
74. Z. Yang, Y. W. Ebright, B. Yu, X. Chen, HEN1 recognizes 21-24 nt small RNA duplexes and deposits a methyl group onto the 2' OH of the 3' terminal nucleotide. *Nucleic Acids Res* **34**, 667-675 (2006).
75. R. H. Huang, Unique 2'-O-methylation by Hen1 in eukaryotic RNA interference and bacterial RNA repair. *Biochemistry* **51**, 4087-4095 (2012).
76. Y. Kirino, Z. Mourelatos, The mouse homolog of HEN1 is a potential methylase for Piwi-interacting RNAs. *Rna* **13**, 1397-1401 (2007).
77. K. Saito *et al.*, Pimet, the *Drosophila* homolog of HEN1, mediates 2'-O-methylation of Piwi- interacting RNAs at their 3' ends. *Genes Dev* **21**, 1603-1608 (2007).
78. Y. Huang *et al.*, Structural insights into mechanisms of the small RNA methyltransferase HEN1. *Faseb Journal* **24**, (2010).
79. S. Baranauske *et al.*, Functional mapping of the plant small RNA methyltransferase: HEN1 physically interacts with HYL1 and DICER-LIKE 1 proteins. *Nucleic Acids Res* **43**, 2802-2812 (2015).
80. T. Iki *et al.*, In vitro assembly of plant RNA-induced silencing complexes facilitated by molecular chaperone HSP90. *Mol Cell* **39**, 282-291 (2010).

81. Y. Cui, X. Fang, Y. Qi, TRANSPORTIN1 Promotes the Association of MicroRNA with ARGONAUTE1 in Arabidopsis. *Plant Cell* **28**, 2576-2585 (2016).
82. W. Wang *et al.*, An importin beta protein negatively regulates MicroRNA activity in Arabidopsis. *Plant Cell* **23**, 3565-3576 (2011).
83. A. L. Eamens, N. A. Smith, S. J. Curtin, M. B. Wang, P. M. Waterhouse, The Arabidopsis thaliana double-stranded RNA binding protein DRB1 directs guide strand selection from microRNA duplexes. *Rna* **15**, 2219-2235 (2009).
84. S. Mi *et al.*, Sorting of small RNAs into Arabidopsis argonaute complexes is directed by the 5' terminal nucleotide. *Cell* **133**, 116-127 (2008).
85. H. Zhu *et al.*, Arabidopsis Argonaute10 specifically sequesters miR166/165 to regulate shoot apical meristem development. *Cell* **145**, 242-256 (2011).
86. G. Ren *et al.*, Methylation protects microRNAs from an AGO1-associated activity that uridylates 5' RNA fragments generated by AGO1 cleavage. *Proc Natl Acad Sci U S A* **111**, 6365-6370 (2014).
87. I. Papp *et al.*, Evidence for nuclear processing of plant micro RNA and short interfering RNA precursors. *Plant Physiol* **132**, 1382-1390 (2003).
88. M. H. Han, S. Goud, L. Song, N. Fedoroff, The Arabidopsis double-stranded RNA-binding protein HYL1 plays a role in microRNA-mediated gene regulation. *Proc Natl Acad Sci U S A* **101**, 1093-1098 (2004).
89. L. Song, M. H. Han, J. Lesicka, N. Fedoroff, Arabidopsis primary microRNA processing proteins HYL1 and DCL1 define a nuclear body distinct from the Cajal body. *Proc Natl Acad Sci U S A* **104**, 5437-5442 (2007).
90. Y. S. Mao, B. Zhang, D. L. Spector, Biogenesis and function of nuclear bodies. *Trends Genet* **27**, 295-306 (2011).
91. T. W. Han *et al.*, Cell-free formation of RNA granules: bound RNAs identify features and components of cellular assemblies. *Cell* **149**, 768-779 (2012).
92. M. Kato *et al.*, Cell-free formation of RNA granules: low complexity sequence domains form dynamic fibers within hydrogels. *Cell* **149**, 753-767 (2012).
93. S. E. Schauer, S. E. Jacobsen, D. W. Meinke, A. Ray, DICER-LIKE1: blind men and elephants in Arabidopsis development. *Trends Plant Sci* **7**, 487-491 (2002).
94. F. Wu *et al.*, The N-terminal double-stranded RNA binding domains of Arabidopsis HYPONASTIC LEAVES1 are sufficient for pre-microRNA processing. *Plant Cell* **19**, 914-925 (2007).

95. A. Kohler, E. Hurt, Exporting RNA from the nucleus to the cytoplasm. *Nat Rev Mol Cell Biol* **8**, 761-773 (2007).
96. E. Lund, S. Guttinger, A. Calado, J. E. Dahlberg, U. Kutay, Nuclear export of microRNA precursors. *Science* **303**, 95-98 (2004).
97. M. Y. Park, G. Wu, A. Gonzalez-Sulser, H. Vaucheret, R. S. Poethig, Nuclear processing and export of microRNAs in Arabidopsis. *Proc Natl Acad Sci U S A* **102**, 3691-3696 (2005).
98. R. Yi, Y. Qin, I. G. Macara, B. R. Cullen, Exportin-5 mediates the nuclear export of pre-microRNAs and short hairpin RNAs. *Genes Dev* **17**, 3011-3016 (2003).
99. M. T. Bohnsack, Exportin 5 is a RanGTP-dependent dsRNA-binding protein that mediates nuclear export of pre-miRNAs. *Rna* **10**, 185-191 (2004).
100. P. Krishna, G. Gloor, The Hsp90 family of proteins in Arabidopsis thaliana. *Cell Stress Chaperones* **6**, 238-246 (2001).
101. B. Zheng *et al.*, Intergenic transcription by RNA polymerase II coordinates Pol IV and Pol V in siRNA-directed transcriptional gene silencing in Arabidopsis. *Genes Dev* **23**, 2850-2860 (2009).
102. Y. J. Kim, X. Chen, The plant Mediator and its role in noncoding RNA production. *Front Biol (Beijing)* **6**, 125-132 (2011).
103. S. Machida, H. Y. Chen, Y. A. Yuan, Molecular insights into miRNA processing by Arabidopsis thaliana SERRATE. *Nucleic Acids Research* **39**, 7828-7836 (2011).
104. C. Speth, E. M. Willing, S. Rausch, K. Schneeberger, S. Laubinger, RACK1 scaffold proteins influence miRNA abundance in Arabidopsis. *Plant J* **76**, 433-445 (2013).
105. C. J. Zhang *et al.*, The splicing machinery promotes RNA-directed DNA methylation and transcriptional silencing in Arabidopsis. *Embo J* **32**, 1128-1140 (2013).

CHAPTER 1

The *Arabidopsis* MOS4-associated complex promotes microRNA biogenesis and precursor messenger RNA splicing

ABSTRACT

In *Arabidopsis thaliana*, the MOS4-ASSOCIATED COMPLEX (MAC) is required for defense and development. The evolutionarily conserved, putative RNA helicase MAC7 is a component of the *Arabidopsis* MAC and the human MAC7 homolog, Aquarius, is implicated in pre-mRNA splicing. Here, we show that *mac7-1*, a partial loss-of-function mutant in *MAC7*, and two other MAC subunit mutants, *mac3a mac3b* and *pleiotropic regulatory locus 1 pleiotropic regulatory locus 2 (prl1 prl2)*, exhibit reduced microRNA (miRNA) levels, indicating that MAC promotes miRNA biogenesis. The *mac7-1* mutant shows reduced primary miRNA (pri-miRNA) levels without affecting miRNA gene (*MIR*) promoter activity or the half-life of pri-miRNA transcripts. As a nuclear protein, MAC7 is not concentrated in dicing bodies, but it affects the localization of HYPONASTIC LEAVES1 (HYL1), a key protein in pri-miRNA processing, to dicing bodies. Immunoprecipitation of HYL1 retrieved eleven known MAC subunits, including MAC7, indicating association between HYL1 and MAC. We propose that MAC7 links *MIR* transcription to pri-miRNA processing. RNA-seq analysis showed that down-regulated genes in MAC subunit mutants are mostly involved in plant defense and stimulus responses, confirming a role of MAC in biotic and abiotic stress responses. We also

discovered global intron retention defects in mutants in three subunits of MAC, thus linking MAC function to splicing in *Arabidopsis*.

INTRODUCTION

MicroRNAs (miRNAs) are a class of small RNAs that are approximately 20 to 24 nucleotides in length and act as post-transcriptional regulators of gene expression in both animals and plants. MiRNAs are processed from hairpin-containing precursors, primary miRNAs (pri-miRNAs), by RNase III family enzymes. A mature miRNA is loaded into an Argonaute protein to form a silencing complex and guides this silencing complex to target RNAs through sequence complementarity with target RNAs to result in their degradation or translational repression (1).

The plant miRNA pathway has been intensively studied in the past decade. Early studies identified key proteins with catalytic activities, including RNA polymerase II (Pol II) that transcribes miRNA genes (*MIR*) (2, 3), DICER-LIKE1 (DCL1), an RNase III family enzyme excising miRNAs from stem-loop precursors, HUA ENHANCER1 (HEN1), a methyltransferase stabilizing miRNAs by 2'-*O*-methylation (4-8), and ARGONAUTE 1 (AGO1), the enzyme mediating miRNA-guided target RNA cleavage (9-12). Many factors that assist in the transcription and processing steps of miRNA biogenesis have been identified (1, 13). Pol II-mediated *MIR* transcription requires the transcriptional coactivator Mediator (14) and the transcription factor NEGATIVE ON TATA LESS 2 (NOT2) (15), and is regulated by CYCLIN-DEPENDENT KINASES (CDKs) (16). In addition, the DNA binding protein CELL DIVISION CYCLE 5 (CDC5) (17) and Elongator (18) also regulate *MIR* transcription by Pol II.

While interacting with Pol II, NOT2, CDC5 and Elongator also interact with DCL1 and several DCL1-interacting proteins, and probably recruit them to pri-miRNAs to facilitate their processing (15, 17, 18). During the processing of miRNA precursors, DCL1 forms a dicing complex with the dsRNA binding domain protein HYPONASTIC LEAVES1 (HYL1/DRB1), and the zinc finger protein SERRATE (SE) (19-23). DCL1 and HYL1 are enriched in subnuclear bodies, referred to as dicing bodies, which are considered to be sites of miRNA precursor processing (20, 24, 25). Many other proteins also interact with DCL1 directly or indirectly in miRNA biogenesis, such as Cap-Binding Proteins (CBPs) (26, 27), the forkhead-associated domain containing protein DAWDLE (DDL) (28), the G-patch domain containing RNA binding protein TOUGH (TGH) (29), and the WD-40 protein PLEIOTROPIC REGULATORY LOCUS 1 (PRL1) (30). In addition, several proteins act in miRNA biogenesis through the regulation of HYL1 phosphorylation, such as C-TERMINAL DOMAIN PHOSPHATASE-LIKE (CPL) proteins and a K homology (KH) domain protein REGULATOR OF CBF GENE EXPRESSION 3 (RCF3) (31, 32). Several other factors, including two core members of the THO/TREX complex THO2 and EMU (33, 34), and the RNA binding protein MODIFIER OF SNC1, 2 (MOS2) (35) do not seem to interact with any known dicing complex components, but still affect miRNA biogenesis.

Among the known miRNA biogenesis factors, CDC5 and PRL1 belong to the same complex, the MOS4-associated complex (MAC). MAC is a highly conserved complex among eukaryotes, with its orthologs known as the NINETEEN COMPLEX (NTC) or

Prp19 complex (Prp19C) in yeast and humans. The *Arabidopsis thaliana* MAC, yeast and human NTC/Prp19C all associate with the spliceosome and are predicted to share conserved functions in splicing in all three systems (36-39). It is also known that the yeast MAC, Prp19C, functions in transcription elongation, as Pol II processivity is decreased in Prp19C subunit mutants. *Arabidopsis* CDC5 is a MYB-related transcription factor and PRL1 is a conserved nuclear WD-40 protein. As core components of MAC, they regulate plant development and immunity through molecular mechanisms that remain unclear (40-42). They both promote miRNA biogenesis but may have distinct molecular functions. CDC5 binds to *MIR* promoters and interacts with DCL1 and SE to enhance miRNA biogenesis (17), while PRL1 may stabilize pri-miRNAs through its RNA binding activity and enhance DCL1 activity (30). Other *Arabidopsis* MAC subunits include two homologous proteins MAC3A and MAC3B (MAC3B was shown to have E3 ligase activity *in vitro*), MAC7/Aquarius (an RNA helicase), and more than ten other members (37, 38, 43, 44).

Arabidopsis MAC7 is a putative RNA helicase, which is highly conserved in eukaryotes. The human MAC7 homolog is the intron-binding protein Aquarius (IBP160/AQR), which has ATPase and RNA helicase activities. It associates with the spliceosome and contributes to efficient precursor-mRNA splicing *in vitro* (45, 46). Recent publications show that EMB-4, the *Caenorhabditis elegans* homolog of MAC7, physically interacts with germline AGOs. It participates in the nuclear RNAi pathway and maintains the homeostasis of germline transcriptome in worms (47, 48). *MAC7* was predicted to be an essential gene for

embryo development (*EMB* gene) in *Arabidopsis* based on sequence similarity with *EMB* genes found in other eukaryotes (49). However, the molecular function of *MAC7* remains unknown.

In this study, we performed a genetic screen using a miRNA activity reporter line, the *pSUC2:amiR-SUL* (*amiR-SUL*) transgenic line (50). The expression of *SUCROSE-PROTON SYMPORTER 2* (*SUC2*) promoter-driven artificial miRNA targeting the *CHLORINA42* (*CH42*) gene creates a bleached phenotype along the leaf veins. This bleaching of mesophyll cells caused by the silencing effects of the artificial miRNA results in an easily scorable phenotype reflecting miRNA activities in plants (50). From this screen, we identified a point mutation in *MAC7*, *mac7-1*, as a miRNA activity suppressor mutant. We showed that *MAC7* affects miRNA accumulation through promotion of pri-miRNA biogenesis in *Arabidopsis*. We found that in other MAC subunit mutants, including *mac3a mac3b* and *prl1 prl2* double mutants, miRNA biogenesis is also compromised, indicating that MAC participates in miRNA biogenesis as a complex. Consistent with this, HYL1 immunoprecipitation mass spectrometry analyses revealed that HYL1 associates with MAC *in vivo*. In addition, we uncovered global intron retention defects in *mac7-1*, *mac3a mac3b* and *prl1 prl2* mutants through RNA-seq analysis. Our molecular characterization of *MAC7* and its associated MAC components revealed their functions in miRNA biogenesis and pre-mRNA splicing, which could possibly explain their roles in plant development and stress responses.

RESULTS

A silencing suppressor mutant exhibits pleiotropic phenotypes and reduced miRNA levels

To identify new players of the miRNA pathway, we performed an ethylmethane sulfonate (EMS) mutagenesis screen using the *amiR-SUL* line (50). From this screen, we identified a mutant, which we named *mac7-1* based on subsequent characterization (*amiR-SUL mac7-1*) with a reduced area of bleaching along the veins, indicating compromised miRNA activity (Figure 1.1A). The mutant has pleiotropic phenotypes, such as pointed leaves, reduced root length, reduced number of lateral roots, smaller plant stature and reduced fertility (Figure 1.1B, C and Figure 1.2). We crossed *amiR-SUL mac7-1* with wild type (Col-0) plants to remove the *amiR-SUL* transgene. We found that the *mac7-1* mutant shows the same range of phenotypes in the Col-0 background as in the *amiR-SUL* background (Figure 1.1B, C and Figure 1.2).

Because the mutant shows compromised amiR-SUL activities, we speculated that reduced miRNA accumulation could be a reason. We performed RNA gel blot analyses to detect amiR-SUL as well as many endogenous miRNAs. We found that amiR-SUL, miR156, miR171, miR390 and many other miRNAs showed reduced accumulation in both inflorescences and seedlings. The reduction in miRNA abundance was small but reproducible in several biological replicates (Figure 1.1D and Figure 1.3). To assess the global influence of the mutation on small RNAs, we also performed small RNA-seq with

Col-0 and *mac7-1* seedlings. We found that 21-nt and 24-nt small RNAs, which represent the two most abundant small RNA size classes, showed a significant global reduction in the *mac7-1* mutant. miRNAs showed a slight global reduction in the *mac7-1* mutant but the reduction was not statistically significant (Figure 1.1E). Most miRNAs that were found to show reduced abundance by RNA gel blotting also had lower levels in *mac7-1* in small RNA-seq, although a few miRNAs did not (e.g. miR156, miR164) (Figure 1.3C). The minor inconsistency between RNA gel blotting and small RNA-seq could be caused by technical limitations in small RNA-seq, e.g. bias in RNA adaptor ligation or PCR amplification. Since small RNA-seq entails more procedures that are prone to bias, we believe that the RNA gel blotting results are more accurate.

We examined the expression of *CH42* and eight known miRNA target genes in *amiR-SUL* and *amiR-SUL mac7-1*. Opposite to the reduced miRNA accumulation, the expression levels of *CH42* and seven endogenous miRNA targets were upregulated in *amiR-SUL mac7-1* (Figure 1.1F).

The reduced bleaching phenotype in *amiR-SUL mac7-1* might be attributed to other factors other than, or in addition to, reduced amiR-SUL activity, e.g. enhanced transcription of *CH42*. We performed RT-qPCR and immunoblotting to examine *CH42* mRNA and protein levels, respectively, in Col-0 and *mac7-1* to determine whether an amiR-SUL-independent effect of the *mac7-1* mutation was present. While *CH42* mRNA and protein levels were both increased in *amiR-SUL mac7-1* in comparison to *amiR-SUL*, no difference was found

between Col-0 and *mac7-1* (Figure 1.4), indicating that the reduced bleaching phenotype and elevated *CH42* expression in *amiR-SUL mac7-1* was most likely caused by changes in amiR-SUL activity only.

A point mutation in *MAC7* is responsible for the suppression of amiR-SUL-induced silencing

Through genome re-sequencing of pooled F2 mutants from a backcross between *amiR-SUL mac7-1* and the parental *amiR-SUL* line, we found a C-to-T nucleotide transition that changes a conserved glutamic acid to lysine (E1131K) in the open reading frame (ORF) of At2g38770 (*MAC7/EMB2765*) (Figure 1.5A). To determine whether this mutation was the causal mutation that suppressed *amiR-SUL*, we first examined the linkage between the mutation and the visible phenotype. In the F2 population of the backcross, we identified 84 plants with the *mac7-1* phenotype, i.e., reduced vein-centered leaf bleaching. These plants were genotyped for the C-to-T mutation in *MAC7*. All 84 plants were found to be homozygous for this mutation, and thus the *amiR-SUL* suppressor phenotype was linked with this mutation.

Next, to confirm that the causal mutation of the suppressor was in *MAC7*, we generated *MAC7* promoter driven *MAC7-GFP* and *MAC7-mCherry* fusion constructs, and introduced them into *mac7-1* in *amiR-SUL* and Col-0 backgrounds. *pMAC7:MAC7-GFP* or *pMAC7:MAC7-mCherry* fully complemented the morphological defects (Figure 1.5B) and *pMAC7:MAC7-mCherry* restored the accumulation of miRNAs of the mutant in both

inflorescences and seedlings (Figure 1.5C, D). Thus, the causal mutation in the suppressor was in *MAC7* and we named this mutation *mac7-1*.

To obtain another *mac7* allele, we ordered a T-DNA insertion line, SALK_129044 (which we named *mac7-2*) (Figure 1.6A). Genotyping *mac7-2* in the progeny of a selfed *mac7-2/+* plant revealed 2:1 segregation between heterozygous and wild-type plants (Table 1.1), suggesting that *mac7-2* homozygous plants were embryo lethal. Unlike wild-type siliques that contained only normal-looking green seeds, siliques from heterozygous (*mac7-2/+*) plants had aborted seeds that appeared pale white (Figure 1.6B). An approximate 3:1 ratio between normal seeds and aborted seeds were found (Table 1.1), which was a strong indication of embryo lethality of the homozygous mutant. In addition, the viability of pollen from *mac7-2/+* plants appeared normal (Figure 1.6C). The above evidence demonstrated that *mac7-2* is a recessive, embryo-lethal mutation. As *mac7-1* homozygous plants are viable, this mutation is likely a partial loss-of-function allele.

The *MAC7* gene encodes an RNA helicase conserved from yeast to animals and plants. The structure of the human *MAC7* ortholog, Aquarius, has been determined (45). Based on homology modeling, we predict that the plant *MAC7* also has a central RNA helicase core which consists of two RecA-like domains, an alpha-helical stalk and a beta-barrel domain. The canonical sequence motifs of two RecA-like domains (RecA1 and RecA2) are conserved, and they form a motor module required for ATP hydrolysis, RNA unwinding and the coupling of these two processes. The *mac7-1* mutation leads to a change of a highly

conserved amino acid in the RecA1 domain, which very likely affects the core functions of this protein in ATP-binding/hydrolysis and nucleic acid unwinding (Figure 1.5A and Figure 1.7). MAC7 is also predicted to have several unique accessory domains which are not found in other helicases, including a long N-terminal domain composed of armadillo repeats, a globular domain and an elongated helical bundle located on the opposite side of the N-terminal domain; they are denoted as the ARM, thumb and pointer, respectively. Based on protein cross-linking coupled with mass spectrometry analysis of human MAC7, the ARM domain cross-links to U2 snRNP proteins, suggesting it has a crucial protein–protein scaffolding role and may be required for spliceosome assembly. The globular thumb domain is composed of a four-stranded beta-sheet packed against an alpha-helix. The pointer domain has an elongated arched shape, which is mainly defined by three long helices. Both thumb and pointer domains do not show similarities with other proteins (Figure 1.5A and Figure 1.7) (45, 51).

MAC7 promotes pri-miRNA levels without affecting *MIR* promoter activities or pri-miRNA transcript half-life

To determine the molecular mechanisms of MAC7 in miRNA biogenesis, we examined the levels of pri-miRNAs. We found that the levels of pri-miRNAs from many *MIR* genes were reduced in the *mac7-1* mutant (Figure 1.8A). This reduction could be attributed to reduced transcription of *MIR* genes or enhanced pri-miRNA degradation or processing. We first examined whether *mac7-1* affects *MIR* promoter activities. We crossed *mac7-1* to a miRNA promoter reporter line, *pMIR167a:GUS*, in which the transgene was inserted into

a single genomic locus. We then examined GUS activity by staining and *GUS* transcripts level by real-time RT-PCR in wild-type and *mac7-1* plants, in which the transgene was homozygous. There was no detectable difference between wild type and *mac7-1* in terms of GUS activity or *GUS* transcripts level, indicating that *mac7-1* did not affect miRNA promoter activities (Figure 1.8B). We also examined whether the reduction of amiR-SUL level in *mac7-1* was due to reduced *SUC2* promoter activity. If this was the case, we would expect the endogenous *SUC2* RNA to be reduced in abundance in *mac7-1*. Real-time RT-PCR revealed that *SUC2* mRNA levels did not change significantly in *mac7-1*, implying that *SUC2* promoter activity was not affected by *mac7-1* and the reduction in amiR-SUL levels was not due to impaired *SUC2* promoter activities in the mutant (Figure 1.9A).

Next, we measured the half-lives of pri-miRNAs in *amiR-SUL* and *mac7-1 amiR-SUL*. Seedlings were treated with the transcription inhibitor cordycepin, and RNAs were isolated at different time points. The levels of pri-amiR-SUL, pri-miR167a, and pri-miR172a, were determined by real-time RT-PCR. Similar half-lives were found for these pri-miRNAs in the two genotypes (Figure 1.8C).

It is also possible that MAC7 affects miRNA biogenesis indirectly through affecting the expression of key miRNA biogenesis factors. To test this possibility, we determined transcript and protein levels of several key miRNA biogenesis factors [i.e., DCL1, HYL1, SE, AGO1, HEN1 (transcript only)] in *amiR-SUL* and *amiR-SUL mac7-1*. No differences were detected for any of the genes in the two genotypes (Figure 1.9B, C, D).

MAC7 is a nuclear protein interacting with other MAC components, which are also required for miRNA biogenesis

To further characterize the molecular functions of the MAC7 protein, we examined the subcellular localization of fluorescent protein tagged MAC7. Transgenic plants expressing *p35S:YFP-MAC7* or *pMAC7:MAC7-GFP* exhibited nuclear GFP signals (Figure 1.10A), which is consistent with the presence of a predicted Nuclear Localization Signal (NLS) in MAC7 (Figure 1.5A). To determine if MAC7 localized in dicing bodies, we transiently expressed YFP-MAC7 and DCL1-YFP in tobacco leaves and compared their expression patterns. While DCL1-YFP showed weak nucleoplasmic signals and strong dicing body signals, YFP-MAC7 exhibited dispersed distribution in the nucleoplasm and was absent from the nucleolus (Figure 1.11A). When YFP-MAC7 was co-expressed with TagRFP-HYL1, signals from the two proteins overlapped in the nucleoplasm while only TagRFP-HYL1 was found concentrated in dicing bodies (Figure 1.11B). Immunoblot analyses confirmed the expression of GFP- or YFP-tagged MAC7 in *Arabidopsis* transgenic lines (Figure 1.11C) and in infiltrated tobacco leaves (Figure 1.11D). These data suggest that MAC7 does not display the dicing body patterns as DCL1 and HYL1 do. The dispersed nucleoplasmic distribution of MAC7 suggests broader roles than miRNA biogenesis.

To uncover interacting partners of MAC7 and to determine whether MAC7 interacts with any known miRNA pathway proteins, we immunoprecipitated (IP) MAC7 and performed mass spectrometry (MS) analyses. In one experiment, IP was performed with Col-0 plants

using anti-MAC7 antibodies with pre-immune IgG as a negative control. In another experiment, IP was performed using Chromotek-RFP-Trap with a *pMAC7:MAC7-mCherry mac7-1* line in which the transgene rescued the mutant phenotypes; IP was performed using the same RFP-Trap with Col-0 as a negative control. The two independent experiments consistently pulled down all major MAC components, demonstrating that MAC7 was part of the MAC (Table 1.2). An overlapping set of proteins was identified from our MAC7 IP-MS and from published MOS4 IP-MS (Figure 1.10B and Table 1.2) (38). Besides the known MAC subunits (38), seven additional proteins were discovered as potential MAC7-associated proteins (Table 1.2), but the association between these proteins and MAC7, or whether they also belong to MAC, needs to be further investigated.

Previous studies showed that the MAC subunits CDC5 and PRL1 promote miRNA biogenesis (17, 30). To further confirm the interactions between MAC7 with these two MAC subunits that are also miRNA biogenesis factors, we performed Bimolecular Fluorescence Complementation (BiFC) analysis, and found that MAC7 interacted with CDC5 and also interacted, albeit weakly, with PRL1 (Figure 1.10C). Similar BiFC studies did not reveal interactions between MAC7 and HYL1 or DCL1 (Figure 1.10C).

These results and our work on MAC7 raise the possibility that MAC plays a role in miRNA biogenesis as a whole. To test this hypothesis, we examined whether core subunits of MAC are required for miRNA biogenesis. *Arabidopsis* MAC3B is a U-box E3 ubiquitin ligase, and is a core MAC component (38). MAC3B has a homolog, MAC3A, which shares 82%

identity with MAC3B at the amino acid level (38). PRL1 is also a core member of MAC. It has a homolog PRL2, which is expressed at a much lower level than PRL1 (52). Previous studies showed that a *prl1* single mutant has reduced miRNA levels but a *mac3b* mutant (SALK_050811) does not (30). Several siRNAs, ta-siRNA255 and miRNA171 were found to show reduced accumulation in the *mac3a mac3b* (SALK_089300 and SALK_050811) double mutant (53). Here we examined the morphological and molecular phenotypes of two double mutants, *mac3a mac3b* and *prl1 prl2*. Similar to other MAC subunit mutants, *mac3a mac3b* and *prl1 prl2* exhibited pleiotropic developmental phenotypes (Figure 1.12A). RNA gel blot analyses showed that miR156, miR166, and miR171 all exhibited reduced accumulation in these double mutants (Figure 1.12B). We also performed small RNA-seq with Col-0 and *prl1 prl2* seedlings to assess the global changes of small RNAs in *prl1 prl2*. Similar to those of *mac7-1*, 21-nt and 24-nt small RNAs showed a global reduction in *prl1 prl2*. Unlike in *mac7-1*, a significant global reduction in miRNA levels was also found in *prl1 prl2*. (Figure 1.12C). This is consistent with *mac7-1* being a weak allele representing a partial compromise in MAC function. RT-qPCR analyses revealed that the levels of pri-miRNAs were also reduced in *mac3a mac3b* and *prl1 prl2* (Figure 1.12D). Thus, the molecular phenotypes of *mac3a mac3b* and *prl1 prl2* mutants were similar to those of *mac7-1*.

MAC interacts with HYL1

As CDC5 and PRL1 interact with dicing complex components (17, 30), it is possible that MAC7, or the entire MAC, associates with the dicing complex. To explore the interactions

between the dicing complex and MAC, we performed HYL1 IP-MS. In one experiment, IP was performed with a *p35S:HYL1-YFP* line and the negative control *p35S:YFP* transgenic line and Col-0 using Chromotek-GFP-Trap. In another experiment, IP was performed with Col-0 and the negative control *hyl1-2* using anti-HYL1 antibodies. Eleven MAC subunits, including MAC7, were found in both IP-MS experiments (Table 1.3), indicating that HYL1 associates with MAC *in vivo*.

The interactions between HYL1 and MAC raised the possibility that MAC7 could be involved in pri-miRNA processing through affecting the dicing body localization of HYL1. We crossed *mac7-1* to a *HYL1-YFP* transgenic line and quantified dicing body numbers in wild type and *mac7-1*. The number of HYL1-YFP-positive dicing bodies was significantly reduced in *mac7-1* compared with wild type plants (Figure 1.13A,B), demonstrating that *MAC7* is required for the proper localization of HYL1 in dicing bodies, which might partially explain the compromised miRNA levels in the *mac7-1* mutant.

