

**INVESTIGATING THE ROLE  
OF MARF1-LIKE (MFL) ENDORIBONUCLEASE GENES  
IN ARABIDOPSIS**

by

Catherine J. Bogdan

A thesis submitted to the Faculty of the University of Delaware in partial fulfillment of the requirements for the degree of Master of Science in Biological Sciences

Summer 2021

© 2021 Catherine J. Bogdan  
All Rights Reserved

**INVESTIGATING THE ROLE  
OF MARF1-LIKE (MFL) ENDORIBONUCLEASE GENES  
IN ARABIDOPSIS**

by

Catherine J. Bogdan

Approved: \_\_\_\_\_  
Pamela J. Green, Ph.D.  
Professor in charge of thesis on behalf of the Advisory Committee

Approved: \_\_\_\_\_  
Velia M. Fowler, Ph.D.  
Chair of the Department of Biological Sciences

Approved: \_\_\_\_\_  
John A. Pelesko, Ph.D.  
Dean of the College of Arts and Sciences

Approved: \_\_\_\_\_  
Louis F. Rossi, Ph.D.  
Vice Provost for Graduate and Professional Education and  
Dean of the Graduate College

## ACKNOWLEDGMENTS

I dedicate this Master's thesis to all of the people in my life who have encouraged and supported me throughout my education.

First and foremost, thank you to my research advisor, Dr. Pamela J. Green, for being an amazing mentor and role model. I hope you know how grateful I am that you gave me the opportunity to work in your lab. I could not have asked for a better graduate research experience, and I certainly could not have asked for a better PI to guide me through it.

Thank you to the other members of the Green lab for welcoming me with open arms and inspiring me to do my best every day. Special thanks to Dr. Vinay Nagarajan for introducing me to this project and for giving me the knowledge that I needed to succeed. You are a fantastic teacher, collaborator, and friend, and I couldn't have done this without you. Thank you to Monica Accerbi, Anna DiBattista, Ardita Gjonbalaj, and Susan Winkelmayer for your help and encouragement throughout my time in the lab. We truly became a family, and I am going to miss you all very much.

Thank you to my committee members, Dr. Erica Selva, Dr. Salil Lachke, and Dr. Ambro van Hoof, for your advice and guidance throughout my graduate career. Your input over the past two years has been crucial to my development as a scientist.

Thank you to my collaborators on this project, Dr. Jeffrey Caplan, Tim Chaya, and Kody Seward, for your assistance in determining the subcellular localization of the AtMFL proteins, and Dr. Renate Wuersig for your assistance in determining the conservation of MFL1-like proteins across plants.

Thank you to my friends, especially Rachel Keown, Kristy Wisser, and Malek Elsayyid, who I met during my time at the University of Delaware. I will never forget our late-night conversations, long study sessions, impromptu coffee runs, and much-needed girls' nights. Your futures are so bright, and I wish you all the success and happiness in the world.

Thank you to my parents and the rest of my family for always believing in me and rooting for my success. I would not be here today without your unwavering support. I hope that over the course of my career and my life, I continue to make you proud.

Thank you to my husband, Skylor Stuart, for always encouraging me to shoot for the stars. I look forward to sharing my life with you and experiencing all of our dreams come true. With you by my side, I know that nothing is impossible.

Thank you to the University of Delaware and the National Science Foundation (grant #1817764 to Pamela J. Green) for my funding. Any opinions, findings, conclusions, or recommendations expressed in this thesis are mine alone and do not necessarily reflect the views of the University of Delaware or the National Science Foundation.

## TABLE OF CONTENTS

|  |      |
|--|------|
| LIST OF TABLES .....   | vii  |
| LIST OF FIGURES .....  | viii |
| ABSTRACT .....   | x    |
| Chapter  |      |
| 1 INTRODUCTION .....   | 1    |
| Canonical eukaryotic mRNA decay .....                                      | 1    |
| The many roles of endoribonucleases .....                                  | 2    |
| Nonsense-mediated mRNA decay (NMD).....                                    | 3    |
| Meiosis regulator and mRNA stability factor 1 (MARF1) .....                | 7    |
| MARF1 homologs in <i>Arabidopsis thaliana</i> .....                        | 8    |
| Global hypothesis, experimental aims, and research goals .....             | 12   |
| 2 MATERIALS AND METHODS .....  | 14   |
| Plant materials and growth conditions .....                                | 14   |
| Plasmid construction .....   | 15   |
| <i>Localization / complementation constructs</i> .....                     | 16   |
| <i>AtMFL1 point mutation construct</i> .....                               | 17   |
| Transient expression and subcellular localization .....                    | 18   |
| RNA analysis.....  | 18   |
| <i>Total RNA extraction and cDNA synthesis</i> .....                       | 18   |
| <i>Reverse transcription (RT)-PCR</i> .....                                | 18   |
| <i>Quantitative (q)-PCR</i> .....  | 19   |
| <i>Northern blots</i> .....  | 19   |
| RNA-seq library construction.....  | 20   |
| Computational approaches .....   | 21   |
| <i>Phylogenetic analysis of MFL1-like proteins in plants</i> .....         | 21   |
| <i>RNA-seq analysis</i> .....  | 21   |
| 3 RESULTS.....   | 25   |
| Organ-specific expression of the <i>AtMFL</i> genes .....                  | 25   |
| Subcellular localization of the <i>AtMFL</i> proteins .....                | 26   |
| <i>AtMFL1</i> -like proteins are conserved across many plant species ..... | 33   |

|          |   |    |
|----------|---|----|
|          | <i>mfl1</i> and <i>mfl1xrn4</i> are RNA null .....                            | 36 |
|          | Certain transcripts are differentially expressed in <i>mfl1</i> mutants ..... | 38 |
|          | Certain 3' fragments appear to be AtMFL1-dependent .....                      | 44 |
|          | D153 may be required for AtMFL1 nuclease activity.....                        | 49 |
| 4        | DISCUSSION.....   | 52 |
|          | Characterization of plant <i>MFL</i> genes .....                              | 52 |
|          | Insights from the impacts of AtMFL1 on gene expression .....                  | 55 |
| 5        | CONCLUSIONS AND FUTURE WORK.....  | 59 |
|          | Conclusions .....   | 59 |
|          | Future Directions .....   | 60 |
|          | REFERENCES .....  | 62 |
| Appendix |   |    |
| A        | <i>ATMFL</i> GENOME SEQUENCES .....   | 74 |
| B        | ATMFL1-LIKE PROTEIN SEQUENCES .....   | 79 |

## LIST OF TABLES

|         |   |    |
|---------|---|----|
| Table 1 | Primers utilized.....   | 22 |
| Table 2 | Alignment statistics of RNA-seq libraries made from Arabidopsis seedlings .....               | 39 |
| Table 3 | Genes significantly upregulated in <i>mfl1</i> and <i>mfl1xrn4</i> compared to controls ..... | 41 |

## LIST OF FIGURES

|           |   |    |
|-----------|---|----|
| Figure 1  | Simplified pathway for certain NMD transcripts in metazoans .....   | 5  |
| Figure 2  | Proposed simplified pathway for certain NMD transcripts in plants .....   | 6  |
| Figure 3  | Bootstrap consensus tree showing relatedness between Arabidopsis MARF1-like proteins and metazoan MARF1 .....   | 10 |
| Figure 4  | Bootstrap consensus tree showing relatedness of relevant MARF1-like proteins and HsMARF1 .....  | 11 |
| Figure 5  | Multiple sequence alignment showing conservation of catalytic residues in AtMFL proteins and metazoan MARF1 .....   | 12 |
| Figure 6  | <i>AtMFL1</i> , <i>AtMFL2</i> , and <i>AtMFL3</i> expression in various Col-0 organs .....  | 26 |
| Figure 7  | Elevated mRNA levels of <i>AtMFL</i> genes in plants expressing <i>CFP::AtMFL</i> fusion constructs .....   | 28 |
| Figure 8  | <i>AtMFL1</i> localizes to distinct foci in the cytosol .....   | 29 |
| Figure 9  | <i>AtMFL2</i> localizes to the cytosol and cytoskeleton .....   | 31 |
| Figure 10 | <i>AtMFL3</i> localizes to the nuclear membrane and forms aggregates in the cytosol.....  | 32 |
| Figure 11 | Multiple sequence alignment showing conservation of catalytic residues in various plant MFL1-like proteins, <i>AtMFL</i> proteins, and metazoan MARF1 ..... | 35 |
| Figure 12 | Strategy for identifying targets of <i>AtMFL1</i> .....   | 37 |
| Figure 13 | <i>mfl1</i> , <i>xrn4</i> , and <i>mfl1xrn4</i> are RNA null .....  | 38 |
| Figure 14 | Impacts of <i>mfl1</i> mutation on gene expression .....  | 40 |
| Figure 15 | Enriched GO categories of genes upregulated in <i>mfl1</i> compared to Col-0 .....  | 42 |
| Figure 16 | Enriched GO categories of genes upregulated in <i>mfl1xrn4</i> compared to Col-0 and <i>xrn4</i> .....  | 43 |
| Figure 17 | Elevated mRNA levels of selected genes in <i>mfl1xrn4</i> .....   | 44 |

|           |   |    |
|-----------|---|----|
| Figure 18 | Computational pipeline used for PARE analysis .....                                   | 46 |
| Figure 19 | AtMFL1-dependent 3' fragments over-accumulate in an <i>xrn4</i> mutant...             | 48 |
| Figure 20 | RNA blots detect 3' RNA fragments produced by AtMFL1 .....                            | 49 |
| Figure 21 | CFP::AtMFL1 fusion protein rescues the presence of the <i>AtPAO2</i> 3' fragment..... | 50 |
| Figure 22 | AtMFL1 cleavage activity may require D153.....  | 51 |

## ABSTRACT

RNA decay in eukaryotes is mainly carried out by exoribonucleases and endoribonucleases, but the latter have been underappreciated. In metazoan nonsense-mediated mRNA decay (NMD), the endoribonuclease SMG6 cleaves certain mRNAs, leaving the 5' and 3' fragments to be degraded by the exosome and XRN1, respectively. Though SMG6 orthologs are absent in plants, recent studies in *Arabidopsis thaliana* indicate that some NMD transcripts are cleaved by an unknown endoribonuclease, and their 3' fragments over-accumulate in the absence of XRN4 (the plant cytosolic homolog of XRN1). Metazoan MARF1 is an endoribonuclease in humans and mice that cleaves targeted mRNAs via its catalytic NYN domain. *Arabidopsis* has several different NYN domain-containing proteins that we predicted to be homologs of metazoan MARF1. Importantly, one of these proteins (AtMFL1) localizes to P-bodies and has an RNA-dependent association with critical NMD factor AtUPF1. I therefore hypothesized that MARF1-like (MFL) endoribonucleases play a role in post-transcriptional control of gene expression in plants.

The major goal of this project was to study the role of selected MFL endoribonucleases in *Arabidopsis*, particularly AtMFL1. Here I report the organ-specific expression of the *AtMFL* genes— *AtMFL1* and *AtMFL2* are ubiquitously expressed in all major organs, while *AtMFL3* is minimally expressed in certain organs. I also report the subcellular localization of the AtMFL proteins— AtMFL1 localizes to cytosolic foci, AtMFL2 localizes to the cytosol and cytoskeleton, and AtMFL3 localizes to the nuclear membrane and cytosolic aggregates. Additionally, I report that AtMFL1-like proteins with catalytically active NYN domains are evolutionarily

conserved across many plant species. Finally, I provide evidence for the role of AtMFL1 as a cytosolic endoribonuclease that is involved in mRNA degradation.

## **Chapter 1**

### **INTRODUCTION**

#### **Canonical eukaryotic mRNA decay**

Cytosolic messenger RNA (mRNA) decay is important for the posttranscriptional modulation of gene expression and for the degradation of unwanted or aberrant mRNA molecules. Eukaryotic mRNAs have a 5' 7-methylguanosine cap and a 3' poly(A) tail that protect them from premature degradation by nuclear and cytosolic exonucleases. Most eukaryotic mRNA degradation is initiated by the shortening of the poly(A) tail by the CCR4-CAF1-NOT1 deadenylase complex or by the poly(A) ribonuclease PARN (Garneau et al., 2007; Nagarajan et al., 2013). Deadened RNA can subsequently be degraded in the 3' to 5' direction by either the multi-subunit exosome complex or the exoribonuclease DIS3L2 (SOV in plants); alternatively, it can be degraded in the 5' to 3' direction by the exoribonuclease XRN1 (XRN4 in plants) after the 5' cap is removed by the DCP1-DCP2-VCS decapping complex (Zhang, Murphy, and Sieburth, 2010; Lubas et al., 2013; Malecki et al., 2017; Nagarajan et al., 2013). Some eukaryotic mRNA degradation is initiated by endoribonucleolytic cleavage or by decapping, but either way the major cytosolic exoribonucleases further degrade the mRNA molecules as previously described (Badis et al., 2004; Conti and Izaurralde, 2005).

## **The many roles of endoribonucleases**

For many years, research was primarily focused on the role of exoribonucleases in RNA processing and decay, but recent discoveries have shown that endoribonucleases play an important role in these processes as well. Notably, endoribonucleases are involved in various RNA maturation pathways, gene silencing by RNA interference, and several different RNA surveillance mechanisms. In the nucleus of eukaryotic cells, CPSF-73 is an endoribonuclease that cleaves the 3' end of mRNA, providing polyA polymerase with a substrate to add the 200–300-nt-long poly(A) tail required for translation and overall transcript stability (Ryan, Calvo, and Manley, 2004; Mandel et al., 2006). In the cytosol of eukaryotic cells, Nob1 is an endoribonuclease that cleaves 20S pre-ribosomal (r)-RNA near its 3' end, which leads to the formation of mature 18S rRNA (Lamanna and Karbstein, 2009; Pertschy et al., 2009). Also in the cytosol, Dicer and Argonaute are endoribonucleases involved in RNA interference (RNAi), a post-transcriptional mechanism of gene silencing in eukaryotes. Dicer, a member of the RNase III family of endoribonucleases, cleaves double-stranded RNA (dsRNA) substrates to produce 21-25-nt-long short interfering RNA (siRNA) molecules (Hammond et al., 2000; Zamore et al., 2000; Bernstein et al., 2001). The siRNA molecules are further processed into two strands—the guide strand and the passenger strand. The guide strand incorporates into the RNA-induced silencing complex (RISC) and base pairs with a complementary sequence on an mRNA transcript, targeting it for cleavage by the PIWI domain-containing endoribonuclease Argonaute (Bohmert et al., 1998; Martinez et al., 2002; Song et al., 2004). In Arabidopsis, the RNS endoribonucleases are essential for the production of both long and short tRNA fragments (tRFs), which go on to play a role in various biological processes (Megel et al., 2018). Also in Arabidopsis, RNase E cleaves RNA

similarly to the *E. coli* enzyme and is predicted to be involved in polyadenylation-stimulated RNA degradation in chloroplasts (Schein et al., 2008). Endoribonucleases have also been shown or predicted to be involved in several different eukaryotic mRNA surveillance pathways, such as nonsense-mediated mRNA decay (NMD), non-stop decay (NSD), and no-go decay (NGD) (Welch and Jacobson, 1999; Doma and Parker, 2007; Shoemaker and Green, 2012; Popp and Maquat, 2014). Ultimately, plants rely on these RNA decay pathways as well as the canonical RNA turnover mechanisms to respond to changes in the cell quickly and efficiently. The endoribonucleases that have been characterized thus far are essential for many different biological processes; however, there are many other many putative endoribonucleases whose roles have yet to be elucidated. In particular, not much is known about plant-specific endoribonucleases that are involved in mRNA turnover.

### **Nonsense-mediated mRNA decay (NMD)**

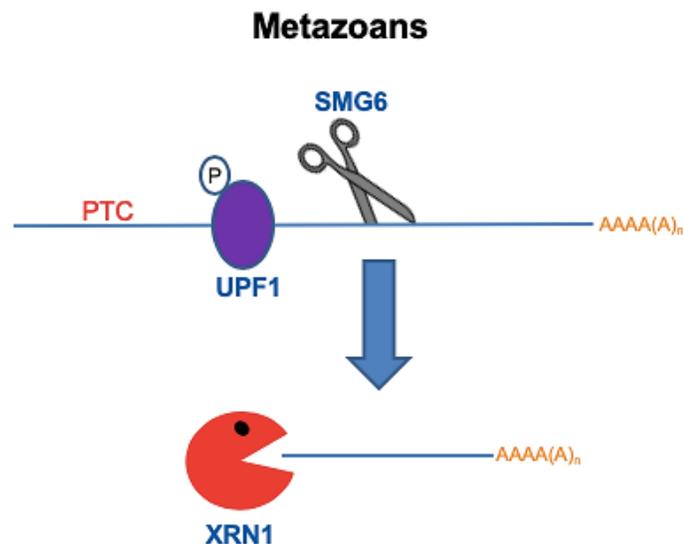
Nonsense-mediated mRNA decay (NMD) is a eukaryotic translation-dependent mRNA surveillance mechanism that targets certain transcripts for rapid degradation. While NMD typically targets transcripts that contain premature termination codons (PTCs) that are  $\geq 50$ – $55$  nucleotides upstream of the exon-junction complex (EJC), NMD can also target normal transcripts with other features such as introns downstream of stop codons, long 3' untranslated regions (UTRs), or upstream ORFs (uORFs) (Popp and Maquat, 2014; He and Jacobson, 2015). Disruption of NMD leads to altered levels of NMD targets in certain disease states, making it crucial to understand this pathway in a variety of eukaryotic organisms (Karam et al., 2013). While there are many proteins involved in the NMD pathway, the most important NMD factor is UPSTREAM FRAMESHIFT 1 (UPF1), an RNA helicase that binds to

targeted transcripts and initiates degradation. It is important to note that without UPF1, NMD fails to occur.

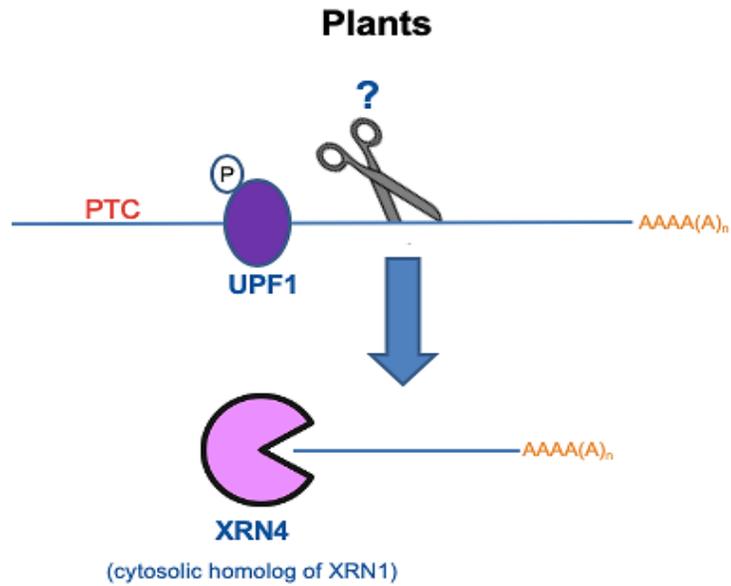
The NMD pathway is highly conserved across all eukaryotes and has been most extensively studied in mammalian cells. In normal mammalian translation termination, close proximity of the termination codon (TC) to the polyA tail allows for interaction of eukaryotic release factor 3 (eRF3) with the eIF4G-PABPC1 complex, which promotes efficient translation termination and ribosome recycling. In aberrant mammalian translation termination, significant distance between the TC and the polyA tail prevents interaction of eRF3 with the eIF4G-PABPC1 complex, resulting in the formation of the SURF complex which consists of eRF1-eRF3, UPF1, and SMG1 (Kashima et al., 2006). Interaction between the SURF complex, UPF2, and UPF3 causes the activation of SMG1, which phosphorylates UPF1 (Kashima et al., 2006). Phosphorylated UPF1 suppresses translation and recruits the endoribonuclease SMG6, which cleaves certain NMD transcripts in a sequence-specific manner via its catalytic PiIT N-terminal (PIN) domain (Glavan et al., 2006; Schmidt et al., 2015; Lykke-Andersen et al., 2014). The 3' products that result from this cleavage are degraded by XRN1 (Fig. 1), and the 5' products that result from this cleavage are presumably degraded by the cytosolic exosome complex (Gatfield & Izaurralde, 2004; Huntzinger et al., 2008). Phosphorylated UPF1 also recruits SMG5, which binds to either SMG7 or PNRC2 and results in decapping followed by 5'-to-3' degradation and/or deadenylation followed by 3'-to-5' degradation. While plants have orthologs of many NMD components (including UPF1), they lack an ortholog of SMG6 (Shaul, 2015).

There are very few PIN domain-containing proteins in Arabidopsis, and most of them are orthologs of endoribonucleases that are known to be involved in rRNA

maturation and/or processing. However, recent RNA degradome studies in *Arabidopsis thaliana* indicate that some NMD transcripts, such as those with conserved peptide uORFs (CPuORFs) and those elevated in mutants of conserved NMD factors (*upf1* and *upf3*), are cleaved by an unknown endoribonuclease, and the 3' fragments that result from this cleavage are degraded by XRN4 (Nagarajan et al., 2019) (Fig. 2). It has been proposed that plant NMD is not initiated by an endonucleolytic cleavage because the 3' fragments of PTC-containing reporter transcripts appeared unchanged in an XRN4-silenced line of *Nicotiana benthamiana* (Méraï et al., 2013). However, because NMD can target transcripts with a range of features, and 3' fragments of endogenous NMD targets have been shown to over-accumulate in *Arabidopsis* lacking XRN4, the search for an endogenous NMD endoribonuclease is justified (Nagarajan et al., 2019).



**Figure 1 Simplified pathway for certain NMD transcripts in metazoans**  
 Phosphorylated UPF1 suppresses translation and recruits the endoribonuclease SMG6 (scissors), which cleaves certain NMD transcripts in a sequence-specific manner via its catalytic PiIT N-terminal (PIN) domain. The 3' products that result from this cleavage are degraded by XRN1 in the 5' to 3' direction.



**Figure 2 Proposed simplified pathway for certain NMD transcripts in plants**  
 Phosphorylated UPF1 suppresses translation and recruits an unknown endoribonuclease (scissors), which cleaves certain NMD transcripts in a similar manner as metazoan SMG6. The 3' products that result from this cleavage are degraded by XRN4 (the plant cytosolic homolog of XRN1) in the 5' to 3' direction.

As previously mentioned, the RNA helicase UPF1 is a key NMD factor that binds to targeted transcripts and initiates degradation. Recent studies by Chicois et al. (2018) have identified a large set of proteins that co-purify with UPF1 in Arabidopsis, either in an RNA-dependent or RNA-independent manner. Notably, the study identified a putative endonuclease (encoded by *AT2G15560*) as one of the RNA-dependent partners of UPF1 (Chicois et al., 2018). The putative endonuclease co-localizes with UPF1 in P-bodies, which are cytosolic foci that accumulate proteins involved in mRNA decay (Aizer et al., 2014; Chicois et al., 2018). While the role of UPF1 in NMD has been very well characterized, UPF1 is also required for a variety of other mRNA decay pathways (reviewed in Kim and Maquat, 2019). One example is

the structure-mediated RNA decay (SRD) pathway that involves the association of UPF1 and G3BP1 with highly structured 3' UTRs (Fischer et al., 2020). Accordingly, it is possible that the putative endonuclease that associates with UPF1 could endonucleolytically cleave NMD targets or other UPF1-associated mRNAs in Arabidopsis.

### **Meiosis regulator and mRNA stability factor 1 (MARF1)**

The protein shown to associate with UPF1, encoded by *AT2G15560*, is an Nedd4-BP1 Bacterial YacP Nuclease (NYN) domain-containing protein. The NYN domain typically consists of four conserved aspartic acid residues that chelate a single divalent cation ( $Mg^{2+}$  or  $Mn^{2+}$ ) in a similar way as the PIN domain. These acidic residues are critical for catalysis, as they likely activate water for a nucleophilic attack on the phosphodiester bond of single-stranded RNA (ssRNA) (Anantharaman and Aravind, 2006). The most well-characterized NYN domain-containing protein is meiosis arrest female 1 (MARF1), also called meiosis regulator and mRNA stability factor 1. MARF1 is a *bona fide* endoribonuclease in humans and mice that is required for the cleavage and subsequent silencing of certain mRNAs (Nishimura et al., 2018; Yao et al., 2018). MARF1 has two isoforms, an oocyte form and a somatic form. The two splice variants have nearly the same sequence, with the only difference being that the somatic isoform has an additional 537-bp at the 3' end of exon 3. The domain architectures of human and mouse MARF1 indicate that they are nearly identical, consisting of an N-terminal NYN domain, two RNA recognition motifs, and a C-terminal tandem repeat of LOTUS/OST-HTH domains. The oocyte form of MARF1 is highly expressed in mouse oocytes and regulates oogenesis by silencing the expression of retrotransposons *Line1* and *Iap* as well as silencing the expression of *Ppp2cb*,

which encodes the catalytic beta subunit of the major cellular phosphatase PP2A (Su et al., 2012). Mutation of the oocyte form of MARF1 causes female-only infertility in mice due to immature germinal vesicle (GV)-stage oocytes being arrested at prophase I (Su et al., 2012). The somatic form of MARF1 promotes neuronal differentiation and cortical neurogenesis both *in vivo* and *in vitro* (Kanemitsu et al., 2017). MARF1 physically interacts with the DCP1:DCP2 mRNA decapping complex via a specific motif within its C-terminal end and preferentially binds to the 3' UTRs of its target mRNAs via its C-terminal LOTUS/OST-HTH domains (Nishimura et al., 2018; Yao et al., 2018; Brothers, 2020). Additionally, MARF1 cleaves target mRNAs via its catalytic NYN domain (Nishimura et al., 2018; Yao et al., 2018).