Down-regulated genes in *mac* mutants are significantly related to stress responses

We also explored whether MAC plays a role in RNA metabolism in general. We performed RNA-seq with Col-0, *mac7-1*, *mac3a mac3b*, and *prl1 prl2* seedlings in two biological replicates. Differentially expressed genes (DEGs) were identified between each mutant and wild type with fold change of 1.5 or more. Among the three mutants, we identified 2,007 and 2,268 down-regulated (hypo-DEGs) and up-regulated (hyper-DEGs) genes, respectively. 189 hypo-DEGs and 222 hyper-DEGs were commonly found among *mac7-*

1, *mac3a mac3b* and *prl1 prl2* (Figure 1.14). The overlap of the DEGs among these three mutants was significant (SuperExactTest, $p = 0$), indicating that these MAC components function as a complex and regulate the same group of genes. The large portion of non-overlapped DEGs indicates that each subunit may also have its own function independent of MAC.

To understand the biological functions of MAC, we examined the Gene Ontology terms enriched in the common DEGs (54). An enrichment of genes involved in stress responses, including stimulus responses, plant defense or immune responses was found in the hypo-DEGs. As for the hyper-DEGs, GO terms in various small molecule biosynthetic and metabolic processes were enriched (Figure 1.14). Many more GO terms related to stress responses were identified from the hypo-DEGs than the hyper-DEGs, implying that MAC tends to activate the expression of stress response genes. The results are consistent with previous findings showing that *mac3a mac3b* and *prl1 prl2* double mutants are more susceptible to pathogen infection (38, 52). To test whether *MAC7* is also required for plant immunity, *mac7-1* was infected with *Pseudomonas syringae* p.v. *maculicola* (*P.s.m.*) strain ES4326 together with Col-0 and *prl1-2*. The *prl1* mutants are more susceptible to pathogen infection and therefore served as a positive control (52). Pathogen growth was assayed three days after bacterial inoculation. While *P.s.m.* ES4326 accumulated to higher levels in *prl1-2* than in wild type, *mac7-1* was similar to wild type in terms of bacterial titer (Figure 1.15A). Perhaps this was due to *mac7-1* being a weak allele, in which the

magnitude of downregulation of the hypo-DEGs was small compared to that in other *mac* mutants (Figure 1.15B).

MAC subunits affect pre-mRNA splicing

Homologs of MAC in yeast and mammals play a critical role in pre-mRNA splicing. In *Arabidopsis*, we lack evidence for a widespread role of MAC in splicing; only the splicing patterns of several genes, such as *SUPPRESSOR OF NPR1-1 CONSTITUTIVE 1 (SNC1)* and *RESISTANCE TO PSEUDOMONAS SYRINGAE4 (RPS4)*, were shown to be altered in several MAC subunit mutants (53, 55).

Since intron retention is a major form of alternative splicing in *Arabidopsis* (56), we examined whether *mac7-1*, *mac3a mac3b*, and *prl1 prl2* mutants had global intron retention defects. The ratio of RNA-seq reads mapping to introns (including 5'/3' splice sites) and those mapping to exons only was used as a measure of intron retention (see Methods for details). All annotated transcripts were considered in sum as long as the read counts passed an abundance filter. All three mutants exhibited significantly higher levels of intron retention compared to Col-0 (Wilcoxon test, $p < 2.2\text{e-}16$) (Figure 1.16A). Next, using the ratio of intron reads vs. exon reads as a measure of intron retention levels, we identified genes with intron retention defects in each mutant as compared to wild type (see Methods for details). 2819, 1466 and 298 genes were found to have intron retention defects in *mac3a mac3b*, *prl1 prl2* and *mac7-1*, respectively (Figure 1.16B, C). Significant overlap was found among the genes with intron retention defects in the three mutants

(SuperExactTest, $p < 0.05$), indicating that MAC works as a complex in pre-mRNA splicing (Figure 1.16C). Two examples of intron retention events in these mutants are shown in Figure 1.16D.

We next examined whether there was any correlation between splicing defects and gene expression. We compared the expression levels of genes with intron retention defects in each mutant vs. wild type. As a reference, all genes that passed a minimum intron read coverage filter (see Figure 1.16 legend) were analyzed. There was no correlation between intron retention defects and the status of their differential expression in the *mac* mutants. Like all analyzed genes, genes with intron retention defects were increased, reduced, or unchanged in expression levels in each mutant as compared to wild type (Figure 1.17A). In addition, we examined intron vs. exon ratios in total genes, up-regulated genes and down-regulated genes in each mutant. Both up-regulated and down-regulated genes showed significant intron retention as total genes (Wilcoxon test, $p < 2.2e-16$), again indicating that the intron vs. exon ratio has no correlation with gene expression changes (Figure 1.17B).

Given the intron retention defects and miRNA accumulation defects in the *mac* mutants, we asked whether intron retention in pri-miRNAs contributed to the miRNA accumulation defects. Many *MIR* genes were shown to have introns, which are spliced out in pri-miRNAs (26, 57, 58). We performed RT-PCR with intron-flanking primers to detect unspliced miRNA precursors, including pri-miR163, pri-miR156, pri-miR166, pri-miR168, and pri-

miR172. Genomic DNA was amplified with the same sets of primers to indicate the sizes of intron-containing fragments. No intron retention was observed in these miRNA precursors in *mac7-1* (Figure 1.18).

Thus, intron retention and differential expression of genes (including *MIR* genes) are not linked. MAC seems to regulate gene expression and RNA splicing independently.

DISCUSSION

MAC7 is an evolutionarily conserved protein across eukaryotes, however little is known about the molecular and biological functions of MAC7 or its orthologs. The human ortholog Aquarius is an RNA helicase with ATPase activity, and it binds to introns to assist intron splicing *in vitro* (45, 46). The *C. elegans* ortholog EMB-4 was reported to act in the nuclear RNAi pathway, where it interacts with nuclear AGOs and functions in germline-specific chromatin remodeling (47, 48, 59). In *Arabidopsis*, MAC7 is involved in plant defense and was predicted to be an essential gene (38, 49). In this study, we showed that a T-DNA insertion mutant of *MAC7* is indeed embryo lethal, and we isolated the first viable *mac7* mutant, *mac7-1*, as a miRNA biogenesis-defective mutant. In this mutant, the artificial miRNA, amiR-SUL, and many endogenous miRNAs show reduced accumulation. MAC7 was identified as a MAC subunit through MOS4 IP-MS (38); our MAC7 IP-MS confirmed this. Two other MAC subunits, CDC5 and PRL1, were previously shown to promote miRNA biogenesis (17, 30). In this study, we showed that MAC3 also

has a similar role. Thus, it is likely that MAC as a complex promotes miRNA biogenesis in general. Based on small RNA-seq analysis, MAC may promote the biogenesis of not only miRNAs but also siRNAs. Both 21-nt and 24-nt small RNAs, the two most abundant small RNA size classes in *Arabidopsis*, are reduced in *mac7-1* and *prl1 prl2* mutants. A role of MAC in promoting siRNA biogenesis is also supported by previous findings on CDC5, MAC3 and PRL1 (17, 30, 53).

How does MAC7 promote miRNA biogenesis? In *mac7-1*, the reduced accumulation of miRNAs correlated with reduced levels of pri-miRNAs. Thus, MAC7 probably acts in miRNA biogenesis by promoting *MIR* transcription or pri-miRNA stability. However, the activity of a promoter driving *MIR* expression was not affected in *mac7-1*, nor was the endogenous *SUC2* promoter activity. The half-lives of pri-miRNAs were not affected either. Although we cannot exclude the possibility that MAC7 promotes pri-miRNA processing into pre- or mature miRNAs and prevents pri-miRNA decay at the same time and therefore pri-miRNA half-lives appeared unchanged in the mutant, we prefer to hypothesize that MAC7 plays a role in transcription elongation and/or maturation of pri-miRNAs, considering the interactions between Pol II and MAC subunits CDC5 and PRL1, as well as the functions of yeast NTC in transcription elongation (17, 30, 60, 61). Given the role of MAC7 in pre-mRNA splicing discovered before and in this study (53, 55), we considered the possibility that MAC7 acts in miRNA biogenesis by promoting the splicing of pri-miRNAs. However, we did not detect intron retention in several intron-containing pri-miRNAs in *mac7-1*, although the mature miRNAs from those pri-miRNAs are reduced.

In addition, *MIR* genes without any introns (e.g. the *pSUC2:amiR-SUL* transgene, *MIR159a*, and *MIR167a*) (50, 62) were also affected in *mac7-1*. Thus, the miRNA biogenesis defects of *mac7-1* could not be attributed to defects in splicing. Intriguingly, we observed that MAC7 affects HYL1 localization to dicing bodies. Little is known about how dicing bodies form. MAC7 may recruit the dicing complex to pri-miRNAs through protein–protein interactions to form dicing bodies. Alternatively, dicing complex proteins are recruited by pri-miRNAs to dicing bodies, and therefore the reduced number of dicing bodies in *mac7-1* could be a consequence of reduced pri-miRNA levels in the mutant. We favor the second hypothesis because MAC7 plays a more general role in transcription and RNA metabolism (such as splicing), while the dicing complex acts more specifically on miRNA precursors and is presumably not recruited to other RNAs that MAC7 may also act on.

CDC5 and PRL1 were shown to interact with dicing complex proteins (e.g. DCL1, SERRATE) *in vivo* through co-IP or BiFC analyses (17, 30). We did not detect interactions between MAC7 and the dicing complex through MAC7 IP-MS or BiFC analyses. However, many MAC components including MAC7 were found in HYL1 IP-MS, which clearly indicates association between the dicing complex and MAC *in vivo*. The possible reasons for the inability of MAC7 to pull down the dicing complex proteins are: 1) The interaction between MAC7 and the dicing complex is indirect or weak, and bridged through CDC5 or PRL1; 2) only a small portion of MAC7 proteins interacts with the dicing complex. The second hypothesis is consistent with MAC7 having broader functions

beyond miRNA biogenesis. We hypothesize that most of HYL1 or the dicing complex is associated with MAC, but not the other way around, which explains the recovery of MAC in HYL1 IP-MS.

MAC has functions beyond miRNA biogenesis in *Arabidopsis*. In our RNA-seq analysis, stress response-related GO terms were significantly enriched in down-regulated genes in three MAC subunit mutants. This is consistent with previous studies showing that many *mac* mutants are more susceptible to pathogen infection (38, 43, 52, 55). Although the *mac7-1* hypo-DEGs are enriched in defense related GO terms, the genes show the smallest reduction magnitude compared to other *mac* mutants, which could be the reason why *mac7-1* is not susceptible to *P. syringae* infection as *prl1-2* and *mac3a mac3b* mutants are (38, 52).

It has been suspected that MAC is also involved in splicing in plants, like its orthologs in human and yeast (36, 37). Indeed, our RNA-seq analyses uncovered intron retention defects in three MAC subunit mutants, thus linking MAC with pre-mRNA splicing. However, there were no significant correlations between intron retention and changes in gene expression. Genes with intron retention were up-regulated, down-regulated, or unchanged in the *mac* mutants. We also did not detect intron retention for intron-containing pri-miRNAs in *mac7-1*. It is likely that MAC has separate roles in RNA splicing and gene expression regulation (including the regulation of *MIR* genes). Thus, MAC has broad functions in nuclear RNA metabolism.

We speculate that a common theme of MAC's role in nuclear RNA metabolism is linking RNA processing to transcription (Figure 1.19). In yeast, the NTC promotes transcription elongation (61, 63). In both yeast and human, NTC or Prp19C associates with spliceosomes *in vivo*, although this association may entail NTC subcomplexes in human (45, 64, 65). In *Arabidopsis*, two MAC subunits, CDC5 and PRL1, have been shown to interact with Pol II *in vivo* (17, 30). IP-MS of an *Arabidopsis* spliceosome subunit also pulled down multiple subunits of MAC (39). Thus, it is possible that MAC promotes co-transcriptional splicing through its interactions with both Pol II and the spliceosome. Similarly, in miRNA biogenesis, MAC may promote co-transcriptional pri-miRNA processing through its interactions with both Pol II and HYL1.

METHODS

Plant materials and growth conditions

The *pSUC2:amiR-SUL* transgenic line is a gift from Dr. Detlef Weigel (50). *mac7-1* is a new allele isolated from our EMS mutagenesis screen with the *pSUC2:amiR-SUL* transgenic line. SALK_120944 (*mac7-2*) is a T-DNA insertion line obtained from the *Arabidopsis* Biological Resource Center (ABRC). The following published transgenic lines or mutants were used: *pMIR167a:GUS*, and *prl1-2* (SALK_008466) (17, 30); *p35S:HYL1-YFP* (66); *p35S:YFP* (30); *mac3a mac3b* (*mac3a* is SALK_089300; *mac3b* is SALK_050811) and *prl1 prl2* (*prl1* is SALK_008466; *prl2* is SALK_075970) double

mutants (38, 52).

In *pMIR167a:GUS* or *p35S:HYL1-YFP* transgenic lines, the transgene was confirmed to be homozygous by Basta selection (Phosphinothricin 25mg/L, Gold Biotechnology) and by examining YFP signals under fluorescence microscopy, respectively, in multiple individuals. The *mac7-1* mutant was crossed into *pMIR167a:GUS* or *p35S:HYL1-YFP*, and F2 plants containing homozygous transgenes were confirmed in the F3 generation.

Genotyping primers used are listed in Table 1.4. Plants were grown in a plant growth chamber at 23°C for 16h light (Cool white fluorescent lamps, 25-watt Sylvania 21942 FO25/741/ECO T8 linear tube) and 8h dark cycles.

Mutagenesis screening and mapping of *MAC7*

The *amiR-SUL mac7-1* M2 mutant was backcrossed with the parental line *pSUC2:amiR-SUL*. Genomic DNA was extracted from 100 pooled F2 mutants and used in library construction. The library was paired-end sequenced on Illumina's HiSeq 2000 at ~30x coverage, and the reads were mapped to the TAIR 10 genome using the Burrows-Wheeler Alignment tool (BWA) (67). SamTools (68) was used to identify EMS-induced single nucleotide polymorphisms (SNPs). The SNP calls generated by SamTools were processed by the Next-Generation EMS mutation mapping (NGM) website tools (69). Two candidate mutations with 100% mutation rate were identified as EMS-typical C:G to T:A transitions that are predicted to cause nonsynonymous substitutions in the coding region of genes.

Because one candidate mutation was only supported by three reads, we focused on the one in At2g38770 (*MAC7/EMB2765*), which was supported by 29 reads. A Derived Cleaved Amplified Polymorphic Sequences (dCAPS) marker was designed to genotype this mutation (see Table 1.4 for primers). The PCR products from wild type can be digested by *EcoR1*, whereas those from *mac7-1* could not. Linkage analysis was performed on 84 individual mutant plants in the F2 population of the backcross using this dCAPS marker to assess linkage between the mutation and the *mac7-1* mutant phenotype.

DNA constructs and complementation

The *MAC7* genomic region without the stop codon was amplified and cloned into pENTR/D-TOPO (Invitrogen) and then introduced to a modified pEarleyGate 301 vector (70) to generate *pMAC7:MAC7-mCherry* via LR reaction. The *MAC7* genomic region without the stop codon was amplified and cloned into the pMDC107 gateway vector (71) to generate *pMAC7:MAC7-GFP* with the ClonTech In-Fusion HD Cloning Kit. The *MAC7* coding region was cloned and then introduced into the pEarleyGate104 vector (70) via pENTR/D-TOPO (Invitrogen) through LR reactions. The above plasmids were used to transform *mac7-1* plants through the *Agrobacterium*-mediated floral dip method. Primers used are listed in Table 1.4.

Small RNA gel blotting, RT-PCR and quantitative RT-PCR

Total RNA from 2 to 3-week old seedlings (aerial part) or inflorescences was extracted with TRI reagent (Molecular Research Center). RNA gel blotting for detection of miRNAs

was performed as described (72). 10 µg total RNA from aerial part of seedlings or inflorescences was used in RNA gel blotting. 5' end ³²P-labelled antisense DNA oligonucleotides were used to detect miRNAs. Oligonucleotide probes used are listed in Table 1.4.

To perform RT-PCR, total RNA was first treated with DNase I (Roche) followed by reverse transcription using RevertAid Reverse Transcriptase (Thermo Fisher Scientific) with oligo-d(T) primers according to manufacturer's instructions. Quantitative RT-PCR was carried out in triplicate using iQ SYBRGreen Supermix (BioRad) on the BioRad CFX96 system. Primers used are listed in Table 1.4.

RNA half-life measurements

RNA half-life measurements were performed as described (73) with minor modifications. 12-day-old Col and *mac7-1* whole seedlings were transferred from Murashige and Skoog medium (PhytoTechnology Lab) agar plates to 6-well-plates with 1/2 MS medium and incubated overnight. The next day, cordycepin (Sigma) was added to a final concentration of 0.6 mM and the seedlings were collected for RNA extraction at 0, 30, 90, and 120 min after cordycepin addition. RT-qPCR was then performed to determine pri-miRNA levels. *UBQ5* was used as an internal control.

Small RNA-seq library construction and data analysis

To construct small RNA libraries, the aerial parts of 12-day-old Col-0, *mac7-1*, and *prl1 prl2* seedlings grown on plates were harvested for total RNA extraction. Two biological replicates were included: plants grown on different plates under the same conditions were collected at the same time into two separate samples for RNA extraction and subsequent procedures. To isolate small RNAs from total RNA, 50 µg of total RNA from each sample was resolved on 15% urea-PAGE gel, and the 18–30-nt region was excised from the gel. Small RNAs were recovered by soaking the smashed gel in 0.3 M NaCl overnight, followed by ethanol precipitation. Small RNA libraries were constructed following instructions from the NEBNext Multiplex Small RNA Library Prep Set for Illumina (E7300). The libraries were sequenced on an Illumina Hiseq 2500 at the UC Riverside Genomics core facility.

Reads from small RNA-seq were first processed to remove the adaptor sequences by cutadapt (sequence: AGATCGGAA) (74). The reads were mapped to the TAIR10 genome using ShortStack with default parameters (75). Normalization was performed by calculating the RPKM (reads per million of 45S rRNA reads) value (76). Only one biological replicate for Col-0 was included in this analysis since the other Col-0 sample had very few reads caused by unknown problems in library construction or sequencing.

RNA-seq library construction and data analysis

Polyadenylated RNA was isolated from total RNA extracted from 12-day-old seedlings

(aerial part) of Col-0, *mac7-1*, *mac3a mac3b*, and *prl1 prl2* using the Magnetic mRNA Isolation Kit (New England Biolabs), with two biological replicates for each genotype. For the biological replicates, plants grown on different plates under the same conditions were collected at the same time into two separate samples for RNA extraction and subsequent procedures. RNA-seq libraries were prepared using NEBNext mRNA Library Prep Reagent Set for Illumina (New England Biolabs) and sequenced on an Illumina HiSeq 2500 platform to generate high-quality single-end reads of 101bp in length. Data analysis was performed with the pRNASeqTools pipeline (<https://github.com/grubbybio/RNASeqTools>). Firstly, the raw reads were aligned to the TAIR10 genome using HISAT2 (77). Secondly, exonic and intronic reads were classified and quantified as follows. Reads with five or more nucleotides overlapping with intronic regions (intronic regions defined in all isoforms, including splicing donor/acceptor sites) were counted as intronic reads, and reads that mapped exclusively to exonic regions were counted as exonic reads. Transcript levels were measured in reads per million total read counts. Differentially expressed genes were identified using DEseq2 with fold change of 1.5 and $p < 0.01$ as the parameters (78). To identify genes that exhibited significantly higher levels of intron retention compared to Col-0, the intron/exon ratio was calculated as $(I1 + I2)/2$ over $(E1+E2)/2$ (I: Intronic reads, E: Exonic reads, 1: biological replicate 1, 2: biological replicate 2) while applying an abundance cutoff (raw intronic read number ≥ 2 and exonic read number ≥ 5). We only considered intron/exon (I/E) ratios between 0 to 1. The genes with intron retention were identified using DEseq2 based on values of $(I_{\text{mutant}}/E_{\text{mutant}})/(I_{\text{Col-0}}/E_{\text{Col-0}})$ and fold changes ≥ 2 and FDR < 0.01 as parameters (78).

SuperExactTest was employed to access the statistical significance of DEGs or intron retention gene overlaps among *mac7-1*, *mac3a mac3b* and *prl1 prl2* (79).

Protein sequence alignment

Database searching of MAC7 homologs was performed at National Center for Biotechnology Information (NCBI) (www.ncbi.nlm.nih.gov/). Alignment of protein sequence was performed with Muscle (80) and the alignments were edited with Jalview (81).

Antibody generation and immunoblotting

To generate anti-MAC7 antibodies, a 5' portion of the coding region of MAC7 corresponding to the first 416 amino acids of the protein was amplified (primers listed in Table 1.4) and inserted into pMCSG7-His-MBP and pSUMO-His vectors, respectively. The constructs were then transformed into the E. coli strain BL21 for protein expression. The recombinant proteins were purified by AKTA fast protein liquid chromatography (GE Healthcare) using the MBP-Trap or His-Trap column. The purified MBP-tagged protein was used as antigens to raise polyclonal antibodies in rabbits as described (82). The anti-serum was affinity-purified using a MAC7-SUMO-His conjugated column. The purified antibodies were used in immunoblotting and IP-MS experiments. Similar approach was employed to generate the anti-CH42/SUL antibody. The full-length CH42 protein fused with SUMO-His tag was expressed and purified to immunize two rabbits. Affinity purified antibodies were used in immunoblotting analysis.

Other antibodies used in immunoblotting experiments include anti-GFP (Roche, Cat. No. 11814460001), anti-AGO1 (Agrisera, AS09 527), anti-HYL1 (Agrisera, AS06 136), anti-SERRATE (Agrisera, AS06 136), anti-DCL1 (Agrisera, AS12 2102), and anti-Tubulin (Sigma, T9026).

Proteomic Analysis

Total proteins from 12-day-old seedlings were extracted and immunoprecipitated with indicated antibodies. The IP products were resolved in SDS-PAGE. The antibody bands were removed and the samples were subjected to mass spectrometry as described (39, 83). Two biological replicates were performed, and the identified interacting proteins were those represented by peptides with high hits from both biological replicates.

Transient expression of fluorescent fusion proteins in tobacco leaf epidermal cells

The CDS of *HYL1* was amplified and cloned into pENTR/D-TOPO (Invitrogen) and then introduced into the pGWB661 gateway vector (84) to generate *p35S:TagRFP-HYL1*. *p35S:DCL1-YFP* in the pEG101 vector was from a published study (17). The generation of *p35S:YFP-MAC7* was described above. *Agrobacterium* GV3101::mp90 transformed with *p35S:YFP-MAC7*, *p35S:DCL1-YFP*, or *p35S:TagRFP-HYL1* was used to infiltrate tobacco leaves as described (85). The expression of fluorescent fusion proteins was observed using a Leica SP5 confocal laser-scanning microscope.

BiFC analysis

BiFC analysis was performed as described (86). Paired cCFP and nVenus constructs were co-infiltrated into *Nicotiana benthamiana* leaves. After 48 h, YFP signals and chlorophyll auto fluorescence signals were excited at 488 nm and detected by Olympus Fluoview 500 confocal microscopy with a narrow band pass filter (BA505–525 nm).

Histochemical GUS assay and Alexander's staining of pollen

GUS staining was performed as described (14). Briefly, 12-day-old seedlings from Col-0 and *mac7-1* harboring a homozygous *pMIR167a:GUS* transgene were vacuum infiltrated for 10 min and then incubated in GUS staining solution at 37°C for several hours until blue color became visible. Tissue clearing was performed with 70% ethanol for 1 to 2 days before imaging.

Alexander's staining of pollen was performed as described (82). The stained pollen grains were observed under a microscope equipped with a charge-coupled device camera (Olympus).

Accession numbers

Genes referred to in this study correspond to the following *Arabidopsis* Genome Initiative locus identifiers: *MAC7/EMB2765*, AT2G38770; *SUL/CHLORINA42*, AT4G18480; *MICRORNA156A*, AT2G25095; *MICRORNA159A*, AT1G73687; *MICRORNA163* AT1G66725; *MICRORNA166A* AT2G46685; *MICRORNA167A*, AT3G22886;

MICRORNA168A, AT4G19395; *MICRORNA171A*, AT3G51375; *MICRORNA172A*, AT2G28056; *MICRORNA390B*, AT5G58465; *MICRORNA394B*, AT1G76135; *MICRORNA396B*, AT5G35407; *MICRORNA397A*, AT4G05105; *ACTIN8*, AT1G49240; *UBIQUITIN5*, AT3G62250; *TUBULIN3*, AT5G62700; *EIF4A1*, AT3G13920; *IPP2*, AT3G02780 ;*CDC5*, AT1G09770; *PRL1*, AT4G15900; *PRL2*, AT3G16650; *MAC3A*, AT1G04510; *MAC3B*, AT2G33340 ; *HYL1*, AT1G09700; *DCL1*, AT1G01040; *SERRATE*, AT2G27100; *AGO1*, AT1G48410; *HEN1*, AT4G20910; *MYB33* AT5G06100; *SCL6-IV* AT4G00150; *SPL10* AT1G27370; *MYB65*, AT3G11440; *ARF17*, AT1G77850; *ARF8*, AT5G37020; *SCL6-III* At3g60630; *CUC2* AT5G53950.

Protein sequences of MAC7 homologs in other species correspond to the following NCBI references: *Homo sapiens* intron-binding protein aquarius, NP_055506.1; *Mus musculus* intron-binding protein aquarius, NP_033832.2; *Danio rerio* intron-binding protein aquarius, NP_956758; *Drosophila melanogaster* CG31368, NP_996198.2; *Caenorhabditis elegans* EMB-4, NP_001256831.1; *Schizosaccharomyces pombe* Cwf11, NP_595360.1.

RNA sequencing data are available from NCBI Gene Expression Omnibus (GEO) under the following reference numbers: Col-0_1, GSM2585832; Col-0_2, GSM2585833; mac7-1_1, GSM2585834; mac7-1_2, GSM2585835; mac3amac3b_1, GSM2585836; mac3amac3b_2, GSM2585837; prl1prl2_1, GSM2585838; prl1prl2_2, GSM2585839; Col-0_1_sRNA GSM2771029; mac7-1_1_sRNA GSM2771030; mac7-1_2_sRNA GSM2771031, prl1 prl2_1_sRNA GSM2771032, prl1 prl2_2_sRNA GSM2771033.

FIGURES

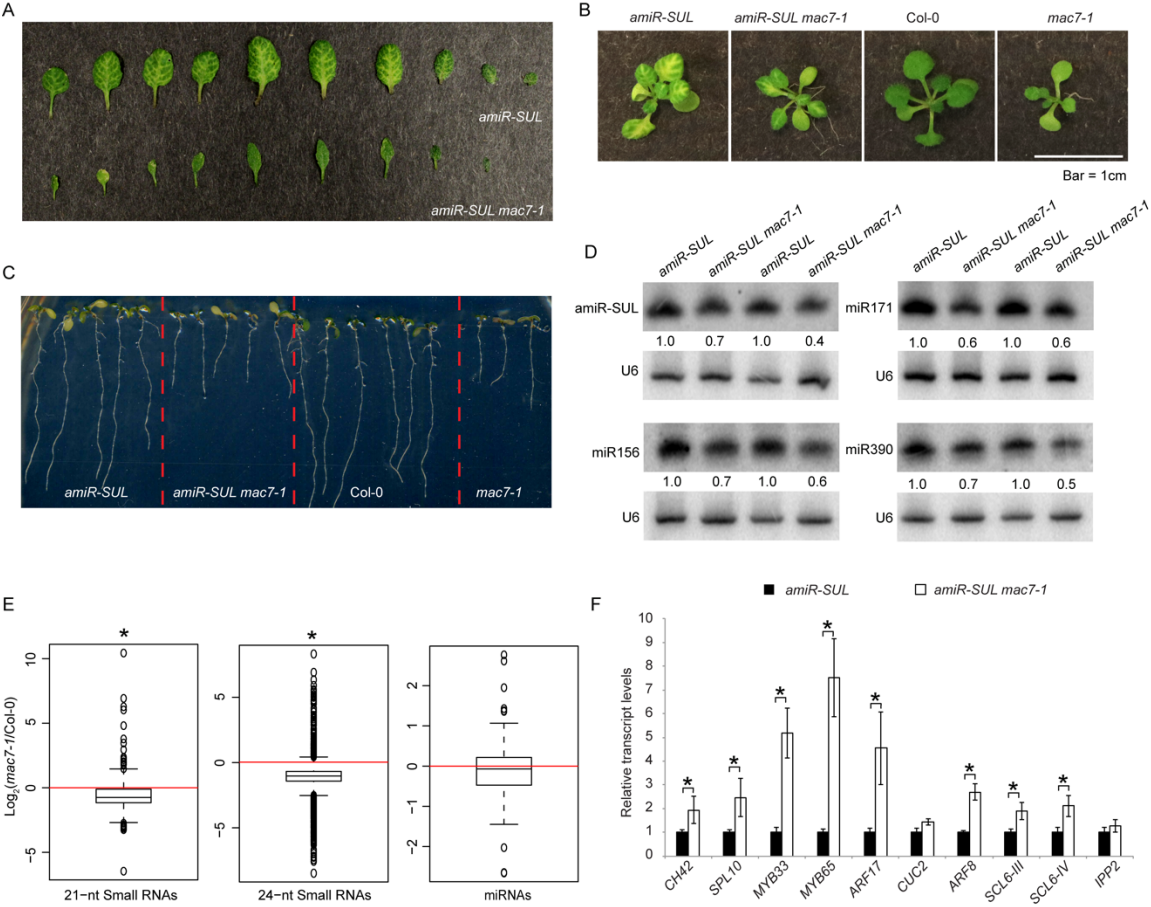


Figure 1.1 The silencing suppressor mutant *mac7-1* exhibits pleiotropic phenotypes and compromised miRNA accumulation.