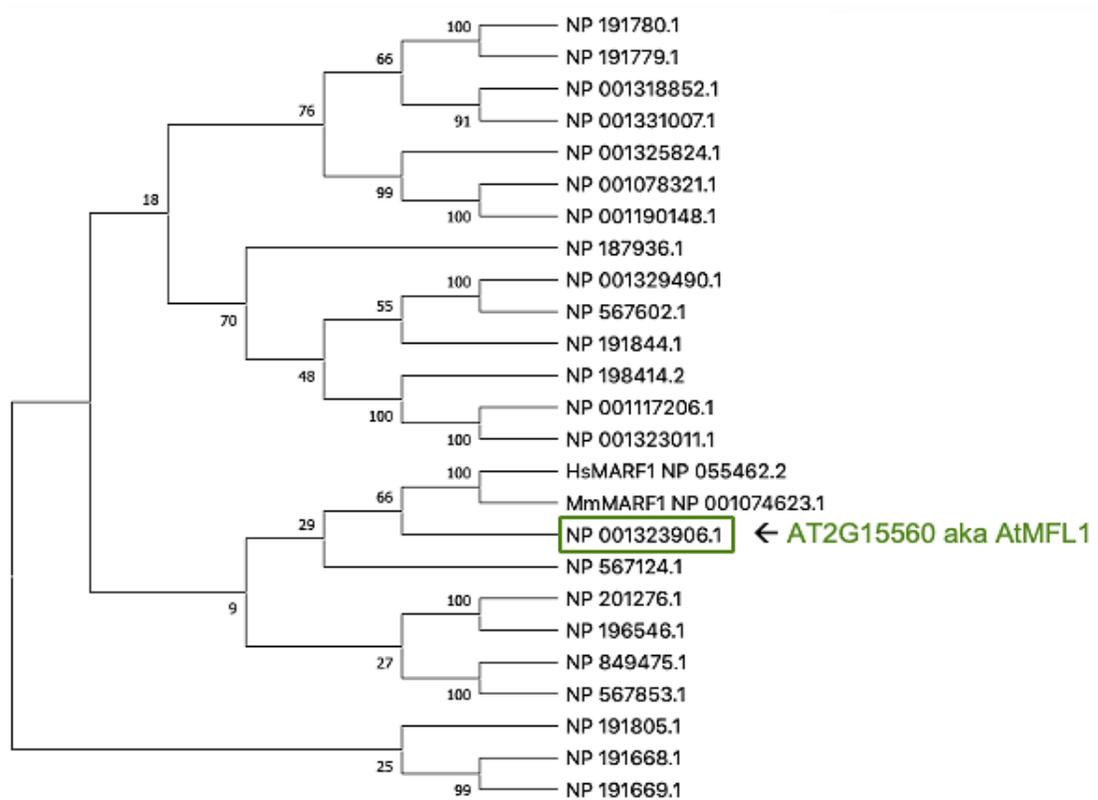
### **MARF1 homologs in *Arabidopsis thaliana***

To determine the relatedness between the protein encoded by *AT2G15560*, other *Arabidopsis* MARF1-like proteins, and metazoan MARF1, a Domain Enhanced Lookup Time Accelerated (DELTA) BLAST search of human MARF1 against all *Arabidopsis* proteins was performed. The search identified 31 *Arabidopsis* proteins related to metazoan MARF1, but only those described as “putative endonuclease or glycosyl hydrolase” were selected for subsequent analyses. A multiple sequence alignment of the remaining 23 *Arabidopsis* proteins as well as human MARF1 (HsMARF1) and mouse MARF1 (MmMARF1) was generated using the BLOSUM62 algorithm within MUSCLE (default settings) (Madeira et al., 2019). Finally, a bootstrap consensus tree was generated from this multiple sequence alignment using the maximum likelihood algorithm within MEGA (default settings) (Kumar et al., 2018) (Fig. 3). Because the first few nodes of this tree have low bootstrap values, there is not much confidence in the initial taxon bipartitions. That being said, the overall

results of this tree verified that the Arabidopsis protein with an accession number of NP\_001323906.1 (encoded by *AT2G15560*) is the most evolutionarily related to metazoan MARF1. Due to its similarity to MARF1, the protein encoded by *AT2G15560* will be referred to as *Arabidopsis thaliana* MARF-like 1 or AtMFL1 going forward. Gobert et al. (2019) also referred to this protein as the “plant MARF1 enzyme.”

Not much is known about NYN domain-containing proteins in Arabidopsis, yet related Arabidopsis proteins may be performing similar functions as AtMFL1 or otherwise be involved in mRNA decay. Therefore, several more AtMFL proteins were chosen for further study within the scope of this project. To this end, a DELTA-BLAST search of AtMFL1 against all Arabidopsis proteins was performed. Only those proteins containing an NYN domain with an alignment score of 80-200 were selected for subsequent analyses. A multiple sequence alignment of the remaining Arabidopsis proteins as well as HsMARF1 was generated using the BLOSUM62 algorithm within MUSCLE (default settings) (Madeira et al., 2019). Finally, a bootstrap consensus tree was generated from this multiple sequence alignment using the maximum likelihood algorithm within MEGA (default settings) (Kumar et al., 2018) (Fig. 4). Two more AtMFL proteins, AtMFL2 (encoded by *AT3G62200*) and AtMFL3 (encoded by *AT3G62210*), were identified based on their RNA expressivity and putative localization according to the ePlant Browser (Fig. 4), with the rationale being that any proteins playing a role in mRNA decay should have detectable expression levels and localize to the cytosol. To determine whether the candidate AtMFL proteins likely exhibit endoribonucleolytic function, a multiple sequence alignment of the NYN domains of all three AtMFL proteins, HsMARF1, and

MmMARF1 was generated using the BLOSUM62 algorithm within MUSCLE (Madeira et al., 2019). Based on the results of Yao et al. (2018), in which four aspartic acid residues within the NYN domain of MmMARF1 were shown to be required for catalytic activity, all three AtMFL proteins contain the conserved aspartic acid residues required for ssRNA cleavage (Fig. 5).



**Figure 3 Bootstrap consensus tree showing relatedness between Arabidopsis MARF1-like proteins and metazoan MARF1**

Full-length amino acid sequences were aligned using the BLOSUM62 algorithm within MUSCLE (default parameters) (Madeira et al., 2019) and bootstrap consensus tree was generated using maximum likelihood algorithm within MEGA (Kumar et al., 2018). Bootstrap values shown are from 1000 iterations. Hs, human; Mm, mouse; NP accession number, Arabidopsis.



**Figure 5 Multiple sequence alignment showing conservation of catalytic residues in AtMFL proteins and metazoan MARF1**

Amino acid sequence of the NYN domains of AtMFL proteins and metazoan MARF1 were aligned using the BLOSUM62 algorithm within MUSCLE (default parameters) (Madeira et al., 2019). Conserved amino acid residues are highlighted in red. Residues required for ssRNase activity as shown by Yao et al. (2018) are indicated by yellow triangles. Hs, human; Mm, mouse; At, Arabidopsis. ! is either I or V; \$ is either L or M; % is either F or Y; # is either N, D, Q, E, D, or Z.

**Global hypothesis, experimental aims, and research goals**

Considering the compelling evidence presented here, I hypothesize that MARF1-like (MFL) endoribonucleases play a role in post-transcriptional control of gene expression in plants. Specifically, I predict that I will identify a subset of transcripts that are cleaved by AtMFL1 *in vivo*. To test this hypothesis, I will accomplish the following experimental aims:

1. Determine the characteristics of the candidate *MFL* genes in *Arabidopsis thaliana*
  - 1.1 Verify that the *MFL* genes are reliably expressed in all major organs as indicated by the ePlant Browser
  - 1.2 Determine the subcellular localization of the MFL proteins using transient and/or stable transformations of *reporter::MFL* fusion constructs
  - 1.3 Determine whether MFL1-like proteins and their NYN domains are evolutionarily conserved across plant species
2. Evaluate the impact of *MFL1* on global RNA abundance in *Arabidopsis thaliana*
  - 2.1 Enable the functional characterization of *MFL1* by isolating knockout mutants
  - 2.2 Identify differentially accumulating transcripts in *mfl1* mutants by conducting transcriptomic analyses using RNA-seq

2.3 Independently validate selected *in vivo* substrates of the putative MFL1 endoribonuclease identified by RNA degradome studies of *mfl1xrn4* double mutants

2.4 Verify the functional dependence of MFL1 on a catalytic residue within its NYN domain by performing genetic complementation

## Chapter 2

### MATERIALS AND METHODS

#### Plant materials and growth conditions

*Arabidopsis* (*Arabidopsis thaliana*) ecotype Col-0 (WT), *xrn4* (*xrn4-5*, Souret et al., 2004) and *mfl1* (SALK\_132521) in the Col-0 background were used for these studies. Seeds were surface-sterilized with 70% (v/v) ethanol followed by 3.25% sodium hypochlorite. After two days of stratification in the dark at 4°C, the surface-sterilized seeds were plated on germination media containing 1X Murashige and Skoog (MS) salts, 1X B5 Vitamin mix, 1.0% (w/v) sucrose, 0.5% (w/v) 2-(N-morpholino)-ethanesulfonic acid (MES) and 1% (w/v) Phyto agar, adjusted to pH 5.7 (Nagarajan et al., 2019). Seedlings germinated and grew on media for 14 days under long-day conditions of 16-h light/8-h dark at 21°C. For RNA extractions, 2-week-old seedlings were harvested directly from the germination media or rosette leaves were harvested from 6-8-week-old plants that had been transferred to soil and grown for an additional 4 weeks in a chamber under the same long-day conditions as described above. For plant transformation, 10-week-old plants were infiltrated with *Agrobacterium tumefaciens* (GV3101) containing the respective binary vectors using the floral dip method as described in Bent (2006). T1 seeds were selected on germination/selection media containing 0.5X MS salts, 1.0% (w/v) sucrose, 0.5% (w/v) MES and 0.7% (w/v) Phyto agar adjusted to pH 5.7 and supplemented with 20 mg/L hygromycin and 75 mg/L cefotaxime (Nagarajan et al., 2019). The seedlings were germinated under constant light for 7 hours before dark treatment for two days. Then seedlings were grown for 12 more days under long-day conditions of 16-h light/8-h dark at 21°C before transferring to soil. Seeds harvested from mature T1

plants were surface-sterilized, as previously described. After two days of stratification in the dark at 4°C, the surface-sterilized T2 seeds were plated on the same germination/selection media as previously described with the following modifications: 1.0% (w/v) Phyto agar and no cefotaxime. T2 seedlings were germinated as previously described and grew vertically for a total of 7-10 days before confocal imaging. Crosses to generate double homozygous mutants were performed as described in Weigel (2002). Briefly, pollen from a homozygous male parent was transferred to the emasculated pistil of a homozygous female parent. Heterozygous plants from the F1 seeds obtained from the cross were identified via genotyping PCR using the following primers: *mfl1* T-DNA insertion: 6961/1557; *AtMFL1* WT gene: 6960/6961; *xrn4* T-DNA insertion: 5100/1521; *AtXRN4* WT gene: 5100/5186. Heterozygous plants were allowed to self-fertilize, and homozygous plants in the F2 generation were identified via genotyping PCR as previously described. All primers used for genotyping are listed in Table 1.

### **Plasmid construction**

Sequence information about the *AtMFL* genes (Appendix A) was obtained from The Arabidopsis Information Resource (TAIR). All PCR reactions were carried out using Phusion Hot Start Flex DNA Polymerase (New England Biolabs, Ipswich, MA). All sub-cloning vectors and final vectors were transformed into XL-1 Blue supercompetent cells or DH5 $\alpha$  chemically competent cells. All DNA fragments were gel eluted using the NucleoSpin Gel and PCR Clean-up Kit (Macherey-Nagel, Düren, Germany). All plasmids were extracted using the NucleoSpin Plasmid Miniprep Kit (Macherey-Nagel, Düren, Germany).

### *Localization / complementation constructs*

The full-length coding region of *AtMFL1* was PCR amplified from WT cDNA using the primer pair 7033/7034. The purified PCR product was treated with GoTaq DNA polymerase to generate A overhangs and subcloned into pGEM-T Easy (Promega, Durham, NC, USA) according to the manufacturer's instructions. Subsequently, the *AtMFL1* ORF was released from the entry clone (p2619) by restriction digest using BsrGI and SpeI (New England Biolabs, Ipswich, MA) and ligated to the *35S::CFP* vector (p2622, built by Vinay Nagarajan) that had been cut with the same enzymes to generate the final vector *35S::CFP::AtMFL1* (p2624, built by Anna DiBattista and Vinay Nagarajan). The full-length coding region of *CFP* was PCR amplified from p2622 using the primer pair 7130/7131; the full-length coding region of *AtMFL2* was PCR amplified from WT cDNA using the primer pair 7132/7133; the purified PCR products with overlapping regions were fused together using the Gibson Assembly Master Mix (New England Biolabs, Ipswich, MA) according to the manufacturer's instructions with the following modifications: the reaction was incubated at 50°C for 30 minutes. The purified PCR product was subcloned into pGEM-T Easy (Promega, Durham, NC, USA) according to the manufacturer's instructions. Subsequently, the *CFP::AtMFL2* gene fusion was released from the entry clone (p2640) by restriction digest using AscI and PacI (New England Biolabs, Ipswich, MA) and ligated to the *35S::CFP* vector (p2622) that had been cut with the same enzymes to generate the final vector *35S::CFP::AtMFL2* (p2649). The full-length genomic region of *AtMFL3* was PCR amplified from WT genomic DNA using the primer pair 7128/7129 and the purified PCR product was subcloned into pGEM-T Easy (Promega, Durham, NC, USA) according to the manufacturer's instructions. Subsequently, the *AtMFL3* genomic region was released from the entry clone (p2639) by restriction digest using

BsrGI and SpeI (New England Biolabs, Ipswich, MA) and ligated to the *35S::CFP* vector (p2622) that had been cut with the same enzymes to generate the final vector *35S::CFP::AtMFL3* (p2642). All final vectors were verified by Sanger sequencing and transformed into *Agrobacterium tumifaciens* (GV3101) electrocompetent cells according to Weigel (2006) using the following settings: voltage: 2000 V; duration: 0.6 ms; number of pulses: 1; electrode gap: 1.0 mm. All primers used for cloning are listed in Table 1.

#### *AtMFL1 point mutation construct*

The 5' end of the coding region of *AtMFL1* was amplified from p2624 using the primer pair 7209/7212; the 3' end of the coding region of *AtMFL1* was PCR amplified from p2624 using the primer pair 7213/7214. The purified PCR products with overlapping regions were fused together using the Gibson Assembly Master Mix (New England Biolabs, Ipswich, MA) as previously described. The purified PCR product was subcloned into pGEM-T Easy (Promega, Durham, NC, USA) according to the manufacturer's instructions. Subsequently, the mutant *AtMFL1* coding region was released from the entry clone (p2673) by restriction digest using PspOMI and SpeI (New England Biolabs, Ipswich, MA) and ligated to *35S::CFP::AtMFL1* vector (p2624) that had been cut with the same enzymes to release WT *AtMFL1*. The final mutant construct (p2676) was verified by Sanger sequencing and transformed into *Agrobacterium tumifaciens* electrocompetent cells as previously described. All primers used for cloning are listed in Table 1.

### **Transient expression and subcellular localization**

Fluorescent fusion proteins were transiently expressed in fully expanded *Nicotiana benthamiana* leaves as described in Caplan et al. (2015) or stably expressed in *Arabidopsis thaliana* Col-0 via the floral dip method as previously described. Subcellular localizations were imaged in *N. benthamiana* leaves 3 days after infiltration or in root epidermal cells of 7-day-old *Arabidopsis* using a LSM880 confocal microscope (Zeiss) with a 40X objective using an excitation wavelength of 458 nm. All images were processed using ImageJ.

### **RNA analysis**

#### *Total RNA extraction and cDNA synthesis*

Total RNA was extracted from 2-week-old seedlings or 6-week-old plants using TRI reagent (Molecular Resource Center, Inc., Cincinnati, OH, USA) according to the manufacturer's instructions, with the addition of an acid phenol/chloroform clean up step at the end. The quality of the RNA was assessed using agarose gel electrophoresis and Nanodrop quantification. Total RNA (2 µg) was DNase-treated prior to oligo(dT) priming and was reverse transcribed using SuperScript II (Thermo Fisher Scientific, Waltham, MA, USA).

#### *Reverse transcription (RT)-PCR*

RT-PCR (30 cycles) was performed on Applied Biosystems MiniAmp Thermal Cycler (Thermo Fisher Scientific, Waltham, MA, USA) using the GoTaq Master Mix (Promega, Durham, NC, USA) and the following primers (0.4 µM, final concentration): *AtMFL1*- 7060/7061, *AtMFL2*- 7112/7113, *AtMFL3*- 7114/7115. Transcript levels were normalized to those of *AtACTIN2* (*AT3G18780*) using the

primer pair 7027/7028. All RT-PCR experiments were validated using three or more biological replicates as per the corresponding figure legends. All primers used for RT-PCR are listed in Table 1.

#### *Quantitative (q)-PCR*

qPCR reactions contained cDNA diluted 1:10, 2X Premix Ex Taq SYBR Green (Clontech Inc.) and the following primers (0.5  $\mu$ M, final concentration): *AtMFL1*- 7060/6959, *AtMFL2*- 7278/7151, *AtMFL3*- 7114/7279, *AtHSPRO2*- 7274/7276, *AtOXS3*- 7280/7282. Relative levels were computed by the  $2^{-\Delta\Delta C_t}$  method of quantification (Livak and Schmittgen, 2001) and normalized to either *AtUBC* (*AT5G25760*) using the primer pair 5095/5096 or *AtACTIN2* using the primer pair 5329/5330. All qPCR experiments were performed with three technical replicates. The qPCR experiments to validate the RNA-seq data were performed using two biological replicates as per the corresponding figure legends. All primers used for qPCR are listed in Table 1.

#### *Northern blots*

Total RNA (15-20  $\mu$ g) was resolved on a MOPS-formaldehyde denaturing gels solidified with 1.5% or 1.8% (w/v) agarose and blotted onto Hybond N+ membranes. Probes were amplified from WT cDNA via PCR using the following primer pairs: *AT2G43020*- 7284/7285; *AT4G00780*- 7300/7301; *AT5G11580*- 7304/7305. The purified PCR products were radiolabeled using the Thermo Scientific DecaLabel DNA Labeling Kit according to the manufacturer's instructions. Hybridization occurred by incubating the membranes in PerfectHyb Plus Hybridization Buffer (Sigma-Aldrich, St. Louis, MO, USA) with the  $^{32}$ P-dCTP labeled DNA probes at 68°C overnight. All

membranes were washed with  $2\times$  SSC, 0.1% SDS for 10 min at 68°C followed by one or two washes of  $0.2\times$  SSC, 0.1% SDS for 10 min at 68°C. Signal intensities were analyzed using the Typhoon system (GE Health Sciences). Membranes were stripped at least twice in boiling 0.1% SDS for 20 min between hybridizations as described in Nagarajan et al. (2019). All northern blot results presented are representative of two biological experiments. All primers used for making probes are listed in Table 1. All images were processed using ImageJ.

### *5' RACE*

A modified RNA ligase-mediated (RLM) 5'RACE was performed using the FirstChoice RLM-RACE kit (Thermo Fisher Scientific) as described in Nagarajan et al (2019). Briefly, DNase-treated total RNA (10  $\mu$ g) was ligated to a 5' RNA adapter (GUUCAGAGUUCUACAGUCCGAC) and reverse transcribed with oligo(dT). After an initial round of PCR amplification (35 cycles) of the cDNA using primers 4402/7285, a second round of PCR amplification (30 cycles) was performed on the resulting dsDNA with internal primers 4404/7289. All primers used for 5' RACE are listed in Table 1.

### **RNA-seq library construction**

PolyA+ RNA was fractionated from seedling total RNA (10  $\mu$ g) using a standard RNA magnetic bead-based oligo(dT) purification. RNA-seq libraries were generated using the NEBNext Ultra II RNA Library Prep Kit (New England Biolabs) following the manufacturer's instructions. Size distribution and concentration of the libraries were estimated using the Fragment Analyzer Automated CE System (Agilent

Technologies, Inc.). Libraries were multiplexed and sequenced as 75 nt single-end reads on Illumina Next-Seq 550.

## Computational approaches

### *Phylogenetic analysis of MFL1-like proteins in plants*

A list of plant species spanning a large evolutionary time scale was generated with the help of Renate Wuersig (Ph.D., Adjunct Assistant Professor, University of Delaware). A DELTA BLAST search was performed of AtMFL1 against all proteins in each of the following plants species: *Chlamydomonas reinhardtii*, *Physcomitrella paten*, *Amborella trichopoda*, *Cinnamomum micranthum*, *Nymphaea colorata*, *Brachypodium distachyon*, *Oryza sativa*, *Zea mays*, *Ananas comosus*, *Musa acuminata*, *Capsella rubella*, *Medicago truncatula*, *Cannabis sativa*, *Solanum lycopersicum*, *Solanum tuberosum*, *Gossypium raimondii*, *Vitis vinifera*, *Phaseolus vulgaris*, *Glycine max*, and *Arabidopsis thaliana*. The domain architecture of the most related protein in each species was determined using the SMART algorithm (<http://smart.embl-heidelberg.de/>). A multiple sequence alignment of the NYN domains of all the AtMFL1-like proteins (as well as all the AtMFL proteins and HsMARF1 and MmMARF1) was performed using the BLOSUM62 algorithm within MUSCLE (default parameters) (Madeira et al., 2019). The residues required for ssRNase activity as determined by Yao et al. (2018) were subsequently annotated within this alignment.

### *RNA-seq analysis*

RNA-seq reads were checked for quality using FastQC v0.11.9 (<http://www.bioinformatics.babraham.ac.uk/projects/fastqc/>). The reads were mapped

to the Arabidopsis TAIR10 reference genome using TOPHAT v2.1.1 (<http://ccb.jhu.edu/software/tophat>) with following parameters: *library type, fr-firststrand; transcript features and junction coordinates from the TAIR10 GTF*. Finally, transcript abundance levels were compared to determine differential expression using Cufflinks/CuffDiff v2.2.1 (<https://github.com/cole-trapnell-lab/cufflinks/>). The criteria for differentially accumulating nuclear-encoded protein-coding transcripts and non-coding RNA were as follows: RPKM  $\geq 2$  in mutant or WT; fold change:  $\geq 1.5$  in either direction ( $\log_2 FC \geq 0.58$ ); and False Discovery Rate (FDR)-adjusted  $P \leq 0.05$ . Three replicates of each library were used in the overall analysis. The DAVID program (<https://david.abcc.ncifcrf.gov/>) was used to identify enriched gene ontology (GO) categories upregulated in the *mfl1* mutants compared to Col-0. The results were filtered to include only those categories with a Benjamini-Hochberg adjusted p-value of  $\leq 0.05$ .

**Table 1 Primers utilized**

| Primer Name        | Sequence (5' to 3')                  | Product Size (nt) |
|--------------------|--------------------------------------|-------------------|
| Genotyping primers |                                      |                   |
| 6960               | GCAAGGAGCTTACATTGCTTG                | 1136              |
| 6961               | GCGACTATCGTTCTCGTATCG                |                   |
| 5100               | GGAAATGGCTTTATATCTACTGACG            | 487               |
| 5186               | AGTTGATGACTGATCCCTCATCC              |                   |
| 1557               | TGGTTCACGTAGTGGGCCATCG               |                   |
| 1521               | TAGCATCTGAATTTTCATAACCAATCTCGATACAC  |                   |
| Cloning primers    |                                      |                   |
| 7033               | GCTGTACAAGgggcccATGATACAAAACGCTATGTC | 1492              |
| 7034               | actagtTTAAACCGGGCTGATGAGTTTC         |                   |
| 7130               | tggcgcgccATGGTGAGCAAGGGCGAGGAGCT     | 768               |

|                 |   |      |
|-----------------|---|------|
| 7131            | CTCGCCGAGTTCGTCGACATcctgtacagctcgtccatgccg  |      |
| 7132            | gcatggacgagctgtacaagATGTCGACGAACTCGGCGAGT   | 2052 |
| 7133            | ccttaattaaTCAGGCAACGGTTTGGATACCTGG          |      |
| 7128            | ggtgtacaagATGAATACGGCGGGGAAAGAAGATTTG       | 1222 |
| 7129            | ggactagtTTAATTTGAGTTCACAGTTGCAAGTATCA       |      |
| 7209            | TTGgggcccATGATACAAAACGCTATGTC               | 498  |
| 7212            | CAATATATGAAGCGCAGGCGCGAAATCAACAgCACCCGATAC  |      |
| 7213            | CCGCCTGCGACTATCGTTCTCGTATCGGGTGcTGTTGATTCGC | 1065 |
| 7214            | GGGAGGCCTGGATCGactagtTTAAACCGGG             |      |
| PCR primers     |   |      |
| 7060            | ACCGAACATTATCTCTCCTTCTTC                    | 406  |
| 7061            | AAACGCACCGTCTGATCTTAT                       |      |
| 7112            | GAAGAAGATGTAGTAGCAGCCATAA                   | 519  |
| 7113            | GCTTGAAGGTTTCAGAATAAGACATAC                 |      |
| 7114            | GTCTGAGACGGTACAATATTCTCTT                   | 534  |
| 7115            | TGAGTTGATTAAGTTCTCCGTAGTC                   |      |
| 7027            | TGCCAATCTACGAGGGTTTC                        | 446  |
| 7028            | GTCAGCGATACCTGAGAACATAG                     |      |
| qPCR primers    |   |      |
| 7060            | ACCGAACATTATCTCTCCTTCTTC                    | 95   |
| 6959            | AGACTCCTCCATTGGTAGTTACC                     |      |
| 7278            | CCATCTGAGTATGTTCAAGGCC                      | 95   |
| 7151            | GAGATATTTGGCTCGGTAGGC                       |      |
| 7114            | GTCTGAGACGGTACAATATTCTCTT                   | 102  |
| 7279            | CGGCTAATAGACTTGTCCAAAGC                     |      |
| 7274            | GATAACGAGATGAGTCCGGTGTAAG                   | 103  |
| 7276            | TCTCGTTTGGTCGGCAAAT                         |      |
| 7280            | GATCCAACAAGGATCCAAGAAGA                     | 102  |
| 7282            | GGAAGAAGATAATGAACAAGATGATGAG                |      |
| 5329            | CAGGTATCGCTGACCGTATGAG                      | 146  |
| 5330            | CATCTGCTGGAATGTGCTGAGG                      |      |
| Northern probes |   |      |
| 7284            | GGAGTGAAGGTAACGACAGAAA                      | 600  |
| 7285            | CAGAACTCGCATCCTACAATCT                      |      |
| 7300            | TGTTGGAGAATGTACCTACTGAGACTTGGA              | 150  |
| 7301            | AACTAATTGCAAGTCTGACTTTGATGAC                |      |
| 7304            | GGATTAGGATTATGCCCCGATGTCAA                  | 490  |

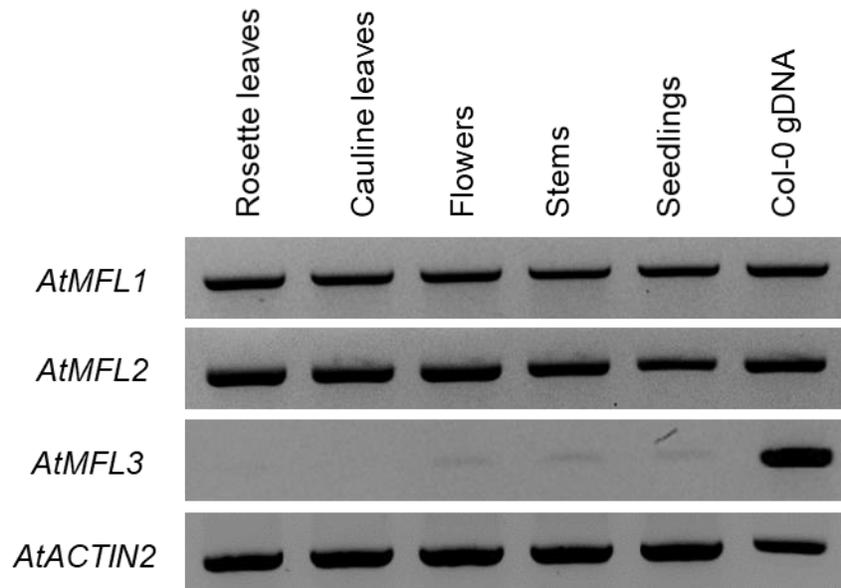
|                 |   |     |
|-----------------|---|-----|
| 7305            | ATCAGCTGACTCTTATCAGCACTCTCCA                  |     |
| 5' RACE primers |   |     |
| 4402            | G TTCAGAGTTCTACAGTCCGAC                       | 706 |
| 7285            | CAGAACTCGCATCCTACAATCT                        |     |
| 4404            | AATGATACGGCGACCACCGACAGG TTCAGAGTTCTACAGTCCGA | 251 |
| 7289            | CCTAGGTCGTTGATTGCTTCT                         |     |

## Chapter 3

### RESULTS

#### Organ-specific expression of the *AtMFL* genes

To determine whether the *AtMFL* genes are reliably expressed, I looked at their relative expression levels in different above ground organs of Arabidopsis. We postulated that if the AtMFL endoribonucleases are involved in global gene regulation in plants, then their expression levels should be detectable in all major organs. Total RNA was extracted from rosette leaves, cauline leaves, flowers, stems, and seedlings of Col-0 (WT) plants, and RT-PCR was performed. The results indicate that *AtMFL1* and *AtMFL2* are reliably expressed in the Arabidopsis organs examined, but *AtMFL3* is not (Fig. 6). In particular, *AtMFL3* seems to be minimally expressed in the flowers, stems, and seedlings. For this reason, I decided that I would not pursue this gene in downstream transcriptome-wide analyses. Further literature search confirmed that mutations resulting in a full knockout of *AtMFL3* are embryonic lethal, and therefore *AtMFL3* has also been termed *EMBRYO SAC DEVELOPMENT ARREST 32 (EDA32)* (Pagnussat et al., 2004).

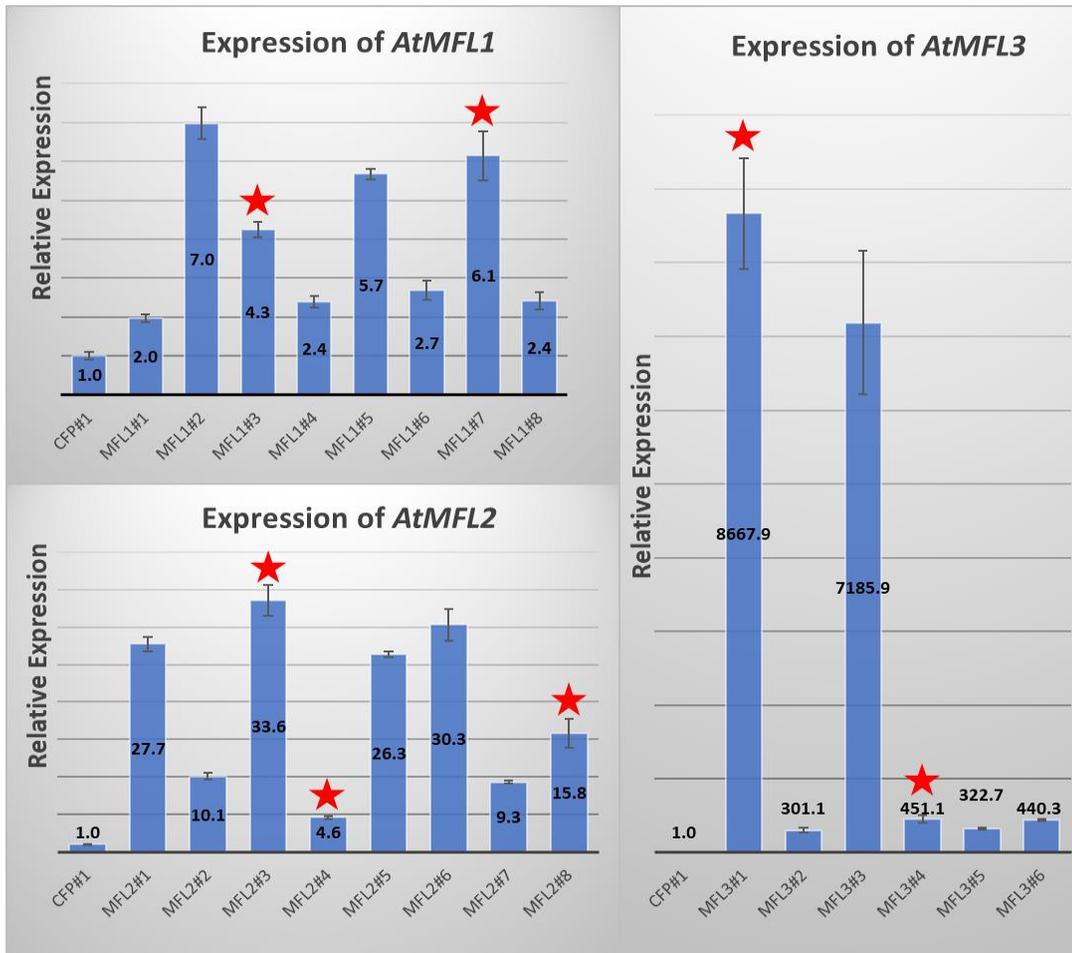


**Figure 6** *AtMFL1*, *AtMFL2*, and *AtMFL3* expression in various Col-0 organs RT-PCR analysis of *AtMFL1* (top), *AtMFL2* (second from top), and *AtMFL3* (second from bottom) using pooled RNA isolated from various organs of WT individuals (Col-0). *AtACTIN2* (bottom) is a loading control and shows uniformity of cDNA synthesis Results are representative of three biological replicates.

### Subcellular localization of the AtMFL proteins

The majority of mRNA decay, specifically NMD, occurs in the cytosol of the cell. It is therefore reasonable to assume that if the AtMFL proteins are playing a role mRNA decay, then they should localize to the cytosol along with AtXRN4, which would degrade the 3' fragments resulting from AtMFL cleavage. AtMFL1 has been previously shown to localize to the cytosol (Chicois et al., 2018). However, the subcellular location of AtMFL2 and AtMFL3 are unknown. To determine the sites of AtMFL action, the full-length coding regions of *AtMFL1* and *AtMFL2* and the genomic region of *AtMFL3* were fused in-frame at the C-terminal end of cyan

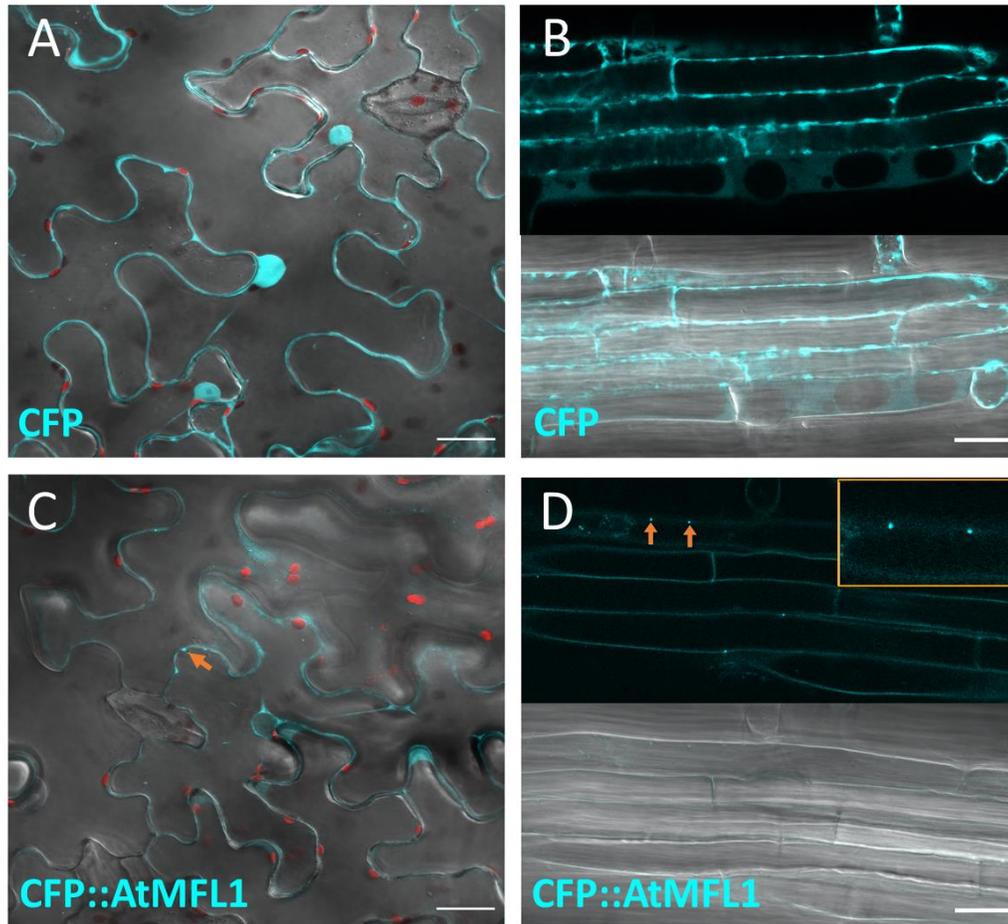
fluorescent protein (CFP) under the control of the constitutive 35S promoter from cauliflower mosaic virus. The coding region of *AtMFL3* was not used to generate the *35S::CFP::AtMFL3* construct because when sub-cloned and transformed into chemically competent *E. coli*, the bacterial cells did not grow. Because of this, I hypothesized that *AtMFL3* could be cytotoxic to bacteria. Therefore, the genomic region of *AtMFL3* was used to generate the *35S::CFP::AtMFL3* construct because the presence of its single intron would hinder bacterial transcription and therefore prevent translation of a potentially cytotoxic protein. Each of the *35S::CFP::AtMFL* constructs as well as the *35S::CFP* control were transformed into *Agrobacterium tumefaciens* and then transiently transformed into *Nicotiana benthamiana* (tobacco) leaves and stably transformed into Arabidopsis. Several T1 plants that arose from the stable transformation were screened for relative expression of the *AtMFL* genes via qPCR (Fig. 7). The results indicated that all of the T1 plants were expressing the respective *AtMFL* genes and therefore the CFP::AtMFL fusion proteins at higher levels than the control (CFP#1). Several T1 plants with varying *CFP::AtMFL* expression were chosen for T2 imaging so that the localization result could be shown in multiple independent lines (Fig. 7). As expected, CFP was uniformly expressed in the cytosol and nucleus of tobacco leaf cells (Fig. 8A, 9A, and 10A) as well as Arabidopsis root cells (Fig. 8B and 10B).



**Figure 7 Elevated mRNA levels of *AtMFL* genes in plants expressing *CFP::AtMFL* fusion constructs**

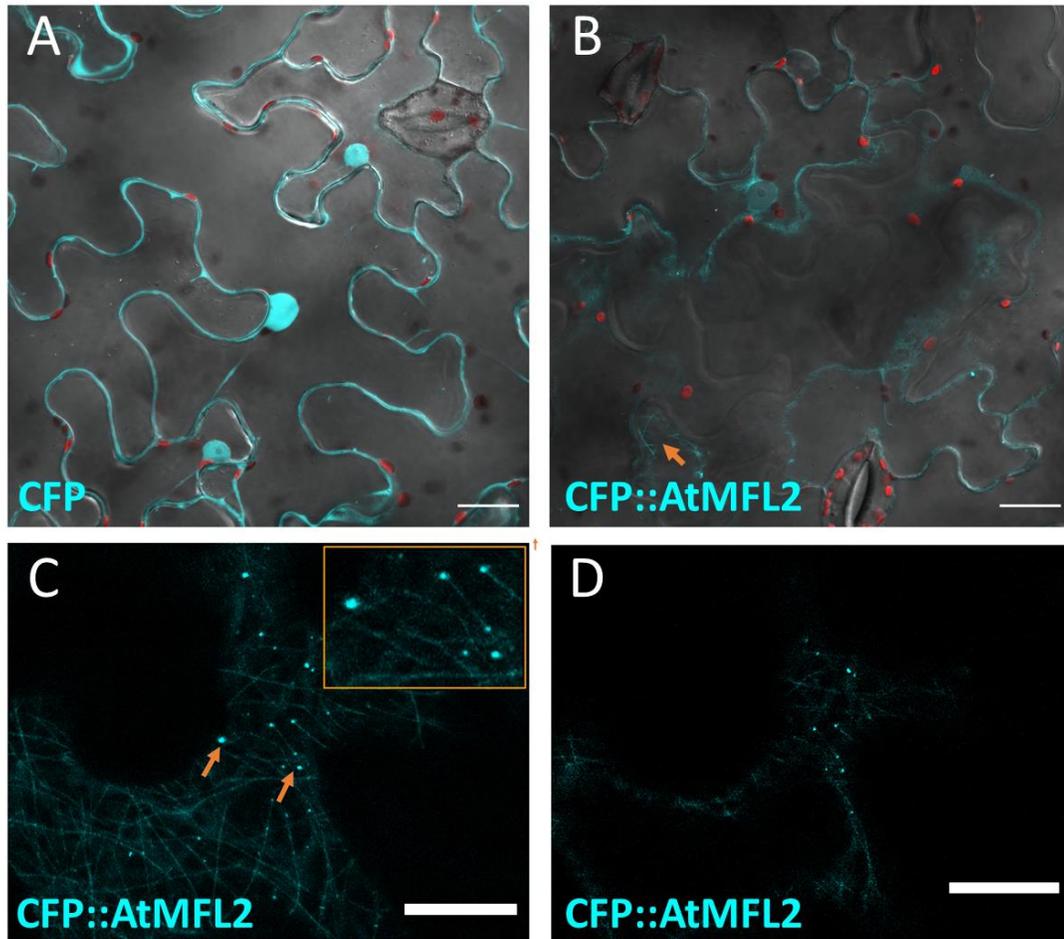
qPCR shows relative levels of *AtMFL1*, *AtMFL2*, and *AtMFL3* in the rosette leaves of 6-week-old plants expressing the respective *CFP::AtMFL* fusion constructs. Levels were normalized to *AtUBC* and bars are mean values + SD from three technical replicates. WT (CFP#1) expression levels are set to 1. Red stars indicate T1 plants whose progeny were used for T2 imaging.

Analysis of transiently transformed tobacco leaves and stably transformed Arabidopsis roots showed that AtMFL1 localizes to the cytosol, particularly to distinct foci (Fig. 8), thus confirming the result of Chicois et al. (2018).



**Figure 8 AtMFL1 localizes to distinct foci in the cytosol**  
(A) Tobacco leaf cells transiently expressing CFP. (B) Arabidopsis root cells stably expressing CFP (top; fluorescence; bottom; fluorescence and DIC). (C) Tobacco leaf cells transiently expressing CFP::AtMFL1. (D) Arabidopsis root cells stably expressing CFP::AtMFL1 (top; fluorescence; bottom; fluorescence and DIC). The orange arrows point to CFP::AtMFL1 in cytosolic foci. Expression was imaged on an LSM880 multiphoton confocal microscope. Images show a single slice from a z-stack maximum intensity projection. Scale bars represent 25  $\mu\text{m}$ . Red autofluorescence is from intact chloroplasts. Images in B and D are representative of two independent lines (indicated in Fig. 7). Imaging collaborators: Kody Seward, Tim Chaya, and Jeffrey Caplan.

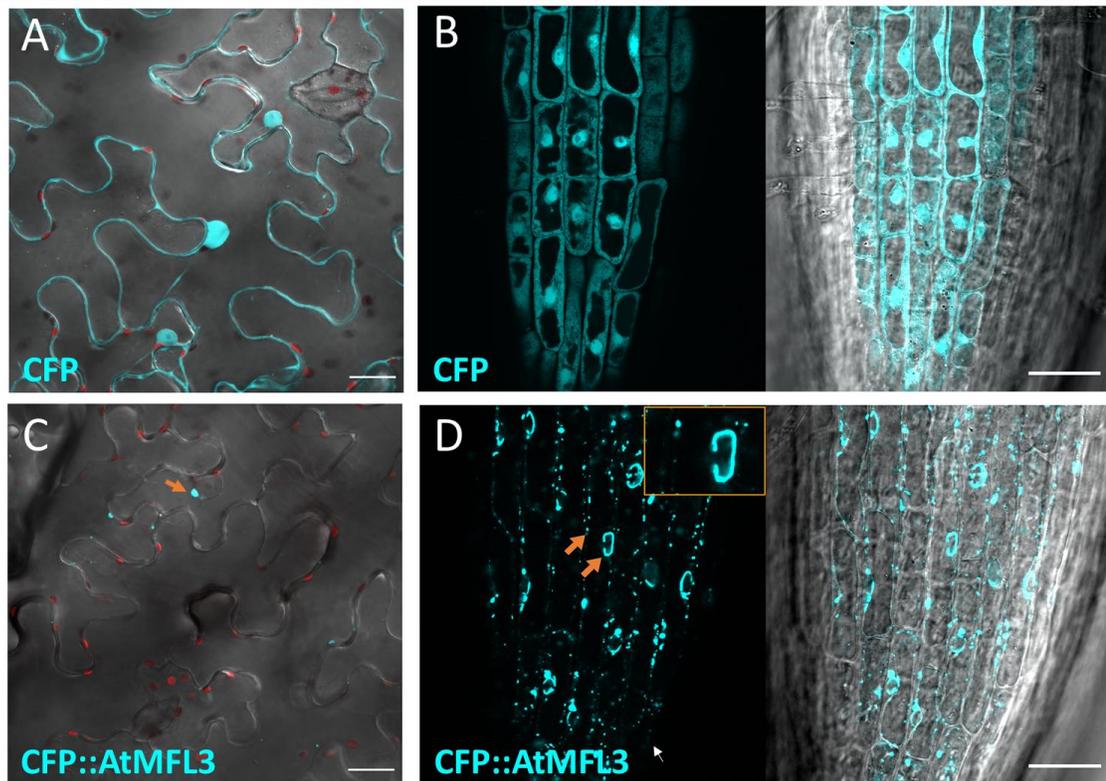
Analysis of transiently transformed tobacco leaves suggested that AtMFL2 localizes to the cytosol and cytoskeleton (Fig. 9). In particular, AtMFL2 seems to form protein complexes that move along the cytoskeletal filaments. However, I was unable to validate this result in stably transformed Arabidopsis because CFP fluorescence was not detectable in the progeny from any of the chosen T1 lines (Fig. 7) even though they were strongly expressing *CFP::AtMFL2*. Lack of AtMFL2 expression could be due to post-transcriptional gene silencing (PTGS), which is a common mechanism of gene regulation in plants. To address this hypothesis, Arabidopsis *rdr6* mutants will be stably transformed with the *35S::CFP::AtMFL2* construct, and their progeny will be imaged as previously described. AtRDR6 is an RNA-dependent RNA polymerase that plays a critical role in generating a dsRNA template for the production of siRNAs that target sense transgenes and other unwanted RNAs (Luo and Chen, 2007). The lack of AtRDR6 would greatly diminish PTGS, so CFP fluorescence may be detectable in *rdr6* mutants expressing the *35S::CFP::AtMFL2* construct.



**Figure 9 AtMFL2 localizes to the cytosol and cytoskeleton**  
 (A) Tobacco leaf cells transiently expressing CFP. (B-D) Tobacco leaf cells transiently expressing CFP::AtMFL2. The orange arrows point to CFP::AtMFL2 in cytosolic foci and on the cytoskeleton. Microscopy was performed as in Figure 8. Scale bars represent 25  $\mu\text{m}$ . Red autofluorescence is from intact chloroplasts. Imaging collaborators: Kody Seward, Tim Chaya, and Jeffrey Caplan.

Analysis of transiently transformed tobacco leaves and stably transformed Arabidopsis roots suggested that AtMFL3 localizes to the nuclear membrane and forms aggregates within the cytosol (Fig. 10). AtMFL3 has a transmembrane helical

region at its C-terminal end, which explains its apparent nuclear membrane localization. Interestingly, AtMFL3 does not seem to localize around the entire nuclear membrane, which could be due to physical barriers such as the nucleolus or the endoplasmic reticulum. This could be tested by colocalizing AtMFL3 with a particular marker gene and/or using specific stains for these structures. Considering that AtMFL3 is lowly expressed in Arabidopsis and appears to be cytotoxic in bacteria, overexpression of this protein in plants might cause the cells to form aggregates of AtMFL3 for rapid degradation.



**Figure 10 AtMFL3 localizes to the nuclear membrane and forms aggregates in the cytosol**

(A) Tobacco leaf cells transiently expressing CFP. (B) Arabidopsis root cells stably expressing CFP (left; fluorescence; right; fluorescence and DIC). (C) Tobacco leaf cells transiently expressing CFP::AtMFL3. (D) Arabidopsis root cells stably expressing CFP::AtMFL3 (left; fluorescence; right; fluorescence and DIC). The

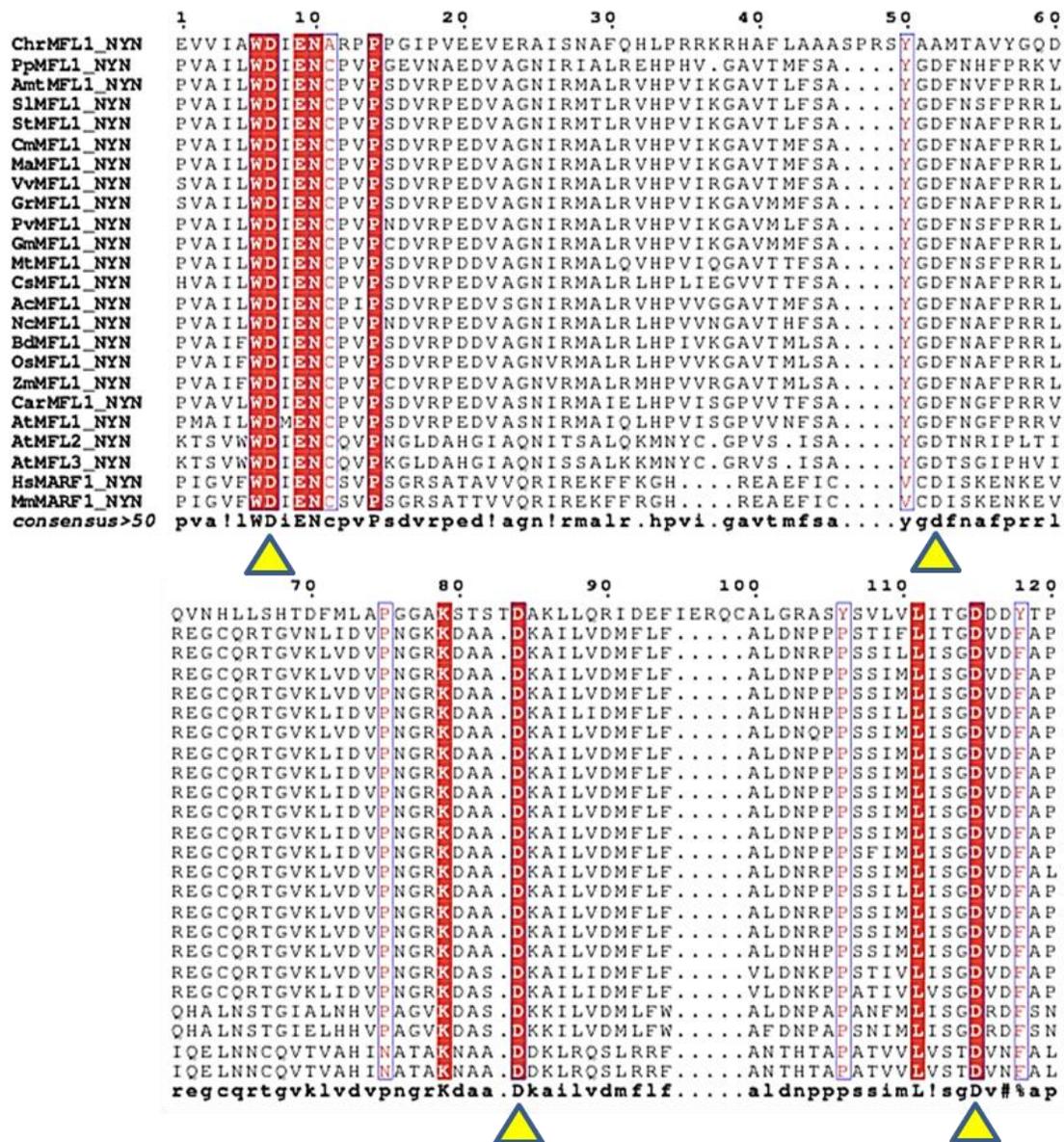
orange arrows point to CFP::AtMFL3 on the nuclear membrane and within cytosolic aggregates. Microscopy was performed as in Figure 8. Scale bars represent 25  $\mu\text{m}$ . Red autofluorescence is from intact chloroplasts. Images in B and D are representative of two independent lines (indicated in Fig. 7). Imaging collaborators: Kody Seward, Tim Chaya, and Jeffrey Caplan.

### **AtMFL1-like proteins are conserved across many plant species**

Because MARF1-like proteins have not been extensively studied in plants, I wanted to determine the conservation of AtMFL1-like proteins across different plant species. For this purpose, a DELTA BLAST search was performed to find AtMFL1-like proteins in various plant species spanning a large evolutionary time scale. According to the NCBI database, the following plant species encode an AtMFL1 ortholog: *C. reinhardtii* (algae), *P. patens* (moss); *A. trichopoda*, *C. micranthum*, *N. colorata* (early diverging angiosperms); *B. distachyon*, *O. sativa*, *Z. mays*, *A. comosus*, *M. acuminata* (monocots); *C. rubella*, *M. truncatula*, *C. sativa*, *S. lycopersicum*, *S. tuberosum*, *G. raimondii*, *V. vinifera*, *P. vulgaris*, *G. max*, and *A. thaliana* (eudicots). On an evolutionary timescale, green algae diverged about one billion years ago, mosses diverged about 600 million years ago, and all flowering plants diverged about 125 million years ago. The domain architectures of the AtMFL1-like proteins were determined using the SMART algorithm. Based on the results, all of the AtMFL1-like proteins examined have an N-terminal NYN domain and two C-terminal LOTUS/OST-HTH domains, with the exception of *C. reinhardtii* MFL1 (*ChrMFL1*) which does not contain any LOTUS domains (Appendix B). To determine whether these proteins likely exhibit endoribonucleolytic function, their NYN domains were aligned using the BLOSUM62 algorithm within MUSCLE (default parameters) (Madeira et al., 2019). Based on the results, all of the AtMFL1-like proteins examined contain the four conserved aspartic acid residues shown in Yao et al. (2018) to be

required for nuclease activity, with the exception of ChrMFL1 which has an alanine residue at the second site (Fig. 11). Since *C. reinhardtii* was the most divergent plant species examined, it seems that ChrMFL1 lost its catalytic activity while the other plant species did not.