(A) Differences in vein-centered bleaching in 3 to 4-week old *pSUC2:amiR-SUL* (*amiR-SUL*) and *amiR-SUL mac7-1* seedlings. (B and C) Morphological phenotypes of *amiR-SUL*, *amiR-SUL mac7-1*, Col-0 and *mac7-1* plants. Images of rosettes and roots were taken from 2 to 3-week old and 1-week old plants, respectively. (D) RNA gel blotting analysis of miRNAs from *amiR-SUL* and *amiR-SUL mac7-1* inflorescences. The miRNA signals were quantified and normalized to those of U6, and values were relative to *amiR-SUL* (arbitrarily set to 1). Two biological replicates of inflorescences collected from plants grown separately but under the same conditions were processed and shown. (E) Global abundance of 21-nt and 24-nt small RNAs and miRNAs in Col-0 and *mac7-1* as determined by small RNA-seq. Small RNA libraries were generated from 12-day-old seedlings growing on MS plates. The normalization of small RNAs was against 45S rRNA reads and abundance was expressed as RPMR (reads per million of 45S rRNA reads), and $\log_2$ ratios of *mac7-1*/Col-0 were plotted. Asterisks indicate that the mean is significantly below 0 (Wilcoxon test, $p < 2.2e-16$). (F) Determination of miRNA target mRNA levels in *amiR-SUL* and *amiR-SUL mac7-1* using 12-day-old seedlings by quantitative RT-PCR (RT-qPCR). The housekeeping gene *IPP2* was included as a control. Expression levels were normalized to those of *UBQUITIN5* (*UBQ5*) and compared with those in *amiR-SUL* (set to 1). Error bars indicate standard deviation from three technical replicates. Asterisks indicate significant difference between Col-0 and *mac7-1* (t-test, $p < 0.05$).

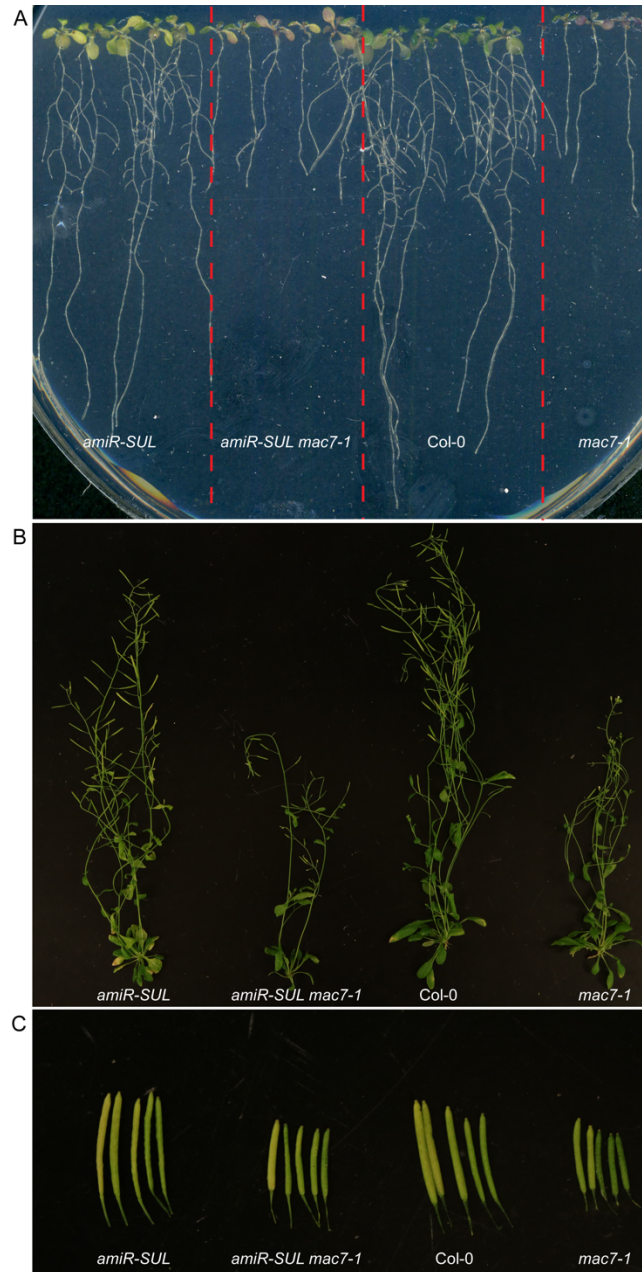


Figure 1.2 The *mac7-1* mutant shows pleiotropic developmental phenotypes.

(A to C) Images of *amiR-SUL*, *amiR-SUL mac7-1*, *Col-0* and *mac7-1* plants at different developmental stages. The plants were grown under long days.

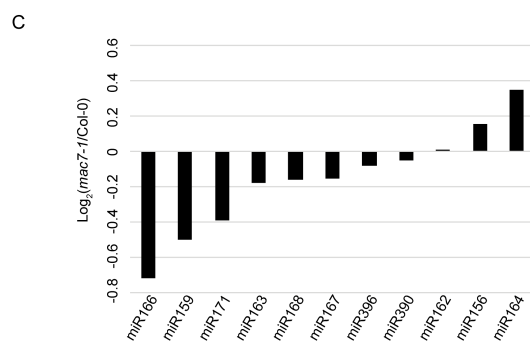
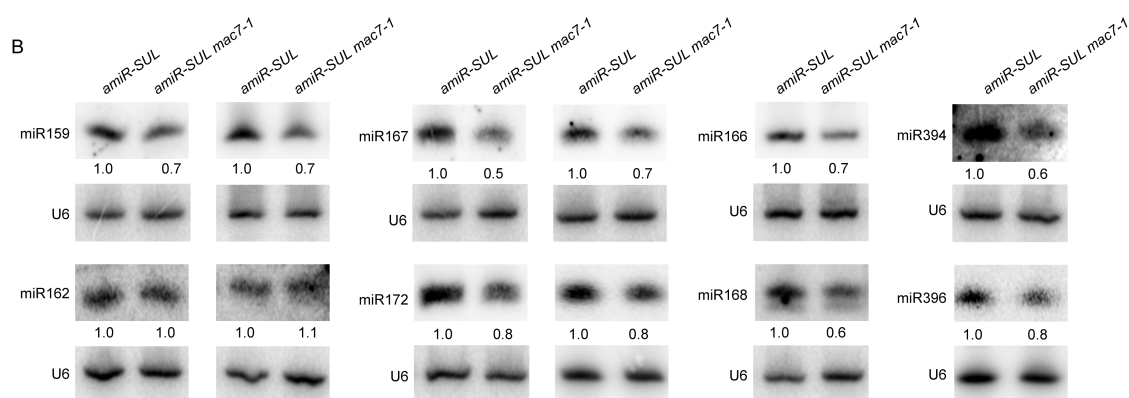
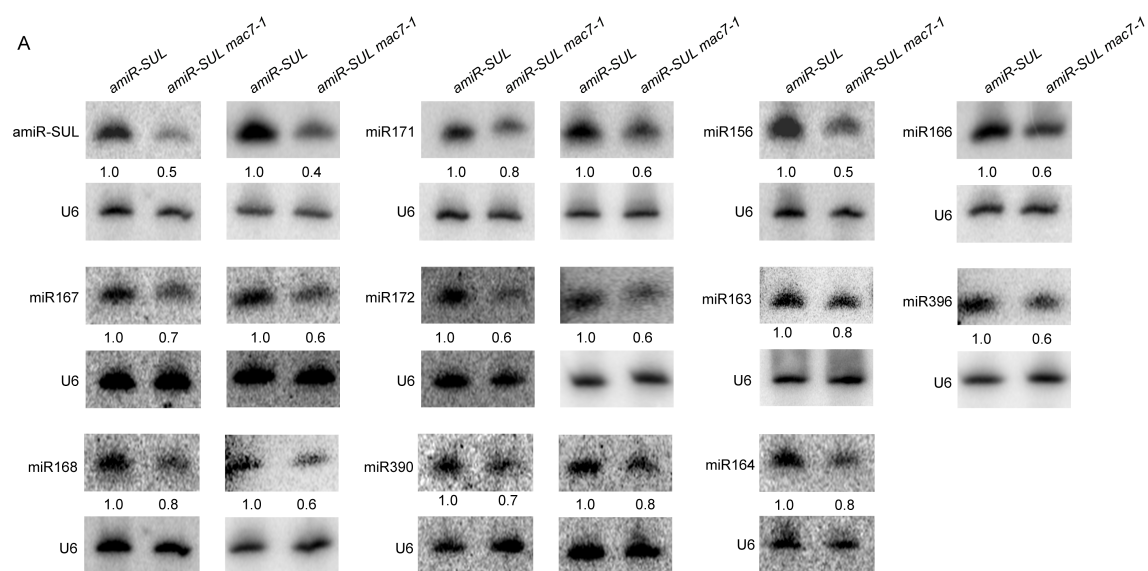


Figure 1.3 Reduced miRNA accumulation in both seedlings and inflorescences in the *mac7-1* mutant.

RNA gel blotting to determine the levels of miRNAs from 3-week-old seedlings (A) or inflorescences (B) of *amiR-SUL* and *amiR-SUL mac7-1*. The miRNA signals were quantified and normalized to those of U6 RNAs. Relative values were calculated by comparison with those in *amiR-SUL* (arbitrarily set to 1.0). When two blots are shown for the same miRNA, they represent biological replicates from independently harvested plant materials. (C) Relative levels of individual miRNAs in Col-0 and *mac7-1* as determined by small RNA-seq. Small RNA libraries were generated from 12-day-old seedlings grown on MS plates. The normalization of small RNAs was against 45S rRNA reads and abundance was expressed as RPKM (reads per million of 45S rRNA reads), and $\log_2$ ratios of *mac7-1*/Col-0 were plotted. Expression levels of different members from the same miRNA family were combined.

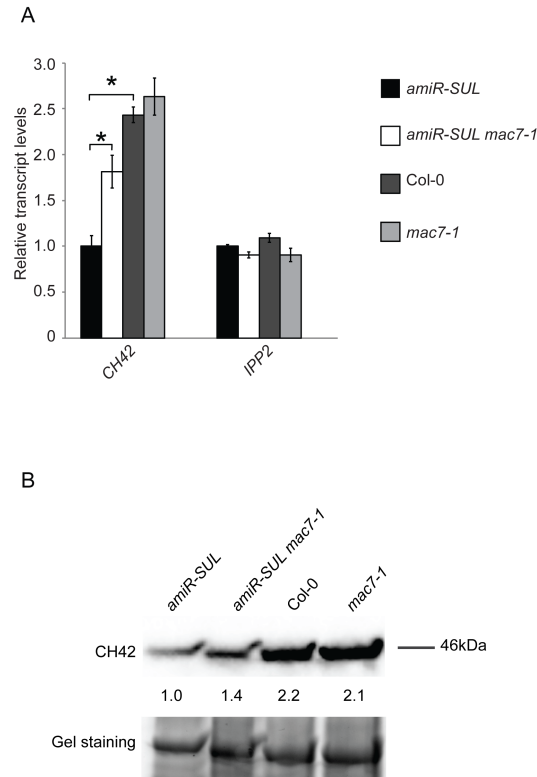


Figure 1.4 The *mac7-1* mutation in the *Col-0* background does not affect *CH42* expression at either the mRNA or protein levels.

(A) Transcript levels of *CH42* in inflorescences of the indicated genotypes as determined by RT-qPCR. The housekeeping gene *IPP2* was included as a control. *UBQ5* was used as an internal control and values in *amiR-SUL* were set to 1. Error bars indicate standard deviation from three technical replicates. Asterisks indicate significant difference between the samples (t-test, $p < 0.05$). (B) *CH42* protein levels in 12-day-old seedlings of the indicated genotypes as determined by immunoblotting blotting using anti-*CH42* antibodies. Note that the reduced levels of *CH42* in *amiR-SUL* relative to *Col-0*, as well as the size of the band, attest to the specificity of the antibodies in recognizing *CH42*.

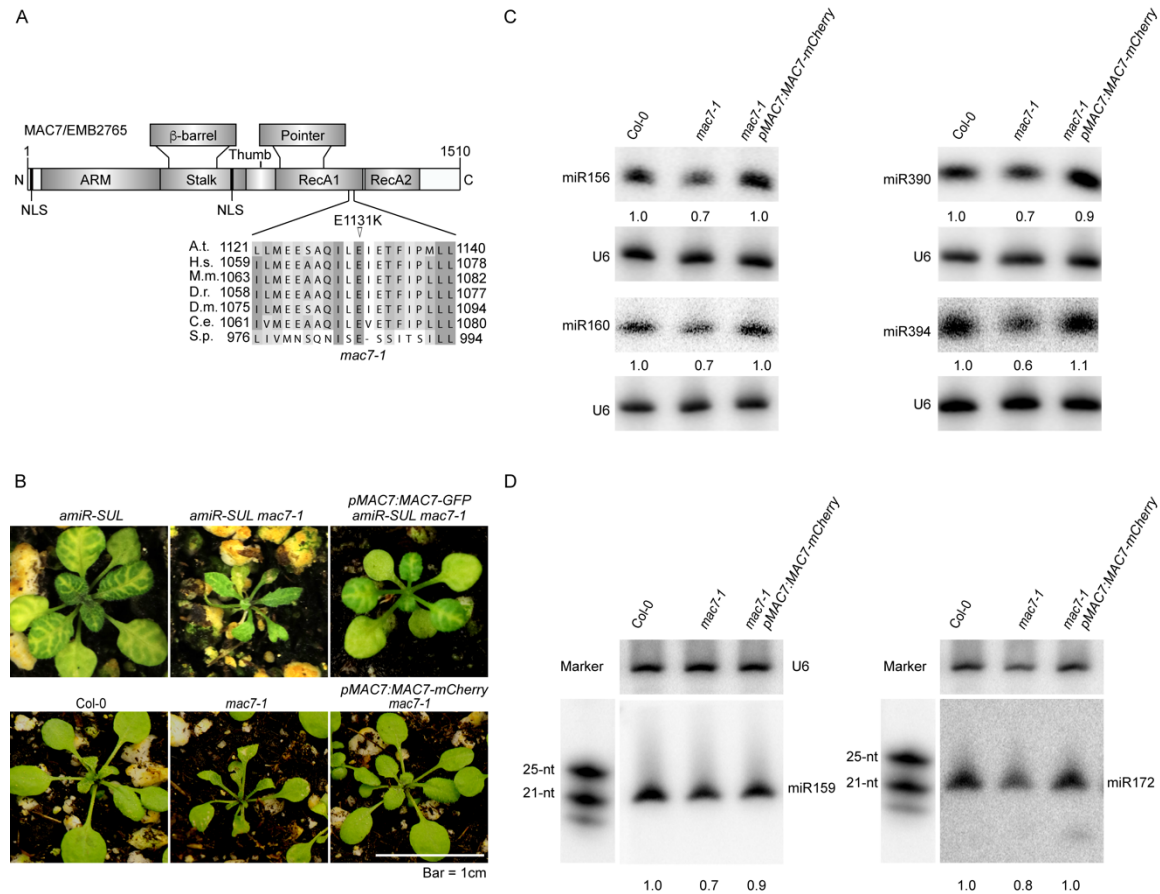


Figure 1.5 A point mutation in *MAC7* is responsible for the morphological and molecular phenotypes in the *mac7-1* mutant.

(A) A diagram of the *MAC7* protein showing various domains, the predicted nuclear localization signal (NLS), and the E1131K mutation in the *mac7-1* mutant. A sequence alignment of *MAC7* and its orthologs in the region containing the E1131K mutation in the *mac7-1* mutant is also shown. Abbreviations for species are as follows: *Arabidopsis thaliana* (A.t.), *Homo sapiens* (H.s.), *Mus musculus* (M.m.), *Danio rerio* (D.r.), *Drosophila melanogaster* (D.m.), *Caenorhabditis elegans* (C.e.), and *Schizosaccharomyces pombe* (S.p.). N: N-terminus; ARM: armadillo domain; RecA: RecA-like domains; C: C-terminus. The point mutation site is labeled by a triangle. (B) Morphological phenotypes of 3 to 4-week old seedlings of the indicated genotypes. *pMAC7:MAC7-GFP* and *pMAC7:MAC7-mCherry* were transformed into *amiR-SUL mac7-1* and *mac7-1*, respectively. (C and D) RNA gel blotting analysis of miRNAs from Col-0, *mac7-1*, and the complementation line *mac7-1 pMAC7:MAC7-mCherry* using inflorescences (C) and 12-day-old seedlings (D). The miRNA signals were quantified as described in Figure 1.1D. RNA markers (NEB, N2102S) shown in (D) were resolved in the same gel as miRNAs and probed separately by a DNA probe complementary to the marker sequences.

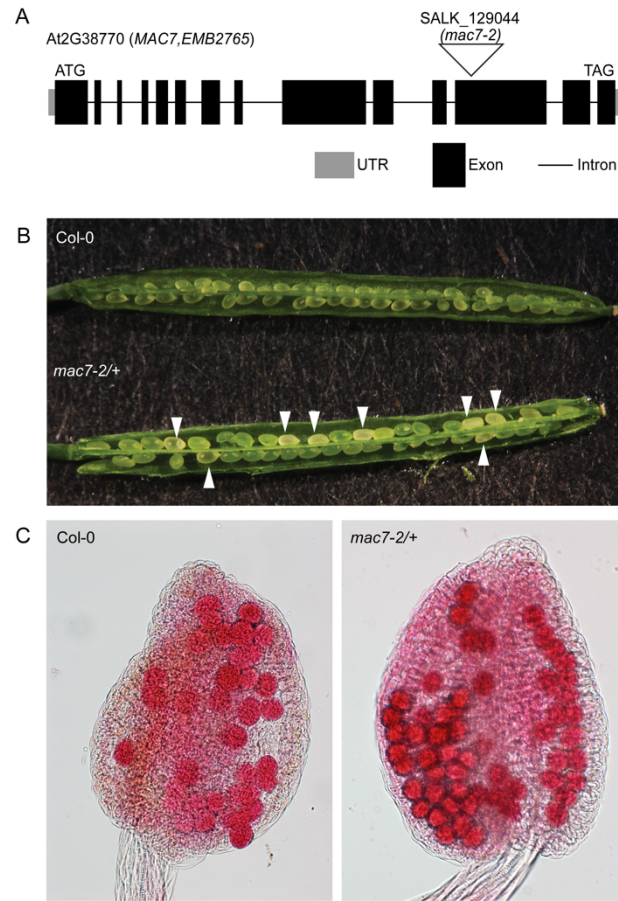


Figure 1.6 The *MAC7* T-DNA insertion line *mac7-2* is a single-locus, recessive, embryo-lethal mutant.

(A) Gene structure of At2g38770 (*MAC7*). The position of the T-DNA insertion (SALK_129044), designated *mac7-2*, is shown as a triangle. The gene structure is adapted from Integrated Genome Browser. (B) Developing seeds in siliques from wild type (Col-0) and *mac7-2/+* plants. The white seeds (dead seeds) are indicated by white arrowheads. (C) Alexander's staining of anthers from Col-0 and *mac7-2/+* plants.