According to the NCBI database, several different plant species across the previously mentioned lineages do not seem to encode an AtMFL1 ortholog. Additionally, several gymnosperms, cycads, and ferns were examined, but none of them seem to encode an an ortholog of AtMFL1 either. This could be due to limitations of the method used and/or gaps or errors in the available genome sequences of these plants. Overall, this analysis shows that AtMFL-like proteins are well-conserved across a range of plant lineages, though they may not be present in all plant species. Importantly, it seems that these proteins have retained catalytically active NYN domains over the course of millions of years of evolution.



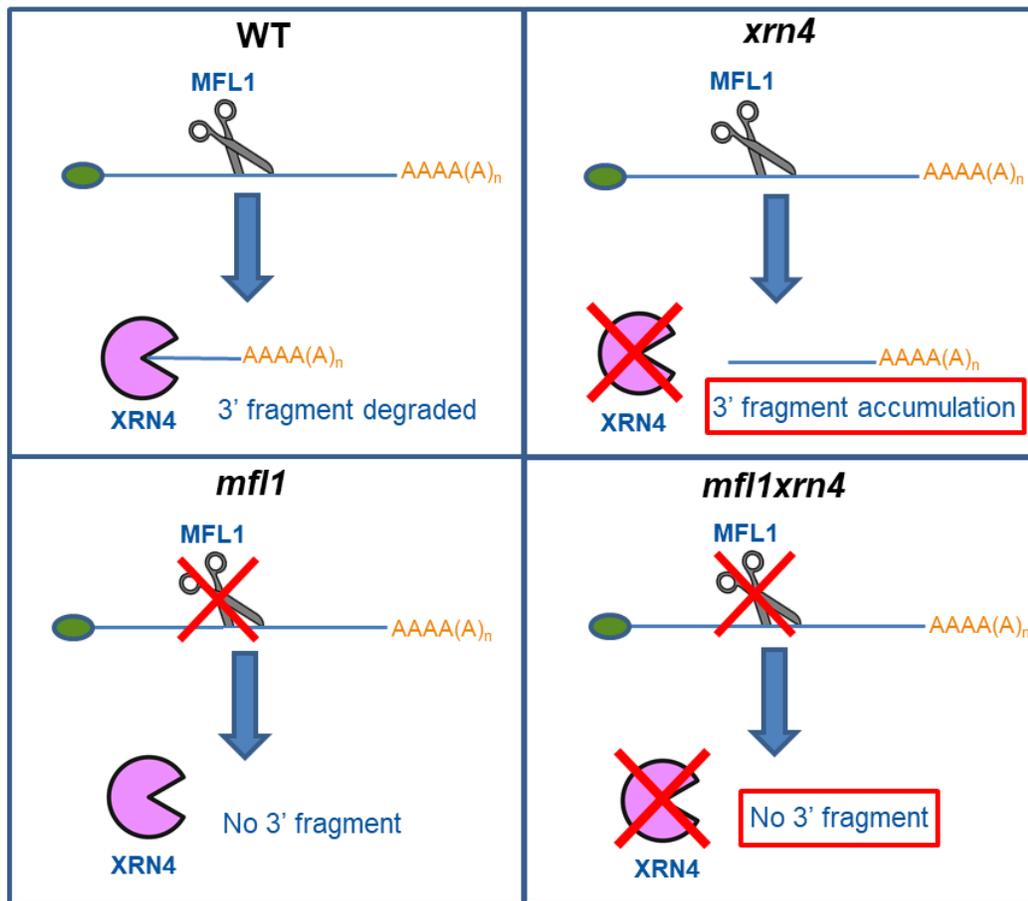
**Figure 11 Multiple sequence alignment showing conservation of catalytic residues in various plant MFL1-like proteins, AtMFL proteins, and metazoan MARF1**

Amino acid sequence of the NYN domains of MFL1-like proteins, AtMFL proteins, and metazoan MARF1 were aligned using the BLOSUM62 algorithm within MUSCLE (default parameters) (Madeira et al., 2019). Conserved amino acid residues are highlighted in red. Residues required for ssRNase activity as shown by Yao et al. (2018) are indicated by yellow triangles. Algae (Chr, *Chlamydomonas reinhardtii*),

Moss (Pp, Physcomitrella patens), Early diverging angiosperms (Amt, Amborella trichopoda; Cm, Cinnamomum micranthum; Nc, Nymphaea colorata), Monocots (Bd, Brachypodium distachyon; Os, Oryza sativa; Zm, Zea mays; Ac, Ananas comosus; Ma, Musa acuminata), Eudicots (Cr, Capsella rubella; Mt, Medicago truncatula; Cs, Cannabis sativa; Sl, Solanum lycopersicum; St, Solanum tuberosum; Gr, Gossypium raimondii; Vv, Vitis vinifera; Pv, Phaseolus vulgaris; Gm, Glycine max; At, Arabidopsis thaliana); Hs, human; Mm, mouse. ! is either I or V; \$ is either L or M; % is either F or Y; # is either N, D, Q, E, D, or Z.

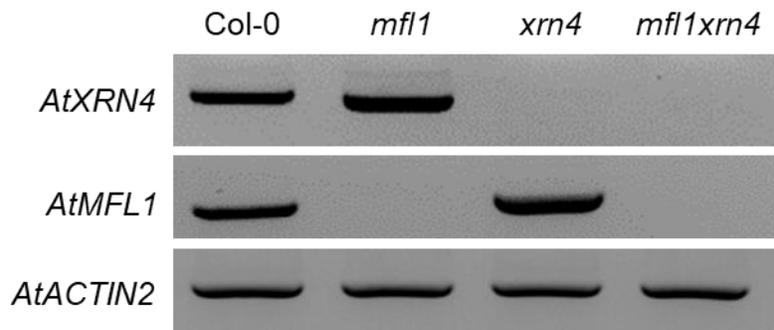
### ***mfl1* and *mfl1xrn4* are RNA null**

Based on the results of Chicois et al. (2018), Nagarajan et al. (2019), and data presented here, I narrowed the scope of my project to focus on AtMFL1. In WT cells, we hypothesize that AtMFL1 cleaves certain mRNA transcripts, and the 3' fragments resulting from this cleavage are degraded by AtXRN4. It has been previously shown that many different 3' fragments accumulate in an *xrn4* mutant (Nagarajan et al., 2019). Therefore, to identify putative targets of AtMFL1, we needed to be able to compare the 3' fragments that accumulate in cells lacking AtXRN4 to the 3' fragments that accumulate in cells lacking both AtMFL1 and AtXRN4 (Fig. 12). Any 3' fragments present in *xrn4* mutants and absent in *mfl1xrn4* double mutants are presumed to be AtMFL1-dependent. To this end, *mfl1* homozygous mutants were crossed to *xrn4* homozygous mutants to generate *mfl1xrn4* double homozygous mutants. The *mfl1* mutants have a T-DNA insertion in exon 1 (position 6788913 of chromosome 2) which should disrupt the expression of *AtMFL1*. The *xrn4* mutants have a T-DNA insertion in exon 18 which has already been shown to disrupt the expression of *AtXRN4* (Souret et al., 2004). To verify that the mutant lines no longer express the respective genes, total RNA was extracted from rosette leaves of Col-0, *mfl1*, *xrn4*, and *mfl1xrn4*, and RT-PCR was performed. The results indicate that *mfl1*, *xrn4*, and *mfl1xrn4* are RNA null and can therefore be used for downstream analyses (Fig. 13).



**Figure 12 Strategy for identifying targets of AtMFL1**

In WT cells, AtMFL1 cleaves certain transcripts and AtXRN4 degrades the 3' fragments produced by this cleavage. In cells lacking AtXRN4, 3' fragments from many transcripts over-accumulate. In cells lacking AtMFL1, certain transcripts are no longer cleaved, but the 3' fragments that result from this cleavage cannot be identified because AtXRN4 is still present. In cells lacking both AtMFL1 and AtXRN4, certain transcripts are no longer cleaved and the 3' fragments that result from this cleavage can be identified because AtXRN4 is absent. The 3' fragments that are present in *xrn4* and absent in *mfl1xrn4* are predicted to be targets of AtMFL1.



**Figure 13 *mfl1*, *xrn4*, and *mfl1xrn4* are RNA null**

RT-PCR analysis of *AtXRN4* (top) and *AtMFL1* (middle) using pooled RNA isolated from WT individuals (Col-0), *mfl1* homozygous mutants, *xrn4* homozygous mutants, and *mfl1xrn4* double homozygous mutants. *AtACTIN2* (bottom) is a loading control and shows uniformity of cDNA synthesis. Results are representative of three biological replicates.

**Certain transcripts are differentially expressed in *mfl1* mutants**

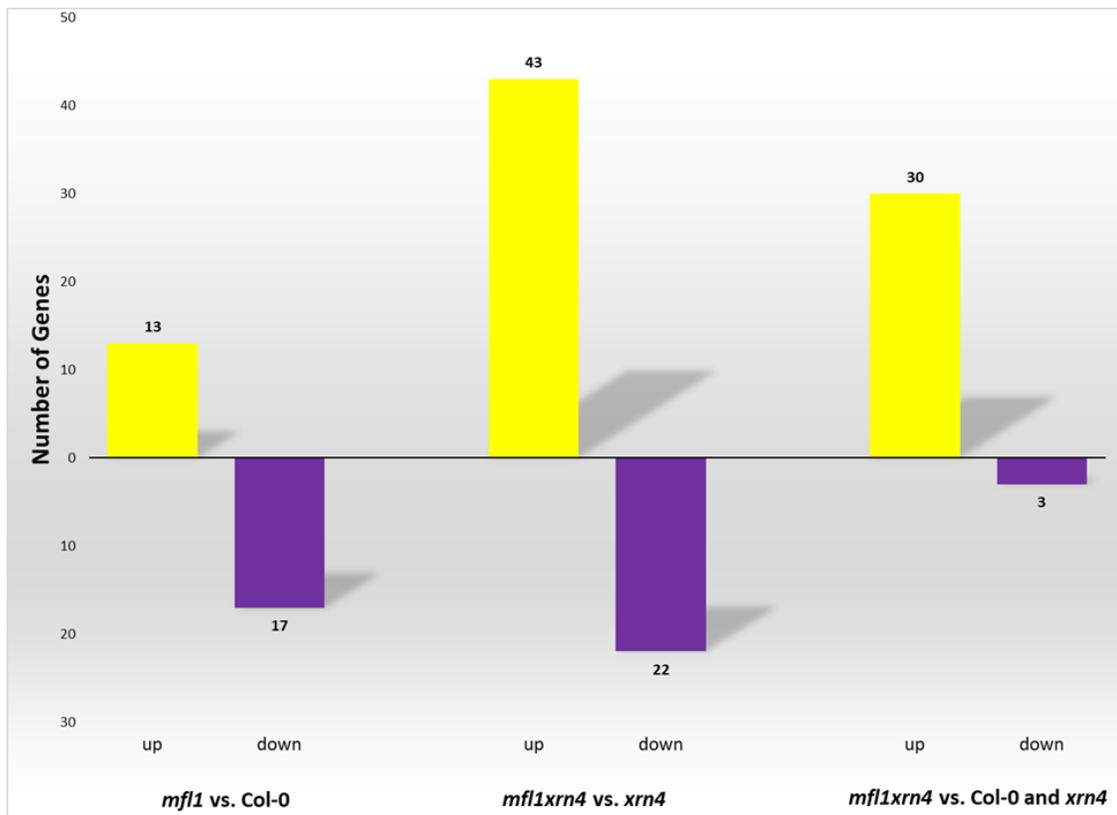
To better understand the impact of AtMFL1 on gene expression in seedlings, transcript abundance was investigated on a global scale. Strand-specific RNA-seq libraries were made from polyA<sup>+</sup> (polyadenylated) RNA fractionated from the total RNA of Col-0 (WT), *mfl1*, *xrn4*, and *mfl1xrn4* seedlings. Though the results of this analysis cannot confirm AtMFL1 substrates directly, any transcripts found to be over-accumulating in *mfl1xrn4* mutants compared to *xrn4* mutants could be investigated as putative targets of AtMFL1. Additionally, any transcripts found to be significantly differentially expressed in *mfl1* mutants compared to WT plants would indicate that AtMFL1 plays a role in gene expression. The RNA-seq data were analyzed using the Tuxedo suite (Trapnell et al., 2012), and the alignment statistics indicated that the libraries were of high quality due to their high depth of coverage and high percentage of genome-matched reads (Table 2).

**Table 2 Alignment statistics of RNA-seq libraries made from Arabidopsis seedlings**

| Library Name | Sample          | RNA    | Replicate | Length (nt) | Total Reads | Genome-matched reads | Percent Genome match |
|--------------|-----------------|--------|-----------|-------------|-------------|----------------------|----------------------|
| ATH1046      | Col-0           | PolyA+ | 1         | 80          | 41513265    | 39541854             | 95.30%               |
| ATH1047      | <i>mfl1</i>     | PolyA+ | 1         | 80          | 44749000    | 42693119             | 95.40%               |
| ATH1048      | <i>xrn4</i>     | PolyA+ | 1         | 80          | 39186115    | 37134824             | 94.80%               |
| ATH1049      | <i>mfl1xrn4</i> | PolyA+ | 1         | 80          | 41876730    | 39853410             | 95.20%               |
| ATH1050      | Col-0           | PolyA+ | 2         | 80          | 46397807    | 44223338             | 95.30%               |
| ATH1051      | <i>mfl1</i>     | PolyA+ | 2         | 80          | 44733523    | 42498788             | 95.00%               |
| ATH1052      | <i>xrn4</i>     | PolyA+ | 2         | 80          | 50500043    | 48042392             | 95.10%               |
| ATH1053      | <i>mfl1xrn4</i> | PolyA+ | 2         | 80          | 45254280    | 43038091             | 95.10%               |
| ATH1054      | Col-0           | PolyA+ | 3         | 80          | 42072897    | 40121339             | 95.40%               |
| ATH1055      | <i>mfl1</i>     | PolyA+ | 3         | 80          | 38719256    | 36987712             | 95.50%               |
| ATH1056      | <i>xrn4</i>     | PolyA+ | 3         | 80          | 65419868    | 62066672             | 94.90%               |
| ATH1057      | <i>mfl1xrn4</i> | PolyA+ | 3         | 80          | 46505424    | 44348533             | 95.40%               |

Differentially accumulating transcripts in *mfl1* compared to the WT and in *mfl1xrn4* compared to *xrn4* were identified. Because very few transcripts deviated from WT expression levels, the  $\log_2$  FC threshold was lowered to 0.6 (1.5 fold change) to account for more modest changes. In *mfl1*, 13 transcripts were significantly upregulated ( $\log_2$  FC  $\geq$  0.6) compared to the WT, and 22 transcripts were significantly downregulated ( $\log_2$  FC  $\leq$  -0.6) compared to the WT (Fig. 14). In *mfl1xrn4*, 43 transcripts were significantly upregulated compared to *xrn4*, and 22 transcripts were significantly downregulated compared to *xrn4* (Fig. 14). To control for differentially expressed transcripts that were coming through the pipeline as an artifact of the *xrn4* mutation, I also compared the levels of these transcripts to the WT. As a result, in *mfl1xrn4*, 30 transcripts were significantly upregulated compared to WT and *xrn4*, and 3 transcripts were significantly downregulated compared to WT and *xrn4* (Fig. 14). There were 10 transcripts significantly upregulated in both *mfl1* and *mfl1xrn4* compared to WT and *xrn4*, respectively (Table 3). Overall, these results indicate that

the loss of both *AtMFL1* and *AtXRN4* has a greater impact on the number of upregulated genes than the loss of *AtMFL1* alone. That being said, the few differentially expressed genes that came through the pipeline did not deviate greatly from WT, indicating that the loss of *AtMFL1* has a marginal effect on global transcript abundance.



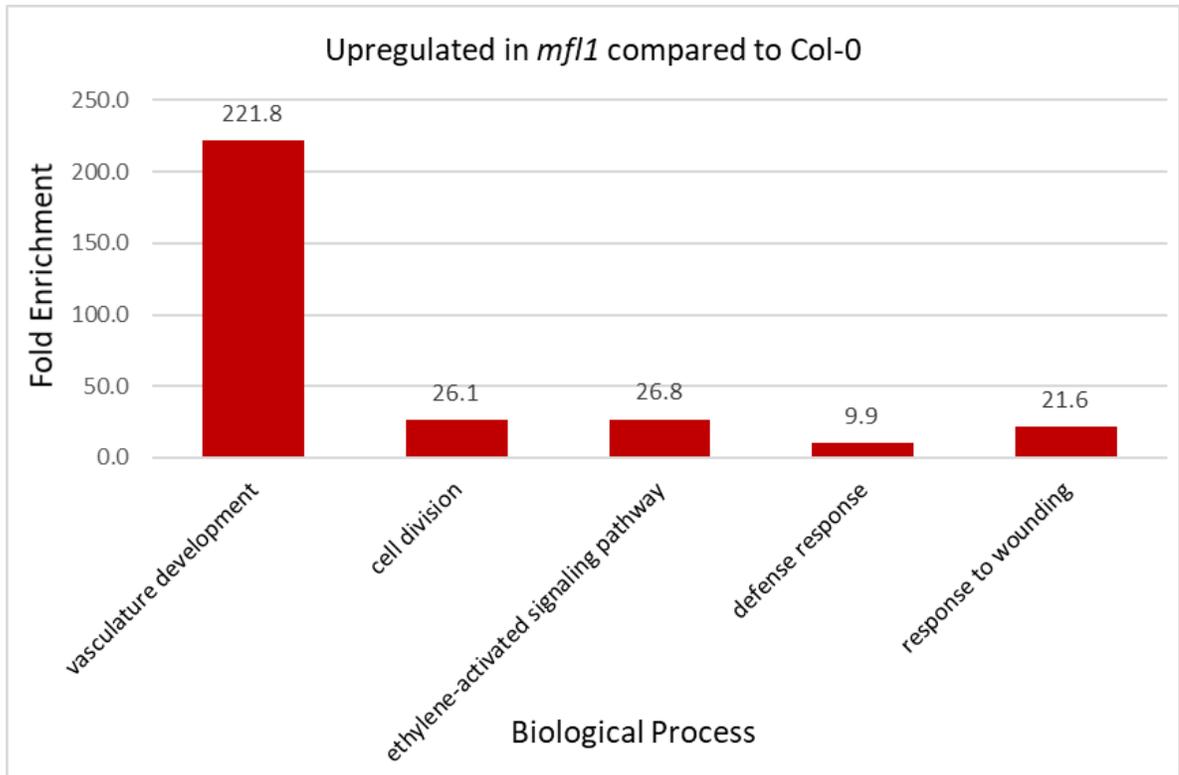
**Figure 14 Impacts of *mfl1* mutation on gene expression**

Total number of differentially expressed genes in *mfl1* compared to Col-0, *mfl1xrn4* compared to *xrn4*, and *mfl1xrn4* compared to both Col-0 and *xrn4*. Yellow bars represent upregulated genes; purple bars represent downregulated genes.

**Table 3 Genes significantly upregulated in *mfl1* and *mfl1xrn4* compared to controls**

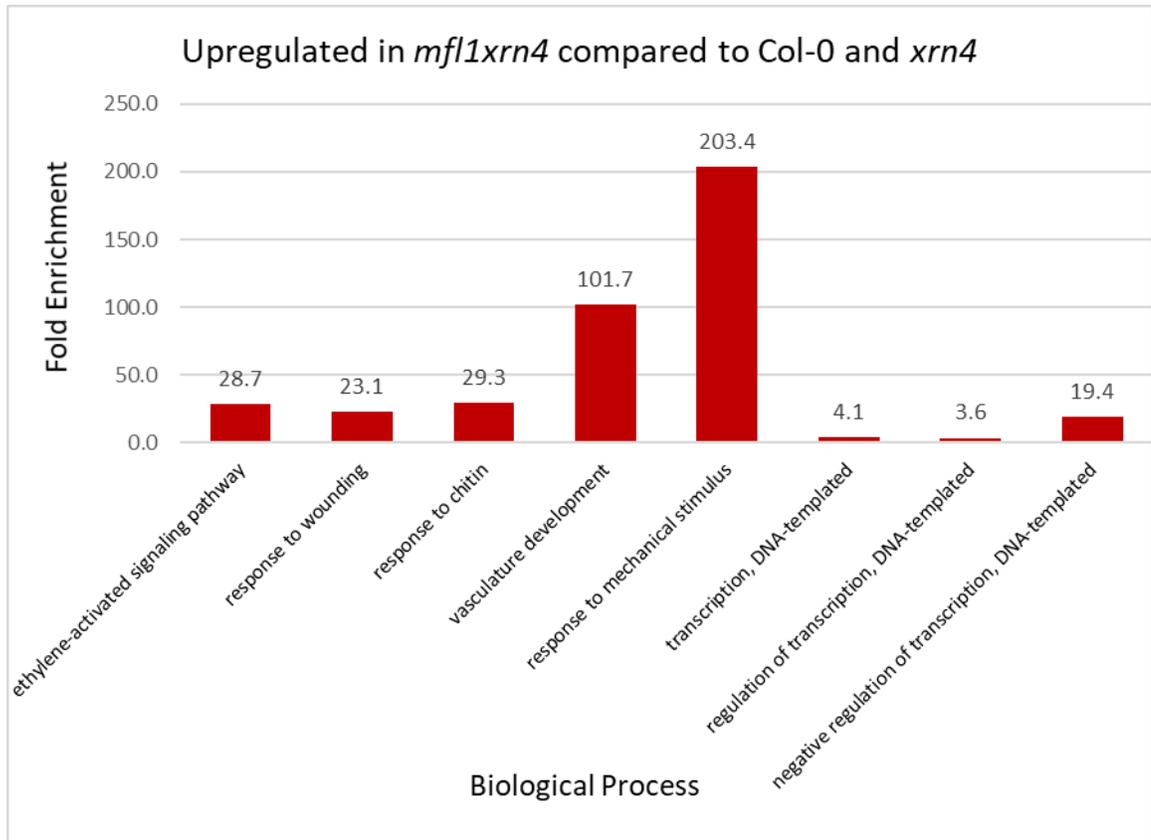
| Gene ID   | Gene Name | Transcript  | Type | Annotation  |
|-----------|-----------|-------------|------|---|
| AT1G72920 | AT1G72920 | AT1G72920.1 | mRNA | Toll-Interleukin-Resistance (TIR) domain family protein     |
| AT2G40000 | HSPRO2    | AT2G40000.1 | mRNA | ortholog of sugar beet HS1 PRO-1 2                          |
| AT3G55980 | SZF1      | AT3G55980.2 | mRNA | salt-inducible zinc finger 1                                |
| AT5G56550 | OXS3      | AT5G56550.1 | mRNA | oxidative stress 3  |
| AT5G61600 | ERF104    | AT5G61600.1 | mRNA | ethylene response factor 104                                |
| AT1G72910 | AT1G72910 | AT1G72910.1 | mRNA | Toll-Interleukin-Resistance (TIR) domain-containing protein |
| AT4G17490 | ATERF6    | AT4G17490.1 | mRNA | ethylene responsive element binding factor 6                |
| AT4G11280 | ACS6      | AT4G11280.1 | mRNA | 1-aminocyclopropane-1-carboxylic acid (acc) synthase 6      |
| AT4G27280 | AT4G27280 | AT4G27280.1 | mRNA | Calcium-binding EF-hand family protein                      |
| AT1G74930 | ORA47     | AT1G74930.1 | mRNA | Integrase-type DNA-binding superfamily protein              |

Gene ontology (GO) analysis was used to identify enriched biological processes that were upregulated in *mfl1* compared to WT (Fig. 15) and in *mfl1xrn4* compared to WT and *xrn4* (Fig. 16). Genes associated with vasculature development, ethylene-activated signaling pathway, and response to wounding were upregulated among both *mfl1* mutants compared to their respective controls. Additionally, various pathways involving regulation of transcription were upregulated in *mfl1xrn4*.



**Figure 15** Enriched GO categories of genes upregulated in *mfl1* compared to Col-0

Categories were identified using the DAVID program (Huang et al., 2009) and had an adjusted p-value of  $\leq 0.05$  determined by the Benjamini-Hochberg test.

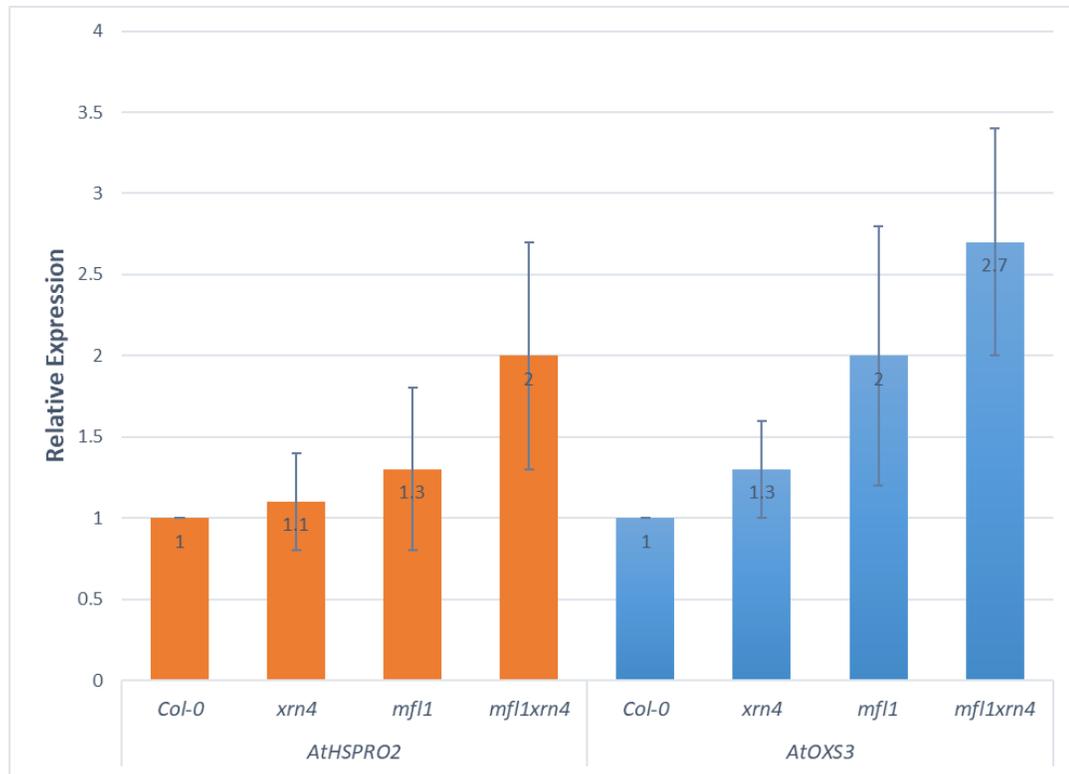


**Figure 16 Enriched GO categories of genes upregulated in *mfl1xrn4* compared to Col-0 and *xrn4***

Categories were identified using the DAVID program (Huang et al., 2009) and had an adjusted p-value of  $\leq 0.05$  determined by the Benjamini-Hochberg test.