Arabidopsis, thaliana/1-1510	1	MTKVVYGTGTDPKRHHVAALVPEELPSHPAAEPLESKPGSMIPSSILITLISQQDRLTKALEESWIKTGGKLPKPDPOPEVVKLIVYTELKVTSGRKPPVLRQVRI	120
Homo_sapiens/1-1485	1	MAAPAPKKIVAP-----TVSQINAEFVTQLACKYVAPHIKKSPDIKIVIEEYKEIKVSRFA-----IRKIMLEFSOYLENVLWNYS82	
Mus_musculus/1-1481	1	MAAPAPKKIVAP-----TVSQINAEFVTQLACKYVAPHIKKSPDIKIVIEEYKEIKVSRFA-----IRKIMLEFSOYLENVLWNYS82	
Danio_rerio/1-1525	1	MQKDSRLVKTSAPE-----VSQINAEFVTQLSNKYVAPHVKNKLPDSKILIEEVQVNEIKSKFS-----IRKIMLEFSOYLENVLWNYS82	
Drosophila/1-1486	1	-----MKRSOTRAEQAVPAKTTIKHFLTLTSSDVIWQLDQVWTPDQKADHPYSAKIIEFYSKFGSDGGGHS-----ARVIMMLEFSOYLENVLWNYS82	
Caenorhabditis_elegans/1-1467	1	-----MVTKRHGEAAV-----TRGAIENDTISAAVAKFWAPFAETAETHENDAKLIOTLVNDEMLKTSFN-----SRKIMMLEFSOYLENVLWNYS82	
Schizosaccharomyces_pombe/1-1284	1	-----MVLKNNKKQIINYANSNW-----GONGEVREOETMLQCLLEAVIVYSKDA-----KQVLSLHELKLFNFVNVN81	
Arabidopsis, thaliana/1-1510	121	PITAAETEEHYMSMLINKLEKRENVAAWICEDHORDQILKGLKGLIKL-----EGROLTIAKTNQIVVHINAQSLEDAKXVEIVLSAGDSWHSLSYGRGOMELDLOPDLI	131
Homo_sapiens/1-1485	83	PEVSSKAYLMSICCMVNEKPRENVPAMEIKKKPDHPFFKFKHILKALAAE-----TDGFSLHQTVLLEFLDHCNSLEVDLIRSOVQQLISLPMMMGQLARLELEKTKPKLR	195
Mus_musculus/1-1481	83	PEVSSKAYLMSICCMVNEKPRENVPAMEIKKKPDHPFFKFKHILKALAAE-----TDGFSLHQTVLLEFLDHCNSLEVDLIRSOVQQLISLPMMMGQLARLELEKTKPKLR	195
Danio_rerio/1-1525	83	PEVSSKAYLMSICCMVNEKPRENVPAMEIKKKPDHPFFKFKHILKALAAE-----TDGFSLHQTVLLEFLDHCNSLEVDLIRSOVQQLISLPMMMGQLARLELEKTKPKLR	195
Drosophila/1-1486	98	RETAETHAHLMSIVIMANEKPREVEVWIVTEKLPDQPAFPHVHLESLCPKAKAKEASTLRBTALEMLNHCCNSMEIEICREDAKRVLSLMMHGOARRRLEKTEFVPEWN	213
Caenorhabditis_elegans/1-1467	82	PEKASKAWMSIVIMANEKPRENRLSDWNCITKSEHFPFKSLQLSL-----QEGLASSEHCALETLVNAAGSVETPIVHKTEKRLVLSIEIWMAGDLSOREDLFKOKKLK	293
Schizosaccharomyces_pombe/1-1284	68	TMSLNHINILTMCLLYKSY-----EYITLDLIDE--NRGOLFEKVEVSL-----SINLSVYVMIQITLQFSNIEKRLKLVQITNISLNSLDNLKVKYLLHDSFSL	173
Arabidopsis, thaliana/1-1510	232	KWK-----RSSKWAAEAKSGKEGFDLSSPFAVNFGRLEEVVEVGHVFADEVEDTASQSLVDSSVLYCE-----RMEFLDMLNOLPTRRYLRPLVADIAVAKKRL	335
Homo_sapiens/1-1485	196	FNNLIKKNDMDPEAREQAYV-----RRFLSQIQKIFISVSKSI-----PLSEPV-TMDKVHYCE-----RFEELMDLEALLPTRRWFTLDDSLHLVHYCL	386
Mus_musculus/1-1481	196	FNNLIKKNDMDPEAREQAYV-----RRFLSQIQKIFISVSKSI-----PLSEPV-TMDKVHYCE-----RFEELMDLEALLPTRRWFTLDDSLHLVHYCL	386
Danio_rerio/1-1525	214	YWKRLKK-----EK-----DSKPELVW-----RHFMONLIDFLHIESI-----PAEGEV-PRNVHYCE-----RLEFLIDLEALLPTRRWFTLDDSLHLVHYCL	386
Drosophila/1-1486	214	YWKRLKK-----EK-----DSKPELVW-----RHFMONLIDFLHIESI-----PAEGEV-PRNVHYCE-----RLEFLIDLEALLPTRRWFTLDDSLHLVHYCL	386
Caenorhabditis_elegans/1-1467	194	WE-----NVROMTAAADANNEF-----RTYLVNLEKIKRVINSLEPNAQSEEGEVRDPSIKYCE-----RFEELMDLEALLPTRRWFTLDDSLHLVHYCL	386
Schizosaccharomyces_pombe/1-1284	174	AFD-----SYKEAPSAIV-----KFFLHNLRSWHS-----LKSISLAQTEKQEAQVTPLLAIIIMLSYLSAFTRFNFAMPIVHSCFYTLR	358
Arabidopsis, thaliana/1-1510	336	ALYKHEGKGLAQVLDIQYKEREIKHNDTQ-----TDDLEALQYDFRMAFLLAKKPKKLODISLANIGSIHKSSDARRISVLSLEDRDVCCKLVLS--RHDPWAS	445
Homo_sapiens/1-1485	287	NLVRREEDGHSQGLDMKRYTQETINDOTNA-----TQENMTTIHYDRITSLQRAAAHFPELYDPALSNVAEVDRESVKFFGPPSSNTQHOVA-SYGLUPTL--PKNED	395
Mus_musculus/1-1481	287	NLVRREEDGHSQGLDMKRYTQETINDOTNA-----TQENMTTIHYDRITSLQRAAAHFPELYDPALSNVAEVDRESVKFFGPPSSNTQHOVA-SYGLUPTL--PKNED	395
Danio_rerio/1-1525	287	NLVRREEDGHSQGLDMKRYTQETINDOTNA-----TQENMTTIHYDRITSLQRAAAHFPELYDPALSNVAEVDRESVKFFGPPSSNTQHOVA-SYGLUPTL--PKNED	395
Drosophila/1-1486	301	PLVRRREEGKGLDMKRYTQETINDOTNA-----TQENMTTIHYDRITSLQRAAAHFPELYDPALSNVAEVDRESVKFFGPPSSNTQHOVA-SYGLUPTL--PKNED	395
Caenorhabditis_elegans/1-1467	291	SLSITDAGSLFOLVLOKAVRFEJDLDSLR-----LTHKEVTSQSVYTRLOKAAHLENTMKHEVYLVSGWOTRRALQDQFDNNNAHVIRFAEYLVHVAFGEDHPH	403
Schizosaccharomyces_pombe/1-1284	255	LYVDSN-----LKKMTDQVYVLEFEDNTREYKEQKINDELVYV-----LQTLISDQKEGDLVFCETQSLQQRKLEETISFSEFNFKSLCKSKCYITTSF--PERVA	365
Arabidopsis, thaliana/1-1510	446	-----KQFTEVVVSSEKQOSKEAIALPLYPNEQIMWDSVPSINSYSGEGCLAPKLNLOFTLHOV-----LNRNF--NIFRESTEYIEODQEAHPHLLAH	541
Homo_sapiens/1-1485	396	TT-----FDKEFLLELLVSRHRRISQIQLNQMPLYPTEKIWDENIYPTTEYSGEGCLAPKLNLOFTLHOV-----LNRNF--NIFRESTEYIEODQEAHPHLLAH	541
Mus_musculus/1-1481	396	TT-----FDKEFLLELLVSRHRRISQIQLNQMPLYPTEKIWDENIYPTTEYSGEGCLAPKLNLOFTLHOV-----LNRNF--NIFRESTEYIEODQEAHPHLLAH	541
Danio_rerio/1-1525	396	TT-----FDKEFLLELLVSRHRRISQIQLNQMPLYPTEKIWDENIYPTTEYSGEGCLAPKLNLOFTLHOV-----LNRNF--NIFRESTEYIEODQEAHPHLLAH	541
Drosophila/1-1486	404	VSPLDWHRVDOERELLLTHKREKSEALKNMPLVYDQVLEENINWVSPSEYSGEGCLAPKLNLOFTLHOV-----LNRNF--NIFRESTEYIEODQEAHPHLLAH	541
Caenorhabditis_elegans/1-1467	403	SL-----LHLVPHQVETITLHCBPQHLQDGLRPLPTEVSVKQKRRAYENIEEAPACEEQVLEVEPHVPTNBPQYNNKRTIQETIEARAGMGPGLTVGGP	602
Schizosaccharomyces_pombe/1-1284	366	IK-----VDFELKNVFINTYDRTRLVNDYOETINFTLKQVGLGERSVMDQENSLTYNPLQNTAIDYISISFFMQQSKAYKLLLSLYAEALNFSQDNRLSKATNKLT--	473
Arabidopsis, thaliana/1-1510	542	INNEGDGTAQGSWRMAQVNOPIAQQKRNIGKSSVIMAMVPSIKS-VTQDSEWNSKSHOVFLCIRKSPFELPPEADRAVQBLGLQVVRGCLIDIDRISGALMND	659
Homo_sapiens/1-1485	496	QSEYGGVVGWARMADQVATVVEAKNIGEMWARRADVTINL-NVRDHKDEWEGRRKHQVCFITVRPNT-PYGTFRDRRRPIEQVGLVYVRGCEIOGMLDKGRVIED	610
Mus_musculus/1-1481	496	QSEYGGVVGWARMADQVATVVEAKNIGEMWARRADVTINL-NVRDHKDEWEGRRKHQVCFITVRPNT-PYGTFRDRRRPIEQVGLVYVRGCEIOGMLDKGRVIED	610
Danio_rerio/1-1525	496	QSEYGGVVGWARMADQVATVVEAKNIGEMWARRADVTINL-NVRDHKDEWEGRRKHQVCFITVRPNT-PYGTFRDRRRPIEQVGLVYVRGCEIOGMLDKGRVIED	610
Drosophila/1-1486	511	KSDGDGVVGWARMALPASVVEAKNIGEMWARRADVTINL-NVRDHKDEWEGRRKHQVCFITVRPNT-PYGTFRDRRRPIEQVGLVYVRGCEIOGMLDKGRVIED	610
Caenorhabditis_elegans/1-1467	506	QHESTRNETVVGWARMALQDNRQISEAKPLVEKSKAVRGVVTINL-GRRQIRQEWENRRKHQVCFIVACRSRKSASGLKFDVRRPFSQVLEIVSRGCEIOGMLDKGRVIED	610
Schizosaccharomyces_pombe/1-1284	474	-----KDNFIS-----LNNKVTISAPQIQVLOFQKCMGLSRGPFHSAALDRKNSIKSPFLCLYISKDMEYKLLHGNAOLDLEGVDTQITACN-----IDVGMFMS	674
Arabidopsis, thaliana/1-1510	660	IRKXKDEWKKPKKEMHIVIVAAQAAHLDVDTIAKGAEDVYQETFNIMRRKPKKNKAVLETIRNMNTCEVVPDWLHDIILGYGDSAHYSKMPNQIATLDFNDTFLSIEHLKA	726
Homo_sapiens/1-1485	611	GPPEPPNLL--RGESRTFRVLPDQNDQDMTNTIQNGAEDVYQETFNIMRRKPKKNKAVLETIRNMNTCEVVPDWLHDIILGYGDSAHYSKMPNQIATLDFNDTFLSIEHLKA	726
Mus_musculus/1-1481	611	GPPEPPNLL--RGESRTFRVLPDQNDQDMTNTIQNGAEDVYQETFNIMRRKPKKNKAVLETIRNMNTCEVVPDWLHDIILGYGDSAHYSKMPNQIATLDFNDTFLSIEHLKA	726
Danio_rerio/1-1525	611	GPPEPPNLL--RGESRTFRVLPDQNDQDMTNTIQNGAEDVYQETFNIMRRKPKKNKAVLETIRNMNTCEVVPDWLHDIILGYGDSAHYSKMPNQIATLDFNDTFLSIEHLKA	726
Drosophila/1-1486	626	GPPEPPNLL--RGESRTFRVLPDQNDQDMTNTIQNGAEDVYQETFNIMRRKPKKNKAVLETIRNMNTCEVVPDWLHDIILGYGDSAHYSKMPNQIATLDFNDTFLSIEHLKA	726
Caenorhabditis_elegans/1-1467	624	TAWKAKKII--PDGRLKRLLDQNDRIHME--GQTKDDIYDTNLSIRDSKTNKAVLETIRNMNTCEVVPDWLHDIILGYGDSAHYSKMPNQIATLDFNDTFLSIEHLKA	726
Schizosaccharomyces_pombe/1-1284	575	QS-----DSDNKSINWISPYFIHSLAGLGEYRPQGL--KENFALVLISREAFKWKYLDNINLVSLNRAKEFKKIFEDLIFGFTIDICAFPNAG--LNSIYARNLINTVEQ	679
Arabidopsis, thaliana/1-1510	780	SFSDYSFVFINAEAGALDPS--PFRKILTPKTLKNGAISGKISVENPANNMMVDAKPKKE--LIVEAYTPDPDPPYPQDQKNSVKETPTVGAISIDPGLTMVVGPGT	892
Homo_sapiens/1-1485	727	SPFGHNKVV--TVS-DPALQIPFRITF--PVRSGOKKKRDAAD--VEEDTEEAKTILVEPHVIPNRGPYPNPKNRTIQETIEARAGMGPGLTVGGP	827
Mus_musculus/1-1481	727	SPFGHNKVV--TVS-DPALQIPFRITF--PVRSGOKKKRDAAD--VEEDTEEAKTILVEPHVIPNRGPYPNPKNRTIQETIEARAGMGPGLTVGGP	827
Danio_rerio/1-1525	727	SPFGHNKVV--TVS-DPALQIPFRITF--PVRSGOKKKRDAAD--VEEDTEEAKTILVEPHVIPNRGPYPNPKNRTIQETIEARAGMGPGLTVGGP	827
Drosophila/1-1486	741	SPFOYNLKC--EVP-VENRQPFRLIFEDV--PIQ--KESDGE--PEEKPEELKTSILVQPYKYEARGPYSDKPKQNSIRTPTOVIEARAGMGPGLTVGGP	840
Caenorhabditis_elegans/1-1467	738	SPFOYNLKC--EVP-VENRQPFRLIFEDV--PIQ--KESDGE--PEEKPEELKTSILVQPYKYEARGPYSDKPKQNSIRTPTOVIEARAGMGPGLTVGGP	840
Schizosaccharomyces_pombe/1-1284	680	VLRNCHV-----ADGDFEKEAVPFKLEKEL--ERR--QD--VEIKPGE--LITLVPLTRKKVTPYSYD--PKNKQVPSQVIAKSMQMDPLTMVVGPGT	831
Arabidopsis, thaliana/1-1510	893	GKTDTAVQINLVLYHHPDORTLIVTHSNQALNOL--EKIMALDERHLRLHG--HGEAELETEKDSRYGRNVYIARRIELEEKRRQKSLGVGGPDYSYTCETAGVLYQMSR	943
Homo_sapiens/1-1485	828	GKTDTAVQINLVLYHHPDORTLIVTHSNQALNOL--EKIMALDERHLRLHG--HGEAELETEKDSRYGRNVYIARRIELEEKRRQKSLGVGGPDYSYTCETAGVLYQMSR	943
Mus_musculus/1-1481	828	GKTDTAVQINLVLYHHPDORTLIVTHSNQALNOL--EKIMALDERHLRLHG--HGEAELETEKDSRYGRNVYIARRIELEEKRRQKSLGVGGPDYSYTCETAGVLYQMSR	943
Danio_rerio/1-1525	828	GKTDTAVQINLVLYHHPDORTLIVTHSNQALNOL--EKIMALDERHLRLHG--HGEAELETEKDSRYGRNVYIARRIELEEKRRQKSLGVGGPDYSYTCETAGVLYQMSR	943
Drosophila/1-1486	841	GKTDTAVQINLVLYHHPDORTLIVTHSNQALNOL--EKIMALDERHLRLHG--HGEAELETEKDSRYGRNVYIARRIELEEKRRQKSLGVGGPDYSYTCETAGVLYQMSR	943
Caenorhabditis_elegans/1-1467	832	GKTDTAVQINLVLYHHPDORTLIVTHSNQALNOL--EKIMALDERHLRLHG--HGEAELETEKDSRYGRNVYIARRIELEEKRRQKSLGVGGPDYSYTCETAGVLYQMSR	943
Schizosaccharomyces_pombe/1-1284	753	GKLVYCKLLEVLDOT--SNDRIVDN--SMTNLTLLLE--ARCFHCHLYE--DEEDKE--TLEVEYTLSSWISKLPGLREIGLAAEIQAFGHDAQSP--LIRFGLDGL	865
Arabidopsis, thaliana/1-1510	1009	WELFLAAC--GNEDNOSF--VRDRPFKDF--SOTPKPVFNSEFEKDMARAKGCFESHKTVQELTECAFELIKSTADANLMTKQAKVAMTCTHAALKRRDFLOLSKRYDN	1102
Homo_sapiens/1-1485	944	WELFLAAC--GNEDNOSF--VRDRPFKDF--SOTPKPVFNSEFEKDMARAKGCFESHKTVQELTECAFELIKSTADANLMTKQAKVAMTCTHAALKRRDFLOLSKRYDN	1102
Mus_musculus/1-1481	944	WELFLAAC--GNEDNOSF--VRDRPFKDF--SOTPKPVFNSEFEKDMARAKGCFESHKTVQELTECAFELIKSTADANLMTKQAKVAMTCTHAALKRRDFLOLSKRYDN	1102
Danio_rerio/1-1525	944	WELFLAAC--GNEDNOSF--VRDRPFKDF--SOTPKPVFNSEFEKDMARAKGCFESHKTVQELTECAFELIKSTADANLMTKQAKVAMTCTHAALKRRDFLOLSKRYDN	1102
Drosophila/1-1486	957	WELFLAAC--GNEDNOSF--VRDRPFKDF--SOTPKPVFNSEFEKDMARAKGCFESHKTVQELTECAFELIKSTADANLMTKQAKVAMTCTHAALKRRDFLOLSKRYDN	1102
Caenorhabditis_elegans/1-1467	948	WELFLAAC--GNEDNOSF--VRDRPFKDF--SOTPKPVFNSEFEKDMARAKGCFESHKTVQELTECAFELIKSTADANLMTKQAKVAMTCTHAALKRRDFLOLSKRYDN	1102
Schizosaccharomyces_pombe/1-1284	866	WELFLAAC--GNEDNOSF--VRDRPFKDF--SOTPKPVFNSEFEKDMARAKGCFESHKTVQELTECAFELIKSTADANLMTKQAKVAMTCTHAALKRRDFLOLSKRYDN	1102
Arabidopsis, thaliana/1-1510	1121	ILMEEASADILEETFIPIPLQNPQ--DGSRKRWIMIGDQHLPPVKNMAQKYSNMNDSLTPRVKKEIPYIEINAGGRAPSLAKVYWRADLDGSLISAP--IFODANAESY	1275
Homo_sapiens/1-1485	1059	ILMEEASADILEETFIPIPLQNPQ--DGSRKRWIMIGDQHLPPVKNMAQKYSNMNDSLTPRVKKEIPYIEINAGGRAPSLAKVYWRADLDGSLISAP--IFODANAESY	1275
Mus_musculus/1-1481	1063	ILMEEASADILEETFIPIPLQNPQ--DGSRKRWIMIGDQHLPPVKNMAQKYSNMNDSLTPRVKKEIPYIEINAGGRAPSLAKVYWRADLDGSLISAP--IFODANAESY	1275
Danio_rerio/1-1525	1063	ILMEEASADILEETFIPIPLQNPQ--DGSRKRWIMIGDQHLPPVKNMAQKYSNMNDSLTPRVKKEIPYIEINAGGRAPSLAKVYWRADLDGSLISAP--IFODANAESY	1275
Drosophila/1-1486	1075	ILMEEASADILEETFIPIPLQNPQ--DGSRKRWIMIGDQHLPPVKNMAQKYSNMNDSLTPRVKKEIPYIEINAGGRAPSLAKVYWRADLDGSLISAP--IFODANAESY	1275
Caenorhabditis_elegans/1-1467	1061	ILMEEASADILEETFIPIPLQNPQ--DGSRKRWIMIGDQHLPPVKNMAQKYSNMNDSLTPRVKKEIPYIEINAGGRAPSLAKVYWRADLDGSLISAP--IFODANAESY	1275
Schizosaccharomyces_pombe/1-1284	976	ILMEEASADILEETFIPIPLQNPQ--DGSRKRWIMIGDQHLPPVKNMAQKYSNMNDSLTPRVKKEIPYIEINAGGRAPSLAKVYWRADLDGSLISAP--IFODANAESY	1275
Arabidopsis, thaliana/1-1510	1238	SYLVNVPDYELRGESTSWIWEYDNOGKAKILVSVLMLHGLVYPAKNSILITTINGKHLRDVYNNREVPYPIIRPASKVITVQKFGQGN--DFILSLSVTRVYGLHDBVRLVVM	1356
Homo_sapiens/1-1485	1176	DFQILNVEDFORVGESENNFYDNLGAEEYVVALFMVCLLGYPAKISILITTINGKHLRDVYNNREVPYPIIRPASKVITVQKFGQGN--DFILSLSVTRVYGLHDBVRLVVM	1356
Mus_musculus/1-1481	1176	DFQILNVEDFORVGESENNFYDNLGAEEYVVALFMVCLLGYPAKISILITTINGKHLRDVYNNREVPYPIIRPASKVITVQKFGQGN--DFILSLSVTRVYGLHDBVRLVVM	1356
Danio_rerio/1-1525	1176	DFQILNVEDFORVGESENNFYDNLGAEEYVVALFMVCLLGYPAKISILITTINGKHLRDVYNNREVPYPIIRPASKVITVQKFGQGN--DFILSLSVTRVYGLHDBVRLVVM	1356
Drosophila/1-1486	1192	DFQILNVEDFORVGESENNFYDNLGAEEYVVALFMVCLLGYPAKISILITTINGKHLRDVYNNREVPYPIIRPASKVITVQKFGQGN--DFILSLSVTRVYGLHDBVRLVVM	1356
Caenorhabditis_elegans/1-1467	1176	DFQILNVEDFORVGESENNFYDNLGAEEYVVALFMVCLLGYPAKISILITTINGKHLRDVYNNREVPYPIIRPASKVITVQKFGQGN--DFILSLSVTRVYGLHDBVRLVVM	1356
Schizosaccharomyces_pombe/1-1284	1085	EVDFINWGAESQOTTEVSGYKNGLEAEYVVALFMVCLLGYPAKISILITTINGKHLRDVYNNREVPYPIIRPASKVITVQKFGQGN--DFILSLSVTRVYGLHDBVRLVVM	1356
Arabidopsis, thaliana/1-1510	1357	SARALGLVYVCRRLSEGVETOPTFOLLQRLGLNFNENTAYTVRVVEEYENS--YLVHDEVEEAMHIDDRMNKIEYQAGGAYEQYNNM-----	1447
Homo_sapiens/1-1485	1299	SARALGLVYVCRRLSEGVETOPTFOLLQRLGLNFNENTAYTVRVVEEYENS--YLVHDEVEEAMHIDDRMNKIEYQAGGAYEQYNNM-----	1447
Mus_musculus/1-1481	1299	SARALGLVYVCRRLSEGVETOPTFOLLQRLGLNFNENTAYTVRVVEEYENS--YLVHDEVEEAMHIDDRMNKIEYQAGGAYEQYNNM-----	1447
Danio_rerio/1-1525	1311	SARALGLVYVCRRLSEGVETOPTFOLLQRLGLNFNENTAYTVRVVEEYENS--YLVHDEVEEAMHIDDRMNKIEYQAGGAYEQYNNM-----	1447
Drosophila/1-1486	1299	SARALGLVYVCRRLSEGVETOPTFOLLQRLGLNFNENTAYTVRVVEEYENS--YLVHDEVEEAMHIDDRMNKIEYQAGGAYEQYNNM-----	1447
Caenorhabditis_elegans/1-1467	1297	SARALGLVYVCRRLSEGVETOPTFOLLQRLGLNFNENTAYTVRVVEEYENS--YLVHDEVEEAMHIDDRMNKIEYQAGGAYEQYNNM-----	1447
Schizosaccharomyces_pombe/1-1284	1295	SARALGLVYVCRRLSEGVETOPTFOLLQRLGLNFNENTAYTVRVVEEYENS--YLVHDEVEEAMHIDDRMNKIEYQAGGAYEQYNNM-----	1447
Arabidopsis, thaliana/1-1510	1448	AQMEQNGHDMESQSVYD-----GDSEKNNMOOLNOSPDIQGLSKSEVGM-----EVDNN-----TGATSTPEAIPALSETPTTVVGAVSAPAEANTPQ--GFSEEN	1503
Homo_sapiens/1-1485	1389	AMVEEGEGEVOGQTELE-----TEEAMTVOADIIIPSTDSQCEPTAF-----QDTDTPE-----ASCEVSAGKNTPTDIAAAE-----PVSAAGATPQ--DAES	1475
Mus_musculus/1-1481	1393	AMVEEGEGEVOGQTELE-----TEEAMTVOADIIIPSTDSQCEPTAF-----QDTDTPE-----ASCEVSAGKNTPTDIAAAE-----PVSAAGATPQ--DAES	1475
Danio_rerio/1-1525	1397	AMVEEGEGEVOGQTELE-----TEEAMTVOADIIIPSTDSQCEPTAF-----QDTDTPE-----ASCEVSAGKNTPTDIAAAE-----PVSAAGATPQ--DAES	1475
Drosophila/1-1486	1388	AMVEEGEGEVOGQTELE-----TEEAMTVOADIIIPSTDSQCEPTAF-----QDTDTPE-----ASCEVSAGKNTPTDIAAAE-----PVSAAGATPQ--DAES	1475
Caenorhabditis_elegans/1-1467	1378	AMRDAVQAMNVEYSQ-----RLNLPDIDETOMDQTEHEKKHREAMEKRRKQEMDMDKEADIIHED-----MDHMEQAPATAA--PAPG	1457
Schizosaccharomyces_pombe/1-1284	1274	-----VEMTKRNLNTN-----	1284
Arabidopsis, thaliana/1-1510	1504	GKADKKN-----	1510
Homo_sapiens/1-1485	1480	APETTK-----	1485
Mus_musculus/1-1481	1476	VPTTE-----	1481
Danio_rerio/1-1525	1517	GDSEEDQE-----	1525
Drosophila/1-1486	1481	APAEES-----	1486
Caenorhabditis_elegans/1-1467	1485	APAVEPPPK-----	1467
Schizosaccharomyces_pombe/1-1284	-----	-----	-----

Figure 1.7 Amino acid sequence alignment of *Arabidopsis* MAC7 and its orthologs in other species.

Sequence alignment of *Arabidopsis thaliana*. MAC7 and its orthologs from *Homo sapiens*, *Mus musculus*, *Danio rerio*, *Drosophila melanogaster*, *Caenorhabditis elegans*, and *Schizosaccharomyces pombe*. Alignment of protein sequences was performed with Muscle and the alignments were edited with Jalview. The applied color was based on percentage of identity of aligned amino acids.

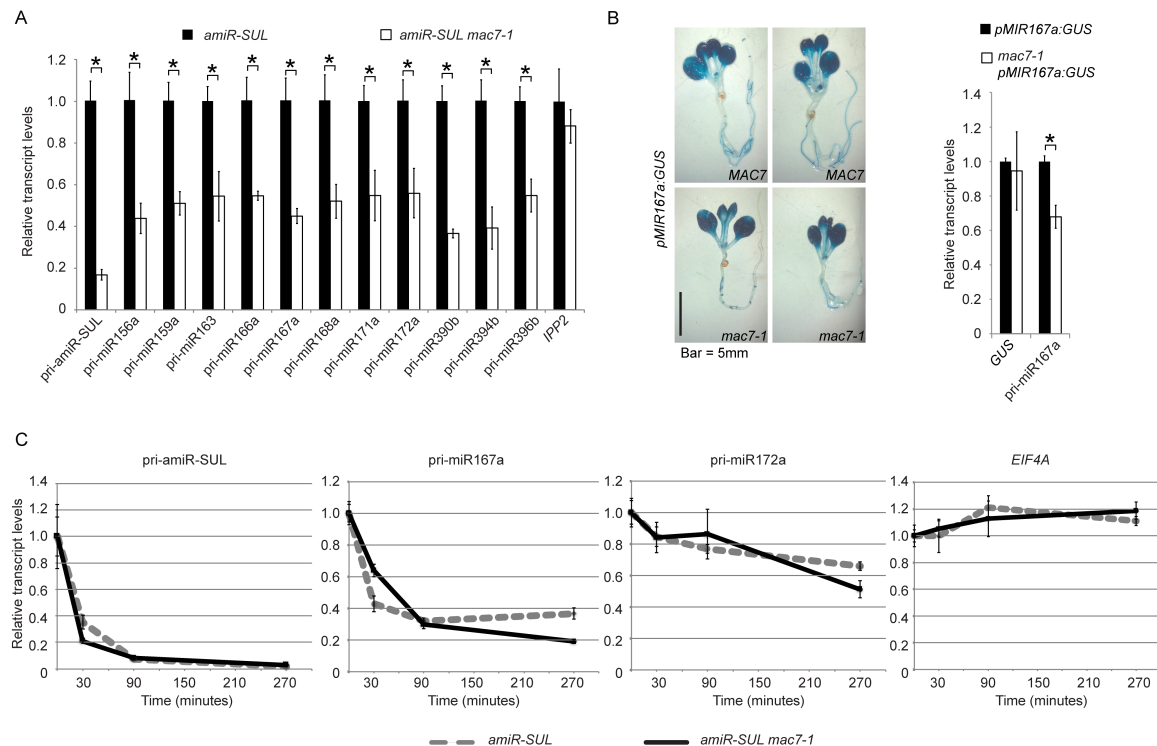


Figure 1.8 MAC7 promotes pri-miRNA production.

(A) Determination of pri-miRNA levels in *amiR-SUL* and *amiR-SUL mac7-1* inflorescences by RT-qPCR. The housekeeping gene *IPP2* was included as a control. Expression levels were normalized to those of *UBQUITIN5 (UBQ5)* and compared with those in *amiR-SUL* (set to 1). Error bars indicate standard deviation from three technical replicates. Asterisks indicate significant difference between *amiR-SUL* and *amiR-SUL mac7-1* (t-test, $p < 0.05$). (B) Representative GUS staining images of *pMIR167a:GUS* and *mac7-1 pMIR167a:GUS* seedlings. The transcript levels of *GUS* and endogenous *pri-miR167a* in *pMIR167a:GUS* and *mac7-1 pMIR167a:GUS* seedlings were determined by RT-qPCR. Expression levels were normalized to those of *UBQUITIN5 (UBQ5)* and compared with those in *pMIR167a:GUS* (set to 1). Error bars indicate standard deviation from three technical replicates. Asterisks, t-test $p < 0.05$. (C) Half-life measurements for pri-amiR-SUL, pri-miR167a, pri-miR172a and *EIF4A* mRNA. Two-week-old *amiR-SUL* and *amiR-SUL mac7-1* seedlings were treated with 0.6 mM cordycepin and harvested at various time points. RT-qPCR was performed to determine the levels of various pri-miRNAs and *EIF4A* mRNA. *UBQ5* served as an internal control. Values at time 0 were set to 1. Error bars indicate standard deviation from three technical replicates. Two biological replicates were performed and similar results were obtained.

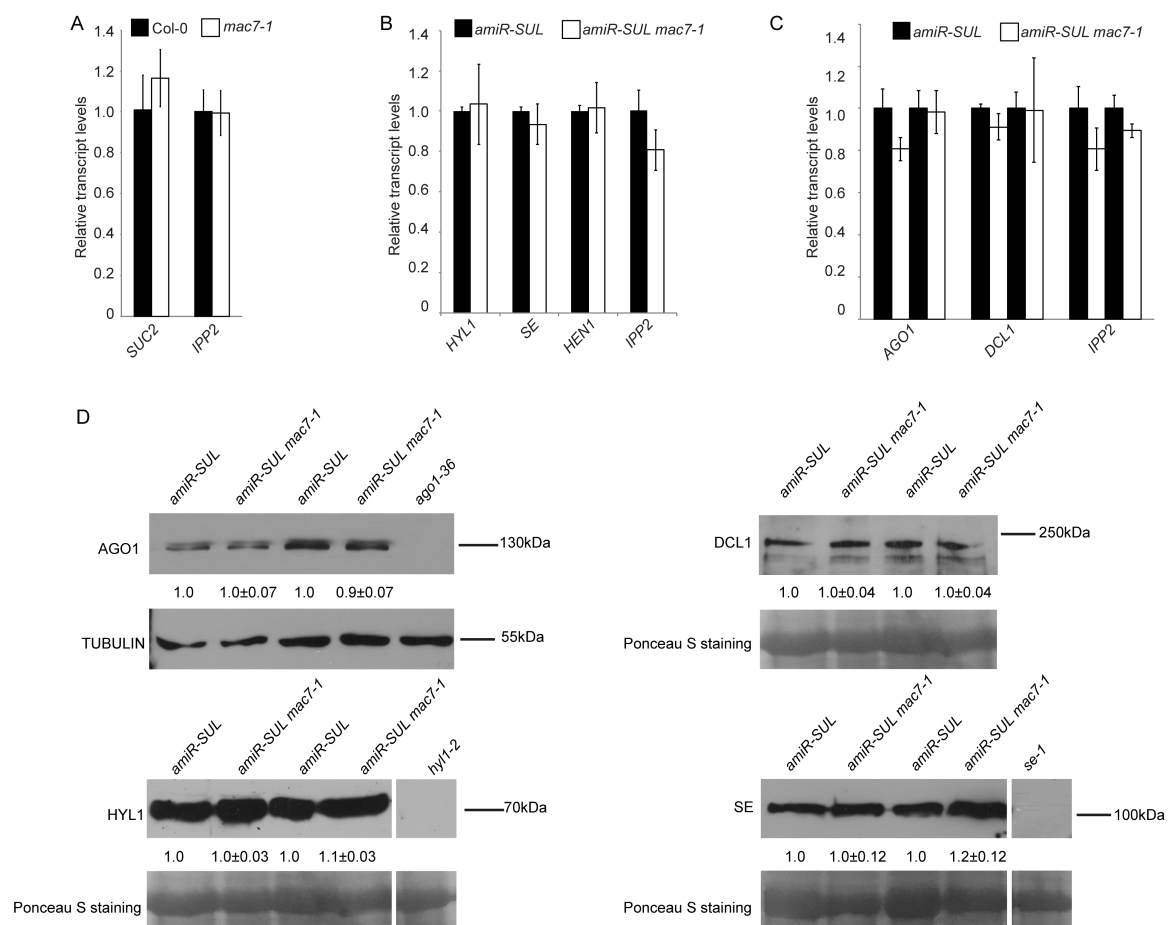


Figure 1.9 MAC7 does not affect the expression of the endogenous *SUC2* gene or key miRNA biogenesis factors.

(A) Transcript levels of *SUC2* and the housekeeping gene *IPP2* in seedlings of Col-0 and *mac7-1* as determined by RT-qPCR. Normalization was performed with *UBQ5* as the internal control and values in Col-0 were set to 1. Error bars indicate standard deviation from three technical replicates. (B and C) Transcript levels of key miRNA biogenesis genes in inflorescences of *amiR-SUL* and *amiR-SUL mac7-1* as determined by RT-qPCR. One biological replicate was included for *HYL1*, *SERRATE* and *HEN1*, and two biological replicates were included for *AGO1* and *DCL1*. The housekeeping gene *IPP2* was included as a control. *UBQ5* was used as the internal control and values in *amiR-SUL* were set to 1. Error bars indicate standard deviation from three technical replicates. (D) Protein levels of AGO1, HYL1, SERRATE (SE), and DCL1 in *amiR-SUL* and *amiR-SUL mac7-1* as determined by immunoblotting blotting. *ago1-36*, *hyl1-2* and *se-1* were included as negative controls. TUBULIN immunoblotting blotting and Ponceau S staining were performed to indicate relative loading. The immunoblotting blot signals were quantified and normalized to TUBULIN or Ponceau S staining. The values in *amiR-SUL* were set to 1. The numbers following $\pm$ represent standard deviation from four biological replicates for AGO1, DCL1 and SE, and two biological replicates for HYL1. 12-day-old seedlings were used for this immunoblotting blotting.

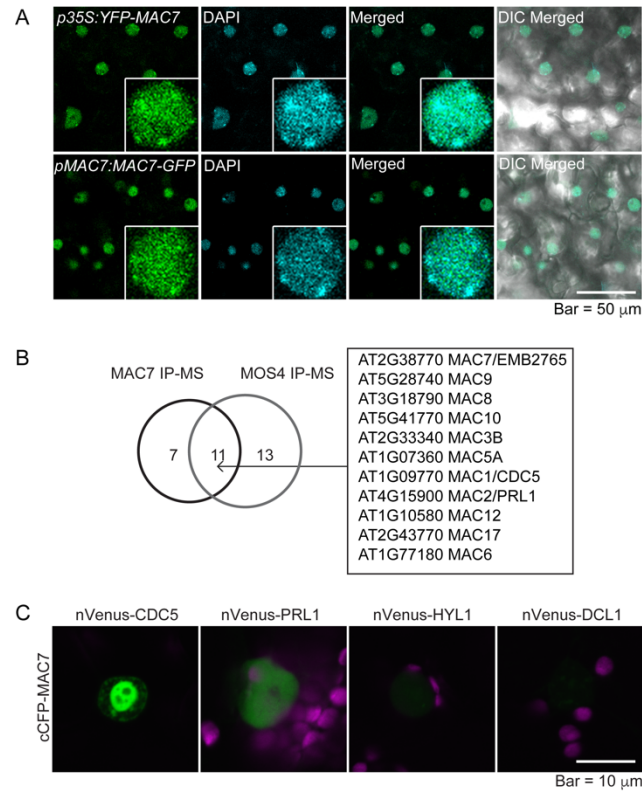


Figure 1.10 MAC7 is a nuclear protein associated with other MAC subunits.

(A) Subcellular localization of N-terminal (*p35S:YFP-MAC7*) or C-terminal (*pMAC7:MAC7-GFP*) fluorescent protein tagged MAC7 in young leaves of *Arabidopsis* transgenic lines. Nuclei were stained with DAPI and pseudo-colored in cyan. Enlarged nuclei are shown in the insets. (B) MAC subunits identified from both MAC7 and MOS4 immunoprecipitation followed by mass spectrometry analysis (IP-MS). (C) BiFC analysis of MAC7 with CDC5, PRL1, HYL1 and DCL1. Paired cCFP- and nVenus-fusion proteins were co-infiltrated into tobacco leaves. The BiFC signal (YFP) was detected at 48 h after infiltration by confocal microscopy, and was pseudo-colored in green. Magenta: auto fluorescence of chlorophyll.