To confirm the differential expression of biologically interesting target genes in *mfl1* and *mfl1xrn4* identified from the RNA-seq data, relative expression levels were measured using qPCR. *AtHSPRO2* and *AtOXS3* were chosen due to their involvement in pathogen response and oxidative stress, respectively. The expression levels of these genes were reported to be significantly elevated in both *mfl1* and *mfl1xrn4* compared to Col-0 and *xrn4* ( $\log_2$  FC  $\geq 0.6$ ). qPCR data showed that *AtHSPRO2* and *AtOXS3* seemed to be slightly more abundant in *mfl1* and *mfl1xrn4*

than in the controls, though this trend was much clearer in *mfl1xrn4* (Fig. 17). This could be due to the fact that the absence of two ribonucleases has an additive effect.

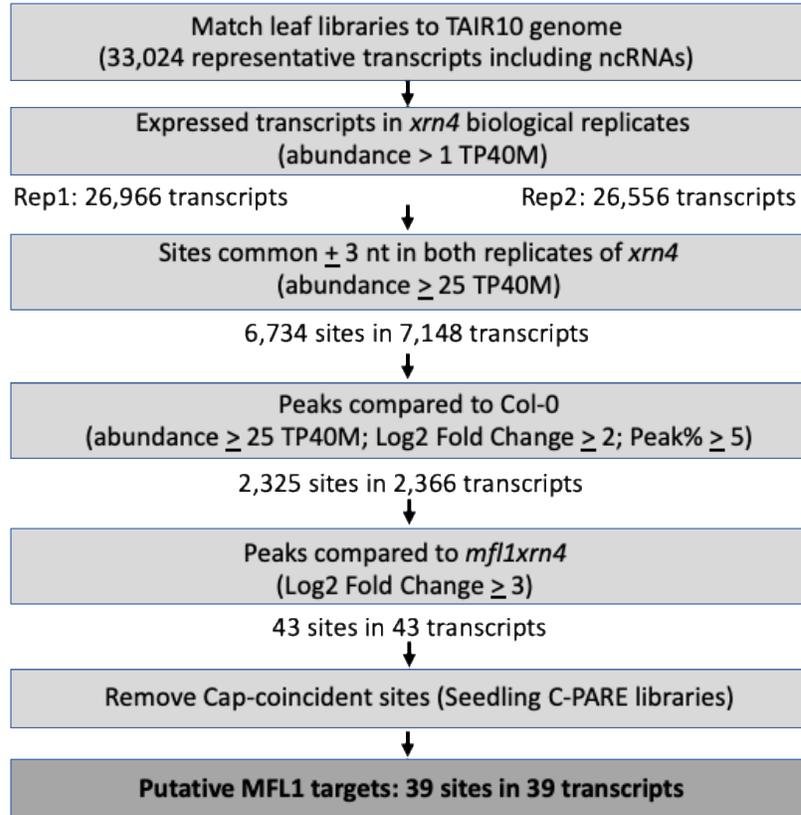


**Figure 17 Elevated mRNA levels of selected genes in *mfl1xrn4***  
qPCR shows relative levels of *AtHSPRO2* and *AtOXS3* in 2-week-old seedlings. Levels were normalized to *AtACTIN2* and bars are mean values + SD from two biological replicates. Col-0 expression levels are set to 1.

### Certain 3' fragments appear to be AtMFL1-dependent

Parallel Analysis of RNA Ends (PARE) captures mRNA decay intermediates that have a 5' monophosphate. PARE has been used to successfully identify and map cleavage sites of known endoribonucleases (Schmidt et al., 2015; Hurtig et al., 2021). To identify AtMFL1-dependent 3' fragments, PARE libraries were generated from

poly(A)<sup>+</sup> RNA isolated from rosette leaves of Col-0, *xrn4* and *mfl1xrn4*. The raw data were filtered through a stringent computational pipeline to identify the best candidates for validation (Fig. 18). We postulated that abundant 3' fragments present in *xrn4* should disappear or be drastically reduced in *mfl1xrn4*, when the endoribonuclease generating the fragments is absent. From two biological replicates, over 2,325 sites that over-accumulate ( $\log_2 \text{FC} \geq 2$ ) in *xrn4* compared to Col-0 were identified. Then, the peaks were filtered for those that showed a  $\log_2 \text{FC} \geq 3$  overaccumulation in *xrn4* compared to *mfl1xrn4*. Transcripts with peak positions coincident with the annotated TSS or experimentally verified cap sites (Nagarajan et al., 2019) were removed from the analysis so that they would not be mistaken as cleavage sites. The pipeline identified 39 putative targets of AtMFL1. Interestingly, 17 of these targets are elevated in at least one NMD mutant (either *upf1*, *upf3*, or *smg7*) based on previously published RNA-seq data (Raxwal et al., 2020; Gloggnitzer et al., 2014). This indicates that AtMFL1 may be involved in the turnover of a subset of NMD targets as well as a subset of non-NMD targets, but further studies must be done to verify this.

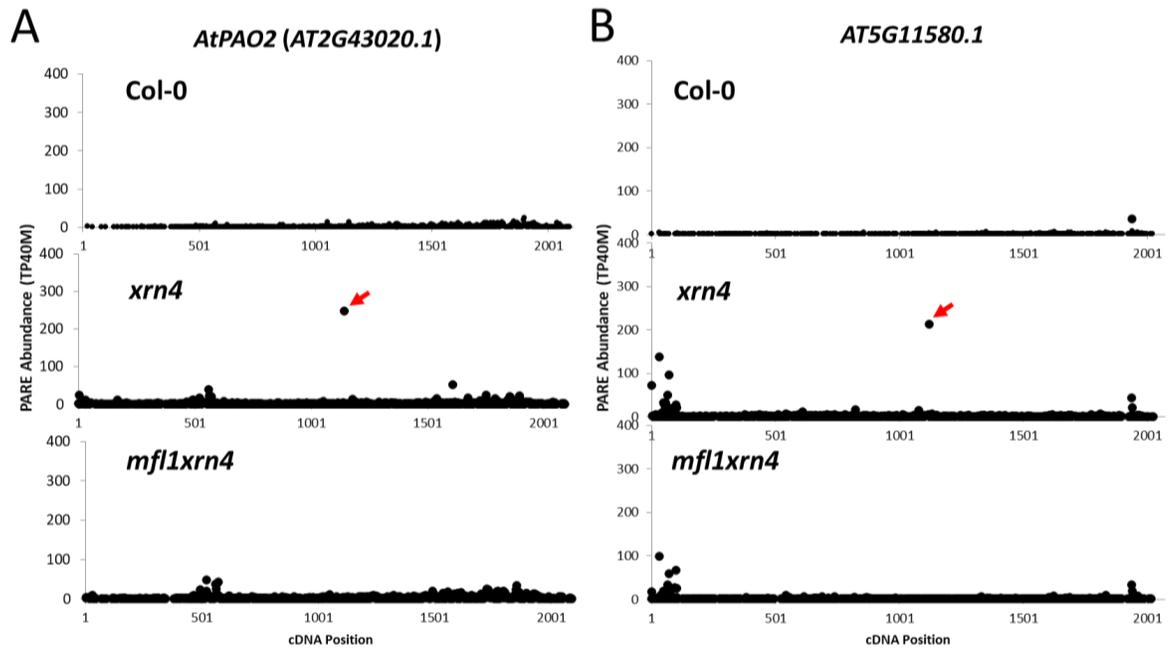


**Figure 18 Computational pipeline used for PARE analysis**

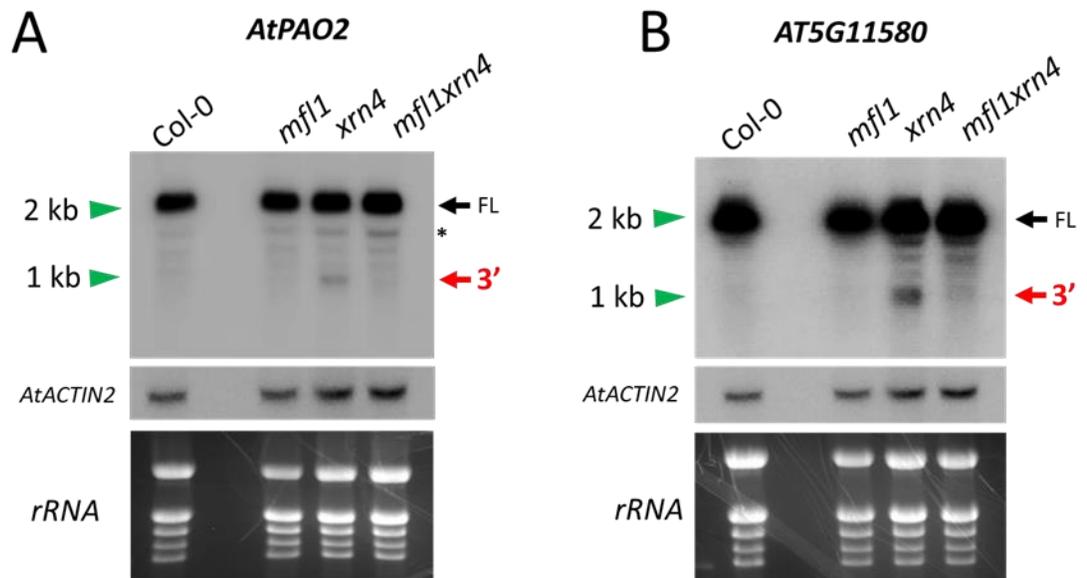
The final outputs of the pipeline highlighted in dark gray are transcripts classified as putative AtMFL1 targets. The data were filtered using criteria as shown. PARE library construction and computational analysis were carried out by Vinay Nagarajan.

Decay plots (D-plots) are used to visualize PARE results of individual transcripts. Plots of AtMFL1-dependent 3' fragments are presented for two transcripts, *AtPAO2* and *AT5G11580* (Fig. 19). In these examples, the peaks (indicated by the red arrows) in *xrn4* are strongly reduced in *mfl1xrn4*. To validate the PARE results, total RNA was extracted from Col-0, *xrn4*, *mfl1*, and *mfl1xrn4* seedlings, and Northern blot analysis was performed. The 3' fragments of *AtPAO2* and *AT5G11580* mRNAs were chosen for validation purposes because they were highly abundant in *xrn4* and their

large size made them easily detectable on a Northern blot. The 3' fragment of *AtPAO2* mRNA had already been identified by PARE analysis in *xrn4* seedlings (Nagarajan et al., 2019), but had not yet been validated. For both transcripts, a probe downstream of the cleavage site detected a 3' RNA fragment approximately 1 kb in size as expected (Fig. 20). Overall, these results validate the efficacy of the computational pipeline used for PARE analysis and indicate that *AtPAO2* and *AT5G11580* mRNAs are targets of AtMFL1. For these examples and the majority of the 39 targets identified by PARE, the AtMFL1 cleavage site is within the CDS or the 3' UTR of the transcripts. *AtPAO2* mRNA has an upstream open reading frame (uORF), which is a characteristic NMD trigger in plants. Additionally, *AtPAO2* is upregulated in a *upf1-5* mutant compared to WT, which is further evidence that it is an NMD target (Degtiar et al., 2015). While *AT5G11580* mRNA does not have any NMD features, it is upregulated in a *upf1-5* mutant compared to WT and in a *upf1-5pad4* double mutant compared to *pad4*. (Degtiar et al., 2015; Raxwal et al., 2020). As previously mentioned, AtUPF1 has been implicated in a variety of mRNA decay pathways, not just NMD. Therefore, *AT5G11580* mRNA may be targeted for degradation by an alternate AtUPF1-dependent decay pathway.



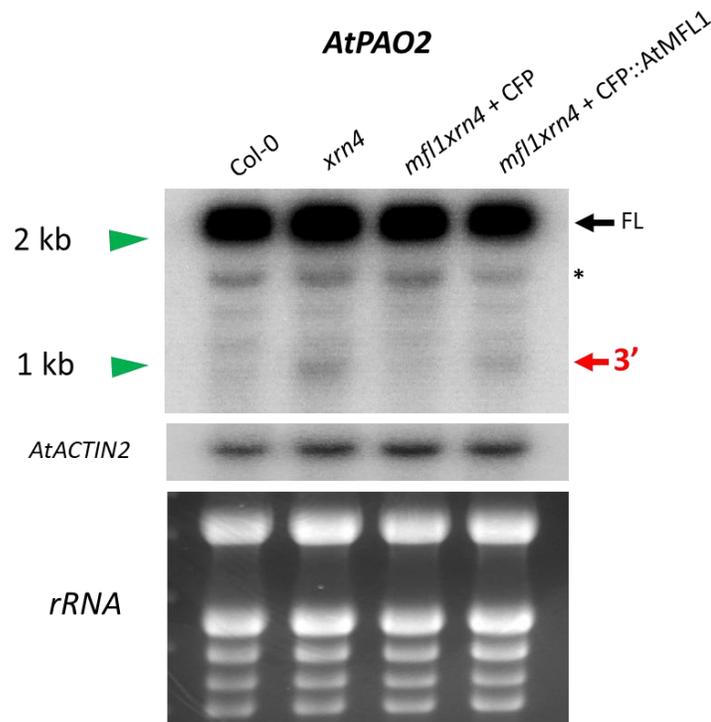
**Figure 19** AtMFL1-dependent 3' fragments over-accumulate in an *xrn4* mutant (A) D-plots showing prominent 3' fragment of *AtPAO2* RNA in WT, *xrn4*, and *mfl1xrn4* polyA+ PARE. (B) D-plots showing prominent 3' fragment of *AT5G11580* RNA in WT, *xrn4*, and *mfl1xrn4* polyA+ PARE. Red arrow, MaxSeq in *xrn4* and its corresponding position (1144 and 1120, respectively) in WT and *mfl1xrn4*. D-plots are representative of two biological replicates.



**Figure 20 RNA blots detect 3' RNA fragments produced by AtMFL1**  
 (A) RNA levels of full-length (FL) and 3' fragment (3') of *AtPAO2* RNA in WT, *mfl1*, *xrn4*, and *mfl1xrn4* seedlings. \*, unknown product. (B) RNA levels of full-length (FL) and 3' fragment (3') of *AT5G11580* RNA in WT, *mfl1*, *xrn4*, and *mfl1xrn4* seedlings. Top and middle, Northern blots. Bottom, ethidium bromide stained gel. Blots are representative of two biological replicates.

### D153 may be required for AtMFL1 nuclease activity

To ensure that the *35S::CFP::AtMFL1* localization construct produces a functional protein *in vivo*, the construct was transformed into *mfl1xrn4* double mutants. Total RNA was extracted from rosette leaves of Col-0, *xrn4*, and pooled T1 plants and used for Northern blot analysis. The same probe as in Figure 20 detected the approximately 1 kb RNA fragment of *AtPAO2* in *mfl1xrn4* stably expressing *CFP::AtMFL1* but not in *mfl1xrn4* stably expressing *CFP* alone (Fig. 21). These results indicate that the *CFP::AtMFL1* fusion protein is catalytically active *in vivo*.

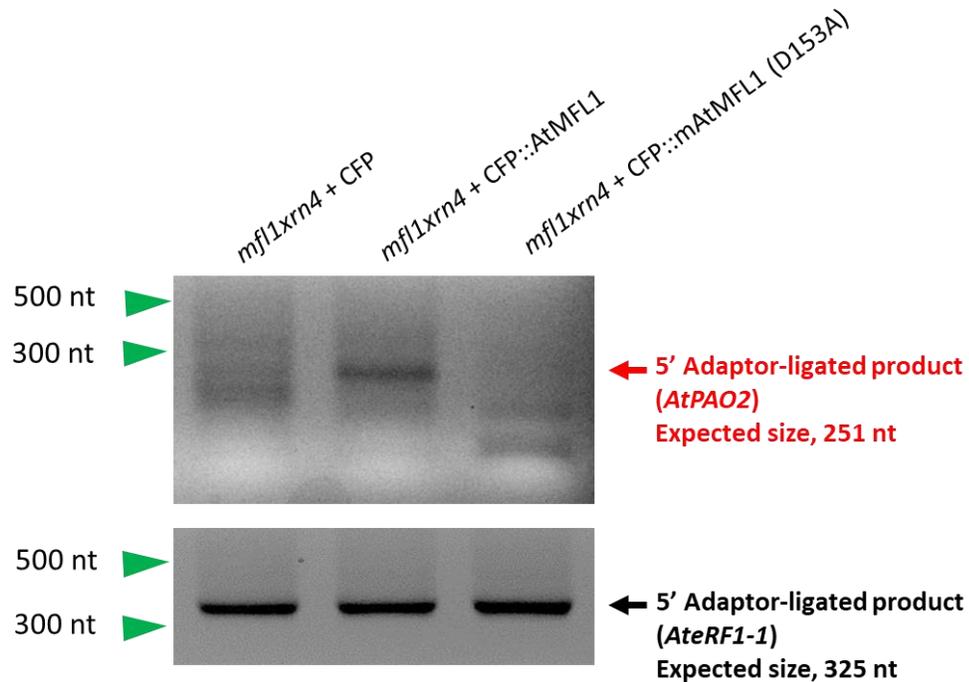


**Figure 21 CFP::AtMFL1 fusion protein rescues the presence of the *AtPAO2* 3' fragment**

RNA levels of full-length (FL) and 3' fragment (3') of *AtPAO2* RNA in WT, *xrn4*, and pooled T1 rosette leaves from the transformation of *mfl1xrn4* with either the *35S::CFP* empty vector or *35S::CFP::AtMFL1* are shown. Top and middle, Northern blots. Bottom, ethidium bromide stained gel. \*, unknown product.

To verify that AtMFL1 cleavage activity is dependent on the aspartic acid residues within its catalytic pocket, the fourth aspartic acid residue (D153) was mutated to alanine in the *35S::CFP::AtMFL1* fusion construct, and the mutant construct was introduced into *mfl1xrn4* double mutants. Total RNA was extracted from Col-0, *xrn4*, and pooled T1 seedlings and used for 5' RACE analysis. Based on the results of the second PCR reaction, part of the major 3' fragment of *AtPAO2* is detectable in *mfl1xrn4* stably expressing *CFP::AtMFL1* but not in *mfl1xrn4* stably expressing *CFP* alone or in *mfl1xrn4* stably expressing *CFP::mAtMFL1* (D153A)

(Fig. 22). These results indicate that D153 may be required for AtMFL1 nuclease activity, which would be consistent with the catalytic activity of metazoan MARF1. Future experiments are planned to further validate the requirement of D153 on AtMFL1 cleavage activity in stable T2 lines.



**Figure 22 AtMFL1 cleavage activity may require D153**

5' RACE PCR-amplified 3' fragment of *AtPAO2* (251 nt) in pooled T1 seedlings from the transformation of *mfl1xrn4* (*dbl*) with either the *35S::CFP* empty vector, *35S::CFP::AtMFL1*, or *35S::CFP::mAtMFL1* (D153A). *AteRF1-1* is a ligation control and shows uniformity of cDNA synthesis.

## Chapter 4

### DISCUSSION

Endoribonucleases play a major role in initiating mRNA decay; however, there is limited information about these proteins, particularly in plants. I decided to investigate three NYN domain-containing endoribonucleases in Arabidopsis that we predicted to be homologs of metazoan MARF1. One of these proteins, AtMFL1, had been previously shown to associate with conserved NMD factor AtUPF1, though nothing was known about its function (Chicois et al., 2018). The other two proteins, AtMFL2 and AtMFL3, had not yet been characterized, other than the observation that AtMFL3 is required for early embryonic development (Pagnussat et al., 2005). Here I report the organ-specific expression of *AtMFL* genes, the subcellular localization of CFP::AtMFL1, CFP::AtMFL2, and CFP::AtMFL3, and the evolutionary conservation of AtMFL1-like proteins in plants. I also report the gene categories and processes impacted by AtMFL1 and provide evidence for its role as a cytosolic endoribonuclease in Arabidopsis.

#### Characterization of plant *MFL* genes

*AtMFL1* is ubiquitously expressed in all major Arabidopsis organs examined (Fig. 6), indicating that it functions from the seedling to the flower stage of plant life. Transient transformation of *N. benthamiana* showed that AtMFL1 localizes to distinct foci, or processing (P)-bodies, within the cytosol (Chicois et al., 2018), and stable transformation of Arabidopsis further validated this result (Fig. 8). P-bodies are cytosolic aggregates that store translationally repressed mRNAs and various mRNA decay factors such as AtDCP1, AtDCP2, and AtXRN4 (Weber et al., 2008). Because AtMFL1 and AtXRN4 both localize to P-bodies, it is reasonable to assume that

AtXRN4 degrades the 3' fragments produced by AtMFL1 cleavage. Indeed, this assumption is supported by the PARE data and Northern blot analysis presented here (Figs. 19 and 20).

Like *AtMFL1*, *AtMFL2* is ubiquitously expressed in all major Arabidopsis organs examined (Fig. 6), indicating that it also functions in all stages of plant life. Interestingly, *AtMFL2* localizes to the cytosol and appears to reside and migrate along cytoskeletal filaments following transient expression (Fig. 9), though this result could not be validated by stable transformation of Arabidopsis. *AtMFL2* has an OST-HTH associated (OHA) domain at C-terminal end. Typically, OHA and OST-HTH domains occur in the same protein, but in the case of *AtMFL2* there is no OST-HTH domain present (Anantharaman and Aravind, 2006). This may be because *AtMFL2* associates with an OST-HTH domain-containing protein to carry out its function. Co-immunoprecipitation followed by mass spectrometry analysis could determine the protein partners of *AtMFL2*. Previous studies have shown that P-bodies migrate along cytoskeletal filaments by direct binding of AtDCP1 to the tails of several different myosin XI family members (Steffens et al., 2014). Colocalization of CFP::*AtMFL2* and RFP::*DCP1* could determine if *AtMFL2* localizes to P-bodies like *AtMFL1*.

Unlike *AtMFL1* and *AtMFL2*, *AtMFL3* is not ubiquitously expressed in the major Arabidopsis organs examined (Fig. 6). Based on RNA-seq profiling of the Arabidopsis developmental transcriptome, it appears that *AtMFL3* is highly expressed in roots but not in any other major organs (Klepikova et al., 2016). This could be validated by performing RT-PCR using total RNA from root tissue. *AtMFL3* is also called *EMBRYO SAC DEVELOPMENT ARREST 32 (EDA32)*, since a Ds transposon insertion disrupting this gene results in unfused polar nuclei in female gametophytes

and a lack of endosperm formation (Pagnussat et al., 2005). Additionally, overexpression of this protein appears to be cytotoxic in bacteria. Taken together, these results indicate that the timing and abundance of *AtMFL3* expression is tightly regulated and critical for plant development. Based on its apparent nuclear membrane localization (Fig. 10), *AtMFL3* might perform a function similar to that of *SWT1*, an endoribonuclease that associates with the nuclear pore complex and is functionally linked to perinuclear mRNP quality control (Skruzny, et al., 2009).

The presence of a catalytically active *AtMFL1*-like protein in the 20 plant species examined (Fig. 11) supports the hypothesis that *AtMFL1*-like proteins are evolutionarily conserved across plants and likely perform a critical function. Regarding the few flowering plants, gymnosperms, cycads, and ferns that lack an *AtMFL1*-like protein, it is possible that this protein was lost because a different protein acquired the ability to perform the same function. Alternatively, it is possible that these species do contain an *AtMFL1*-like protein, but I was unable to identify it in my analysis. Overall, the apparent lack of an *AtMFL1*-like protein does not appear to be specific to a particular plant lineage, and future studies could determine the evolutionary history of this protein in more detail. It is interesting to note that the SMART algorithm identified an NYN domain within *ChrMFL1* even though it lacks one of the required catalytic residues. This discrepancy illustrates the fact that homology does not always confer function, so it is crucial to examine key functional residues when identifying homologs in different species. In general, the *AtMFL1*-like proteins examined are relatively uncharacterized. Further studies could validate the function of these proteins in different plant species, though I hypothesize that they are performing a function similar to that of *AtMFL1*.

### **Insights from the impacts of AtMFL1 on gene expression**

To determine the impact of AtMFL1 on global transcript abundance, RNA-seq analysis was performed. Overall, very few genes were differentially expressed in *mfl1* compared to the WT (Fig. 14). I initially postulated that any transcripts found to be over-accumulating in *mfl1xrn4* compared to *xrn4* may be putative targets of AtMFL1. However, of the few transcripts that came through the RNA-seq pipeline, none were related to the AtMFL1 targets identified by PARE. This could be because only a small percentage of full-length AtMFL1 targets are actually cleaved by the enzyme, as shown by Northern blot analysis (Fig. 20). Since the cleaved products are much less than 50% of the full-length products, these transcripts would not have met the cut-off for differential expression (1.5 fold). This discrepancy is also true for substrates of AtXRN4 (Nagarajan et al., 2019). For our purposes, the RNA-seq data were more important for determining the indirect effects of *mfl1* mutations, which seem to be marginal. Slightly more genes were differentially expressed in *mfl1xrn4* compared to the WT and *xrn4*. This is most likely due to the fact that the loss of two genes involved in mRNA decay has a greater effect on gene expression than the loss of one alone. That being said, the loss of *AtXRN4* has a much greater effect on gene expression than the loss of *AtMFL1* (Nagarajan et al., 2019 and data presented here).