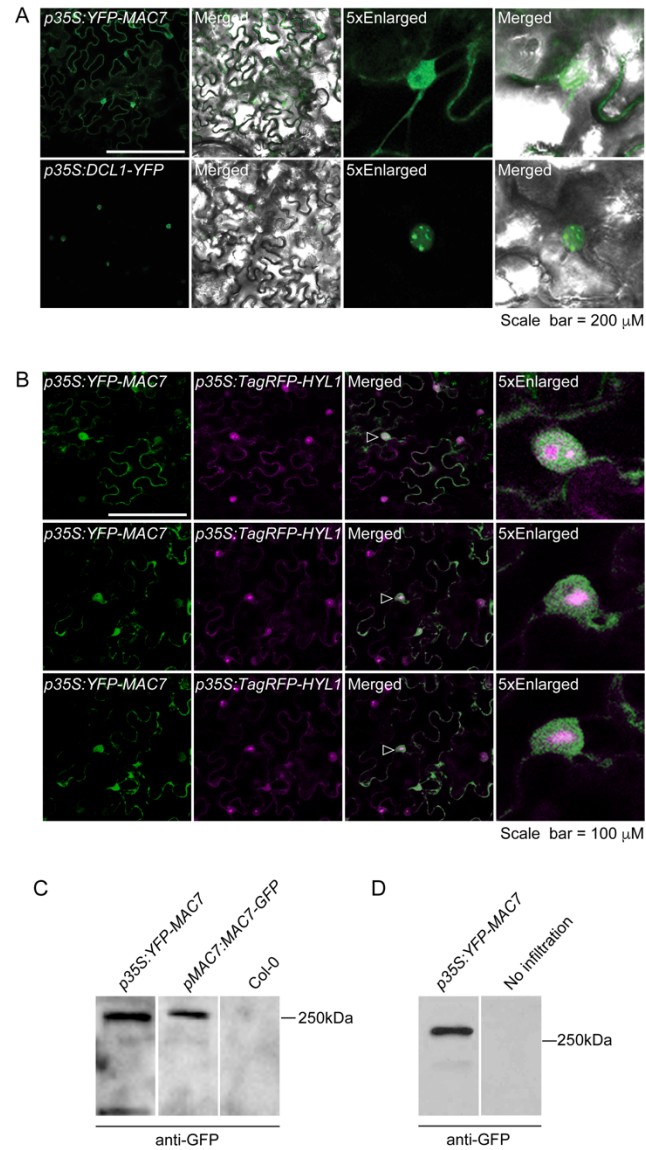


Figure 1.11 MAC7 shows dispersed distribution in the nucleoplasm, while DCL1 and HYL1 concentrate in dicing bodies in the nucleus.

(A) Subcellular localization of YFP-MAC7 and DCL1-YFP in *Agrobacterium*-infiltrated tobacco leaves. (B) Localization of YFP-MAC7 and TagRFP-HYL1 in co-infiltrated tobacco leaves. Arrowheads indicate the nuclei that were enlarged five times in the images to the right. Expression of *p35S::YFP-MAC7* and *pMAC7::MAC7-GFP* in transgenic plants (C) and infiltrated tobacco leaves (D) was shown by immunoblotting blotting.

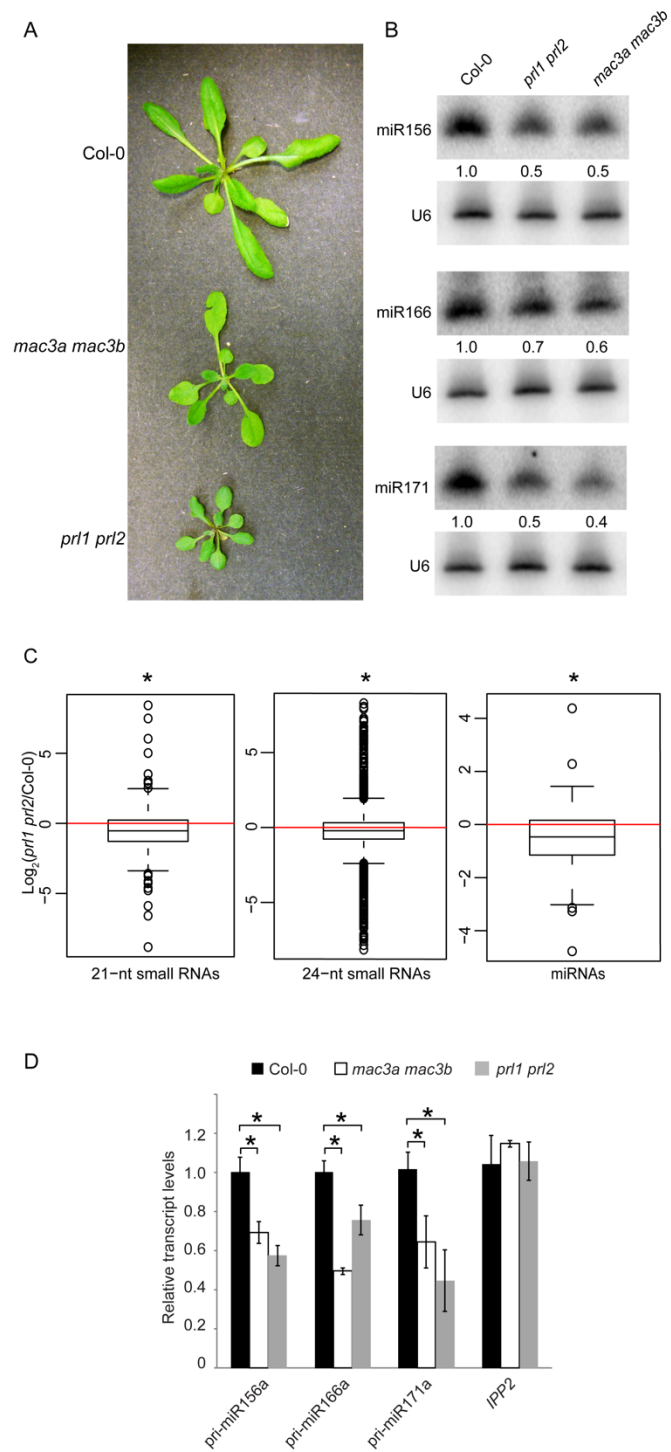


Figure 1.12 The MAC subunit genes *MAC3a*, *MAC3b*, *PRL1* and *PRL2* also promote miRNA biogenesis.

(A) Morphological phenotypes of 4 to 5-week old plants of the indicated genotypes. (B) MiRNA levels in wild type (Col-0) and indicated mutants as determined by northern blotting. The miRNA signals were quantified and normalized to those of U6, and values were relative to Col-0 (arbitrarily set to 1). The RNA used in northern blotting was extracted from the aerial parts of 12-day-old seedlings growing on MS agar plates. (C) Global abundance of 21-nt and 24-nt small RNAs and miRNAs in Col-0 and *prl1 prl2* as determined by small RNA-seq. Small RNA libraries were generated from 12-day-old seedlings growing on MS plates. The normalization of small RNAs was against 45S rRNA reads and abundance was expressed as RPMR (reads per million of 45S rRNA reads), and $\log_2$ ratios of *prl1 prl2*/Col-0 were plotted. Asterisks indicate that the mean is significantly below 0 (Wilcoxon test, $p < 2.2e-16$). (D) Pri-miRNA levels in plants of the indicated genotypes as determined by RT-qPCR. The housekeeping gene *IPP2* was included as a control. *UBQ5* was used as an internal control and values in Col-0 were set to 1. Error bars indicate standard deviation from three technical replicates, and asterisks indicate significant difference between Col-0 and the mutants (t-test, $p < 0.05$).

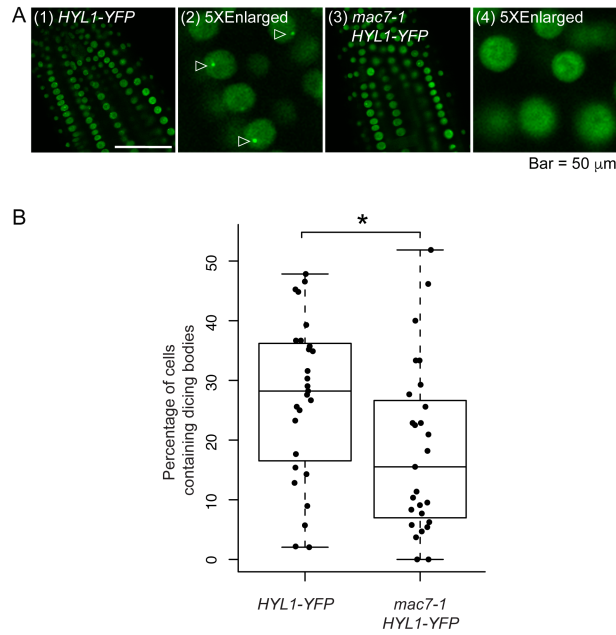


Figure 1.13 MAC7 affects HYL1 subcellular localization.

(A) Images of nuclei in root cells of 7 to 10-day-old seedlings of the indicated genotypes. Images (2) and (4) show five times enlarged images cropped from (1) and (3), respectively. Dicing bodies are indicated by triangles. (B) The percentage of cells containing HYL1-positive dicing bodies in wild type and *mac7-1*. The quantification was performed by observing more than 1000 cells from 27 roots for each genotype. The asterisk indicates significant difference between the samples (t-test, $p < 0.05$).

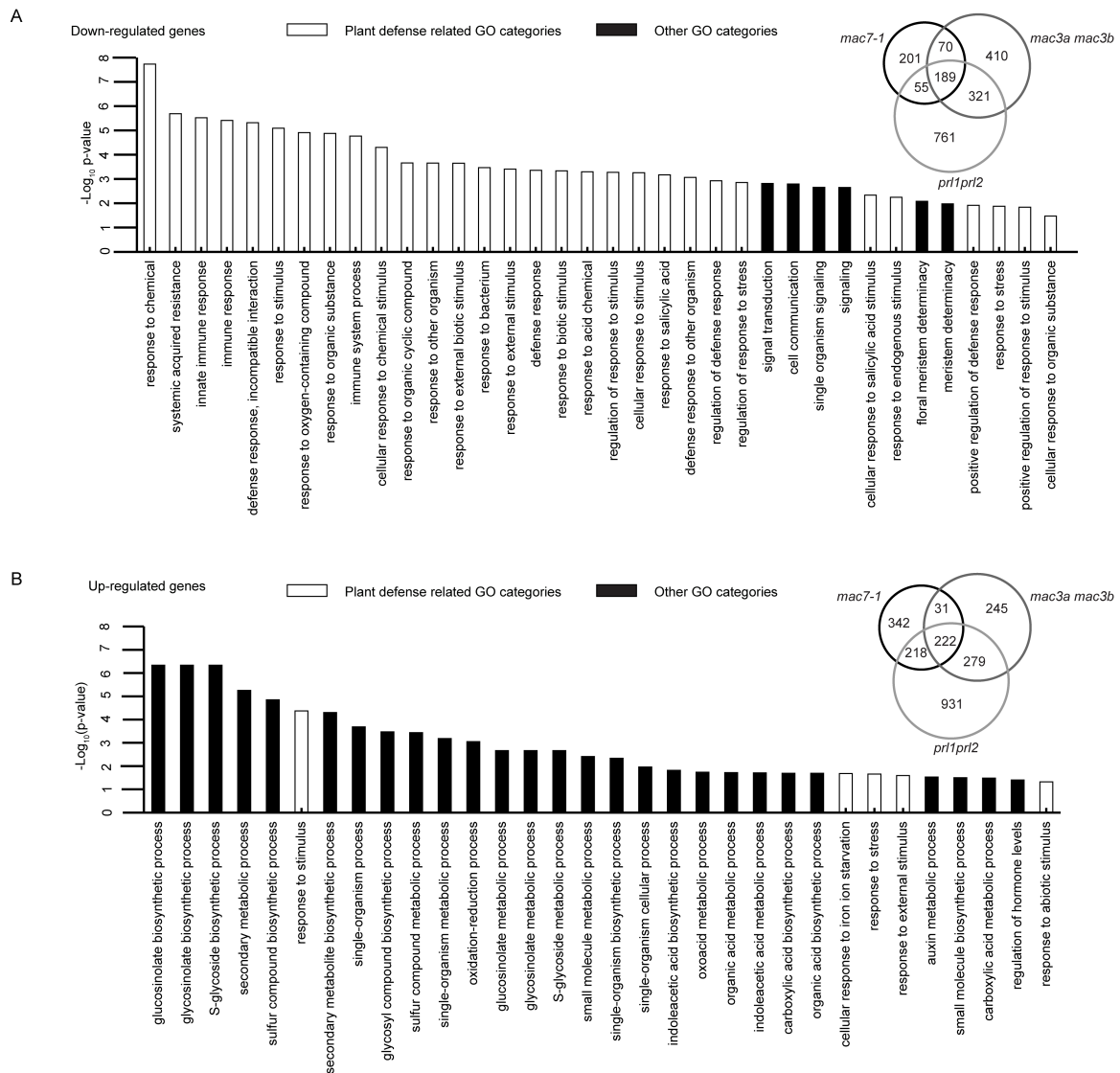


Figure 1.14 Down-regulated genes in the *mac* mutants are significantly related to stress responses, while up-regulated genes are involved in various biosynthetic and metabolic processes.

(A) GO enrichment analysis of 189 commonly down-regulated genes in *mac7-1*, *mac3a mac3b* and *prl1 prl2*. The degree of overlap among down-regulated genes in *mac7-1*, *mac3a mac3b* and *prl1 prl2* is shown in the Venn diagram. (B) GO enrichment analysis of 222 commonly up-regulated genes *mac7-1*, *mac3a mac3b* and *prl1 prl2*. The degree of overlap among up-regulated genes in *mac7-1*, *mac3a mac3b* and *prl1 prl2* is shown in the Venn diagram.

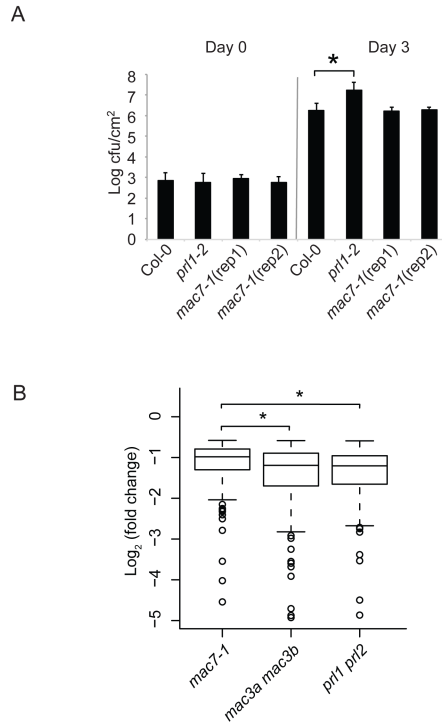


Figure 1.15 Growth of virulent *P.s.m. ES4326* at 0 and 3 days post-inoculation.

(A) Leaves of 5-week-old Col-0, *prl1-2*, *mac7-1* (rep1) and *mac7-1* (rep2) (two biological replicates of *mac7-1* were included in each experiment) plants were infiltrated with a bacterial suspension at $OD_{600} = 0.0001$. Leaf discs within the infected area were taken at day 0 and day 3 to quantify colony-forming units (cfu). Error bars represent standard deviations from five technical replicates. The asterisk indicates significant difference between Col-0 and *prl1-2* (t-test, $p < 0.05$). The experiments were performed four times and similar results were obtained. (B) Box plot of fold changes of overlapped hypo-DEGs in indicated mutants in comparison with Col-0. Asterisks indicate significant difference between the mutants (Wilcoxon test, $p < 2.2e-16$). The degree of downregulation of defense-related genes in *mac7-1* was not as large as the other *mac* mutants.

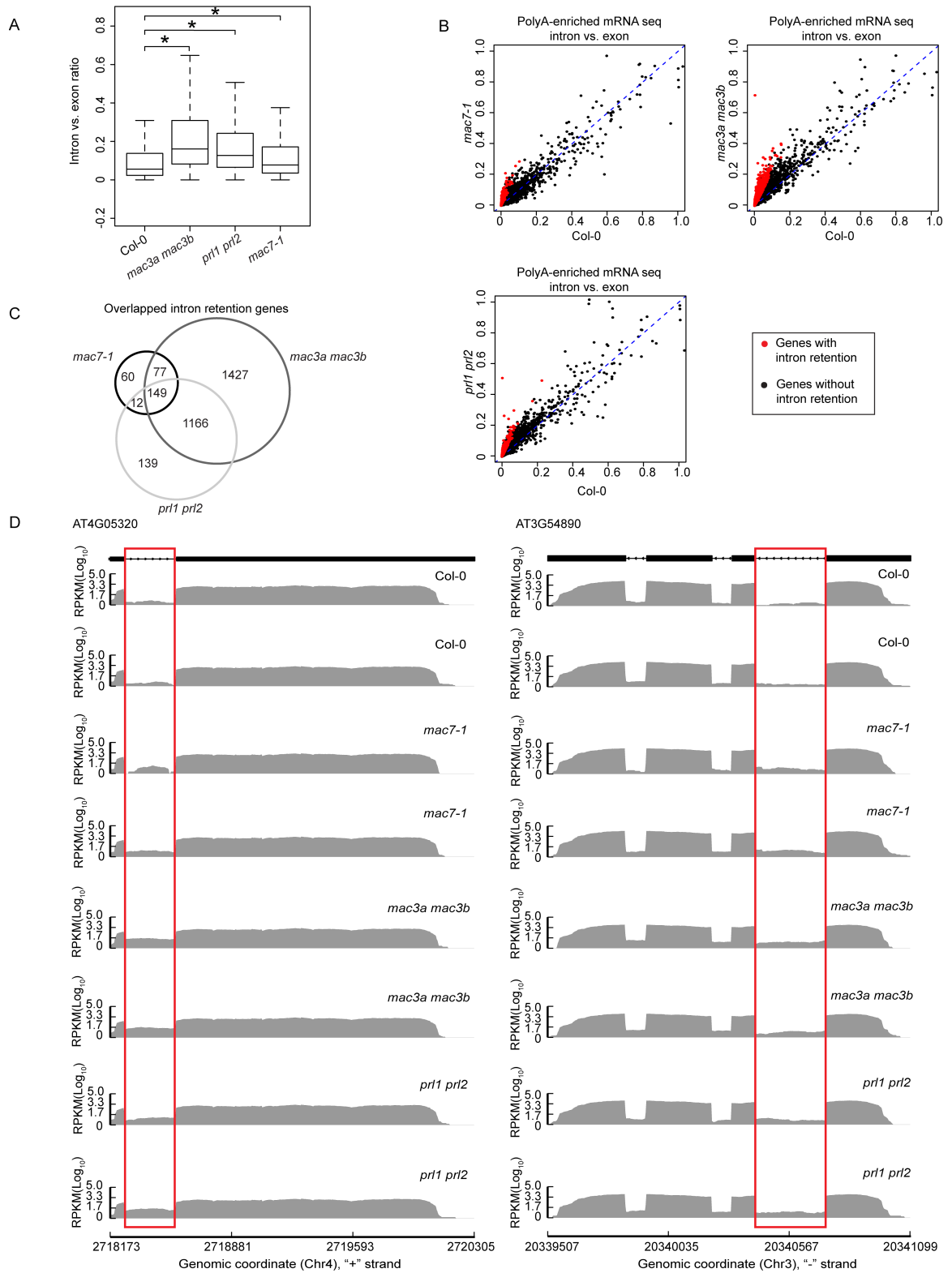


Figure 1.16 MAC subunits affect pre-mRNA splicing.

(A) Box plot of intron/exon ratios in the indicated genotypes. All genes that pass an abundance filter in expression were included in this analysis. Asterisks indicate significant difference between the mutant and wild type (Col-0) (Wilcoxon test, $p < 2.2e-16$). (B) The intron/exon ratio per gene in Col-0 vs. the indicated mutants. Black dots represent all genes with raw intronic read number ≥ 2 , exonic read number ≥ 5 , and final intron/exon ratio between 0 to 1. Red dots represent genes with significantly higher intron/exon ratio relative to wild type (fold change ≥ 2 and FDR < 0.01). (C) A Venn diagram showing the degree of overlap among genes with intron retention defects in *mac7-1*, *mac3a mac3b* and *prl1 prl2*. (D) Examples of genes with intron retention defects. RPKM: Reads Per Kilobase per Million mapped reads. Two biological replicates for each genotype are shown. The rectangles mark introns with higher retention in the mutants.

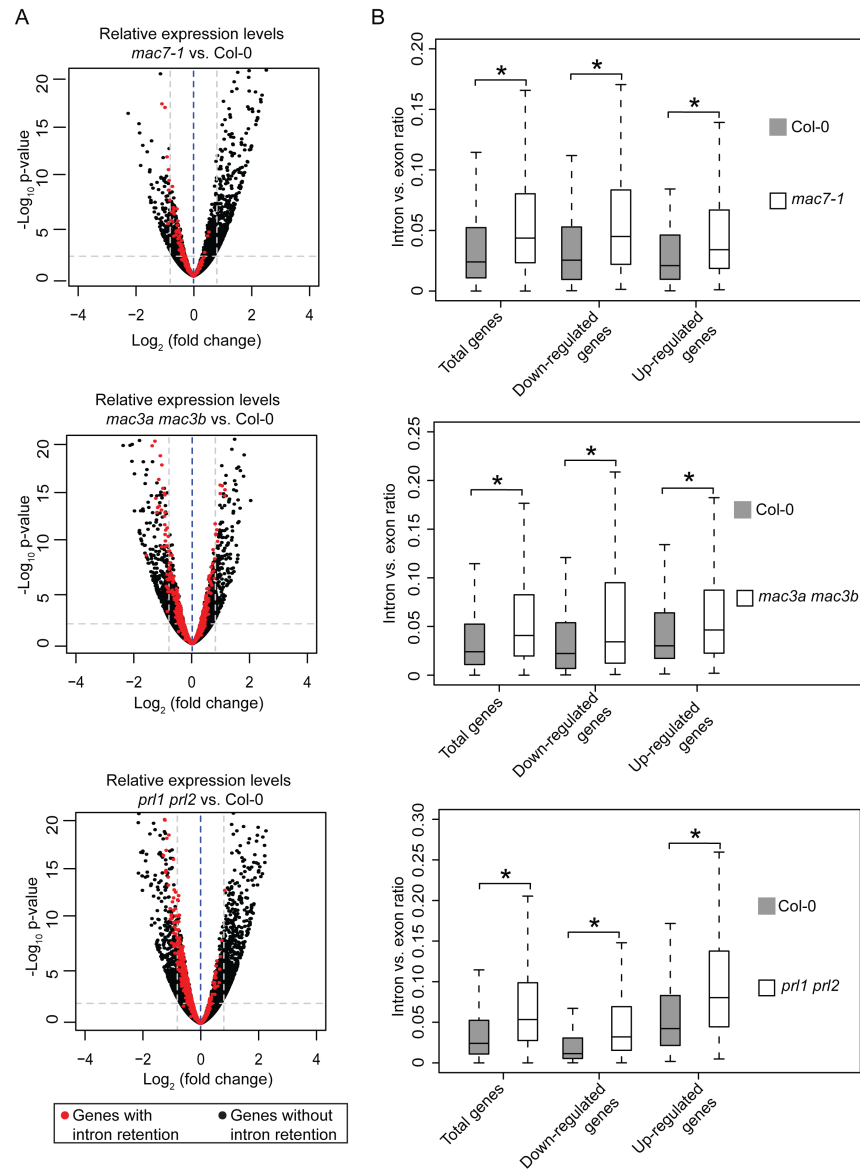


Figure 1.17 MAC affects pre-mRNA splicing and gene expression independently.

(A) Volcano plots illustrating fold changes of gene expression levels in the indicated mutants compared to Col-0. Thresholds for fold change of 1.5 and $p \leq 0.01$ are shown in the plots as gray dashed lines. Black dots and red dots represent the same genes as in Figure 1.16B. (B) Box plots of intron/exon ratios in total genes, up-regulated and down-regulated genes in the indicated genotypes. Only genes that pass an abundance filter (raw intronic read number ≥ 2 and exonic read number ≥ 5) were included in this analysis. Asterisks indicate significant difference between the mutant and wild type (Col-0) (Wilcoxon test, $p < 2.2e-16$).

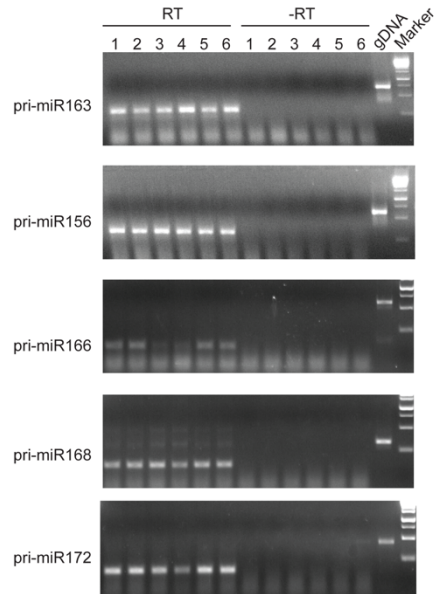


Figure 1.18 Intron-containing pri-miRNAs show similar splicing patterns in Col-0 and *mac7-1*.

Determination of intron splicing patterns of pri-miRNAs by RT-PCR with intron-flanking primers using RNA from inflorescences. Reverse transcription was performed with an oligodT primer. The numbers represent different genotypes. 1 and 2: Col-0; 3 and 4: *mac7-1*; 5 and 6: *mac7-1 pMAC7:MAC7-mCherry*. Genomic DNA (gDNA) from Col-0 was included to indicate the expected sizes of intron-containing pri-miRNAs. The two samples of each genotype were biological replicates (RNA was isolated independently from the samples).

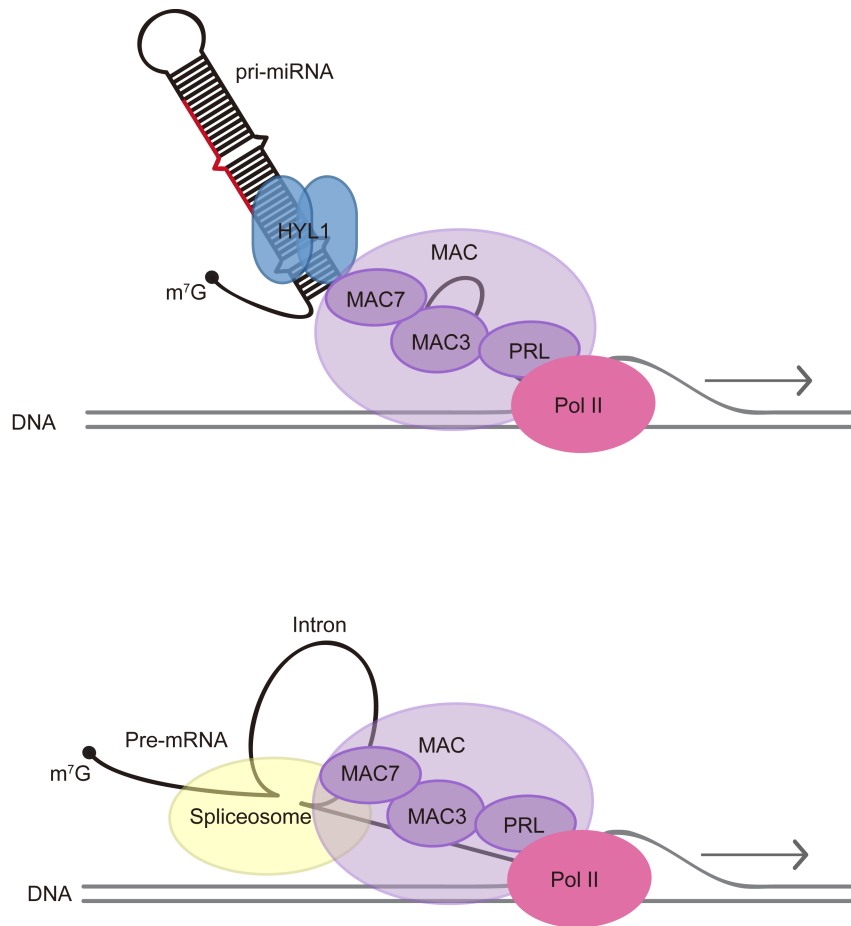


Figure 1.19 A model for MAC functions in miRNA biogenesis and pre-mRNA processing.

MAC affects miRNA biogenesis through influencing Pol II transcription and interacting with the pri-miRNA processing factor HYL1. MAC also plays a role in pre-mRNA splicing through interactions with the spliceosome. MAC seems to have separate roles in miRNA biogenesis and RNA splicing but a common theme appears to be co-transcriptional RNA processing.

TABLES

Table 1.1 Phenotypic and genotypic segregation in the progeny of selfed heterozygous *mac7-2* plants. The *mac7-2* allele harbors a T-DNA insertion (SALK_129044).

Seed abortion analysis*				
	Normal Seed	White Seeds (Aborted Seeds)	Expected Normal Seeds : Aborted Seeds	P-Value from Chi-square Test
Col-0	314	3	1:0	0.87
<i>mac7-2/+</i>	240	75	3:1	0.63

* Seeds from siliques of wild type (Col-0) and *mac7-2/+* plants were analyzed.

Genotypes of progeny of <i>mac7-2/+</i> plants				
+/+	<i>mac7-2/+</i>	<i>mac7-2/mac7-2</i>	Expected +/+: <i>mac7-2/+</i> : <i>mac7-2/mac7-2</i>	P-value from Chi-square Test
10	20	0	1:2:0	1

Table 1.2 MAC7-associated proteins identified by immunoprecipitation followed by mass spectrometry.