Many of the genes that were upregulated in *mfl1xrn4* are involved in various stress responses. *HSPRO2*, or *ORTHOLOG OF SUGAR BEET HS1PRO-1*, encodes a leucine rich repeat (LRR) protein that confers resistance to the nematode *Heterodera schachtii* in sugar beets (Cai et al., 1997). In Arabidopsis, HSPRO2 is a positive regulator of basal resistance to *Pseudomonas syringae* pv. *tomato* and appears to function downstream of salicylic acid (Murray et al., 2007). *ATIG72910* and *ATIG72920* encode toll-interleukin-resistance (TIR) domain-containing nucleotide-

binding site leucine-rich repeat (NLR) proteins known to be associated with pathogen response and disease resistance (McHale et al., 2006). Interestingly, the vast majority of NMD-targeted transcripts in Arabidopsis are associated with pathogen response, and elevation of some of these transcripts in NMD mutants has been shown to confer partial resistance to *Pseudomonas syringae* (Rayson et al., 2012). Recent studies suggest that *AT1G72910* is turned over by NMD in a temperature-dependent manner and plays an important role in regulating plant immunity at lower temperatures (Nasim et al., 2020). Because *AtHSPRO2*, *AT1G72910* and *AT1G72920* are upregulated in *mfl1xrn4*, I hypothesize that mutants with defects in these genes are more resistant to *Pseudomonas syringae* and perhaps other bacterial pathogens. Additionally, the upregulation of these genes provides further evidence that AtMFL1 is involved in plant NMD.

*OXS3*, or *OXIDATIVE STRESS 3*, encodes a chromatin-associated factor that confers enhanced tolerance to a range of metals and oxidizing chemicals in *Schizosaccharomyces pombe* (Blanvillain et al., 2009). In Arabidopsis, OXS3 functions as a negative regulator of several abscisic acid (ABA) responsive genes, most notably that of *ABI4*. Because *AtOXS3* is upregulated in *mfl1xrn4*, I hypothesize that these mutants are less sensitive to ABA and perhaps more tolerant to oxidative stress. Future studies could address how various stress responses are affected in *mfl1xrn4* and perhaps characterize an altered visible phenotype compared to WT or the single mutants.

As previously mentioned, PARE analysis by Vinay Nagarajan identified 39 putative targets of AtMFL1, 17 of which have been reported to be elevated in at least one NMD mutant (Raxwal et al., 2020; Gloggnitzer et al., 2014). The 3' fragments of

*AtPAO2* (AT2G43020) and *AT5G11580* were selected for validation via Northern blot based on their size and their abundance in *xrn4* (Fig. 19 and 20). *AtPAO2*, or *POLYAMINE OXIDASE 2*, encodes an enzyme that is responsible for polyamine catabolism (Takahashi et al. 2010). Specifically, *AtPAO2* is involved in excess spermidine catabolism during germination and early seedling development (Takahashi et al. 2019). It has been previously reported that the upstream open reading frame (uORF) within the 5' UTR of *AtPAO2* is responsible for translational repression of the main ORF (Guerrero-González et al., 2014). I propose that *AtMFL1* cleavage of the transcript is one mechanism by which translation of *AtPAO2* is repressed. *AT5G11580* encodes a previously uncharacterized regulator of chromosome condensation (RCC1) family protein. Unlike the mRNA of *AtPAO2*, the mRNA of *AT5G11580* does not contain any NMD features; however, this transcript may still be turned over by NMD since it is upregulated in a *upf1-5pad4* double mutant (Raxwal et al., 2020). Conversely, *AT5G11580* mRNA may be targeted for degradation by an alternative *AtUPF1*-dependent mechanism, given that *AtUPF1* is engaged in many functionally diverse mRNA decay pathways (reviewed in Kim and Maquat, 2019).

Based on its RNA-dependent association with *AtUPF1* (Chicois et al., 2018), we hypothesized that *AtMFL1* could be involved in plant NMD and potentially perform the same function as metazoan SMG6. Future experiments are planned to determine if *AtMFL1* is involved in the NMD pathway. It is clear from our results that even if *AtMFL1* cleaves some NMD transcripts, it does not cleave all of them. For example, Nagarajan et al (2019) showed that the NMD sensitive *eRF1-1* mRNA is internally cleaved by an unknown endoribonuclease and that the 3' fragment is detectable in plants lacking *AtXRN4* but not in plants lacking *AtUPF1*. We have

already determined that *eRF1-1* mRNA is not dependent on AtMFL1 for cleavage (Nagarajan, Bogdan, and Green, unpublished). It is likely that this transcript and others are cleaved by endoribonucleases that have yet to be characterized, such as the additional NYN domain-containing proteins in Arabidopsis. Previous studies suggest that two of these proteins, MNU1 and MNU2, are involved in the 5' processing of plant mitochondrial transcripts (Stoll and Binder, 2016). Future studies could elucidate the roles of these proteins as well as other putative endoribonucleases in Arabidopsis. Because AtMFL2 is closely related to AtMFL1, it could cleave its own set of transcripts or it could act in a redundant manner; however, the latter is less likely since we have identified targets of AtMFL1. Alternatively, the C-terminal OHA domain of AtMFL2 might associate with the C-terminal OST-HTH domains of AtMFL1, and the two proteins could work together to cleave a larger subset of targets. The previously mentioned co-IP experiments would be one way to test this. If AtMFL1 and AtMFL2 are in any way redundant, or if they work more efficiently in a complex, then future studies could investigate *mfl1mfl2xrn4* triple mutants using a similar strategy as was presented here.

## Chapter 5

### CONCLUSIONS AND FUTURE WORK

#### Conclusions

Canonical mRNA decay pathways as well as various mRNA surveillance pathways are required to maintain cellular homeostasis and remove unwanted or aberrant transcripts. Several different homologs of metazoan MARF1, an endoribonuclease that cleaves certain transcripts via its catalytic NYN domain, were identified in Arabidopsis. One of these proteins (AtMFL1) had previously been shown to localize to P-bodies and have an RNA-dependent association with UPF1, indicating that it may play a role in plant NMD. The goal of this project was to study the role of selected MARF1-like (MFL) endoribonucleases in Arabidopsis and hopefully uncover a novel mechanism of endoribonuclease-initiated RNA decay in plants.

Overall, my research determined the evolutionary conservation of AtMFL1-like proteins across plants, the organ-specific expression of *AtMFL* genes, and the subcellular localization of AtMFL proteins. My research also determined the impacts of AtMFL1 on global transcript abundance and provided evidence for its role as a cytosolic endoribonuclease that may or may not be involved in plant NMD. The RNA-seq and PARE data generated over the course of this project will provide other scientists with the information necessary for future studies of AtMFL1. Additionally, the strategy used to determine targets of AtMFL1 can be used to determine the targets of known and unknown endoribonucleases in different cell types. Overall, this work improves our understanding of mRNA decay in plants and provides a foundation for future studies of endoribonucleases in Arabidopsis and other systems.

## Future Directions

Additional studies could be performed to validate my conclusions as well as expand on the results of this research. To determine the phylogeny of AtMFL1, evolutionary biologists could perform an in-depth analysis that includes many more plant species. As mentioned previously, AtMFL1-like proteins do not seem to be present in the few gymnosperms, cycads, and ferns that were examined. It would be interesting to determine the exact point of divergence in evolutionary history and perhaps explain why certain plants do not have an AtMFL1-like protein. The targets of AtMFL1 could be investigated in other plant species to determine if they are cleaved by another endoribonuclease. I hypothesize that ChrMFL1 does not have endoribonuclease activity because it is missing one of the four aspartic acid residues required for function. To test this, ChrMFL1 could be purified and incubated with radioactively labeled ssRNA *in vitro* as described for MmMARF1 by Yao et al. (2018).

My project was mostly focused on AtMFL1 due to its association with AtUPF1. As a result, I did not determine the impacts of AtMFL2 or AtMFL3 on global RNA abundance. The characteristics of these putative endoribonucleases make them interesting to investigate, even if they are not involved in plant NMD. I have already identified *mfl2* homozygous mutants and generated *mfl2xrn4* double homozygous mutants. RNA-seq and PARE libraries could be made from the RNA of these mutants, similar to the strategy adopted for AtMFL1. If any of the targets identified by PARE can be experimentally validated, then the functional dependence of AtMFL2 on an aspartic acid residue within its NYN domain could be determined by performing genetic complementation. Further, if AtMFL2 targets overlap with those of AtMFL1, then the *mfl1mfl2xrn4* triple mutant could be investigated as

proposed in the Discussion. Because AtMFL3 is required for early embryonic development, *mfl3* homozygous mutants cannot be generated. Instead, expression of *AtMFL3* could be conditionally knocked down by RNAi using the B-estradiol-induced XVE system described in Zuo et al. (2020). AtMFL3 could be characterized in the same manner as AtMFL1 and AtMFL2 by knocking down expression of *AtMFL3* in an *xrn4* mutant background.

To determine whether the AtMFL1 functions in the NMD pathway, AtMFL1 targets with NMD features could be analyzed by performing Northern blot analysis and/or PARE analysis in *upf1xrn4* and *upf3xrn4* double mutants. Presence of the 3' fragment in both double mutants would indicate that AtMFL1 functions independently of AtUPF1 and is not involved in plant NMD. Absence of a 3' fragment in *upf1xrn4* alone would indicate that AtMFL1 is AtUPF1-dependent but not involved in plant NMD. Absence of the 3' fragment in both double mutants would indicate that AtMFL1 is AtUPF1-dependent and involved in plant NMD. Even if AtMFL1 is not involved in plant NMD, this work identified a new mechanism of mRNA decay in Arabidopsis involving the endoribonuclease activity of AtMFL1.

## REFERENCES

- Aizer, A., Kalo, A., Kafri, P., Shraga, A., Ben-Yishay, R., Jacob, A., ... Shav-Tal, Y. (2014). Quantifying mRNA targeting to P-bodies in living human cells reveals their dual role in mRNA decay and storage. *Journal of Cell Science*, *127*(20), 4443–4456. <https://doi.org/10.1242/jcs.152975>
- Anantharaman, V., & Aravind, L. (n.d.). The NYN Domains Novel Predicted RNAses with a PIN Domain-Like Fold. In *RNA Biology* (Vol. 3). Retrieved from <http://www.landesbioscience.com/journals/rnabiology/abstract.php?id=2548>
- Badis, G., Saveanu, C., Fromont-Racine, M., & Jacquier, A. (2004). Targeted mRNA degradation by deadenylation-independent decapping. *Molecular Cell*, *15*(1), 5–15. <https://doi.org/10.1016/j.molcel.2004.06.028>
- Bent, A. (2006). Arabidopsis thaliana floral dip transformation method. *Methods in Molecular Biology (Clifton, N.J.)*, *343*(1), 87–103. <https://doi.org/10.1385/1-59745-130-4:87>
- Bernstein, E., Caudy, A. A., Hammond, S. M., & Hannon, G. J. (2001). Role for a bidentate ribonuclease in the. *Nature*, *409*(2001), 363–366.
- Blanvillain, R., Kim, J. H., Wu, S., Lima, A., & Ow, D. W. (2009). OXIDATIVE STRESS 3 is a chromatin-associated factor involved in tolerance to heavy metals and oxidative stress. *Plant Journal*, *57*(4), 654–665. <https://doi.org/10.1111/j.1365-313X.2008.03717.x>
- Brothers, W. R., Hebert, S., Kleinman, C. L., & Fabian, M. R. (2020). A non-canonical role for the EDC4 decapping factor in regulating MARF1-mediated mRNA decay. *ELife*, *9*, 1–23. <https://doi.org/10.7554/eLife.54995>

- Cai, D., Kleine, M., Kifle, S., Harloff, H. J., Sandal, N. N., Marcker, K. A., ... Jung, C. (1997). Positional cloning of a gene for nematode resistance in sugar beet. *Science*, 275(5301), 832–834. <https://doi.org/10.1126/science.275.5301.832>
- Caplan, J. L., Kumar, A. S., Park, E., Padmanabhan, M. S., Hoban, K., Modla, S., ... Dinesh-Kumar, S. P. (2015). Chloroplast Stromules Function during Innate Immunity. *Developmental Cell*, 34(1), 45–57. <https://doi.org/10.1016/j.devcel.2015.05.011>
- Chicois, C., Scheer, H., Garcia, S., Zuber, H., Mutterer, J., Chicher, J., ... Garcia, D. (2018). The UPF1 interactome reveals interaction networks between RNA degradation and translation repression factors in Arabidopsis. *Plant Journal*, 96(1), 119–132. <https://doi.org/10.1111/tpj.14022>
- Conti, E., & Izaurralde, E. (2005). Nonsense-mediated mRNA decay: Molecular insights and mechanistic variations across species. *Current Opinion in Cell Biology*, 17(3), 316–325. <https://doi.org/10.1016/j.ceb.2005.04.005>
- Degtiar, E., Fridman, A., Gottlieb, D., Vexler, K., Berezin, I., Farhi, R., ... Shaul, O. (2015). The feedback control of UPF3 is crucial for RNA surveillance in plants. *Nucleic Acids Research*, 43(8), 4219–4235. <https://doi.org/10.1093/nar/gkv237>
- Doma, M. K., & Parker, R. (2007). RNA Quality Control in Eukaryotes. *Cell*, 131(4), 660–668. <https://doi.org/10.1016/j.cell.2007.10.041>
- Fischer, J. W., Busa, V. F., Shao, Y., & Leung, A. K. L. (2020). Structure-Mediated RNA Decay by UPF1 and G3BP1. *Molecular Cell*, 78(1), 70-84.e6. <https://doi.org/10.1016/j.molcel.2020.01.021>

- Garneau, N. L., Wilusz, J., & Wilusz, C. J. (2007). The highways and byways of mRNA decay. *Nature Reviews Molecular Cell Biology*, 8(2), 113–126. <https://doi.org/10.1038/nrm2104>
- Gatfield, D., & Izaurralde, E. (2004). Nonsense-mediated messenger RNA decay is initiated by endonucleolytic cleavage in *Drosophila*. *Nature*, 429(6991), 575–578. <https://doi.org/10.1038/nature02559>
- Glavan, F., Behm-Ansmant, I., Izaurralde, E., & Conti, E. (2006). Structures of the PIN domains of SMG6 and SMG5 reveal a nuclease within the mRNA surveillance complex. *EMBO Journal*, 25(21), 5117–5125. <https://doi.org/10.1038/sj.emboj.7601377>
- Gloggnitzer, J., Akimcheva, S., Srinivasan, A., Kusenda, B., Riehs, N., Stampfl, H., ... Riha, K. (2014). Nonsense-mediated mRNA decay modulates immune receptor levels to regulate plant antibacterial defense. *Cell Host and Microbe*, 16(3), 376–390. <https://doi.org/10.1016/j.chom.2014.08.010>
- Guerrero-González, M. L., Rodríguez-Kessler, M., & Jiménez-Bremont, J. F. (2014). UORF, a regulatory mechanism of the Arabidopsis polyamine oxidase 2. *Molecular Biology Reports*, 41(4), 2427–2443. <https://doi.org/10.1007/s11033-014-3098-5>
- Hammond, S. M., Bernstein, E., Beach, D., & Hannon, G. J. (2000). An RNA-directed nuclease mediates post-transcriptional gene silencing in *Drosophila* cells. *Nature*, 404(6775), 293–296. <https://doi.org/10.1038/35005107>
- He, F., & Jacobson, A. (2015). Nonsense-Mediated mRNA Decay: Degradation of Defective Transcripts Is Only Part of the Story. *Annual Review of Genetics*, 49(1), 339–366. <https://doi.org/10.1146/annurev-genet-112414-054639>

- Huang, D. W., Sherman, B. T., & Lempicki, R. A. (2009). Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nature Protocols*, 4(1), 44–57. <https://doi.org/10.1038/nprot.2008.211>
- Huntzinger, E., Kashima, I., Fauser, M., Saulière, J., & Izaurralde, E. (2008). SMG6 is the catalytic endonuclease that cleaves mRNAs containing nonsense codons in metazoan. *Rna*, 14(12), 2609–2617. <https://doi.org/10.1261/rna.1386208>
- Hurtig, J. E., Steiger, M. A., Nagarajan, V. K., Li, T., Chao, T. C., Tsai, K. L., & Van Hoof, A. (2021). Comparative parallel analysis of RNA ends identifies mRNA substrates of a tRNA splicing endonuclease-initiated mRNA decay pathway. *Proceedings of the National Academy of Sciences of the United States of America*, 118(10). <https://doi.org/10.1073/pnas.2020429118>
- Kanemitsu, Y., Fujitani, M., Fujita, Y., Zhang, S., Su, Y. Q., Kawahara, Y., & Yamashita, T. (2017). The RNA-binding protein MARF1 promotes cortical neurogenesis through its RNase activity domain. *Scientific Reports*, 7(1), 1–9. <https://doi.org/10.1038/s41598-017-01317-y>
- Karam, R., Wengrod, J., Gardner, L. B., & Wilkinson, M. F. (2013, June). Regulation of nonsense-mediated mRNA decay: Implications for physiology and disease. *Biochimica et Biophysica Acta - Gene Regulatory Mechanisms*, Vol. 1829, pp. 624–633. <https://doi.org/10.1016/j.bbagr.2013.03.002>
- Kashima, I., Yamashita, A., Izumi, N., Kataoka, N., Morishita, R., Hoshino, S., ... Ohno, S. (2006). Binding of a novel SMG-1-Upf1-eRF1-eRF3 complex (SURF) to the exon junction complex triggers Upf1 phosphorylation and nonsense-mediated mRNA decay. *Genes and Development*, 20(3), 355–367. <https://doi.org/10.1101/gad.1389006>

- Kim, Y. K. I., & Maquat, L. E. (2019). UPFRONT and center in RNA decay: UPF1 in nonsense-mediated mRNA decay and beyond. *Rna*, 25(4), 407–422. <https://doi.org/10.1261/rna.070136.118>
- Klepikova, A. V., Kasianov, A. S., Gerasimov, E. S., Logacheva, M. D., & Penin, A. A. (2016). A high resolution map of the Arabidopsis thaliana developmental transcriptome based on RNA-seq profiling. *Plant Journal*, 88(6), 1058–1070. <https://doi.org/10.1111/tpj.13312>
- Kumar, S., Stecher, G., Li, M., Knyaz, C., & Tamura, K. (2018). MEGA X: Molecular evolutionary genetics analysis across computing platforms. *Molecular Biology and Evolution*, 35(6), 1547–1549. <https://doi.org/10.1093/molbev/msy096>
- Lamanna, A. C., & Karbsteina, K. (2009). Nob1 binds the single-stranded cleavage site D at the 3'-end of 18S rRNA with its PIN domain. *Proceedings of the National Academy of Sciences of the United States of America*, 106(34), 14259–14264. <https://doi.org/10.1073/pnas.0905403106>
- Lebreton, A., Tomecki, R., Dziembowski, A., & Séraphin, B. (2008). Endonucleolytic RNA cleavage by a eukaryotic exosome. *Nature*, 456(7224), 993–996. <https://doi.org/10.1038/nature07480>
- Livak, K. J., & Schmittgen, T. D. (2001). Analysis of relative gene expression data using real-time quantitative PCR and the 2- $\Delta\Delta$ CT method. *Methods*, 25(4), 402–408. <https://doi.org/10.1006/meth.2001.1262>
- Lubas, M., Damgaard, C. K., Tomecki, R., Cysewski, D., Jensen, T. H., & Dziembowski, A. (2013). Exonuclease hDIS3L2 specifies an exosome-independent 3'-5' degradation pathway of human cytoplasmic mRNA. *EMBO Journal*, 32(13), 1855–1868. <https://doi.org/10.1038/emboj.2013.135>

- Luo, Z., & Chen, Z. (2007). Improperly terminated, unpolyadenylated mRNA of sense transgenes is targeted by RDR6-mediated RNA silencing in Arabidopsis. *Plant Cell*, *19*(3), 943–958. <https://doi.org/10.1105/tpc.106.045724>
- Lykke-Andersen, S., Chen, Y., Ardal, B. R., Lilje, B., Waage, J., Sandelin, A., & Jensen, T. H. (2014). Human nonsense-mediated RNA decay initiates widely by endonucleolysis and targets snoRNA host genes. *Genes and Development*, *28*(22), 2498–2517. <https://doi.org/10.1101/gad.246538.114>
- Madeira, F., Park, Y. M., Lee, J., Buso, N., Gur, T., Madhusoodanan, N., ... Lopez, R. (2019). The EMBL-EBI search and sequence analysis tools APIs in 2019. *Nucleic Acids Research*, *47*(W1), W636–W641. <https://doi.org/10.1093/nar/gkz268>
- Malecki, M., Viegas, S. C., Carneiro, T., Golik, P., Dressaire, C., Ferreira, M. G., & Arraiano, C. M. (2013). The exoribonuclease Dis3L2 defines a novel eukaryotic RNA degradation pathway. *EMBO Journal*, *32*(13), 1842–1854. <https://doi.org/10.1038/emboj.2013.63>
- Mandel, C. R., Kaneko, S., Zhang, H., Gebauer, D., Vethantham, V., Manley, J. L., & Tong, L. (2006). Polyadenylation factor CPSF-73 is the pre-mRNA 3'-end-processing endonuclease. *Nature*, *444*(7121), 953–956. <https://doi.org/10.1038/nature05363>
- Martinez, J., Patkaniowska, A., Urlaub, H., Lührmann, R., & Tuschl, T. (2002). Single-stranded antisense siRNAs guide target RNA cleavage in RNAi. *Cell*, *110*(5), 563–574. [https://doi.org/10.1016/S0092-8674\(02\)00908-X](https://doi.org/10.1016/S0092-8674(02)00908-X)
- McHale, L., Tan, X., Koehl, P., & Michelmore, R. W. (2006). Plant NBS-LRR proteins: adaptable guards. *Genome biology*, *7*(4), 212. <https://doi.org/10.1186/gb-2006-7-4-212>

- Megel, C., Hummel, G., Lalande, S., Ubrig, E., Cognat, V., Morelle, G., ... Maréchal-Drouard, L. (2019). Plant RNases T2, but not Dicer-like proteins, are major players of tRNA-derived fragments biogenesis. *Nucleic Acids Research*, *47*(2), 941–952. <https://doi.org/10.1093/nar/gky1156>
- Merai, Z., Benkovics, A. H., Nyiko, T., Debreczeny, M., Hiripi, L., Kerényi, Z., ... Silhavy, D. (2013). The late steps of plant nonsense-mediated mRNA decay. *Plant Journal*, *73*(1), 50–62. <https://doi.org/10.1111/tpj.12015>
- Murray, S. L., Ingle, R. A., Petersen, L. N., & Denby, K. J. (2007). Basal resistance against *Pseudomonas syringae* in *Arabidopsis* involves WRKY53 and a protein with homology to a nematode resistance protein. *Molecular Plant-Microbe Interactions*, *20*(11), 1431–1438. <https://doi.org/10.1094/MPMI-20-11-1431>
- Nagarajan, V. K., Jones, C. I., Newbury, S. F., & Green, P. J. (2013, June). XRN 5'→3' exoribonucleases: Structure, mechanisms and functions. *Biochimica et Biophysica Acta - Gene Regulatory Mechanisms*, Vol. 1829, pp. 590–603. <https://doi.org/10.1016/j.bbagrm.2013.03.005>
- Nagarajan, V. K., Kukulich, P. M., Von Hagel, B., & Green, P. J. (2019). RNA degradomes reveal substrates and importance for dark and nitrogen stress responses of *Arabidopsis* XRN4. *Nucleic Acids Research*, *47*(17), 9216–9230. <https://doi.org/10.1093/nar/gkz712>
- Nishimura, T., Fakim, H., Brandmann, T., Youn, J. Y., Gingras, A. C., Jinek, M., & Fabian, M. R. (2018). Human MARF1 is an endoribonuclease that interacts with the DCP1:2 decapping complex and degrades target mRNAs. *Nucleic Acids Research*, *46*(22), 12008–12021. <https://doi.org/10.1093/nar/gky1011>
- Nasim, Z., Fahim, M., Gawarecka, K., Susila, H., Jin, S., Youn, G., & Ahn, J. H. (2020). Role of *AT1G72910*, *AT1G72940*, and *ADR1-LIKE2* in Plant Immunity

under Nonsense-Mediated mRNA Decay-Compromised Conditions at Low Temperatures. *International journal of molecular sciences*, 21(21), 7986. <https://doi.org/10.3390/ijms21217986>

Pagnussat, G. C., Yu, H. J., Ngo, Q. A., Rajani, S., Mayalagu, S., Johnson, C. S., ... Sundaresan, V. (2005). Genetic and molecular identification of genes required for female gametophyte development and function in Arabidopsis. *Development*, 132(3), 603–614. <https://doi.org/10.1242/dev.01595>

Pertschy, B., Schneider, C., Gnädig, M., Schäfer, T., Tollervey, D., & Hurt, E. (2009). RNA helicase Prp43 and its co-factor Pfa1 promote 20 to 18 S rRNA processing catalyzed by the endonuclease Nob1. *Journal of Biological Chemistry*, 284(50), 35079–35091. <https://doi.org/10.1074/jbc.M109.040774>

Popp, M. W. L., & Maquat, L. E. (2014). The dharma of nonsense-mediated mRNA decay in mammalian cells. *Molecules and Cells*, 37(1), 1–8. <https://doi.org/10.14348/molcells.2014.2193>

Raxwal, V. K., Simpson, C. G., Gloggnitzer, J., Entinze, J. C., Guo, W., Zhang, R., ... Riha, K. (2020). Nonsense-mediated RNA decay factor UPF1 is critical for posttranscriptional and translational gene regulation in arabidopsis. *Plant Cell*, 32(9), 2725–2741. <https://doi.org/10.1105/TPC.20.00244>

Rayson, S., Arciga-Reyes, L., Wootton, L., De Torres Zabala, M., Truman, W., Graham, N., Grant, M., & Davies, B. (2012). A role for nonsense-mediated mRNA decay in plants: pathogen responses are induced in Arabidopsis thaliana NMD mutants. *PloS one*, 7(2), e31917. <https://doi.org/10.1371/journal.pone.0031917>

- Ryan, K., Calvo, O., & Manley, J. L. (2004). Evidence that polyadenylation factor CPSF-73 is the mRNA 3' processing endonuclease. *Rna*, *10*(4), 565–573.  
<https://doi.org/10.1261/rna.5214404>
- Schein, A., Sheffy-Levin, S., Glaser, F., & Schuster, G. (2008). The RNase E/G-type endoribonuclease of higher plants is located in the chloroplast and cleaves RNA similarly to the E. coli enzyme. *Rna*, *14*(6), 1057–1068.  
<https://doi.org/10.1261/rna.907608>
- Schmidt, S. A., Foley, P. L., Jeong, D. H., Rymarquis, L. A., Doyle, F., Tenenbaum, S. A., ... Green, P. J. (2015). Identification of SMG6 cleavage sites and a preferred RNA cleavage motif by global analysis of endogenous NMD targets in human cells. *Nucleic Acids Research*, *43*(1), 309–323.  
<https://doi.org/10.1093/nar/gku1258>
- Shaul O. (2015). Unique Aspects of Plant Nonsense-Mediated mRNA Decay. *Trends in plant science*, *20*(11), 767–779. <https://doi.org/10.1016/j.tplants.2015.08.011>
- Shoemaker, C. J., & Green, R. (2012). Translation drives mRNA quality control. *Nature Structural and Molecular Biology*, *19*(6), 594–601.  
<https://doi.org/10.1038/nsmb.2301>
- Skružný, M., Schneider, C., Rácz, A., Weng, J., Tollervey, D., & Hurt, E. (2009). An endoribonuclease functionally linked to perinuclear mRNP quality control associates with the nuclear pore complexes. *PLoS Biology*, *7*(1).  
<https://doi.org/10.1371/journal.pbio.1000008>
- Song, J. J., Smith, S. K., Hannon, G. J., & Joshua-Tor, L. (2004). Crystal structure of argonaute and its implications for RISC slicer activity. *Science*, *305*(5689), 1434–1437. <https://doi.org/10.1126/science.1102514>

- Souret, F. F., Kastenmayer, J. P., & Green, P. J. (2004). AtXRN4 degrades mRNA in Arabidopsis and its substrates include selected miRNA targets. *Molecular Cell*, *15*(2), 173–183. <https://doi.org/10.1016/j.molcel.2004.06.006>
- Steffens, A., Jaegle, B., Tresch, A., Hülskamp, M., & Jakoby, M. (2014). Processing-body movement in Arabidopsis depends on an interaction between myosins and DECAPPING PROTEIN1. *Plant physiology*, *164*(4), 1879–1892. <https://doi.org/10.1104/pp.113.233031>
- Stoll, B., & Binder, S. (2016). Two NYN domain containing putative nucleases are involved in transcript maturation in Arabidopsis mitochondria. *The Plant journal :for cell and molecular biology*, *85*(2), 278–288. <https://doi.org/10.1111/tpj.13111>
- Su, Y., Sugiura, K., Sun, F., Pendola, J. K., Cox, G. A., Handel, A., ... Harbor, B. (2013). *NIH Public Access*. *335*(6075), 1496–1499. <https://doi.org/10.1126/science.1214680.MARF1>
- Takahashi, Y., Cong, R., Sagor, G. H. M., Niitsu, M., Berberich, T., & Kusano, T. (2010). Characterization of five polyamine oxidase isoforms in Arabidopsis thaliana. *Plant Cell Reports*, *29*(9), 955–965. <https://doi.org/10.1007/s00299-010-0881-1>
- Takahashi, Y., Uemura, T., & Teshima, Y. (2019). Polyamine oxidase 2 is involved in regulating excess spermidine contents during seed germination and early seedling development in Arabidopsis thaliana. *Biochemical and Biophysical Research Communications*, *516*(4), 1248–1251. <https://doi.org/10.1016/j.bbrc.2019.07.022>
- Trapnell, C., Roberts, A., Goff, L., Pertea, G., Kim, D., Kelley, D. R., ... Pachter, L. (2012). Differential gene and transcript expression analysis of RNA-seq

experiments with TopHat and Cufflinks. *Nature Protocols*, 7(3), 562–578.  
<https://doi.org/10.1038/nprot.2012.016>

Weber, C., Nover, L., & Fauth, M. (2008). Plant stress granules and mRNA processing bodies are distinct from heat stress granules. *The Plant journal : for cell and molecular biology*, 56(4), 517–530. <https://doi.org/10.1111/j.1365-313X.2008.03623.x>

Welch, E. M., & Jacobson, A. (1999). An internal open reading frame triggers nonsense-mediated decay of the yeast SPT10 mRNA. *EMBO Journal*, 18(21), 6134–6145. <https://doi.org/10.1093/emboj/18.21.6134>

Xiao, S., Jiang, L., Wang, C., & Ow, D. W. (2021). Arabidopsis OXS3 family proteins repress ABA signaling through interactions with AFP1 in the regulation of ABI4 expression. *Journal of Experimental Botany*.  
<https://doi.org/10.1093/jxb/erab237>

Yao, Q., Cao, G., Li, M., Wu, B., Zhang, X., Zhang, T., ... Su, Y. Q. (2018). Ribonuclease activity of MARF1 controls oocyte RNA homeostasis and genome integrity in mice. *Proceedings of the National Academy of Sciences of the United States of America*, 115(44), 11250–11255.  
<https://doi.org/10.1073/pnas.1809744115>

Zamore, P. D., Tuschl, T., Sharp, P. A., & Bartel, D. P. (2000). RNAi: Double-stranded RNA directs the ATP-dependent cleavage of mRNA at 21 to 23 nucleotide intervals. *Cell*, 101(1), 25–33. [https://doi.org/10.1016/s0092-8674\(00\)80620-0](https://doi.org/10.1016/s0092-8674(00)80620-0)

Zhang, W., Murphy, C., & Sieburth, L. E. (2010). Conserved RNaseII domain protein functions in cytoplasmic mRNA decay and suppresses Arabidopsis decapping mutant phenotypes. *Proceedings of the National Academy of Sciences of the*

*United States of America*, 107(36), 15981–15985.

<https://doi.org/10.1073/pnas.1007060107>

Zuo, J., Niu, Q. W., & Chua, N. H. (2000). Technical advance: An estrogen receptor-based transactivator XVE mediates highly inducible gene expression in transgenic plants. *The Plant journal : for cell and molecular biology*, 24(2), 265–273. <https://doi.org/10.1046/j.1365-313x.2000.00868.x>

## Appendix A

### ATMFL GENOME SEQUENCES

**ATG** = Translational Start/Stop      **atgc** = UTR  
**ATGC** = Exon                              **atgc** = Intron  
*atgc* = Annotation on other strand

#### *AtMFL1 (AT2G15560)*

aagaatttgcaattccacctatgaacctcggctctcggatttcattagacgaaaagaagaagttgatagaagtactgattta  
gtttggtttcgtttccacgagactcaatccagccggttagttctggggaatcgttccaccgactcatgtcccttattcattct  
cgtcggtttatgtttataatattttgtgtgacattcaataaaggttttcagtaaaaataaaataaaaaatacagagcaagagtaaac  
aatatttgctattttgactatataaaatgggaaaaatcgtttaatatatgaatttaaaaaatagctatttaatatataattttgt  
atataatcaattaatataatgtaaaaatcgtaattaatttttacacaaattaatcgttgaccaggttaaaattcacacgacgttate  
tgacgttaacagattgataacaaccgttaactaatcttcattagttccaatacagatgtcattttgattatagcaaaaacgggaaa  
aatgtcatttaattccaattttcaattatataccatttaaccatcaacttcacatatgaccattttatacatgaataaacgctaac  
caacagtttaaaaaataagaatgtgtgaccaggccaaaaaacatactgctatctaagcgtagatattcgaattttgttgat  
ttggttgatttattcagatttagataattcaggtaaagagtttatttccgttaaaattttatgactatttttctggtttctgttagaat  
atgtatcatattttaatcgaattacataaaaaataaaaaaaaattgttaaatattttttgatatctaagttatttagaaaaaatgt  
cgtttaatattttctatttaaatatataatataatataatataatattttattttatttaagtgttagaatcaatatttcatcagattta  
gtaaaattgatattttcaaaagaattcttttaaaaaaacattttctgcaaaaatctatttgccccaagcattacaatacgaacag  
atcggattatccttttcacatatatgcaactattctcttcaaaaagatcaaaacatttattattggatattcttttggcaaaaataaaa  
actagtttaataaccaccgactctctttatattgcttctctggacaaacgcaaaaactttttagaacctaaaaattcccaaaatc  
cgtcggagaagaagatcgagaagaatcaacaactaatctgaagaattttcaaatccgtcttcgtatcgtctacgagatcctt  
atctctcccctgaatctgtaaaaattggttcgaaatgacgaaatcaatagattctactttctcctacatcgaattttcgaactt  
cctttcataatctcactgatctgtagtaattgtttatattggaatctctttgttaattccggttcaggaacctttg**ATG**ATA  
CAAAACGCTATGTCAAGTTATTATAATCCATCAGCAGCAACAACCTTTAGTC  
TCTATTGATTCCGATGAACAAAGAAGAATCAAATACCTAGCGGATTTAAA  
CATGATTGCAACACAAAGACATTCTTCTACAGATGGTCCTATGGCCATTCT  
CTGGGATATGGAGAATTGTCCCGTTCCCTAGCGATGTACGTCCTGAAGATGT  
AGCTAGTAACATAAGAATGGCTATTACAGTTACATCCTGTGATATCCGGTCC  
CGTTGTAACTTCTCGGCTTACGGGGATTTCAATGGTTTCCCTCGTCGGGTT  
CGAGAAGGTTGTCAAAGAACC GGTTGAAGCTTATTGATGTACCAATGG  
TAGGAAAGATGCGTCGGATAAAGCGATTTTGATTGATATGTTCTTGTTTGT  
GCTTGATAATAAGCCGCTGCGACTATCGTTCTCGTATCGGGTGATGTTGA  
TTTCGCGCCTGCGCTTCATATATTGGGTCAGCGTGGGTATACTGTGATTCTT  
GTTATACCTTCTAGTGTGTATGTGAATTCAGCTTTGTCTAATGCTGGTAAGT  
TTGTTTGGGATTGGCATAGTATTGTTACGGTGAAGGCTTTGTGCCGCGAT  
GTAAACCTCGTGTGTTCCGTATCTTATGGGGTGTAATATTGGTGATAACA  
GTAACATGGATGGTTTGAATGAAGATGAACTATTCTCTATAGAGGTAAC  
GTTATAGTAGTGATCCAAGAGAGTCTTCTTCTTGATGGTTTCGCAGTTTC

GTAATGAGTATAGTAGCGGTGTAATGTCGTGTTGGCCATCTAATTCAGGCG  
AGTCTATGGCATGTCCTCCTTCAGGTCACCTTGAGTCCACCATGTGGGTAG  
CGCCGGGAGATTTAAACGGTTTGAAGGGGCAGCTCGTGAAGCTGCTAGAG  
CTTTCAGGTGGATGTATTCTCTTATGCGTGTTCCTTCTGAATACCAACGGA  
AATTCAGTAAACCGCTTTTTGTATCGGATTATGGGGTGGCCAAGCTTGTGG  
ATCTGTTCAAAAAGATGAGTGATGTGATTGTAGTTGATGGTAAAGGCAAC  
AAGAGATTTGTTTACCTGCGCAACTCGAAACCGAACATTATCTCTCCTTCT  
TCCCCTGTGGTTCTACTGAGAAGAGAGAGGAAAGGAAAAGAGCCTAACGG  
GGTAACTACCAATGGAGGAGTCTCCTCTGATGAAATGTCAGACACAGGCT  
CGGTTCAAGAGCGAGAGGAACCTAGAAGAGTTCAAATTCGAGCTACAAGAC  
ATTTTGGTGAGCTATTGTTGTCAGGTACAGATGGATTGTTTCGAAGCGATA  
TACAAGCTAAGGTACAAAAGGCCATTGGCCTACACAAATATGGGCGTGAA  
TCACTTGGAGCAGCTTTTGACAAGTTAAGGGATGTTGTTGCCATACATGA  
AGATCCTGCTACAGGGAGGAAACTCATCAGCCCGGTTTAAaggagcctaactcaga  
gtcatataagatcagacgggctgttgtagctgtttataagcgttctctactcaggacaagcaatgaagctccttgccta  
agaagctctcagtattaatgtatatgctctactcttctcctgggacaaaatgtttcgcggaagctatgattctgcattgcaagc  
attcacttctgttactttcagcaattggaaataagtcagacggagcgtttatgagtcattgactgtatgtgtgttctctgttgg  
gaataagccagtatcattgtatatgctctactcttctcattaggacaaaaggagccttcttcttactcaatcgtttagaacgtt  
tgaagagctcaatagcagctcactccttctgacttttaattcacgaatcaatataagttatatactcattgagtttctacaaac  
cttgtcttctgttgggtcgtctgttcttatacattcagtcaccaccaccaaaggagaaacattcattctcacttcaaat  
gcttctgcaccaccttaaacgttacttcaactctcattgagagctgtaaagatcattctcagtgagtcacaaagaaagtc  
aattcacgagtagaaggtatcttgggttcggatccgggtcaggtaaagccttctcctcctcaatagtttcaaatctta  
atthtttatattacaatcctcctcacttatgtagtgcctctactactcggacattattgcttacttgattaattaccttttc  
gataaaatagggccttgaattgtaaaaaaataattgtagacttctcattctccgacacattagtagcaatagtgtcatat  
gccacatcagcataatacatt

***AtMFL2 (AT3G62200)***

taaccatccatatattttcttgaatcctaattcatcatcgacaagtttagcacaagggaaactcttttaagtttagctacaaaca  
ttcatattcgataatgccagaagtaattgtacattgtcactcaagaacaagaagtggtacattagtgattatgcgtattatataat  
acttttaacaacataatgccacaatcaatgattaaagtagaaaatttctgtagaattatctcatttataaaaaggacggcttcc  
gaaatattaatttaaccaactattaagatgctgacaataatcattgtttatcaactgttcttcttctactaccacgttaataa  
ataggcttagcttaggaaatttgaaaaacaaaattcaccatttcaaatatttatatgtaaaattgaaatcaacaaaaagtatata  
gaaaaaacattgactactgtattagataaaaaaaaaaagattcaagggaagtaggacgtaaacgctgtgacagatggaa  
aacagagagaagtttctcgtgaaaaacaaaacggtagtcttttctggtgaagttgaagtagaagtcattcatggtcaacgt  
caactacaagagaacgGACGTAAACGAAACGGTGCACGAGAGCGTCCACTAAACC  
AACCACAGTATCACACCACTACTCAACCGATGCCAACAAAGGCATAAACCA  
CGACTCTCCACCAATCAAAAACGCTTCACTTTTTCTTTCTTAACTACCTTC  
CCCTCGCCGATACTGTAGGGACAAAATTGTCAAAGAACACTCGCTTCATCA  
TCAGGGAAATATAAAGGGAAGAACAgaaaaacctttcggcctaataattgaaaaatcccttattcga  
agattccttggcgatcgtagtagtagcttacggcggcggaagatcgATGTCGACGAACTCGGCGAGT  
GATGGGGATTTCGGAACATAATTCGGTGCCGGCAGAGATGGCTGAGGCTCA

GTACGTGAGAGCGAAGACTTCAGTGTGGTGGGACATAGAGAATTGCCAAG  
TCCCTAACGGTCTTGATGCTCATGGAATTGCTCAGAATATCACTTCGGCGC  
TTCAGAAGATGAACTATTGTGGTCCTGTTTCTATCTCTGCTTATGGTGATAC  
CAATCGCATCCCTTTGACCATTCAACATGCTCTTAATTCCACCGGAATTGC  
GCTCAACCACGTCCCCGCCGgtaactccccctttaaaactatctctattacttagtttctgtatctttatctgct  
tggttagattcaattgtgaattgattgtttctttcatgtgatggttactgtttcagGGGTGAAAGATGCGAGT  
GATAAGAAGATACTAGTGGACATGTTGTTTTGGGCGTTAGACAATCCTGCA  
CCTGCGAATTTTATGTTAATATCAGGAGACAGAGACTTCTCCAATGCTCTT  
CACGGATTGAGAATGAGACGTTACAATGTTCTTTTGGCACAGCCTCTCAA  
GCATCTGTGCCACTTGTTTCATGCTGCAAAGACTGTGTGGCTTTGGACTAGC  
CTATCTGCTGGTGGAAATCCCTCTTACACGAGCCGAAAGCTTACAACCTTGTT  
GCCAACCAAACGACCCCAAAGCCTGGTTCAGAGATTCCATCAAGCCAACC  
TCTGGATTCAAATTCTGACTCCAGGAGAGTCTTTGACAACAAATCCAAAGT  
CAAATATGTTCCAAAACCATCTAATCATCAGCCTAACAATAATTACCGTCA  
GCAACAACAGAATACTCAAGGGAAACAGTTTAAGAAAGCCCCACATGAGT  
TCTTTGGCACTAGCGAGCCATCTGTTTCCACTAGCAGACCTCCTCCTCCCA  
ATTTACCCTCTAGCAATGTTAACACTTTCCCTGGTAATGTCATGACTAATCC  
TCAGAATCAGAATCAGTACACTTATCCTCCTCGCCCTGGTCCATTCCCTCCT  
AGACAACCTTACCCTAACACTGATCCTTCTTGGAATAATGGAAACAGCATC  
CCAAACCATGCTCAGAACTATTATCCCAATGCTGCTAGGCCAGGTGCTGCC  
ACTATGCGGCCACCTTATGGTAATGTTTTCCGTCCTTACCGCCCAGAGAAT  
CTGAATCCTCCAGTTGGTAATGGTTTCCGTCCAATGCAGCACCCGCGAAAT  
GACGGGCTAGATTTCCGTCTCCACCTCTCTTGACTCCACTTGATATCAGC  
AATCTAAGTGTGTCTCAATATCCTAGTCAAACCTCAAACCGTCCCTAATTC  
AATCCTCAAGTTAGACAAGAATTCAGACCAAAGATGGAATCCTCATATAC  
TCATAATGGTCCAAATAAGAGTTATATTCCACGGTGTAGTAGTGCCCCGGT  
AACTCAATCTACGACAACCACTGCCACACCTACCCTTCCAGTCCCTGGAGT  
TCCACCTTCTCAGCCTCCAATGGTTACTGGATCTGGTTCTAGCAATGATAG  
GTGGGGAACACAAGAGTGTCCACCACCATCTGAGTATGTTCAAGGCCTTA  
TAGGCGTTATCTTACATGCTCTACACATCTTGAAAACCTGAAAAAGTTATGC  
CTACCGAGCCAAATATCTCGGATTGCATTCAGTATGGAGATCCTAAGCATC  
ACGGTACTGACGTAAAGAAGGCTTTGGAAAGTGCTTTGGAGCATCATATG  
ATCATGATGACAAATGTAGGTAAACTAAAGCTCTACATTGGCAAAAATGA  
AGCTCTATGGAACCTGTGTAAACCCTTTAGGGGCAAACGCTAAGCAATACC  
CAAAGAAACTTGGGATAGAATACAACAGTTTCTAACCTCATCTTCTGGGC  
GTGTGGAGTTTACGGCAACTACGTGCAGgtagaatgtttgttacttaaggatctatagttccttct  
taccatgaaaatgttataacatcctttctctctttgatctaataatagGTATGAAGCTGCACAAGTTC  
TGAAAAAGGAATGTCTAAAAGAGTTTACTTTAGGTGACATACTACAGATC  
TTGAATATAACAGCAACCACAAAGAAATGGATTACCCATCATCAAACAGG  
ATGGAAACCAATTACTATTAGCCTTGCTGCAGAGACCACAAACGAGACAG  
CTACTGAAGCGGATCCAGGTATCCAAACCGTTGCC**TGA**atctaactttgttcaatgaaat  
accgaagaagatgtagtagcagccataaacagagcatatatgggtactaccttgaatatcaatatagagaacactaagaca  
tggaanaatagggtttccatatctaaaactcagaatgagatgattctgaagtctaagtactagttccttagtagaagcagcac

ttttttttaattttactttcagtcggttcagatgaaggagttaaggctaaaagtttgggtagttaagtttgacctaataatgtca  
tgactagatcatgattaggttcgtaaaaaccttagtggtccaactgccatcatcaagaagatgcacaaattcatgaggatg  
tgattgggaatctctgcaccaccaaagtttctgcattagttacggcaatctgtttcttttgcttactatagaggttacagttatatt  
cgtgatattcctaattgctggtgggaaaaatctgtaattctacctttttttttacctacgtatgtcttattctgaaccttcaagc  
tttcctttcttttaaagacgtttatgtttatgtatgaaagtaatttctactggactttcgtttatcatattgtttaacataacaggtg  
tagaaatgatgatctgaaagttgtagcgaaatcttgcgcaatcattctataatttctaaattttaaataaagctatc gatcctgagat  
acaagtttctaacaaggtccaagtagattcattcgtacaatacaaggttacaatccaatgattcaaaagttgtgtgaaag  
agactttgaa

*AtMFL3 (AT3G62210)*

actattattcagccaaatacgaagatggtgtaacggtaataattgattgtttataaaagaaatggaaaaagaaataattaa  
attacaaaatgatacgaaaactaacgaattttggatcttatcattgatatagacatttgatcgttaagcgtatctaataattcaac  
ttgttttatacaacaagaacacatcgttcatagaattctgatggtgatagtaattgattgaccactaaactatcctgtttaaacta  
aaacaagtgggtgtaattgaaatctttctaaaataaatttaaattgaaacaagtgaggagtaattattacattgttatagatgatt  
tcttagaaattgtttttcttattattagataaatagtataatatttatattcctattgatcaactgttactgttattgactattttatcaa  
acataccaaaaaaacatgaatattctgtaaaaaattgtggaaaattgaatgtcaatgaagaaatgagtttgaagaggtcgtg  
tgacataagtataatacaaaactcgtataagaaaacggaaaggaagttcgtcgggctaacaagcattgcttcgtgctcactta  
gaaagctctgttgagctaatgatgagttcaaatccctttttttttgtctgtgcctagtaaacaacatttctctgtaaagaagt  
actagttgatgtaatatatttaaccttttacgtaaaataaaataaagtttcatcgtggaaagtggaaacactgacgagtgaggag  
tcttgctgtaactaggagattgagattccttgaagaagattgaagagaaagaagtggtttcgttgataacagaaagtcctt  
caaaacaaaaattccttttcgcatcgcgATGAATACGGCGGGGAAAGAAGATTTGGGAA  
CGGCGGATACGGCTGAGGCTCAGTACGTGATGGCGAAGACCTCAGTATGG  
TGGGATATAGAGAATTGTCAAGTCCCTAAGGGTCTTGATGCACATGGGATT  
GCTCAGAATATTAGTTCGGCGCTTAAGAAAATGAACTATTGTGGTCCGGGTC  
TCAATCTCCGCTTATGGTGATACGAGTGGTATCCCTCATGTCATCCAACAT  
GCTCTTAATTCCACCGGAATCGAGCTCCACCATGTCCC GGCCGgtgaattccttaa  
cccttttatggttttctaaatctagatctactctcttttctttctttttattgtttttttctgataataaagatctgctagtgagctg  
aatgtgtgatcgttgatgatgccttatgtttctttcattctgtttttgtaacaactaacgttgtaggcatttataatcattggtg  
tattttcacttctcaatcgaatcgttttagggtttatactggttggttgctgtgattttgactacctgatcgtgtattgatca  
aagagtgttaataacaaagtggttcgtttcctttttttgtgtgtgtc gatgtgtatttcagGGGTGAAAGATG  
CGAGTGACAAGAAGATTCTAGTGGACATGTTGTTTTGGGCATTTGACAATC  
CTGCGCCTTCAAATATTATGTTAATTT CAGGAGATAGAGACTTTTCCAATG  
CTCTTCACAAATTGAGTCTGAGACGGTACAATATTCTCTTAGCACACCCTC  
CTAAAGCATCAGCGCCGCTAAGTCAAGCTGCAACAACAGTATGGCTTTGG  
ACAAGTCTATTAGCCGGAGGAAACCCTCTCATAACGAGGCAAAGTTAAAAC  
CTCACAACCTTGTTGCCAATGCGTCAACTTCTAGTAATGTCATGAGTAGTCC  
TCCTCACAATCAGTTTCCTGATCCTCCTAGATCAGGTCCTTTGCATGCTCGT  
CAGCCTTACCTAAACCCTGATCCTTTTGTGAATAACAGAGATCCCAATGCT  
GCCAGACCAGGACCTTCTAATATGCGGCCACTTTGTCTAATGCTATCCGT  
CGTCATCGCCAAGAGAAGCTGGAACGCGCTCTCCCTCTACTGATTTTATTG  
GTCTTTATGTTTATGATACTTGCAACTGTGAACTCAAATTAAtgctgcatgaccat  
tttcataataatgagctgttttggctgttccgtgaagatgattgcttggagctgcagactacggagaacttaactcaga

agttaaatgctgcagccactatgtctttcttttattgattctttctttctcagcttgtgttttcgagcttagttgtgtgagtattacc  
tgatgtgattatacgaattacgcattgtttatctccatgggagaaatcacgatttgaattccgacgaatgattttattgtatata  
ttttatataatttgtttaataattaaccatccatatatctttctgtaaatcctaattcatcatcgacaagtttagcacaagggactt  
ctttaagtttagctacaacattcatatcgataatgccagaagtaatgtacattgctactcaagaacaagaagtgggtacat  
tagtggtatgcgtattatataatacttttaacaacataatgccacaatcaatgattaaagtagaaaatttgctagaattatctcatt  
tattaaaaaggacggccttccgaaatattaatttaaccaactattaagatgctgacaataatcattgtttatcaactgtttctgt  
ttcactactaccacgtaataaataggcttagcttaggaaattgaaaaacaaaattcaccatttcaa

## Appendix B

### ATMFL1-LIKE PROTEIN SEQUENCES

#### **NYN domain**

#### **LOTUS domain**

##### **>ChrMFL1**

MK SIGATALANLKRLQLRGRDGPEQKEVVIAWDIENARPPPGIPVEEVERAISN  
AFQHLPRRKRHAFLAAASPRSYAAMTAVYGQDQVNHLLSHTDFMLAPGGAK  
STSTDAKLLQRIDEFIERQCALGRASYSVLVLITGDDDYTPRVKRAMHAGMD  
VELLHPAGITSRSLMATLEGRANVWHAEWQQFLEYWINANSSNSGSAGPSGG  
GQASSRQPQGGPRASESASGSGKPAAAAAAGGNKGTQPPPPPQQRTPQQQQ  
QAPQPAAAASSRGSDDSRKNGTTT SKERASAVLALSSLPVVRQSCDPDRALA  
AVCGWLASERRDNSAFYQSLHAFASECGCKLVVDRDLVTA AEGGGELTLT  
CVDDDERHPQRAVQSARDRLRNEINEVLLGDPHVLLLAAMRQQQQQQQQQQ  
QQQSQWQQWQWQQSPRQQQPFFQPQPQPPHNRPQQAQQQAPA  
QRAQQQAAPQVPPVPPAHGVGSSSSSRV GKERVSTTLRLPDLDFVRQSDCP  
EAALAMVLGWLDAARTDNSAFYQSLHALASECGCKLVVGRDLVAARGELTL  
ACVDDDERHPQRAVQSARDRLRLALDAELRQPGHVLGAKLGSVVVNATKKL  
CDWLFSPGAAATGGGAGGGRRGGGGAAPPCVEDLEALLRQRGFCAAAAPP  
WVWPAREEQLRAAAKAVVLVEEGLGLAPLPLPDGRALTARPFEPAAAGGR  
AADRFPEFEVAAVAAMLRSLRQPNELEAQVWVWHA AICDAAASLGPGG  
WISGGGRRGGGGSHSSSSSSSSSHYDQQVRGLAAALWLRHCDGAGLCPPLQ  
QAPQPA AAAAAASTAAASKPAAAAAGAAASSVAAPRVSDAADAGSTDARS  
SSTPASPATTAPTA AATPLAALPYPALAALLRHGHDAAVLTD PQLYGTAF  
LLWLDARAAAAASTAACTAAGGTASGAAGVGVGGTAAGR DGGDGGGGG  
PGGVAAAPQVGRLPDGGGWVRVVLPTAFAGTAAARA AVAQVLGSGLS  
CGELVAALRGLGAAPPPGSEKSEIAGLLAELVIAAAAAAAS