AGI	Protein	MAC subunit	Antibodies (Ab) used for IP, plant genotypes and counts of peptides			
			MAC 7 Ab	Pre-immune IgG	RFP-Trap	RF P-Trap
			Col-0	Col-0	MAC7-mCherry*	Col-0
Known MAC components						
AT2G38770	Embryo defective 2765 (EMB2765)	MAC7	151	0	1338	1
AT5G28740	Tetratricopeptide repeat (TPR)-like superfamily protein	MAC9	80	0	253	0
AT3G18790	Unknown protein	MAC8	17	0	94	0
AT5G41770	Crooked neck protein, putative / cell cycle protein, putative	MAC10	19	0	8	0
AT2G33340	MOS4-associated complex 3B (MAC3B)	MAC3B	17	0	17	0
AT1G07360	CCCH-type zinc finger family protein with RNA-binding domain	MAC5A	20	0	8	1
AT1G09770	Cell division cycle 5 (CDC5)	MAC1	10	0	7	0
AT4G15900	Pleiotropic regulatory locus 1 (PRL1)	MAC2	14	0	8	0
AT1G10580	Transducin/WD40 repeat-like superfamily protein	MAC12	15	0	4	0

AT2G43770	Transducin/WD40 repeat-like superfamily protein	MAC1 7	15	0	6	2
AT1G77180	SKIP	MAC6	13	0	10	0

AGI	Protein	MAC subunit	Antibodies (Ab) used for IP, plant genotypes and counts of peptides			
			MAC7 Ab	Pre-immune IgG	RFP-Trap	RFP-Trap
			Col-0	Col-0	MAC7-mCherry*	Col-0
Newly identified, potential MAC7-associated proteins						
AT1G13690	ATPase E1 (ATE1)		23	0	2	0
AT2G36130	Cyclophilin-like peptidyl-prolyl cis-trans isomerase family protein		18	0	4	0
AT5G38830	Cysteiny1-tRNA synthetase, class Ia family protein		13	0	37	0
AT2G37230	Tetratricopeptide repeat (TPR)-like superfamily protein		17	0	2	1
AT3G43300	HOPM interactor 7 (ATMIN7)		13	0	1	0
AT1G64790	ILITYHIA (ILA)		12	0	6	0
AT1G65440	Global transcription factor group B1 (GTB1)		9	0	2	0

* The genotype of the plants was *pMAC7:MAC7-mCherry*.

Table 1.3 HYL1-associated MAC subunits identified by immunoprecipitation followed by mass spectrometry.

AGI	Protein	MAC subunit	Antibodies (Ab) used for IP, plant genotypes and counts of peptides				
			GFP-Trap	GFP-Trap	GFP-Trap	HYL1 Ab	HYL1 Ab
			HYL1-YFP*	YFP**	Col-0	Col-0	<i>hyl1-2</i>
AT1G09700	HYL1, Double-stranded RNA-binding protein 1		977	0	0	398	0
AT1G06220	Ribosomal protein S5/Elongation factor G/III/V family protein	MAC11	142	2	0	16	0
AT2G38770	EMBRYO DEFECTIVE 2765 (EMB2765)	MAC7	119	0	0	8	0
AT1G09770	Cell division cycle 5 (CDC5)	MAC1	106	0	0	6	0
AT5G28740	Tetratricopeptide repeat (TPR)-like superfamily protein	MAC9	99	2	0	2	0
AT1G20960	U5 small nuclear ribonucleoprotein helicase, putative	MAC14	95	0	0	2	0
AT5G41770	Crooked neck protein, putative / cell cycle protein, putative	MAC10	85	0	0	4	0
AT1G77180	SKIP; FUNCTIONS IN: transcription activator activity	MAC6	41	0	0	8	0
AT3G18790	Unknown protein	MAC8	37	0	0	1	0
AT2G43770	Transducin/WD40 repeat-like superfamily protein	MAC17	33	0	0	3	0
AT3G18165	MOS4	MOS4	36	0	0	1	0

AGI	Protein	MAC subunit	Antibodies (Ab) used for IP, plant genotypes and counts of peptides				
			GFP-Trap	GFP-Trap	GFP-Trap	HYL1 Ab	HYL1 Ab
			HYL1-YFP*	YFP**	Col-0	Col-0	<i>hyl1-2</i>
AT5G64270	Splicing factor, putative	MAC18	17	1	0	1	0
	GFP or YFP		750	664	N/A	N/A	N/A

* The genotype of the plants was *p35S:HYL1-YFP*.

** The genotype of the plants was *p35S:YFP*

Table 1.4 Oligonucleotide sequences.

Cloning	Sequence (5' to 3')
AC19 genome FP1 cacc	CACC CTTGGGAAAAC TCCCCATGTAGAAAC
AC19 genome RP1 NSC	ATTCTTCTCATCAGCCTTTCCATTCTC
AC19 genome FP1 C-fusion (SmaI)	GAA TTC CTG CAG CCC GGG CTTGGGAAAAC TCCCCATGTAGAAAC
AC19 genome RP1 C-fusion NSC (SmaI)	ACT AGT GGA TCC CCC GGG ATTCTTCTCATCAGCCTTTCCATTCTC
AC19 CDS FP1 cacc	CACC ATGACGAAGGTCTATGGAAGTGG
AC19 CDS RP2 SC	CTAATTCTTCTCATCAGCCTTTCCA
ARM LIC F	TACTTCCAATCCAATGCG ATGACGAAGGTCTATGGAAGTGG
ARM LIC R	TTATCCACTTCCAATGCGCTA TTA CAATCTCCTCCGAAGATCAGAACT
ARM pSUMO InFus F	ACAGATTGGTGGATCCATGACGAAGGTCTATG GAACTGG
ARM pSUMO InFus R	GGTGGTGGTGGCTCGAGTTACAATCTCCTCCGAA GATCAGAACT
cacc HYL1 F	CACC ATGACCTCCACTGATGTTTCCT
NSC HYL1 R	TGCGTGGCTTGCTTCT
Northern blot probes	Sequence (5' to 3')
amiR-SUL probe	AGGGATTTCGTGACACTTAA
miR156 probe	GTGCTCACTCTCTTCTGTCA
miR159 probe	TAGAGCTCCCTTCAATCCAAA
miR160 probe	TGGCATAACAGGGAGCCAGGCA
miR164 probe	TGCACGTGCCCTGCTTCTCCA
miR166 probe	GGGGAATGAAGCCTGGTCCGA
miR167 probe	TAGATCATGCTGGCAGCTTCA
mir168 probe	TTCCCGACCTGCACCAAGCGA
miR171 probe	GATATTGGCGCGGCTCAATCA
miR172 probe	ATGCAGCATCATCAAGATTCT
miR390 probe	GGCGCTATCCCTCCTGAGCTT
miR394 probe	GGAGGTGGACAGAATGCCAA
U6 probe	AGGGGCCATGCTAATCTTCTCTG
Small RNA marker probe	AAATCTCAACCAGCCACTGCT

RT-qPCR primers	Sequence (5' to 3')
qPri-amiR-SUL F	TCAATTAGCTTCCGACTCATTCA
qPri-amiR-SUL R	AGGGATTTCCTGACACTTAA
qPri-156a F	GAAAGAGTTGGGACAAGAGAAACG
qPri-156a R	AGAGAACGAAGACAGGCCAAAGA
qPri-159a F	TCTTTACAGTTTGCTTATGTCAGATCCA
qPri-159a R	ACCCTGCTCAACTCATGTTTGAA
qPri-miR166a F	AGATATATATTCAGAAACCCTAG
qPri-miR166a R	GGTTCATTCACTGGATCTGAAAC
qPri-167a F	TGATCTGCTACGGTGAAGTCTATGG
qPri-167a R	GAAACTGCGAACATGATCTAATCGA
qPri-168a F	ATTCGCTTGGTGCAGGTC
qPri-168a R	TCCGATTCACTTGATGCAAG
qPri-171a F	TGGCCTGGTTCCTCAGATCTTA
qPri-171a R	ACAAGACCACAAAGTCCAAAATAGAGA
qPri-172a F	CTGTTGATGGACGGTGGTGAT
qPri-172a R	CCTGAAGAAGATCTGGATGGAATC
qPri-390b F	CCTATGGATGTAAATTCAGATTGAG
qPri-390b R	CGAAGGAGGGAATGAAGTAGG
qPri-394b F	TCGACAGAAAGGAAATGAGTGA
qPri-394b R	TCCTCTTACGAAACACACCGTAGATAC
qPri-396a F	ACATGACCCTCTCTGTATTCTTCCA
qPri-396a R	CAATCGAGCAGAGATATGAAGAAAATC
qPri-397a F	GCGTTGATGTAATTTTCGTTTTGTTT
qPri-397a R	ACGATCCGCATACCTGTTTAAGTG
qTubulin F	ATCCGTGAAGAGTACCCAGAT
qTubulin R	AAGAACCATGCACTCATCAGC
qACTIN8 F	CACATGCTATCCTCCGTCTC
qACTIN8 R	CAATGCCTGGACCTGCTT
qEIF4A F	TCTTGGTGAAGCGTGATGAG
qEIF4A R	GCTGAGTTGGGAGATCGAAG
qUBQ5 F	GTAAACGTAGGTGAGTCCA
qUBQ5 R	GACGCTTCATCTCGTCC
qGUS F	TAATGTTCTGCGACGCTCACA
qGUS R	ATACCTGTTACCCGACGACG

RT-qPCR primers	Sequence (5' to 3')
qDCL1 uncut F	TGGGTTTCATCTGGAAAGGA
qDCL1 uncut R	CCCTTGAAAGCAAGAGATGC
qHYL1 F	CAGTAAAGAGGAGGAAGTCG
qHYL1 R	ACACCAGAGGAAACATCAGT
qSE F	AACCAGCCACGCAGCAATCT
qSE R	CGTCCACCTCTATCATCCCTCA
qAGO1 uncut F	AAGGAGGTCGAGGAGGGTATGG
qAGO1 uncut R	CAAATTGC TGAGCCAGAACAGTAGG
qHEN1 F	GTTGACATCGGCACTTGCTTA
qHEN1 R	TGGGTTTCTGGTGTAGACCG
qIPP2 F	GTATGAGTTGCTTCTCCAGCAAAG
qIPP2 R	GAGGATGGCTGCAACAAGTGT
CH42 F	GATCCAAAGATTGGTGGTGTATG
CH42 R	AACTTGCTCTCCTTCTCAACTCT
MYB33 F	AGTTGTTGTATCCTGGGTGTAGCA
MYB33 R	CCGTTGGTGGTGGTGGAGAC
SCL6-IV F	ACTCAAGACAACCTCAAGCA
SCL6-IV R	GATAGATGCTTCACGAAACG
SPL10 F	GTGGGAGAATGCTCAGGAGGC
SPL10 R	GAGTGTGTTTGATCCCTTGTGAATCC
MYB65 F	GAT GGTTCCCTGATAGCCATACAGTTAC
MYB65 R	TAGGCATCAACAGAGTCAAGGAGATC
ARF17 F	AGCACCTGATCCAAGTCCTTCTATG
ARF17 R	TGGTGAATAGCTGGGGAGGATTTC
ARF8 F	AGATGTTTGCTATCGAAGGGTTGTTG
ARF8 R	CCATGGGTCATCACCAAGGAGAAG
SCL6-III F	ACCAAGACC AGTCAGCGGTAATC
SCL6-III R	AGTGTCGTCGTTGTTGTTGTTA AGG
CUC2 F	GCA CCAACACAACCGTCACAG
CUC2 R	GAATGAGTTAACGTCTAAGC CCAAGG

RT-PCR primers	
Pri-163F-Intron-F	GAGAGTGAGAAAAATAAAGAG
Pri-163R-Intron-R	AGGATGTTGACACGTGTAAAC
Pri-156a-Intron-F	ATTGTTCACTCTCAAATCTCAAG
Pri-156a-Intron-R	AAGAGATCAGCACCGGAATCTG
Pri-166-Intron-F	CCCAATTGTTGCTCCCTGTTTAC
Pri-166-Intron-R	GGAGAAAGAGGTCCAAATTCAAGC
Pri-168-Intron-F	CGGAAAATGAGTGGTCGTAACC
Pri-168-Intron-R	GAGGATCCGATTCAGTTGATGC
Pri-172-Intron-F	ATTGAAGTCTGGATCGATCTTTG
Pri-172-Intron-R	TGTCTGGTCTCTGGACGAACTAT
Genotyping primers	Sequence (5' to 3')
gt AC19 SALK_129044.2 F	TTTGCTCTTTGGAAGATGGG
gt AC19 SALK_129044.2 R	GTGGTTGGTCCACCAGGTAC
AC19 GT F	CTCTCATCTCAAGACTGTGTTC
AC19 GT R	TGAAGCAGCATTGGTATGAAAGTCTCGAATT

REFERENCES

1. K. Rogers, X. Chen, Biogenesis, turnover, and mode of action of plant microRNAs. *Plant Cell* **25**, 2383-2399 (2013).
2. Z. Xie *et al.*, Expression of Arabidopsis MIRNA genes. *Plant Physiol* **138**, 2145-2154 (2005).
3. B. Zheng *et al.*, Intergenic transcription by RNA polymerase II coordinates Pol IV and Pol V in siRNA-directed transcriptional gene silencing in Arabidopsis. *Genes Dev* **23**, 2850-2860 (2009).
4. W. Park, J. Li, R. Song, J. Messing, X. Chen, CARPEL FACTORY, a Dicer homolog, and HEN1, a novel protein, act in microRNA metabolism in Arabidopsis thaliana. *Curr Biol* **12**, 1484-1495 (2002).
5. J. Li, Z. Yang, B. Yu, J. Liu, X. Chen, Methylation protects miRNAs and siRNAs from a 3'-end uridylation activity in Arabidopsis. *Curr Biol* **15**, 1501-1507 (2005).
6. B. Yu *et al.*, Methylation as a crucial step in plant microRNA biogenesis. *Science* **307**, 932-935 (2005).
7. Z. Yang, Y. W. Ebright, B. Yu, X. Chen, HEN1 recognizes 21-24 nt small RNA duplexes and deposits a methyl group onto the 2' OH of the 3' terminal nucleotide. *Nucleic Acids Res* **34**, 667-675 (2006).
8. B. Yu *et al.*, siRNAs compete with miRNAs for methylation by HEN1 in Arabidopsis. *Nucleic Acids Res.* **38**, 5844-5850 (2010).
9. N. Baumberger, D. C. Baulcombe, Arabidopsis ARGONAUTE1 is an RNA Slicer that selectively recruits microRNAs and short interfering RNAs. *Proc Natl Acad Sci U S A* **102**, 11928-11933 (2005).
10. L. Arribas-Hernandez *et al.*, The Slicer Activity of ARGONAUTE1 Is Required Specifically for the Phasing, Not Production, of Trans-Acting Short Interfering RNAs in Arabidopsis. *Plant Cell* **28**, 1563-1580 (2016).
11. A. Carbonell *et al.*, Functional analysis of three Arabidopsis ARGONAUTES using slicer-defective mutants. *Plant Cell* **24**, 3613-3629 (2012).
12. L. Ji *et al.*, ARGONAUTE10 and ARGONAUTE1 regulate the termination of floral stem cells through two microRNAs in Arabidopsis. *PLoS Genet* **7**, e1001358 (2011).

13. N. P. Achkar, D. A. Cambiagno, P. A. Manavella, miRNA Biogenesis: A Dynamic Pathway. *Trends Plant Sci* **21**, 1034-1044 (2016).
14. Y. J. Kim *et al.*, The role of Mediator in small and long noncoding RNA production in *Arabidopsis thaliana*. *Embo J* **30**, 814-822 (2011).
15. L. Wang *et al.*, NOT2 proteins promote polymerase II-dependent transcription and interact with multiple MicroRNA biogenesis factors in *Arabidopsis*. *Plant Cell* **25**, 715-727 (2013).
16. M. Hajheidari, S. Farrona, B. Huettel, Z. Koncz, C. Koncz, CDKF;1 and CDKD protein kinases regulate phosphorylation of serine residues in the C-terminal domain of *Arabidopsis* RNA polymerase II. *Plant Cell* **24**, 1626-1642 (2012).
17. S. Zhang, M. Xie, G. Ren, B. Yu, CDC5, a DNA binding protein, positively regulates posttranscriptional processing and/or transcription of primary microRNA transcripts. *Proc Natl Acad Sci U S A* **110**, 17588-17593 (2013).
18. X. Fang, Y. Cui, Y. Li, Y. Qi, Transcription and processing of primary microRNAs are coupled by Elongator complex in *Arabidopsis*. *Nat Plants* **1**, 15075 (2015).
19. Z. Dong, M. H. Han, N. Fedoroff, The RNA-binding proteins HYL1 and SE promote accurate in vitro processing of pri-miRNA by DCL1. *Proc Natl Acad Sci U S A* **105**, 9970-9975 (2008).
20. M. H. Han, S. Goud, L. Song, N. Fedoroff, The *Arabidopsis* double-stranded RNA-binding protein HYL1 plays a role in microRNA-mediated gene regulation. *Proc Natl Acad Sci U S A* **101**, 1093-1098 (2004).
21. Y. Kurihara, Y. Takashi, Y. Watanabe, The interaction between DCL1 and HYL1 is important for efficient and precise processing of pri-miRNA in plant microRNA biogenesis. *Rna* **12**, 206-212 (2006).
22. Y. Kurihara, Y. Watanabe, *Arabidopsis* micro-RNA biogenesis through Dicer-like 1 protein functions. *Proc Natl Acad Sci U S A* **101**, 12753-12758 (2004).
23. L. Yang, Z. Liu, F. Lu, A. Dong, H. Huang, SERRATE is a novel nuclear regulator in primary microRNA processing in *Arabidopsis*. *Plant J* **47**, 841-850 (2006).
24. Y. Fang, D. L. Spector, Identification of nuclear dicing bodies containing proteins for microRNA biogenesis in living *Arabidopsis* plants. *Curr Biol* **17**, 818-823 (2007).

25. L. Song, M. H. Han, J. Lesicka, N. Fedoroff, Arabidopsis primary microRNA processing proteins HYL1 and DCL1 define a nuclear body distinct from the Cajal body. *Proc Natl Acad Sci U S A* **104**, 5437-5442 (2007).
26. S. Laubinger *et al.*, Dual roles of the nuclear cap-binding complex and SERRATE in pre-mRNA splicing and microRNA processing in Arabidopsis thaliana. *Proc Natl Acad Sci U S A* **105**, 8795-8800 (2008).
27. K. D. Raczynska *et al.*, The SERRATE protein is involved in alternative splicing in Arabidopsis thaliana. *Nucleic Acids Res* **42**, 1224-1244 (2014).
28. B. Yu *et al.*, The FHA domain proteins DAWDLE in Arabidopsis and SNIP1 in humans act in small RNA biogenesis. *Proc Natl Acad Sci U S A* **105**, 10073-10078 (2008).
29. G. Ren *et al.*, Regulation of miRNA abundance by RNA binding protein TOUGH in Arabidopsis. *Proc Natl Acad Sci U S A* **109**, 12817-12821 (2012).
30. S. Zhang, Y. Liu, B. Yu, PRL1, an RNA-binding protein, positively regulates the accumulation of miRNAs and siRNAs in Arabidopsis. *PLoS Genet* **10**, e1004841 (2014).
31. P. A. Manavella *et al.*, Fast-forward genetics identifies plant CPL phosphatases as regulators of miRNA processing factor HYL1. *Cell* **151**, 859-870 (2012).
32. P. Karlsson *et al.*, KH domain protein RCF3 is a tissue-biased regulator of the plant miRNA biogenesis cofactor HYL1. *Proc Natl Acad Sci U S A* **112**, 14096-14101 (2015).
33. A. G. Francisco-Mangilet *et al.*, THO2, a core member of the THO/TREX complex, is required for microRNA production in Arabidopsis. *Plant J* **82**, 1018-1029 (2015).
34. C. Furumizu, H. Tsukaya, Y. Komeda, Characterization of EMU, the Arabidopsis homolog of the yeast THO complex member HPR1. *Rna* **16**, 1809-1817 (2010).
35. X. Wu *et al.*, A role for the RNA-binding protein MOS2 in microRNA maturation in Arabidopsis. *Cell Res* **23**, 645-657 (2013).
36. K. C. M. Johnson, O. X. Dong, X. Li, The evolutionarily conserved MOS4-associated complex. *Central European Journal of Biology* **6**, 776-784 (2011).
37. C. Koncz, F. Dejong, N. Villacorta, D. Szakonyi, Z. Koncz, The spliceosome-activating complex: molecular mechanisms underlying the function of a pleiotropic regulator. *Front Plant Sci* **3**, 9 (2012).

38. J. Monaghan *et al.*, Two Prp19-like U-box proteins in the MOS4-associated complex play redundant roles in plant innate immunity. *PLoS Pathog* **5**, e1000526 (2009).
39. X. Deng *et al.*, Recruitment of the NineTeen Complex to the activated spliceosome requires AtPRMT5. *Proc Natl Acad Sci U S A* **113**, 5447-5452 (2016).
40. Z. Lin *et al.*, AtCDC5 regulates the G2 to M transition of the cell cycle and is critical for the function of Arabidopsis shoot apical meristem. *Cell Res.* **17**, 815-828 (2007).
41. K. Palma *et al.*, Regulation of plant innate immunity by three proteins in a complex conserved across the plant and animal kingdoms. *Genes Dev.* **21**, 1484-1493 (2007).
42. K. Nemeth *et al.*, Pleiotropic control of glucose and hormone responses by PRL1, a nuclear WD protein, in Arabidopsis. *Genes Dev.* **12**, 3059-3073 (1998).
43. J. Monaghan, F. Xu, S. Xu, Y. Zhang, X. Li, Two putative RNA-binding proteins function with unequal genetic redundancy in the MOS4-associated complex. *Plant Physiol.* **154**, 1783-1793 (2010).
44. J. Wiborg, C. O'Shea, K. Skriver, Biochemical function of typical and variant Arabidopsis thaliana U-box E3 ubiquitin-protein ligases. *Biochem. J.* **413**, 447-457 (2008).
45. I. De *et al.*, The RNA helicase Aquarius exhibits structural adaptations mediating its recruitment to spliceosomes. *Nat. Struct. Mol. Biol.* **22**, 138-144 (2015).
46. T. Hirose *et al.*, A spliceosomal intron binding protein, IBP160, links position-dependent assembly of intron-encoded box C/D snoRNP to pre-mRNA splicing. *Mol. Cell* **23**, 673-684 (2006).
47. A. Akay *et al.*, The Helicase Aquarius/EMB-4 Is Required to Overcome Intronic Barriers to Allow Nuclear RNAi Pathways to Heritably Silence Transcription. *Dev. Cell* **42**, 241-255 e246 (2017).
48. K. M. Tyc *et al.*, The Conserved Intron Binding Protein EMB-4 Plays Differential Roles in Germline Small RNA Pathways of *C. elegans*. *Dev. Cell* **42**, 256-270 e256 (2017).
49. I. Tzafrir *et al.*, Identification of genes required for embryo development in Arabidopsis. *Plant Physiol.* **135**, 1206-1220 (2004).

50. F. F. de Felippes, F. Ott, D. Weigel, Comparative analysis of non-autonomous effects of tasiRNAs and miRNAs in *Arabidopsis thaliana*. *Nucleic Acids Res* **39**, 2880-2889 (2011).
51. S. Ozgur *et al.*, The conformational plasticity of eukaryotic RNA-dependent ATPases. *FEBS J.* **282**, 850-863 (2015).
52. T. Weihmann, K. Palma, Y. Nitta, X. Li, Pleiotropic regulatory locus 2 exhibits unequal genetic redundancy with its homolog PRL1. *Plant Cell Physiol.* **53**, 1617-1626 (2012).
53. C. J. Zhang *et al.*, The splicing machinery promotes RNA-directed DNA methylation and transcriptional silencing in *Arabidopsis*. *Embo J* **32**, 1128-1140 (2013).
54. C. Gene Ontology, Gene Ontology Consortium: going forward. *Nucleic Acids Res.* **43**, D1049-1056 (2015).
55. F. Xu, S. Xu, M. Wiermer, Y. Zhang, X. Li, The cyclin L homolog MOS12 and the MOS4-associated complex are required for the proper splicing of plant resistance genes. *Plant J.* **70**, 916-928 (2012).
56. H. Ner-Gaon *et al.*, Intron retention is a major phenomenon in alternative splicing in *Arabidopsis*. *Plant J.* **39**, 877-885 (2004).
57. A. Zielezinski *et al.*, mirEX 2.0 - an integrated environment for expression profiling of plant microRNAs. *BMC Plant Biol* **15**, 144 (2015).
58. X. Zhan *et al.*, *Arabidopsis* proline-rich protein important for development and abiotic stress tolerance is involved in microRNA biogenesis. *Proc Natl Acad Sci U S A* **109**, 18198-18203 (2012).
59. P. M. Checchi, W. G. Kelly, *emb-4* is a conserved gene required for efficient germline-specific chromatin remodeling during *Caenorhabditis elegans* embryogenesis. *Genetics* **174**, 1895-1906 (2006).
60. I. Kuraoka *et al.*, Isolation of XAB2 complex involved in pre-mRNA splicing, transcription, and transcription-coupled repair. *J. Biol. Chem.* **283**, 940-950 (2008).
61. S. Chanarat, K. Strasser, Splicing and beyond: the many faces of the Prp19 complex. *Biochim. Biophys. Acta* **1833**, 2126-2134 (2013).
62. B. Szarzynska *et al.*, Gene structures and processing of *Arabidopsis thaliana* HYL1-dependent pri-miRNAs. *Nucleic Acids Res* **37**, 3083-3093 (2009).

63. S. Chanarat, M. Seizl, K. Strasser, The Prp19 complex is a novel transcription elongation factor required for TREX occupancy at transcribed genes. *Genes Dev.* **25**, 1147-1158 (2011).
64. C. Yan *et al.*, Structure of a yeast spliceosome at 3.6-angstrom resolution. *Science* **349**, 1182-1191 (2015).
65. S. Chanarat, C. Burkert-Kautzsch, D. M. Meinel, K. Strasser, Prp19C and TREX: interacting to promote transcription elongation and mRNA export. *Transcription* **3**, 8-12 (2012).
66. Y. Qiao, J. Shi, Y. Zhai, Y. Hou, W. Ma, Phytophthora effector targets a novel component of small RNA pathway in plants to promote infection. *Proc Natl Acad Sci U S A* **112**, 5850-5855 (2015).
67. H. Li, R. Durbin, Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* **25**, 1754-1760 (2009).
68. H. Li *et al.*, The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**, 2078-2079 (2009).
69. H. Li, J. Ruan, R. Durbin, Mapping short DNA sequencing reads and calling variants using mapping quality scores. *Genome Res.* **18**, 1851-1858 (2008).
70. K. W. Earley *et al.*, Gateway-compatible vectors for plant functional genomics and proteomics. *Plant J.* **45**, 616-629 (2006).
71. M. D. Curtis, U. Grossniklaus, A gateway cloning vector set for high-throughput functional analysis of genes in planta. *Plant Physiol.* **133**, 462-469 (2003).
72. G. S. Pall, A. J. Hamilton, Improved northern blot method for enhanced detection of small RNA. *Nat Protoc* **3**, 1077-1084 (2008).
73. P. Lidder, R. A. Gutierrez, P. A. Salome, C. R. McClung, P. J. Green, Circadian control of messenger RNA stability. Association with a sequence-specific messenger RNA decay pathway. *Plant Physiol.* **138**, 2374-2385 (2005).
74. M. Martin, Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet.journal* **17**, 10-12 (2011).
75. N. R. Johnson, J. M. Yeoh, C. Coruh, M. J. Axtell, Improved Placement of Multi-mapping Small RNAs. *G3 (Bethesda)* **6**, 2103-2111 (2016).
76. S. Li *et al.*, Biogenesis of phased siRNAs on membrane-bound polysomes in Arabidopsis. *Elife* **5**, (2016).

77. D. Kim, B. Langmead, S. L. Salzberg, HISAT: a fast spliced aligner with low memory requirements. *Nat. Methods* **12**, 357-360 (2015).
78. M. I. Love, W. Huber, S. Anders, Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol* **15**, 550 (2014).
79. M. Wang, Y. Zhao, B. Zhang, Efficient Test and Visualization of Multi-Set Intersections. *Sci Rep* **5**, 16923 (2015).
80. R. C. Edgar, MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res.* **32**, 1792-1797 (2004).
81. A. M. Waterhouse, J. B. Procter, D. M. Martin, M. Clamp, G. J. Barton, Jalview Version 2--a multiple sequence alignment editor and analysis workbench. *Bioinformatics* **25**, 1189-1191 (2009).
82. R. Peterson, J. P. Slovin, C. Chen, A simplified method for differential staining of aborted and non-aborted pollen grains. *Int J Plant Biol* **1:e13**, 66-69 (2010).
83. D. E. Sleat, H. Zheng, M. Qian, P. Lobel, Identification of sites of mannose 6-phosphorylation on lysosomal proteins. *Mol. Cell. Proteomics* **5**, 686-701 (2006).
84. S. Nakamura *et al.*, Gateway binary vectors with the bialaphos resistance gene, bar, as a selection marker for plant transformation. *Biosci Biotechnol Biochem* **74**, 1315-1319 (2010).
85. I. A. Sparkes, J. Runions, A. Kearns, C. Hawes, Rapid, transient expression of fluorescent fusion proteins in tobacco plants and generation of stably transformed plants. *Nat Protoc* **1**, 2019-2025 (2006).
86. M. Walter *et al.*, Visualization of protein interactions in living plant cells using bimolecular fluorescence complementation. *Plant J.* **40**, 428-438 (2004).

CHAPTER 2

Other potential functions of MAC in *Arabidopsis*

ABSTRACT

The MOS4-associated complex (MAC) is a multi-protein complex highly conserved across eukaryotes, including yeast, animals and *Arabidopsis*. Previous studies showed that this complex has multiple functions in eukaryotes, including roles in transcription elongation, mRNA precursor (pre-mRNA) splicing, DNA repair, RNA export and the ubiquitin-proteasome pathway. The functions of *Arabidopsis* MAC in plant immunity, miRNA biogenesis and pre-mRNA splicing have been characterized before. Here we explore other potential functions of MAC and the MAC7 subunit in particular in *Arabidopsis*. We found that MAC7 has a potential role in photosynthetic pathways, based on the finding that many photosynthesis-related genes were down-regulated in a nuclear RNA-seq analysis of the *mac7-1* mutant. In addition, MAC7 and PRL proteins are likely to be involved in ribosomal RNA (rRNA) biogenesis, as many rRNA precursors accumulated aberrantly in the *mac7-1* and *prl1 prl2* mutants.