##### **>PpMFL1**

MASYRSLTPSRQPPGMDSAAIARAQQQHRANNMQQQLRAPPPLQPTSCPVAI  
LWDIENCVPGEVNAEDVAGNIRIALREHPHVGAVTMFSAYGDFNHFPKVR  
EGCQRTGVNLIDVPNGKKDAADKAILVDMFLFALDNPPPSTIFLITGDVDFAP  
ALHKLGRGYVVVLVIPDGVGVSSALRGAGRFVYDWPCLCRGEG LQNPQRQ  
GRQFNLCPVDDR NQGRVPTSIYTDVDRIPVSQFQDSYPEYMRPLYTEESDPRI  
DDVYGEDILAWGGITNPIYNVQPV DYGHHQQYVMPGQVQPQGPRISGYRTSH  
SASSGQLRNQVSGGFVDSRGRSGISSDSGMQPGESE EPTSPGGSQPLWVQPG  
DLTGLKRQLVQLLNNGGQMMLSKVPAEYNKLFGRPLYLSEYNAQRLVHLI  
DKMKDALLIKGDGTNKT LHV LKKEAGRVTNKYK PSSNRASAPKGD KDKDTS  
EDSEPEGSVPPSKELEKDVAKIEFATDVLSKTSLLEVDKVTEKVPFSEPV EAKR

YLLGMDMDENMDETIVVQSSDWIPVPVDVKNNVDARKYEPVVDVEEDQIKV  
EDFITPDVNLTHQTNLQTFRRQLQELLVVYSYQIPYENFIRVYQQRYARTLDLS  
SFGVESVNSLLAKVKDVAVIKDVRGNFFLTAV

>MpMFL1

MRATSTKELLSTRREVSSVMRAQQQQQIVLTGPAAILWDIENCPVPSEVKAE  
DVAGNIRMALRLHPVIEGAVTLFSAYGDFNHFPKRVREGCQRTGVNLIDVPN  
GKKDASDKAILVDLFLFALDNPPPCTIFLISGDVDFAPALHKLQGRGYTVVLAI  
PSGVPVSSALCSAGRFWYWPVSARGEGLVPAKSFVYRVGPEEKNCIESSVC  
QAWSPSDDSDPASDELVYGEEQIVWKSSKTQHENVADGSASRMVPSALYGS  
MSIQVTSQVTSQACAVGPVPVYWGENTSTQGSYVYGPSGAPHPSGYVKPSQG  
AASREGNTGAVSREGSTAARHLAESVEADCRGSNAAGPASLGGLLNLREQLV  
KLISFHGGKLELVRVPPEYSRHYGRPLYLAEYGYTKLVPMLEKMRWEMYIRG  
EGTYKTLHLTKTGFRVAAKLRDARVSRVAVTGEETEVSVCQKIPSESIYAEK  
IPSESIYAEKIPSESIYADNEVVLEDGVAACPTKDIQTFTEKDEPSAKTIVAD  
LPSDVEALIPTYGWRYSYDSDGDEKRTDSQEEKDVRVDDASSELVDITNEILEDTN  
SEDLIVEASLSEDPCEIRLQVFRQELQELLVSHACKICMASFLALYEQRYSR  
ALDCPSFGVQELESLEKVNDAVLMEDPGSKFKYVVASCVN

>AmtMFL1

MASPSNFPSKATNPSPENSNETSPTDPKTGITSPRASNQQRVLSGPAAILWDI  
ENCPVPSDVRPEDVAGNIRMALRVHPVIKGAVTLSAYGDFNVFPRRLREGCQ  
RTGVKLVDPNGRDKAADKAILVDMFLFALDNRPSSILLISGDVDFAPALHIL  
GQGRGYTVILVIPSGVGSSALSANAGRFWWDWPVSARGEGLVPPKSFISRNPEA  
SSLTAFLSLYEPSEAQHEEESIVYRGLAQNEFTSSHINQTFRFNSSHVSWDHQN  
RASQSLTEYSASNLMAVSALPSSRSLSPGLNEVPSSERNVPEWVQPGDIQG  
LKGQLVKLLELFGGSLPLVRVPPEYQKFFGRPLYLSEYGSCKLVHLVKKMAD  
KLTIEGKGHAKFLSLKGSSQRDKASVGGTPGKRENKKGKAQIESCGMNENE  
EAHYCHPNVGCSSDDGSSEEDGNHENIFREAGLVGQIKDFSRELELLVSYAG  
RILLDSFELLYKQRYKKVLDYRSLGVGGLEELIEKVREFAILHEEMGTNRKFFV  
VASCLGRTQ

>CmMFL1

MQILRPRFLSLPLGLQVLGPHPLPPPPPPRPTLFLIFTYNIASVSPPLFSLFLSHG  
DHPSSKEDNSMLVALHSFQSLIAIVAVFIIFPSSVEKACTHHHPNMMAASPNS  
SRALSLESSYTSANTMNPNMRRPPHAPNQSRALNEPAAILWDIENCPVPSDV  
RPEDVAGNIRMALRVHPVIKGAVTMFSAYGDFNAFPRRLREGCQRTGVKLV  
VPNGRDKAADKAILDMFLFALDNHPPSSILLISGDVDFAPALHILGQGRGYTIIL  
VIPAGVGSSALSANAGRFWWDWPVSARGEGLVPPKAVMSRGPGITRYLMGC  
HLGENPDAQNEAIVYRGISQNESSTDANINQMYSFSSSLVSRDSSRMPNSFSDY  
RSNLIGASYFPASICQSLPSTSNVLTGAFAGNVSETQEETLWVQP GDFQGLKG  
QLVRLLELAGGSLPLVRVPADYIKIYGRPLYVAEYGEFKLVNLLKKMADSLT  
VEGSGNRRFVCLRNFLLRPESVGPSARWGTGSKNNKEIQEENMDVVANVDM

VGSSDDFSDDERVIGQDERNTETDSGPAHEFNKQIERFKQELQELLVSYSKIL  
LSSFEALYEQRYKKSLDYNSFRVDELEELIEKVRDVLHHEEQSGRKFVAN  
CFNG

>NcMFL1

MTSAASSSLVSCPDCSSPSLSQIADFNMAPSIANNQQSKVPQGPVAILWDIENC  
PVPNDVRPEDVAGNIRMALRLHPVVGAVTHFSA YGDFNAFPRRLREGCQRT  
GVKLVDPNGRKAADKAILVDMFLFALDNPPPSSILLISGDVDFAPALHILG  
QRGYTIVLAIPAKVSVSSALSSAGRFVWDWPSVARGEFGMPAKSHTVYGAD  
VLHALPSHLTENPDSQIEDESIYQGLPNSGCTMRSNFSQVYYLDASCHVSRDV  
IRAAHSLSEYSSNAISLPQCSGSRSQLPVLNDMPMAASSDLNELQHDHWVQ  
PGDIQGLKGQIVKLELSAGSLQLVRLPSEYQKMFGRPLVYSEYGSCKLVDLL  
KKMADAVTVEGKGNRKFTLGSSAERSQRFLNPSAGSSRRERKGGKILREENI  
SCDGNELLCAGFQNMGCYSDESTDDERGGQDDRNLIPVGNVGTIKAVAD  
ACDRFKEELQELLVSYSCRILVGSFEALYEQRYKKSLLNYQSGVNGLEELIK  
KVKDVAEICEEQTKRRFLMVSGSGGLRSV

>BdMFL1

MVTMRTVFSFTGSVMSFEDKAVASRIASPPKAVVSESDLRMINATSNMEQPQ  
ANSQANAVVGPVAIFWDIENCVPSPDVRPDDVAGNIRMALRLHPVVKGAVTM  
LSAYGDFNAFPRRLREGCQRTGVKLVDPNGRKAADKAILVDMFLFALDN  
RPPSSIMLISGDVDFAPALHILGQRGYTIVLAIPSSVTVSSALSGAGSFVWDW  
SLARGVGSVAPRSLGHRAAEPGSLNSVTLGKYPDTEEEAIVYMGTSRNEYG  
GRTPGNQMYYYNSSQITGEPCKAFYTVEDGNCGTGTSSRSHNLSCGLNEVPEI  
DQGFTGEHSWWVRPGDLQGLKGQLIRLFELSGGSVPLVRVPSEYLKLFGRHL  
YVAEYGAVKLVHLFEKLAESFVVMGQQRKIICLRNSGDRNVKKYSSTPIILK  
NEKRLSATLGESAVGTCQQLSSSSDDFSEDEQNISPDVDGAYVFD EHLARCRT  
EIQDLLVCYKCRPLCDFESLYEQRYKKVLDYQSGVNGLEELAELKLDVV  
ELQLDEVSNMKLIKAKSAK

>OsMFL1

MSFEDKAMASRVASPPKSMASESDPSMMLAITSNMEHSQANNQSVSVLGPV  
AIFWDIENCVPSPDVRPEDVAGNVRMALRLHPVVKGAVTMLSAYGDFNAFPR  
RLREGCQRTGVKLVDPNGRKAADKAILVDMFLFALDNRPPSSIMLISGDV  
DFAPALHILGQRGYTIVLAIPSSVTVSSALSSAGSFVWDWPSLARGEIVAPRS  
IGRRFADPPGYQHGGNFGSFPDQNEEEAIVYMGTSRNECSGRTTSNQIYCYN  
SSQTTREPSKAFYTVTDGNCGTSSRSHNLACSLNEGPDVDQGLPDERSWWVR  
PGDLQGLKGQLLIRLFELSGGSVPLVRVPSEYLKLFGRHLYVSEYGAVKLIHLF  
EKLADSFVIGKGRKVICLRNSGDSNLKKYSTTPILKKNRGGSSILDEST  
IGTGQQLGSSSDDFSEDERNINPDVDGAYAFDSHLDNFRQEIQELLVCYSCPVP  
LGNFKSLYEQRYKKTLIYESFGVDGLEELVEKVKDVVELCEDQTSKRKYLIAN  
YRS

>ZmMFL1

MSSSEDKAMANRVTSHPKAMVSEPDNRMMMFRSNMEYSQTNGQTNAVLG  
PVAIFWDIENCPVPCDVRPEDVAGNVRMALRMHPVVRGAVTMLSAYGDFNA  
FPRRLREGCQRTGVKLVDPNGRKDAADKAILVDMFLFALDNHPPSSIMLISG  
DVDFAPALHILGQRGYTIVIAIPSSVTVSSALSSAGSFVWDWPSLARGEVVIP  
RSLVRRRLADYPCYVNSGNMGQFPDNQNEEEAIVYTGTSRNEYGGRRPPVNQM  
YCYNTSQTREPSKALYTVEDVNCGTSSRPHNLSRQSESPETDQGLTDEQSW  
WVRPGDLQGLKGLRIRLFELSGGCVPLVRIPSEYKLFGRHLYVSEYGAVKLV  
HLFEKLADSFVVVGKGHKMICLNSGDRNLKNYPSTPIILTKKNRGN SALEE  
STIGACQQLGSSSEDEPIINPDINVACVFNEHLDSFRREIQELLVCYSCPVP  
LGNFESLYEQRYKKIIDYESLGVGTGLEELVEKVKDVVDLHEDQTSKSKFLIAN  
YTTG

>CarMFL1

MIQKAMSSYYCNPSAAIVSVDSSSEQRMNYSSDLNMIANQRRSVTDGPVAVLW  
DIENCPVPSDVRPEDVASNIRMAIELHPVISGPVVTFSAYGDFNGFPRRVREGC  
QRTGVKLVDPNGRKDAADKAILIDMFLFVLDNKPPSTIVLISGDVDFAPALHI  
LGQRGYTVILVIPSGVYVNSALSSGGRFVWDWHSIVHGEFVVPKPRVVPYL  
MGCNVNDNSHLDGMNEDETILYRGSCCSNARESSVMVSQFQNECSGSMSC  
WPSNLRESLVCPPPGQLESTMWVAPGDLNGLKGLVKLLELSGGCIPLMRIPS  
EYQRNFSKPLFVSDYGVAKLVDLFKKMGDVITVDGKGNKRIVYLRNSKPNVI  
SPSSPVLLRREKKGKEPNEEMTYRGVSSDDLSDTGSVHSEARNLEEFKVELQD  
ILVSYCCRVLDCFEAIYKLRYPKPLDYKKMGVNHLEQLFDKVKDVVAIYED  
PATGSKLIGAF

>GrMFL1

MANLNASTSLVPSESSEQRGTVLVDSNVNVLPPPMNQKRTSSDGSVAIWLWDI  
ENCPVPSDVRPEDVAGNIRMALRVHPVIKGAVMMFSA YGDFNAFPRRLREGC  
QRTGVKLVDPNGRKDAADKAILVDMFLFALDNPPPSSIMLISGDVDFAPALHI  
LGQRGYIVILVIPAGVGVSSALCSAGKFVWDWPSVARGEGFVPPSKALMPPRG  
GPVDVARCFMGCHISDNPDGQNEEEAIVYKYGASQSCYNPRDFSTVSRSLAEYT  
SNPSVCLPSYPATFRSQSLPCGLNEASGCPGYDQNDTMWVQP GDINGLKGQ  
LVKLELSGGCMPLTRVPAEYQKIFGRPLYVAEYGALKLVNLFKKMGDTLAI  
DGKGHKKFVYLRNWKANPSAPPLVLRKDKKGGKIHEEITDVTAGAGSSDE  
FSDEERVVDERDQRKTDNLEQFKYELQEILVSYSCRIFLGCFEIYQQRYSK  
KTLDYRKLKSVKLEELFDKVRDVVVLHEEPVSKRKFCAVGS

>MtMFL1

MANPSRNSTQTMSLDSMEQKGNITESNASISQCPLSQHRNSLDGPVAILWD  
IENCPVPSDVRPDDVAGNIRMALQVHPVIQGA VTTFSAYGDFNSFPRRLREGC  
QRTGVKLVDPNGRKDAADKAILVDMFLFALDNPPPSSIMLISGDVDFAPALHI  
LGQRGYTIILVIPAGVGVSSALCNAGKFVWDWPIVARGEGFVPPSKALVAPRG  
GSVELAGYLMGCHINNDNTEGQNEEEAIVYRGMSQSYNSREFSMVSQSLSE

YNSPYMPCCLPTSMRSYSLPSGVNDVAGGPIPYSDNTECQMWWQP GDLNGLKG  
QLVRLLELSGGCVPLVRIPA EYQKIYGKTLFISDYGAFKLVDLFKKMDDAISVE  
GKGARRFVY LKNRKGGPSAPQVSLAKKDKK GKRELEENANVVNGSCSSDEL  
SDEERVVVEEHDERSFTGKGNKGRAVRCEID GSVEQFKCELQEILVSYSRILL  
SCFEAVYQQRYYKKQLEYQRYGVDKLEDLLEKVSDDVVTLHEEPVSKRKF LAA  
VDA

>CsMFL1

MAYSNTSASLVRLDSYEQTAHIAGSSTNTLEAPFNQPSRSSSDG HVAILWDIEN  
CPVPSDVRPEDVAGNIRMALRLHPLIEGVTTFSAYGDFNAFPRRLREGCQRT  
GVKLIDVPNGRKDAADKAILVDMFLFALDNPPPSFIMLISGDVDFAPALHILGQ  
RGYTVILVIPSGVGVSSALCNAGKFWWDWPSVARGEGFAPPSRALIPPRAGPA  
GYFMGCQINDNVQNEEEAIVHRGVSQSYYSSRDLSIVSQSLTEYNSSYITPYN  
SALRSNSVPSGLNEASAGPAISADLSDSTLWVQP GDLNGLKGQLVKLELSGG  
CLPLNRVPADYHKVFG RPLYVSEYGVV KLVNLFNKMSDTIAVDGKGRKQK FV  
YLWNWKTGPSAPPMILTRKDNKKGKASQEECTDNVTGNVSSDELSD EERIVT  
EEHDERLNQEKTNLETMTQCDND YQLEQFKYELQEILVSYSRIFVGC FEEIYE  
QRYKKPLDFQKFGVDQLEELFEKLDV VVLHEEPVSKNKFLAAVGG

>SIMFL1

MELPPTSSFGLSSESVEQSGKLMANTNAKALATFSLQSPNSSQGPVAILWDIEN  
CPVPSDVRPEDVAGNIRMTLRVHPVIKGA VTLFSA YGDFNSFPRRLREGCQRT  
GVKLVDVPNGRKDAADKAILVDMFLFALDNPPPSIMLISGDVDFAPALHILG  
QRGYTVILVIPSGVGVSSALCNAGR FVWDW PSVVRGEGFVPPAKALMPCRGG  
VSDITGILMGCCQINDNPDGQEQE DEAILYRGLS QSYYNSREFSMISHSLSEYNT  
TAISMPCYPTGMRTHSLPSGLNEVSAGGSSSHEQSDLTWVQPGDINGLKGQLV  
KLELSGGCLPLTRVPAEYQKIYGRPLYISEYGA AKLVNLLKKMSDAISVGGK  
GQKKFVYLHNSCAVPSAPPITILKRDNKGKGTQEGNADVVTG VGVSSDEFSD  
ERGPIKEHGGSC EKS NMVEKSLENFKYELQEILVSYSRIFLGC FDAIYQQR YK  
RQLDYESFGVVELEQLLAKVKDIVIVQE EPVSKRKF LAAVGA

>StMFL1

MELPPTSSFGLSSESSESEQSGKLMANTNAKALATFSLQSPNSSQGPVAILWDIEN  
CPVPSDVRPEDVAGNIRMTLRVHPVIKGA VTLFSA YGDFNAFPRRLREGCQRT  
GVKLVDVPNGRKDAADKAILVDMFLFALDNPPPSIMLISGDVDFAPALHILG  
QRGYTVILVIPSGVGVSSALCNAGR FVWDW PSVVRGEGFVPPAKALMPCRGG  
VSDITGILMGCCQINDNSDGQHEDEAILYRGLS QSYYNAREFSMISHSLAEYNT  
TAISMPCYPTGMRTHSLPSGLNEVSAGGPSSHDQSDLTWVQPGDINGLKGQL  
VKLELSGGCLPLTRVPAEYQKIYGRPLYVSEYGA AKLVNLLKKMSDAISV G  
GKGQKKFVYLRNSCAVPSAPPITILKRDNKGKGTQEGNADIVTGVSSDEFSD  
DERGPIKEHGGSC EKS NMVEKSLENFKYELQEILVSYSRIFLGC FDAIYQQR Y  
KRQLDYKIFGVVELEQLLAKVKDVVIVQE EPVSKRKF LAAVGA

>PvMFL1

MATPSKSTPITSSDTFEQKGKTITESNATILQCPLNQQ LHSSLDGPVAILWDIEN  
CPVPNDVRPEDVAGNIRMALRVHPVIKGA VMLFSA YGDFNSFPRRLREGCQR  
TGVKLIDVPNGRKDAADKAILVDMFLFALDNPPPSSIMLISGDVDFAPALHILG  
QRGYTVILVIPAGVGVSSALCNAGKFVWD WPSVARGEGFVPPSKALMPPRGG  
PIELAGYLMGCHINDYSEGLNEEEAIVYRGMSQSYNSREFSLVSQSLSEYNC  
GTSNGACLPTTMRSHSLPSVFNDVQGGSMMPSGDPNECQLWVQP GDLNGLK  
GQLVRLLELFGGSLPLARVPAEYQKIYGRPLYVTEYGAIKLVNLFKMGNAM  
TVEGKGNRKVFYLRNWKAGPSAPPLSLAKKDKKGGTQEENVNVVTGGCSS  
DEFSDEERVVIEEDDERNFTGKGNQGTAAKCEVDD CVIEKFKYELQEILVSY  
CRISLGC FEDIYQQR YKKQLEYKRFVVDKLEDLIEKVSDVVLHEEPVSRSKF  
LAAVGG

>GmMFL1

MATPSKSTPIIASDTLEQKGKTMTESNANILQCPLNQQ LYSSLDGPVAILWDIE  
NCPVPCDVRPEDVAGNIRMALRVHPVIKGA VMMFSA YGDFNAFPRRLREGC  
QRTGVKLIDVPNGRKDAADKAILVDMFLFALDNPPPSSIMLISGDVDFAPALHI  
LGQRGYTVILVIPANVGVSSALCNAGKFVWD WPSVVRGEGFVPPSKALVPPR  
GGPVELAGHFMGCHINDNSDGQNEEEAIVYRGMSQSYNSREFSMVVSQSLSE  
YNNGTLTNTTCLPTTMRTHSLPSVLNDVPGGSMMPSSDLNECQLWVQP GDLN  
GLKGQLVRLLELFGGCLPLARVPAEYQKLYGRPLYVSEYGAIKLVNLFKMG  
DALAVEGKGNRKVFYLRNWKAGPSAPPLSLAKKDKKKEAQEENVNIVTGG  
CSSDEFSDEERVVIEEHDERS CIGKGNQGRAAKCEIDDRFIEQFKYELQEILVSY  
SCRIFLGC FEAIYQQR YKKQLEYQRFGVDKLEDLFEKVSDVVLHEEPVSKKK  
FLAAVGG

>VvMFL1

MADPNTSALIVSSESSEQSGTHMMNPSANPLQPLL NQQGRTSPHGSVAILWDI  
ENCPVPSDVRPEDVAGNIRMALRVHPVIRGAVTMFSA YGDFNAFPRRLREGC  
QRTGVKLIDVPNGRKDAADKAILVDMFLFALDNPPPSSIMLISGDVDFAPALHI  
LGQRGYTVILVIPSGVGVASALCNAGR FVWD WPSVARGEGFVPPTKVLIPPRG  
GTADIAGCLMGCHINDNPDGQNEEEAIVYRGMSQGYSTRDFSISQSLSEFNS  
TSSITMSCFPPTLRSQSLPSGLNEASAGPISYGEQNESTLWVQP GDLNGLKAQL  
VKLLELSGGCLPLARIPSDYQKLFGRPLYVSEYGAFKLVNLFKMGADTLAVE  
GKGHRKLVYLRNSKAGPSAPPLIMARKEKKGKGIQEENMDNITGCGSSDEFS  
DDERVVVEEHDERRREEKFGLLASRCEINDQNIEQFKHELQEILVSYSCRIFL  
GC FEAIYQQR YKKPLDYRKFGVNELEGLFDKVKDVVLHEEPVTKRKF LDAVGG  
G

>AcMFL1

MSYEDKPMASHCPSPPKAVAVELDSSKLIPIGSNMEHSQAPKDQGRAVHCPV  
AILWDIENCPISDVRPEDVSGNIRMALRVHPVVGGA VTMFSA YGDFNAFPRR  
LREGCQRTGVKLVDVPNGRKDAADKAILVDMFLFALDNRPSSIMLISGDVDF

ALALHILGQRGYTVILVIPSGVRVSSALS NAGRFVWDWPSVARGEGFVPPKAV  
MPYADNQNEEEIITYRGISQNEYGTTTRGSVNQMYHCNLSYSSRELSRASEYSN  
SCSNFILEPYFNSSRSHSLPSGVSEGLMGGDQNSTQEHSWWVRP **GDVQGLKG**  
**QFIRLLEMSGGSLPLVRVPSEY LKVFGRPLYMAEYGVYKLVNLIK KMTDALIV**  
**VGKGHKMLCLRNTDDRQNIKKCPCPSTPILKREEKKGKSVVVEENSESTD**  
**EFSDDERNNATYENN** **DHLESFKQEVQELLVCYSCPVPISALESLYEQRYKKAL**  
**DYQSGVDGLEELIEKLRDVVELREDWVSKRKF LASC**

**>MaMFL1**

MAYEDKAMASRSPSPKIITPNISNSICTSSETEPVQTQNNQSGTLHG **PVAILWDI**  
**ENCPVPSDVRPEDVAGNIRMALRVHPVIKGAVTMFSAYGDFNAFPRRLREGC**  
**QRTGVKLVDPNGRDKAADKAILVDMFLFALDNQPPSSIMLISGDVDFAPAL**  
**HILGQRGYTVILVIPAGVGVSSALS NAGRFVWD** **WPSVARGEGFVPAKAYMP**  
RIPDHAGCLSNCNIGANLDFQNEEEAIVYKGVSQNEYGGENNDRGYCYNSNS  
MRRGLDRISYSMTEYSGSNSVNGPYFTSSRSQSLPSGLSEGAGMDQNV TQE Q  
AWWVRP **GDLQGLKGQIVRLLEMSGGSMPLVRVPSEY LKFFGRPLYVAEYGV**  
**CKLVNLI RKMADALVVVGKGHKLLCLRNAAYFDAKKYPGTSALVRKDKK**  
**GKQVLEENMDISISPHSGCSSDELSDDGKNDDPALGA ADEFED** **QFENVRHEVQ**  
**ELLVCYSCPVPLGTFEALYKQRYKKDLDYRSYGVDSLEELMEKLRDIVE LCV**  
**DQGSKRKFVLVLSSES**

**>AtMFL1**

MIQNAMSSYYNPSAATTLVSIDSDEQRRIKYLADLNMIATQRHSSTDG **PMAIL**  
**WDMENCPVPSDVRPEDVASNIRMAIQLHPVISGPVNFSA YGDFNGFPRRVRE**  
**GCQRTGVKLVDPNGRDKASDKAILIDMFLFVLDNKPPATIVLVSGDVDFAPA**  
**LHILGQRGYTVILVIPSSVYVNSALS NAGKFWWDWHSIVHGEGFVPRCKPRVV**  
**PYLMGCNIGDNSNMDGLNEDETILYRGHLESTMWVAPGDLNGLKGQLVKLL**  
**ELSGGCIPLMRVPSEYQRKFSKPLFVSDYGVAKLVDLFKKMSDVIVVDGKGN**  
**KRFVYLRNSKPNIISSPVLLRRERK GKPEPNGVTTNGGVSSDEMSDTG SVQS**  
**ERNLEEFK FELQDILVSYCCQVQMDCFEAIYKLR YKRPLAYTNMGVNHLEQL**  
**FDKLRDVVAIHEDPATGRKLISPV**