INTRODUCTION

The MOS4-associated complex (MAC) is a multi-protein complex highly conserved across eukaryotes, including yeast, animals and *Arabidopsis*. The complex is known as the NineTeen Complex (NTC) or Prp19 complex (Prp19C) in yeast (1, 2), in which it was discovered as a spliceosome-associated complex through mutant screening and biochemical approaches. Prp19C was shown to play a critical role in pre-mRNA splicing (3, 4). Human Prp19C, also known as the CELL DIVISION CYCLE 5-LIKE-SENESCENCE EVASION FACTOR (CDC5L) complex, was similarly discovered as a spliceosome-associated complex in early studies utilizing mass spectrometry analysis (5, 6). The study of the *Arabidopsis* MAC began with a genetic screen for suppressor mutants of *snc1* (*suppressor of npr1-1, constitutive 1*). The *snc1* mutant has pleiotropic developmental defects and enhanced resistance to many pathogens. The suppressor screen identified many genes suppressing the *snc1* autoimmune phenotype, including *MOS4* (*MODIFIER OF SNC1*, 4) (7). MOS4-immunoprecipitation followed by mass spectrometry analysis identified many MOS4-associated proteins that comprise the MAC (8). Many MAC components, including MOS4, MAC3, MAC5, PRL and CDC5, play a positive role in plant defense and affect resistance (R) gene splicing (7-12). Recent studies have shown that *Arabidopsis* MAC interacts with spliceosome subunits, affects global gene splicing and contributes to small RNA biogenesis through interactions with key small RNA biogenesis factors (13-16).

MAC/Prp19C has many functions in eukaryotes in addition to those in splicing and small RNA biogenesis. Other key functions of this complex include RNA transcription elongation and export from the nucleus to the cytoplasm (2). Yeast Prp19C affects transcription elongation through interaction with RNA polymerase II (Pol II) and by ensuring the occupancy of the TREX complex at transcribed genes. TREX is another conserved multi-subunit protein complex that couples transcription to nuclear mRNA export in eukaryotes (17, 18). Human Prp19C also interacts with Pol II (19), and the knockdown of Prp19C components in human cells leads to the nuclear retention of intronless mRNAs, indicating that Prp19C links transcription to nuclear mRNA export (20). *Arabidopsis* MAC components also interact with Pol II and probably function in miRNA transcription elongation (13, 14). However, the roles of *Arabidopsis* MAC in RNA transcription elongation remain unclear, and its involvement in nuclear RNA export is completely unknown.

One particularly intriguing function of MAC/Prp19C is in genome maintenance. Yeast Prp19, the core component of Prp19C, was initially identified from a genetic screen for mutants with increased sensitivity to DNA damage (21, 22). Human Prp19 (hPrp19) was shown to interact with a terminal deoxynucleotidyl transferase, which is involved in cell survival after DNA damage. In addition, knockdown of hPrp19 leads to increased DNA double-strand breaks (23). XAB2, a key component of human Prp19C homologous to *Arabidopsis* MAC9, also plays a critical role in transcription-coupled DNA repair. XAB2 knockdown increases the UV sensitivity of cells, indicating defects in DNA repair (24, 25).

In addition, in both human and yeast cells, Prp19C interacts with the TREX complex, which has known roles in genome maintenance (20, 26, 27). Whether *Arabidopsis* MAC functions in genome maintenance is completely unknown, but several lines of evidence have shown that it functions in the small RNA pathway, including miRNA and siRNA biogenesis (13, 14). This may be a clue for connecting *Arabidopsis* MAC with genome maintenance, since siRNAs are required for DNA double-strand break repairs in plants (28).

MAC/Prp19C has a potential role in protein degradation by the proteasome. Prp19 in both human and yeast and the Prp19 homolog in *Arabidopsis*, MAC3, all exhibit E3 ligase activity, so it is tempting to speculate that they are involved in the ubiquitin-proteasome pathway (29, 30). Human and yeast Prp19 are known to interact with a component of the 20S catalytic core of the proteasome (31) and may therefore function in the recruitment of ubiquitylated proteins to the proteasome for degradation, although the substrates of Prp19 remain unclear. Whether *Arabidopsis* MAC3 functions in ubiquitin-proteasome-mediated protein degradation is unknown.

One conserved MAC/Prp19C subunit is known as Aquarius in vertebrates, EMB-4 in *C. elegans*, CG31368 in *Drosophila* and MAC7 in *Arabidopsis* (32). The human Aquarius protein is an RNA helicase with ATPase and RNA-unwinding activity *in vitro*. Aquarius also interacts with the U2 spliceosome and functions in pre-mRNA splicing *in vitro* (32). The *C. elegans* EMB-4 protein physically interacts with the nuclear Argonaute proteins HRDE-1 and CSR-1, and its functions include the removal of intronic sequence barriers to

facilitate nuclear RNAi, maintaining the homeostasis of germline small RNA and mRNA transcriptomes, and resetting germline H3K4me2 chromatin marks (33, 34). Similarly, *Drosophila* CG31368 interacts with Piwi protein, the germline Argonaute protein in *Drosophila* (35). In *Arabidopsis*, it was recently found that MAC7 is required for miRNA biogenesis and pre-mRNA splicing (see Chapter 1). Thus, MAC7 and its homologs have various functions in pre-mRNA splicing, germline nuclear RNAi, chromatin remodeling and small RNA biogenesis. However, the molecular action of this protein in these processes requires further investigation.

Here we focused on other potential functions of MAC and MAC7 in *Arabidopsis*. Since both human and yeast MAC homologs affect nuclear mRNA export, we performed total RNA-seq using the nuclear and cytoplasmic fractions, and compared the nuclear RNA export efficiency between Col-0 and a weak mutant allele of *MAC7*, *mac7-1*. Although we did not observe global differences in the cytoplasmic to nuclear RNA ratios in *mac7-1* compared to Col-0, we found that many photosynthesis-related genes are down-regulated in the mutant through the nuclear fraction RNA-seq. Furthermore, we also observed that *mac7-1* has pale-green cotyledons compared to Col-0 and the *mac7-1* complementation line, further indicating that MAC7 may modulate photosynthetic pathways. In addition, we found that many rRNA precursors (pre-rRNA) were up-regulated in the *mac7-1* and *prl1 prl2* mutants based on pre-rRNA northern blots, therefore MAC7 and PRL proteins might play a role in ribosomal RNA (rRNA) biogenesis.

RESULTS

MAC7 has a potential role in photosynthetic pathways

mac7-1 was isolated as a miRNA biogenesis defective mutant, and the role of MAC7 in pre-mRNA splicing was confirmed through polyA RNA-seq analysis (see Chapter 1). Here we found that *mac7-1* in the Col-0 background had pale-green cotyledons compared to Col-0 grown under the same conditions (Figure 2.1A). This pale-green cotyledon phenotype was not observed in other *mac* mutants, including the *mac3a mac3b* and *prl1 prl2* double mutants (data not shown). This unique phenotype of *mac7-1* indicates that MAC7 might play a role in regulating photosynthetic pathways in addition to its functions associated with the MAC. Based on *mac7-1* and Col-0 nuclear RNA-seq, the gene ontology (GO) analysis revealed an enrichment of genes involved in photosynthetic pathways among the down-regulated genes (hypo-DEGs) (Figure 2.1B). Among up-regulated genes (hyper-DEGs), no outstanding functional categories were found through GO analysis (Figure 2.1C). While these results suggest the potential involvement of MAC7 in regulating photosynthetic pathways, more experimental evidence is required to confirm this function and to characterize the molecular mechanisms of MAC7 in photosynthetic pathways.

***mac7-1* is hyposensitive to red light compared to Col-0**

We examined the light responses of the *mac7-1* and *mac3a mac3b* mutants by measuring hypocotyl length under dark, red light and blue light conditions. Both *mac7-1* and *mac3a mac3b* had shorter absolute hypocotyl length under all conditions, which may be

attributable to the developmental defects of these *mac* mutants (Figure 2.2 A, B). Interestingly, *mac7-1* appeared to be hyposensitive to red light compared to Col-0 and *mac3a mac3b* based on its relatively longer hypocotyl length under red light (Figure 2.2A). The response to blue light was not altered in the *mac* mutants compared to Col-0 (Figure 2.2B). For this analysis, the red light receptor mutant *phyB-9* and blue light receptor mutant *cry1 cry2* served as the hyposensitive controls. It remains unclear if *mac7-1* hyposensitivity to red light affects photomorphogenesis and therefore results in the down-regulation of photosynthetic genes and pale-green cotyledons.

Global histone methylation levels were not altered in *mac7-1*

Several lines of evidence have shown that EMB-4, the *C. elegans* homolog of MAC7, is required for germline-specific chromatin remodeling, specifically, the removal of H3K4me2 activation marks. The global methylation levels of H3K4me3, H3K9me2, H3K27me1 and H3K27me3 were not altered in *mac7-1* compared to Col-0 (Figure 2.3). Chromatin remodeling at individual gene loci in specific tissue types or growth stages needs to be further examined.

MAC functions in rRNA precursor processing and mRNA export

MAC has broad functions in RNA metabolism, including small RNA biogenesis and pre-mRNA splicing. Here we investigated other potential functions of MAC in RNA biogenesis. We found that many rRNA precursors were up-regulated in the *mac7-1* and *prl1 prl2* mutants, but not in *cdc5-2* or the *mac3a mac3b* double mutant (Figure 2.4).

Specifically, *mac7-1* showed up-regulation of 35SP, 27S and P-A3. Similarly, *prl1 prl2* showed significant upregulation of 35SP, P-A3, 18S-A3 and 27S (Figure 2.4). These results indicated the potential functions of MAC in rRNA precursor processing and suggested that different MAC subunits may function independently in this process.

Prp19C in both yeast and human has potential functions in RNA export from the nucleus to the cytoplasm (20, 26). To investigate MAC7 involvement in nuclear mRNA export in *Arabidopsis*, we separated cell nuclei from the cytoplasm, isolated RNA from the nuclear and cytoplasmic extracts and constructed total RNA libraries using the RiboMinus method. We did not observe global differences in the ratios of nuclear RNA to cytoplasmic RNA per gene in *mac7-1* compared to Col-0 (Figure 2.5). One possible interpretation is that MAC7 is not involved in RNA export in *Arabidopsis*; alternatively, the finding may reflect the fact that *mac7-1* is a weak allele. Before any definitive conclusions can be drawn, other *mac* mutants need to be included in this nuclear RNA export analysis.

DISCUSSION

MAC/Prp19C components are highly conserved across different organisms. Most of the subunits are encoded by single-copy genes in yeast and human, but many MAC subunits have paralogs or functional redundancy in *Arabidopsis* (Table 2.1). The best-characterized function of this complex is pre-mRNA splicing (1). Here we explored other potential functions of *Arabidopsis* MAC and MAC7 beyond splicing and small RNA biogenesis.

We found that MAC7 has potential roles in photosynthetic pathways based on cotyledon color phenotypes and nuclear RNA-seq analysis. In addition, MAC7 and PRL proteins are likely to play a role in rRNA biogenesis, which would be consistent with a broad role of MAC in RNA metabolism.

Further experiments are required to confirm the above-mentioned functions of MAC and MAC7 in *Arabidopsis* and to gain mechanistic insight into their complex functions in different biological processes.

We observed the following: *mac7-1* seedlings have pale-green cotyledons, many photosynthesis-related genes are down-regulated in the *mac7-1* mutant based on nuclear RNA-seq analysis, and *mac7-1* is hyposensitive to red light compared to Col-0. We do not know if the above three findings are connected, but the down-regulation of photosynthesis genes may underlie the pale-green cotyledon phenotype and the observed hyposensitivity to red light. Nevertheless, leaf chlorophyll levels need to be measured accurately to confirm this result. Electron microscopy imaging can also be used to examine chloroplast development in the cotyledons. Since MAC7 is a potential RNA helicase, RNA immunoprecipitation followed by RNA-seq (RIP-seq) could be performed using anti-MAC7 antibodies to examine if the down-regulated photosynthesis-related gene transcripts are bound by MAC7. If that were observed, it would indicate that MAC7 directly regulates photosynthesis gene expression. For future nuclear RNA-seq and other experiments, it will be critical to include a *mac7-1* complementation line as well as other *mac* mutants. Since

the pale-green cotyledon phenotype was unique to *mac7-1* among all the *mac* mutants we observed, nuclear RNA-seq of other *mac* mutants could serve as the negative controls. If enrichment of photosynthesis genes were also observed among the hypo-DEGs of other *mac* mutants, this result could be an artifact of the nuclear RNA-seq experiment. Otherwise, we would conclude that MAC7 functions in modulating photosynthetic pathways in addition to its functions associated with MAC. It is also necessary to confirm the subcellular localization of MAC7 in *Arabidopsis*. Although the major subcellular localization of MAC7 is the nucleus, we cannot exclude the possibility that MAC7 also localizes to other subcellular sites, such as the chloroplast, given that MAC7 was found to co-purify with chloroplast proteins (36). If MAC7 also localizes in the chloroplast, it would be an additional line of evidence supporting MAC7 function in photosynthetic pathways.

MAC has broad functions in gene transcriptional, co-transcriptional and potentially post-transcriptional regulations. It will be interesting to explore the multiple roles of MAC and MAC7 in different layers of gene expression regulation. Although we did not observe global changes for several histone methylation marks, we cannot exclude the possibility that MAC7 contributes to transcriptional gene silencing through small RNA-mediated chromatin remodeling, given that *C. elegans* EMB-4 interacts with nuclear AGOs and is required for the loss of the H3K4me2 mark in germline cells (34, 37). MAC is also involved in the co-transcriptional processing of various RNA precursors, such as pre-mRNA splicing, primary miRNA biogenesis and rRNA precursor processing. It remains unknown if MAC has a common function in RNA precursor processing or if it has unique roles in

different pathways. Although we did not observe any defects in RNA nuclear export in *mac7-1*, the MAC may nevertheless play a role in this process, given the conserved functions of MAC in RNA export in yeast and human cells and the fact that *mac7-1* is a weak allele. Other *mac* mutants or other experimental strategies, such as *in situ* hybridization of mRNAs, may help uncover the functions of MAC in RNA export in *Arabidopsis*. Finally, given the close connection between MAC7 and the small RNA pathway, MAC7 may play a role in post-transcriptional gene silencing through RISC-mediated transcript cleavage or translational repression.

MAC is a conserved protein complex with critical functions for the survival of organisms. Many questions remain: What are the specific molecular functions of each MAC subunit? How do multiple MAC subunits coordinate to achieve the functions of this complex? Is MAC really involved in so many biological processes, or does it instead have one primary function, with the other observed phenotypes all due to secondary effects? Do multiple MAC subcomplexes play different roles in different processes? The answers to all of these questions will require further investigation.

METHODS

Nuclear and cytoplasmic RNA-seq

Nuclei isolation was performed as described (38). Briefly, 12-day-old *Arabidopsis* seedlings grown on MS plates were ground in Honda buffer. The lysate was centrifuged at

3000g. The pellet was washed several times then collected as the nuclear fraction. The supernatant was collected as the cytoplasmic fraction. RNAs were extracted with TRI reagent (Molecular Research Center). The RiboMinus kit was used to remove rRNAs before RNA library construction. RNA-seq libraries were prepared using NEBNext® mRNA Library Prep Reagent Set for Illumina (New England Biolabs) and sequenced on an Illumina HiSeq 2500 platform to generate high-quality single-end reads 101 bp in length. Data analysis was performed with the pRNASeqTools pipeline (<https://github.com/grubbybio/RNASeqTools>). GO analysis of DEGs was performed with agriGO (39).

Fluence rate-response experiments

Arabidopsis was grown on MS plates and maintained in the dark at 4°C for 4 days for stratification. Germination was induced with 4 hours white light treatment. For hypocotyl measurement, plants were grown at 22°C under specific light conditions for 4 days. The seedlings were scanned using a scanner, and hypocotyls were measured from the images. For relative hypocotyl length, hypocotyl length in the dark was used for normalization. All hypocotyl lengths in the dark were set to 1 (40).

Western blotting

Twelve-day-old *Arabidopsis* seedlings grown on MS plates were used for protein extraction followed by western blotting analyses. The antibodies used for the western

blotting experiments include anti-H3K4me3 (Abcam, ab8580), anti-H3K9me2 (Abcam, ab1220), anti-H3K27me1 (Millipore, 07-448) and anti-H3K27me3 (Millipore, 07-449).

Northern blotting for rRNA precursors

Total RNA from 2- to 3-week-old seedlings (aerial tissues) or inflorescences was extracted with TRI reagent (Molecular Research Center). Northern blotting for detection of rRNA precursors was performed as previously described (41).

FIGURES

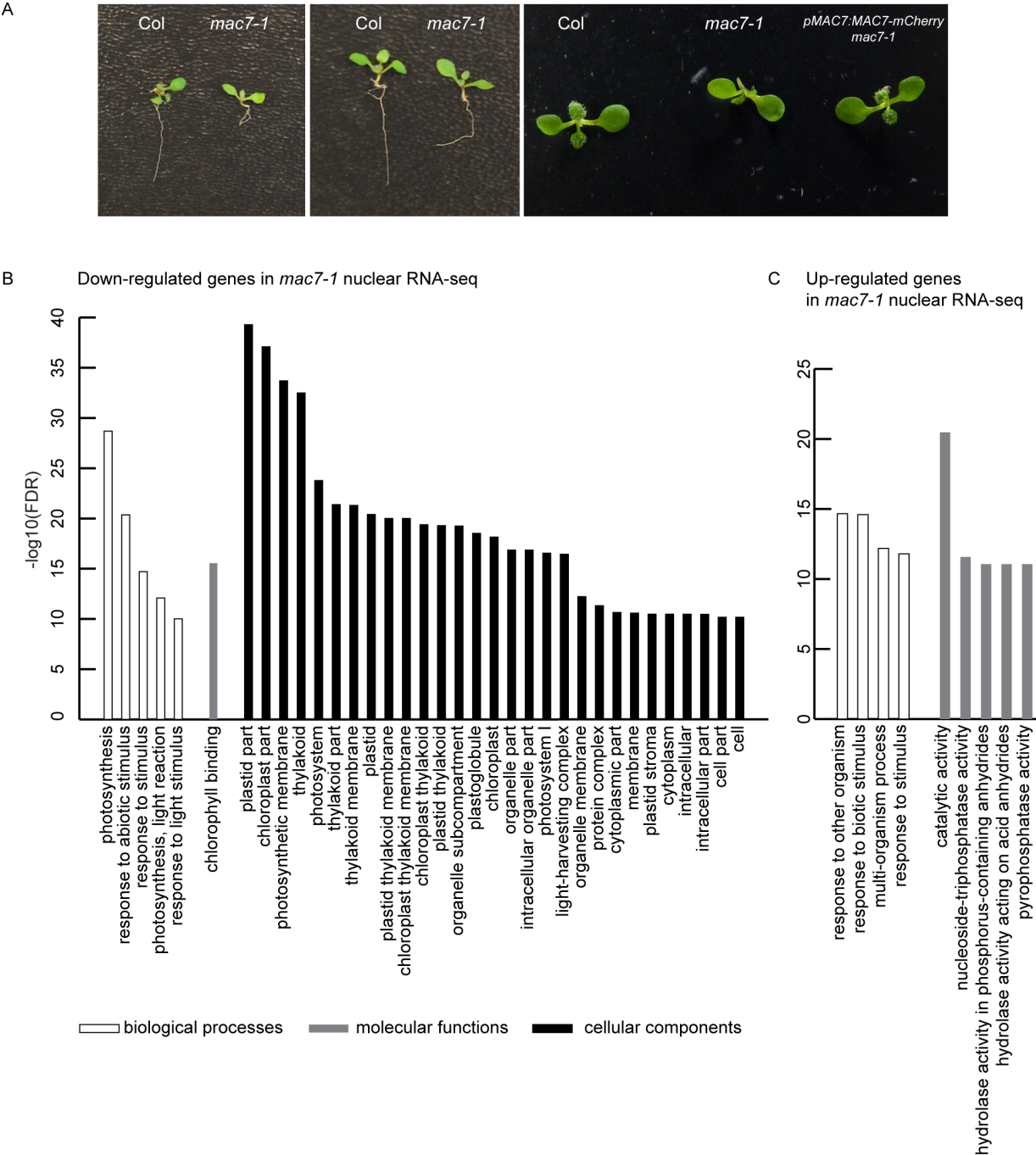


Figure 2.1 MAC7 has a potential role in photosynthetic pathways.

(A) Cotyledon color of 7- to 10-day-old Col-0, *mac7-1*, and *mac7-1 pMAC7:MAC7-mCherry* seedlings grown on the same MS plate. (B and C) GO enrichment analysis of down-regulated genes (B) and up-regulated genes (C) in *mac7-1* compared to Col-0. Three biological replicates were included in the nuclear RNA-seq for each genotype. Only GO categories with a false discovery rate (FDR) $\leq 1.0 \times 10^{-10}$ were included in the bar charts.

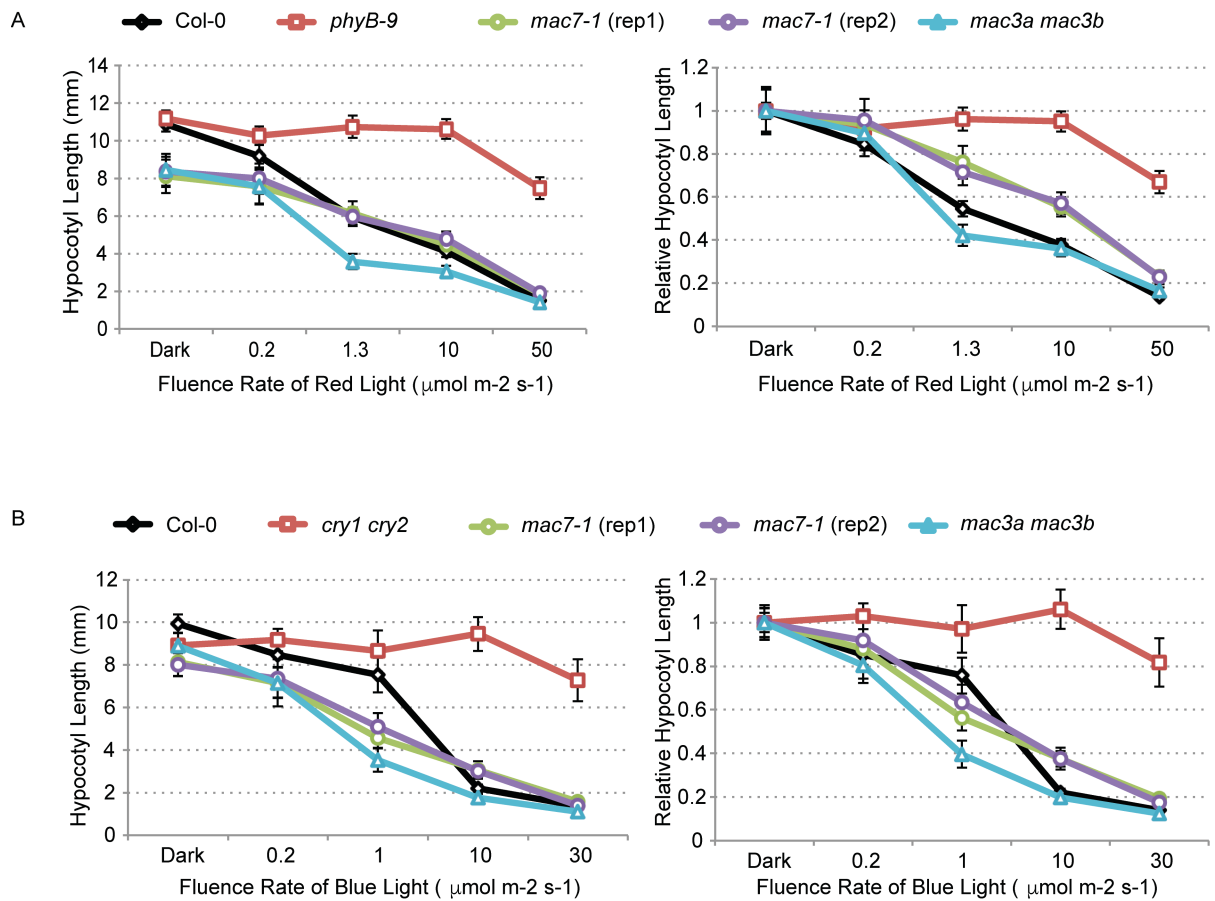


Figure 2.2 *mac7-1* is hyposensitive to red light compared to Col-0.

Hypocotyl length measurements of 4-day-old seedlings grown under different fluence rates of red light (A) or blue light (B). To determine the relative hypocotyl length, the measured length was normalized to the hypocotyl length in the dark (set to 1) for all genotypes. Error bars represent standard error.

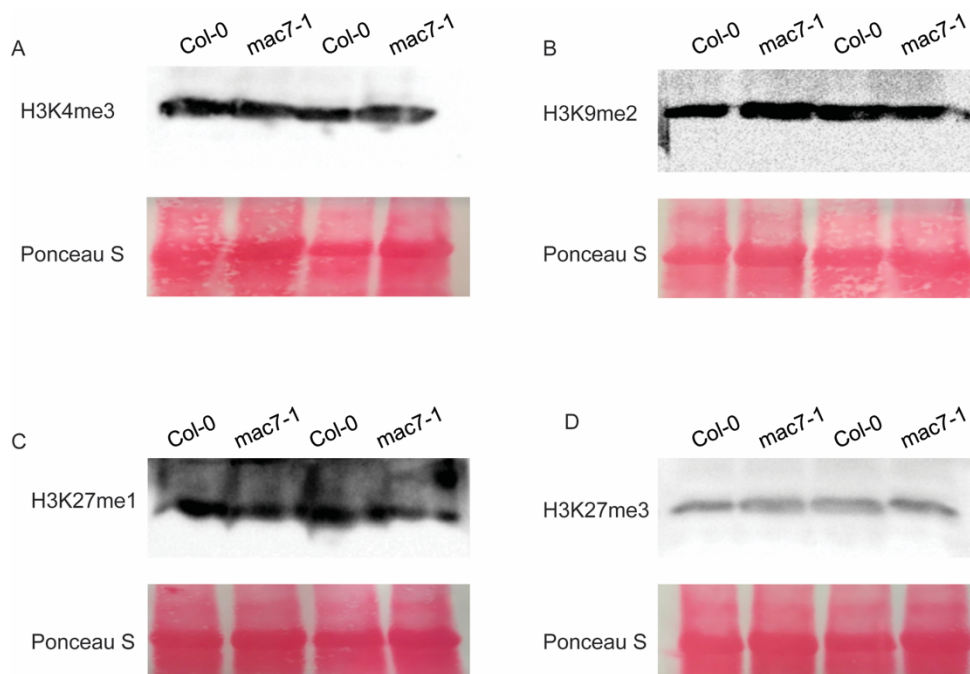


Figure 2.3 Global histone methylation levels were not altered in *mac7-1*.

Protein levels of H3K4me3 (A), H3K9me2 (B), H3K27me1 (C) and H3K27me3 (D) in Col-0 and *mac7-1* as determined by western blotting. Twelve-day-old seedlings grown on MS plates were used for protein extraction. Two biological replicates for each genotype were included. Ponceau S staining of the membrane served as the loading control.

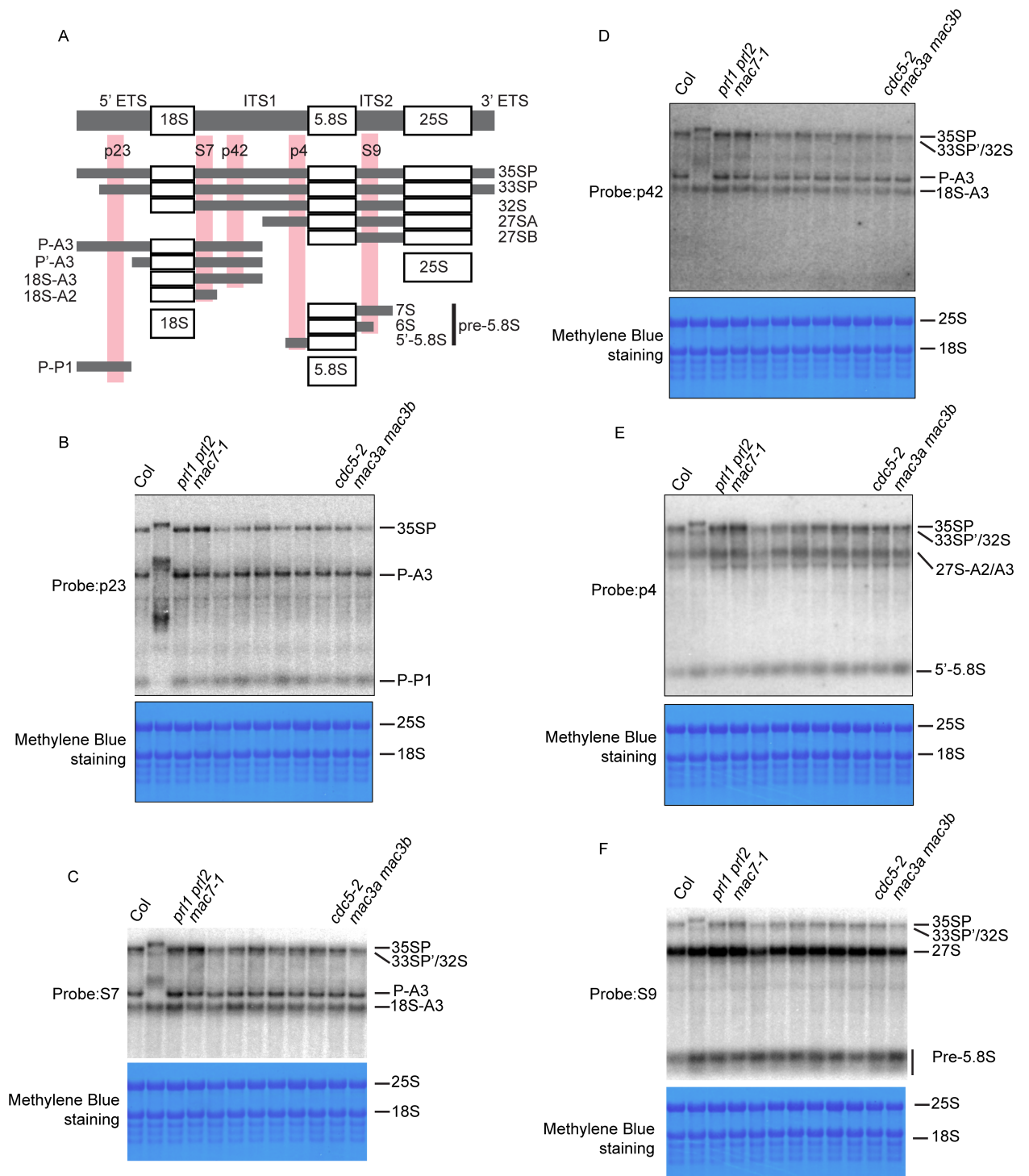


Figure 2.4 Ribosomal RNA precursors accumulated aberrantly in *mac7-1* and *prl1 prl2*.

(A) Diagram illustrating the various pre-rRNA processing intermediates detected by Northern blotting. Probe positions are indicated by red rectangles. (B to F) rRNA precursor levels in Col-0 and the indicated mutants as determined by northern blotting using labeled probes. Aberrant accumulations of 35SP, P-A3, 18S-A3, 27S-A2/A3 and pre-5.8S were observed in *mac7-1* and *prl1 prl2*. Methylene blue staining showing 25S and 18S rRNAs served as the loading control.

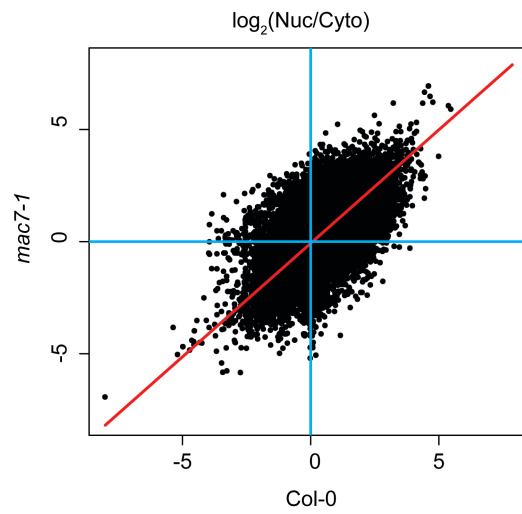


Figure 2.5 *mac7-1* did not cause global change in RNA export from the nucleus to the cytoplasm.

Log₂ ratios of nuclear RNA level/cytoplasmic RNA level per gene in Col-0 vs. *mac7-1*.

TABLES

Table 2.1 Conserved MAC components in *Arabidopsis*, human and yeast complexes.

* Parentheses indicate factors that have not been confirmed in purified complexes from the corresponding organisms.

AGI	MAC No.	Protein Name	Molecular Functions	Paralogs in <i>A. thaliana</i>	Homologs in <i>H. sapiens</i>*	Homologs in Yeast*
AT1G09770	MAC1	Cell division cycle 5 (CDC5)	R2R3 Myb transcription factor	--	CDC5L	Cef1/NTC85
AT4G15900	MAC2	Pleiotropic regulatory locus 1 (PRL1)	WD-40 domain containing protein; Scaffold protein	AT3G16650 (PRL2)	PLRG1/PRL1	Prp46/NTC50/Cwcl
AT2G33340	MAC3B	MOS4-associated complex 3B	E3 ligase	AT1G04510 (MAC3A)	hPRP19SNEV/PSO4	Prp19
AT1G07360	MAC5A	MOS4-associated complex 5A	CCCH-type zinc finger family protein with RNA-binding domain	AT2G29580 (MAC5B)	RMB22	Ecm2
AT1G77180	MAC6	AtSKIP	Transcriptional activation	--	SKIP1/SNW1	Prp45
AT2G38770	MAC7	Embryo defective	RNA helicase	--	Aquarius/AQR	(Cwf11p)

		2765 (EMB2765)				
AT3G18790	MAC8	AtISY1	Protein binding	--	ISY1/KIAA1160/SAP33	Isy1/NTC30
AT5G28740	MAC9	Tetratricopeptide repeat (TPR)-like superfamily protein/AtSYF1	Protein binding	--	XAB2	Syf1/NTC90
At5G41770	MAC10	AtCRN1C	Protein binding	At5g45900/AtCRN1A; At3g13210/AtCRN1B; At3g51110/AtCRN2	CRNKL1/SYF3	Clf1/Syf3/NTC77
At1G06220	MAC11	CLO; GFA1; MEE5; VAJ	U5 snRNP; translation elongation; nucleic acid binding	At5g25230/AtU5-116-1B; At1g56070/AtU5-116-2; At3g22980/AtU5-116-3	(U5-116/EFTUD2)	(Snu114)
AT1G10580	MAC12	AtPRP17-1	Transducin/W D40 repeat-like superfamily protein; Protein binding	At5g54520/AtPRP17-2	CDC40	Prp17/CDC40/Slu4
At1G15200	MAC13	AtPININ	Protein binding	--	--	--

At1G209 60	MAC1 4	EMB1507	U5 snRNP; DNA/RNA helicase	At5g61140/AtU5-200- 1; At2g42270/AtU5- 200-2B; At3g27739/AtU5-200- 3	(U5- 200/SNRNP200)	(Brr2)
At1G324 90	MAC1 5	EMB2733; ESP3	DEAH RNA helicase	At2g35340/MEE29/At PRP2-1B; At4g16680/AtPRP2-2	(DHX16)	(Prp2)
At1G800 70	MAC1 6	SUS2; EMB177	U5 snRNP, Protein binding	At4g38780/AtPRP8B	(U5-220)	(Prp8)
At2G437 70	MAC1 7	AtU5-40	U5 snRNP, Protein binding	--	(SNRNP40)	(Cwf17)
At5G642 70	MAC1 8	AtSAP155	U2 snRNP, Protein binding	--	(SF3B1)	(Prp10)
At3G181 65	MOS4	Modifier of snc1,4	Protein binding	--	BCAS2/SPF27	Snt309/NTC25

REFERENCES

1. C. Koncz, F. Dejong, N. Villacorta, D. Szakonyi, Z. Koncz, The spliceosome-activating complex: molecular mechanisms underlying the function of a pleiotropic regulator. *Front Plant Sci* **3**, 9 (2012).
2. S. Chanarat, K. Strasser, Splicing and beyond: the many faces of the Prp19 complex. *Biochim. Biophys. Acta* **1833**, 2126-2134 (2013).
3. S. C. Cheng, W. Y. Tarn, T. Y. Tsao, J. Abelson, PRP19: a novel spliceosomal component. *Mol. Cell. Biol.* **13**, 1876-1882 (1993).
4. M. D. Ohi *et al.*, Proteomics analysis reveals stable multiprotein complexes in both fission and budding yeasts containing Myb-related Cdc5p/Cef1p, novel pre-mRNA splicing factors, and snRNAs. *Mol. Cell. Biol.* **22**, 2011-2024 (2002).
5. O. V. Makarova *et al.*, A subset of human 35S U5 proteins, including Prp19, function prior to catalytic step 1 of splicing. *EMBO J.* **23**, 2381-2391 (2004).
6. M. Grote *et al.*, Molecular architecture of the human Prp19/CDC5L complex. *Mol. Cell. Biol.* **30**, 2105-2119 (2010).
7. K. Palma *et al.*, Regulation of plant innate immunity by three proteins in a complex conserved across the plant and animal kingdoms. *Genes Dev.* **21**, 1484-1493 (2007).
8. J. Monaghan *et al.*, Two Prp19-like U-box proteins in the MOS4-associated complex play redundant roles in plant innate immunity. *PLoS Pathog* **5**, e1000526 (2009).
9. F. Xu, S. Xu, M. Wiermer, Y. Zhang, X. Li, The cyclin L homolog MOS12 and the MOS4-associated complex are required for the proper splicing of plant resistance genes. *Plant J.* **70**, 916-928 (2012).
10. T. Weihmann, K. Palma, Y. Nitta, X. Li, Pleiotropic regulatory locus 2 exhibits unequal genetic redundancy with its homolog PRL1. *Plant Cell Physiol.* **53**, 1617-1626 (2012).
11. K. C. M. Johnson, O. X. Dong, X. Li, The evolutionarily conserved MOS4-associated complex. *Central European Journal of Biology* **6**, 776-784 (2011).
12. J. Monaghan, F. Xu, S. Xu, Y. Zhang, X. Li, Two putative RNA-binding proteins function with unequal genetic redundancy in the MOS4-associated complex. *Plant Physiol.* **154**, 1783-1793 (2010).

13. S. Zhang, M. Xie, G. Ren, B. Yu, CDC5, a DNA binding protein, positively regulates posttranscriptional processing and/or transcription of primary microRNA transcripts. *Proc Natl Acad Sci U S A* **110**, 17588-17593 (2013).
14. S. Zhang, Y. Liu, B. Yu, PRL1, an RNA-binding protein, positively regulates the accumulation of miRNAs and siRNAs in Arabidopsis. *PLoS Genet* **10**, e1004841 (2014).
15. X. Deng *et al.*, Recruitment of the NineTeen Complex to the activated spliceosome requires AtPRMT5. *Proc Natl Acad Sci U S A* **113**, 5447-5452 (2016).
16. C. J. Zhang *et al.*, The splicing machinery promotes RNA-directed DNA methylation and transcriptional silencing in Arabidopsis. *Embo J* **32**, 1128-1140 (2013).
17. S. Chanarat, M. Seizl, K. Strasser, The Prp19 complex is a novel transcription elongation factor required for TREX occupancy at transcribed genes. *Genes Dev.* **25**, 1147-1158 (2011).
18. K. Strasser *et al.*, TREX is a conserved complex coupling transcription with messenger RNA export. *Nature* **417**, 304-308 (2002).
19. I. Kuraoka *et al.*, Isolation of XAB2 complex involved in pre-mRNA splicing, transcription, and transcription-coupled repair. *J. Biol. Chem.* **283**, 940-950 (2008).
20. H. Lei, B. Zhai, S. Yin, S. Gygi, R. Reed, Evidence that a consensus element found in naturally intronless mRNAs promotes mRNA export. *Nucleic Acids Res.* **41**, 2517-2525 (2013).
21. M. Grey, A. Dusterhoft, J. A. Henriques, M. Brendel, Allelism of PSO4 and PRP19 links pre-mRNA processing with recombination and error-prone DNA repair in *Saccharomyces cerevisiae*. *Nucleic Acids Res.* **24**, 4009-4014 (1996).
22. J. A. Henriques, E. J. Vicente, K. V. Leandro da Silva, A. C. Schenberg, PSO4: a novel gene involved in error-prone repair in *Saccharomyces cerevisiae*. *Mutat. Res.* **218**, 111-124 (1989).
23. K. N. Mahajan, B. S. Mitchell, Role of human Pso4 in mammalian DNA repair and association with terminal deoxynucleotidyl transferase. *Proc Natl Acad Sci U S A* **100**, 10746-10751 (2003).
24. Y. Nakatsu *et al.*, XAB2, a novel tetratricopeptide repeat protein involved in transcription-coupled DNA repair and transcription. *J. Biol. Chem.* **275**, 34931-34937 (2000).

25. R. Yonemasu *et al.*, Disruption of mouse XAB2 gene involved in pre-mRNA splicing, transcription and transcription-coupled DNA repair results in preimplantation lethality. *DNA Repair (Amst)* **4**, 479-491 (2005).
26. S. Chanarat, M. Seizl, K. Strasser, The Prp19 complex is a novel transcription elongation factor required for TREX occupancy at transcribed genes. *Genes Dev.* **25**, 1147-1158 (2011).
27. S. Jimeno, A. G. Rondon, R. Luna, A. Aguilera, The yeast THO complex and mRNA export factors link RNA metabolism with transcription and genome instability. *EMBO J.* **21**, 3526-3535 (2002).
28. W. Wei *et al.*, A role for small RNAs in DNA double-strand break repair. *Cell* **149**, 101-112 (2012).
29. J. Wiborg, C. O'Shea, K. Skriver, Biochemical function of typical and variant *Arabidopsis thaliana* U-box E3 ubiquitin-protein ligases. *Biochem. J.* **413**, 447-457 (2008).
30. M. D. Ohi, C. W. Vander Kooi, J. A. Rosenberg, W. J. Chazin, K. L. Gould, Structural insights into the U-box, a domain associated with multi-ubiquitination. *Nat. Struct. Biol.* **10**, 250-255 (2003).
31. M. Loscher *et al.*, Interaction of U-box E3 ligase SNEV with PSMB4, the beta7 subunit of the 20 S proteasome. *Biochem. J.* **388**, 593-603 (2005).
32. I. De *et al.*, The RNA helicase Aquarius exhibits structural adaptations mediating its recruitment to spliceosomes. *Nat. Struct. Mol. Biol.* **22**, 138-144 (2015).
33. A. Akay *et al.*, The Helicase Aquarius/EMB-4 Is Required to Overcome Intronic Barriers to Allow Nuclear RNAi Pathways to Heritably Silence Transcription. *Dev. Cell* **42**, 241-255 e246 (2017).
34. K. M. Tyc *et al.*, The Conserved Intron Binding Protein EMB-4 Plays Differential Roles in Germline Small RNA Pathways of *C. elegans*. *Dev. Cell* **42**, 256-270 e256 (2017).
35. A. Le Thomas *et al.*, Piwi induces piRNA-guided transcriptional silencing and establishment of a repressive chromatin state. *Genes Dev.* **27**, 390-399 (2013).
36. I. Tzafrir *et al.*, Identification of genes required for embryo development in *Arabidopsis*. *Plant Physiol.* **135**, 1206-1220 (2004).

37. P. M. Checchi, W. G. Kelly, *emb-4* is a conserved gene required for efficient germline-specific chromatin remodeling during *Caenorhabditis elegans* embryogenesis. *Genetics* **174**, 1895-1906 (2006).
38. H. Zhang *et al.*, Protocol: a beginner's guide to the analysis of RNA-directed DNA methylation in plants. *Plant Methods* **10**, 18 (2014).
39. T. Tian *et al.*, agriGO v2.0: a GO analysis toolkit for the agricultural community, 2017 update. *Nucleic Acids Res.*, (2017).
40. M. Chen *et al.*, Arabidopsis HEMERA/pTAC12 initiates photomorphogenesis by phytochromes. *Cell* **141**, 1230-1240 (2010).
41. R. Hang *et al.*, Arabidopsis protein arginine methyltransferase 3 is required for ribosome biogenesis by affecting precursor ribosomal RNA processing. *Proc Natl Acad Sci U S A* **111**, 16190-16195 (2014).

Appendix A

MiRNA movement

INTRODUCTION

In addition to the aforementioned miRNA biogenesis pathways, another intriguing research topic is the non-cell-autonomous effect of many miRNAs. It has long been known that the movement of hormones, proteins and many small molecules serves as a mode of signaling in multicellular organisms. However, the cell-to-cell movement of RNAs was not discovered until 20 years ago, perhaps reflecting the fact that RNAs are very unstable and easily degraded by cellular RNase. The proposal that RNAs could serve as mobile signals was based on the observation that the movable silencing signals in question were nucleotide-sequence dependent (1-3). Small RNAs are now known to move locally from cell to cell, systemically from source tissue to sink tissue and even between organisms (4, 5). Here, we focus specifically on findings regarding the movement of miRNAs.

Local movement: cell-to-cell communication

Unlike many siRNAs, particularly 24-nt siRNAs, which promote systemic silencing through long-range trafficking, miRNAs are best known for short-range intercellular trafficking largely through the symplastic pathway via plasmodesmata (PD) and regulate plant development by establishing miRNA activity gradients (6, 7).

The current model posits that mRNAs, viral RNAs, siRNAs and miRNAs all travel through PD to achieve intercellular communication (8-11). Because plant cells are encased in rigid cell walls, PD, which connect the cytoplasm of adjacent cells, are the major channel for cell-to-cell communication. The structure of PD is notable in this regard. There is an appressed endoplasmic reticulum (ER) in PD, ensuring the continuity of endomembrane between neighboring cells. The sizes of PD vary among different tissues and can be altered through callose deposition. PD can also dilate significantly, permitting the passage of proteins >20 kDa between cells (9, 12, 13) (Figure A.1A).

The molecular mechanisms of small RNA trafficking through PD remain elusive. It is unclear if single-stranded small RNAs, double-stranded small RNAs or small RNA precursors are the mobile molecules. Whether they are bound or protected by any proteins is also unknown, but the prevalence of cellular RNase and its ability to readily degrade RNAs make it seem more probable that small RNAs travel with conjugated proteins (Figure A.1A).

There are several lines of evidence demonstrating that plant miRNAs can move short range across several cell layers and act non-cell-autonomously to regulate plant development.

The first discovery was published in 2009 and concerned the movement of tasiR-ARF small RNAs (14). TasiR-ARF small RNAs are not technically miRNAs due to their

differences in biogenesis, but they do function similarly. TasiR-ARFs are loaded into AGO1 like miRNAs, and they target and mediate the repression of *ARF3*. The study found that even though tasiR-ARF biogenesis factors (*AGO7* and *TAS3A*) are restricted to the adaxial cell layers of leaves, the mature tasiR-ARFs form a gradient from the adaxial side to the abaxial side, sharply restricting *ARF3* expression to the abaxial side. In the shoot apical meristem (SAM) region, tasiR-ARF biogenesis factor *AGO7* is expressed beneath the SAM, while tasiR-ARFs apparently act non-cell-autonomously within the SAM itself. The authors also found that miR390 promoter activity and precursors are similarly restricted beneath the SAM, whereas mature miR390 exists throughout the SAM, implying the movement of mature miR390 into the SAM region (14). Despite these critical lines of evidence for the non-cell-autonomous effects of tasiR-ARFs and miR390, most of the key evidence is based only on *in situ* hybridization, with limited signal quality and resolution. Further direct evidence is still required to establish how tasiR-ARFs and miR390 move from cell to cell (e.g., through symplastic or apoplastic pathways) (Figure A.1B).

Another example of miRNA movement is the ability of miR165/6 to direct root cell patterning non-cell-autonomously (15). The promoter activities of miR165/6 are restricted to the root endodermis, but mature miR165/6 is distributed in the ground tissue (layers outside of endodermis) and stele (layers inside of endodermis), suggesting bi-directional movement of miR165/6 in the root tissue. This study established the importance of the miR165/6 gradient for root development, specifically protoxylem and metaxylem patterning. The higher levels of miR165/6 in cell layers close to the endodermis repress the

expression of *HD-ZIP III* transcription factor genes, particularly *PHABULOSA (PHB)*, and promote the formation of protoxylem rather than metaxylem (15) (Figure A.1C).

One additional example of miRNA local movement in *Arabidopsis* is miR394. During *Arabidopsis* SAM differentiation, miR394 promoter activity is restricted to the surface layer (the L1 layer), but mature miR394 is detected in the underlying L2 and L3 layers and represses the expression of *LEAF CURLING RESPONSIVENESS (LCR)* post-transcriptionally. This non-cell-autonomous effect of miR394 in the L2 and L3 layers is necessary for stem cell maintenance during shoot meristem formation (16) (Figure A.1D).

Despite the clearly demonstrated non-cell-autonomous effects of miR165/6 and miR394 in root patterning and SAM development, respectively, the nature of the mobile miRNAs or miRNA precursors remains unknown. Another open question is whether the movement is actively regulated or achieved by passive diffusion.

In addition to the above examples of mobile endogenous miRNAs, transgenic artificial miRNAs (amiR) have also been used to investigate non-cell-autonomous miRNA functions. In one study, amiR-SUL targeting an *Arabidopsis* homolog of tobacco *SULPHUR (SUL)* was expressed from a construct driven by the *SUCROSE-PROTON SYMPORTER 2 (SUC2)* promoter as an easily scorable marker for miRNA activity *in vivo* (17). While *SUC2:3xYFP* expression was restricted to the vasculature, the bleaching phenotype of *SUC2:amiR-SUL* spread 10 to 15 cell layers beyond the vasculature, indicating movement of the artificial

miRNAs or miRNA precursors. Moreover, artificial tasiRNA atasiR-SUL had phenotypic effects indicative of enhanced mobility compared to amiR-SUL, suggesting distinct intercellular movement pathways employed by miRNAs and tasiRNAs (17) (Figure A1.E).

Systemic movement

There are plentiful examples of systemic movement of small RNA-mediated silencing signals (1, 2, 7, 11, 18, 19). Several grafting experiments have illustrated that the systemic silencing signals are transmitted from source to sink, implying that systemic siRNA movement occurs largely through phloem (1, 7, 11) (Figure A.1F).

There are fewer examples of miRNA systemic movement. *Arabidopsis* miR399 was found to serve as a phloem-mobile signal that moves from shoots to roots to regulate phosphate homeostasis (20). This study relied primarily on qRT-PCR detection of miR399, its precursor and its target *PHO2* in grafted shoots and roots. Other strategies, such as a silencing reporter line to detect miR399 activities, may therefore provide further evidence for the systemic movement of miR399 (Figure A.1F).

A study that identified several miRNA and siRNA species from pumpkin (*Cucurbita maxima*) phloem sap also identified Phloem SMALL RNA BINDING PROTEIN1 (CmPSRP1) as a factor that potentially mediates the trafficking of small RNAs (19). This was the first direct evidence of systemic movement of small RNAs or miRNAs in plants. Through direct analysis of phloem sap instead of artificial grafting assays to assess small

RNA movement, the mobile molecule was characterized specifically as a protein-bound small single-stranded RNA moving through the phloem (19). Although this has not yet been observed in the model organism *Arabidopsis*, it would be interesting to determine whether *Arabidopsis* small RNAs or miRNAs can bind to some RNA binding proteins and move through phloem sap (Figure A.1F).

Compared to local cell-to-cell movement, miRNA systemic movement is less well established. Further investigation is needed to identify additional examples and to better understand the biological significance of miRNA systemic movement.

FIGURES

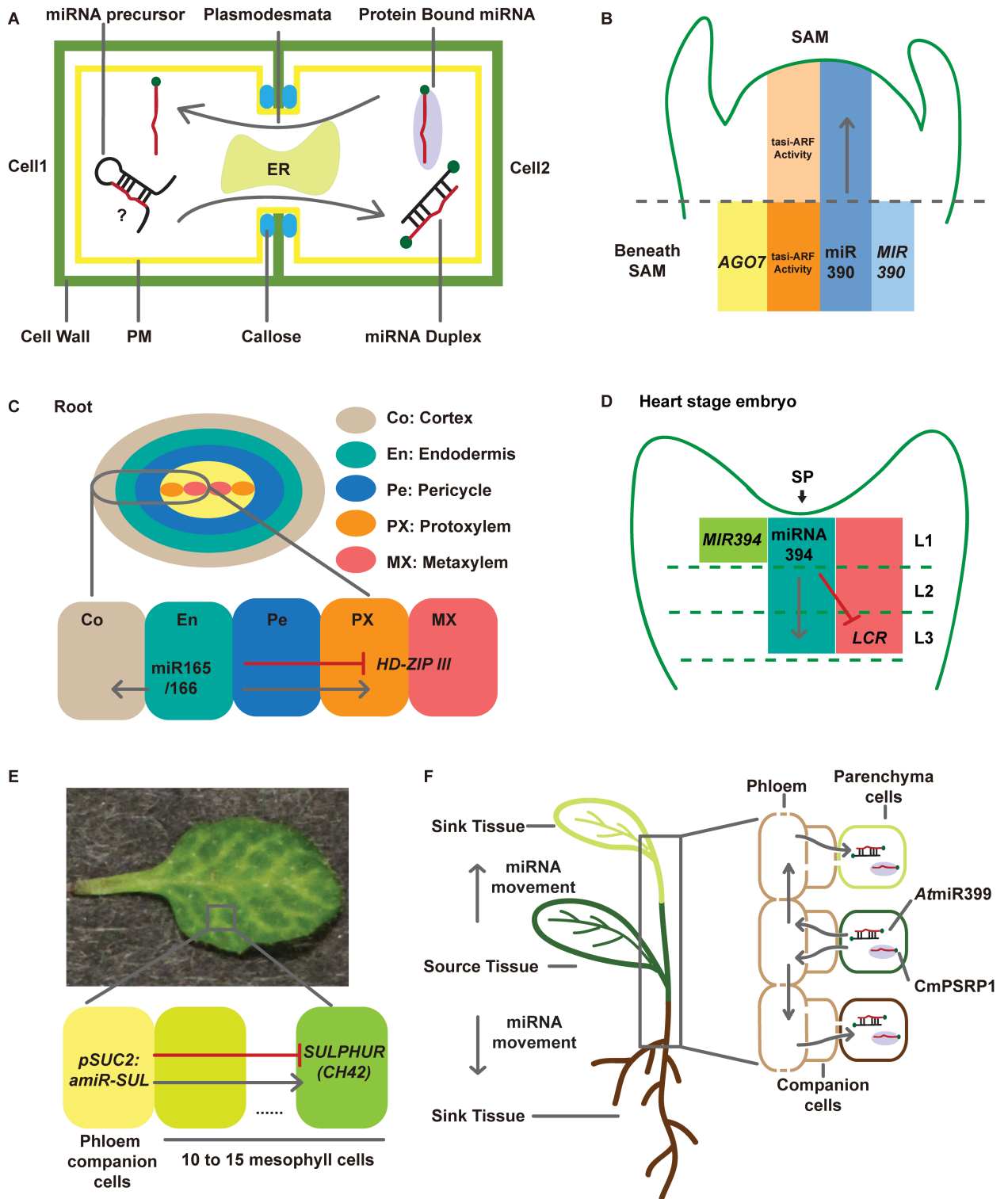


Figure A.1 Illustrations of miRNA cell-to-cell and systemic movement in plants.

(A) Cell-to-cell movement of miRNAs or miRNA precursors through plasmodesmata. (B) Non-cell-autonomous effects of tasiR-ARFs and miR390. (C) MiRNA165/166 movement between root cell layers. (D) Non-cell-autonomous effects of miR394 in stem cell maintenance during shoot meristem formation. (E) Movement of the artificial miRNA amiR-SUL from phloem companion cells to mesophyll cells. (F) Examples of systemic miRNA movement in plants.

REFERENCES

1. J. C. Palauqui, T. Elmayan, J. M. Pollien, H. Vaucheret, Systemic acquired silencing: transgene-specific post-transcriptional silencing is transmitted by grafting from silenced stocks to non-silenced scions. *Embo J* **16**, 4738-4745 (1997).
2. O. Voinnet, D. C. Baulcombe, Systemic signalling in gene silencing. *Nature* **389**, 553 (1997).
3. C. W. Melnyk, A. Molnar, D. C. Baulcombe, Intercellular and systemic movement of RNA silencing signals. *Embo J* **30**, 3553-3563 (2011).
4. P. Sarkies, E. A. Miska, Small RNAs break out: the molecular cell biology of mobile small RNAs. *Nat Rev Mol Cell Biol* **15**, 525-535 (2014).
5. M. Knip, M. E. Constantin, H. Thordal-Christensen, Trans-kingdom cross-talk: small RNAs on the move. *PLoS Genet* **10**, e1004602 (2014).
6. N. R. Gursansky, I. R. Searle, B. J. Carroll, Mobile microRNAs hit the target. *Traffic* **12**, 1475-1482 (2011).
7. A. Molnar *et al.*, Small silencing RNAs in plants are mobile and direct epigenetic modification in recipient cells. *Science* **328**, 872-875 (2010).
8. J. Kehr, A. Buhtz, Long distance transport and movement of RNA through the phloem. *J Exp Bot* **59**, 85-92 (2008).
9. W. J. Lucas, B. C. Yoo, F. Kragler, RNA as a long-distance information macromolecule in plants. *Nat Rev Mol Cell Biol* **2**, 849-857 (2001).
10. B. Ding, A. Itaya, X. Zhong, Viroid trafficking: a small RNA makes a big move. *Curr Opin Plant Biol* **8**, 606-612 (2005).
11. C. W. Melnyk, A. Molnar, A. Bassett, D. C. Baulcombe, Mobile 24 nt small RNAs direct transcriptional gene silencing in the root meristems of *Arabidopsis thaliana*. *Curr Biol* **21**, 1678-1683 (2011).
12. W. J. Lucas, J. Y. Lee, Plasmodesmata as a supracellular control network in plants. *Nat Rev Mol Cell Biol* **5**, 712-726 (2004).
13. J. Y. Lee, Plasmodesmata: a signaling hub at the cellular boundary. *Curr Opin Plant Biol* **27**, 133-140 (2015).
14. D. H. Chitwood *et al.*, Pattern formation via small RNA mobility. *Genes Dev* **23**, 549-554 (2009).

15. A. Carlsbecker *et al.*, Cell signalling by microRNA165/6 directs gene dose-dependent root cell fate. *Nature* **465**, 316-321 (2010).
16. S. Knauer *et al.*, A protodermal miR394 signal defines a region of stem cell competence in the Arabidopsis shoot meristem. *Dev Cell* **24**, 125-132 (2013).
17. F. F. de Felippes, F. Ott, D. Weigel, Comparative analysis of non-autonomous effects of tasiRNAs and miRNAs in Arabidopsis thaliana. *Nucleic Acids Res* **39**, 2880-2889 (2011).
18. B. K. Ham, G. Li, W. Jia, J. A. Leary, W. J. Lucas, Systemic delivery of siRNA in pumpkin by a plant PHLOEM SMALL RNA-BINDING PROTEIN 1-ribonucleoprotein complex. *Plant J* **80**, 683-694 (2014).
19. B. C. Yoo *et al.*, A systemic small RNA signaling system in plants. *Plant Cell* **16**, 1979-2000 (2004).
20. B. D. Pant, A. Buhtz, J. Kehr, W. R. Scheible, MicroRNA399 is a long-distance signal for the regulation of plant phosphate homeostasis. *Plant J* **53**, 731-738 (2008).