The characterisation of SMG1 in the nonsense-mediated mRNA decay pathway of *Physcomitrella patens*

James Peter Brook Lloyd

Submitted in accordance with the requirements for the degree of PhD

The University of Leeds

School of Biology

September 2013

The candidate confirms that the work submitted is his own, except where work which has formed part of jointly authored publications has been included. The contribution of the candidate and the other authors to this work has been explicitly indicated below. The candidate confirms that appropriate credit has been given within the thesis where reference has been made to the work of others.

The third chapter of this thesis is based on work from the jointly authored publication: **Lloyd, J.P.B. and Davies, B.** (2013) SMG1 is an ancient nonsense-mediated mRNA decay effector. *The Plant Journal*, 'Accepted Article', doi: 10.1111/tpj.12329. All the experimental work performed in this publication was performed by the PhD candidate. Experimental designed and the manuscript are the joint work of the candidate and the co-author of the manuscript (Brendan Davies) who was the candidates PhD supervisor.

Acknowledgements

Thanks first and foremost to Prof Brendan Davies for giving me the opportunity to work on this project in his lab and for his valuable comments and suggestions with the experiments and the write-up of this thesis. Thanks are owed to the whole Davies lab, past and present for their help and for making the Davies lab a pleasure to be a part of. In particular, I would like to thank Dr Samantha Rayson for he help introducing me to faculty life and the NMD work in the lab, Dr Barry Causier for encouraging me to take part in Friday-Football in the Centre For Plant Sciences and for giving advice on many experimental techniques, especially yeast two-hybrid and finally, Mary McKay for helping with moss tissue culture media and the sub-culturing of moss. Special thanks also need to go to my co-supervisor Dr Andrew Cuming for his advice and help, in particular with moss experiments and my assessor Dr Christopher West for discussions regarding my research. Finally, I am indebted Dr Yasuko Kamisugi for her kind help teaching me how to culture and transform moss and the ongoing patience this must have taken.

This research in Chapter 5 has been carried out by a team which has included Dr Daniel Lang and Dr Andreas Zimmer of Ralf Reski's lab group at the University of Freiburg. The other members of the group and their contributions have been as follows, Dr Daniel Lang and Dr Andreas Zimmer mapped the RNA-seq reads, performed analysis including fold-change and significance testing as well as functional clustering of these genes and Ralf Reski supervised their work. My own contributions, fully and explicitly indicated in the

thesis, have been to evaluate the RNA-seq data analysis and to analyse it for biological significance and to prioritise and perform downstream analysis such as phenotype analysis and conformation of expression changes (Chapter 5).

I would also like to thank my girlfriend Jo Franklin for kindness and support, especially during the write-up. Additionally, I would like to thank my friends in the faculty for the many good times and jokes, in particular the 'Luncheon group': Rupesh Paudyal, Helen Beeston, Nicki Ooi, Dr Lizzie Glennon and Dr Laura White and the 'Newbies *et al.*': Laura McFarlane, Seb Eves-van den Akker, Jess Marvin, Sean Stevenson, Jamie Kelly, Suruchi Roychoudhry, Dr Pete Thorpe and many others who have made life in the faculty amazing. I would also like to thank Anna Perrett for proof-reading my results chapters.

I would also like to thank Lord David Sainsbury and his charity, The Gatsby Charitable Foundation for funding and supporting me throughout my PhD and the Gatsby mentors: Prof Jane Langdale, Prof Ottoline Leyser, Prof Liam Dolan and Prof Nick Talbot for their constructive criticism and career advice.

Abstract

The regulation of gene expression is not simply confined to the activity of a promoter but can occur at many stages, including mRNA degradation. Nonsense-mediated mRNA decay (NMD) is a eukaryotic mRNA decay pathway. It was first characterised as a pathway degrading transcripts with premature stop codons arising from mutations or alternative splicing. However, NMD also directly targets many 'non-aberrant' transcripts and is important for normal growth and development. For example, NMD is needed for a normal response to pathogens in *Arabidopsis thaliana* and NMD is regulated during mammalian brain development. In animals, it is well known that the kinase SMG1 activates the NMD pathway when a premature stop codon is recognised but no NMD-associated kinase has been characterised outside the animal kingdom. Reported here is that SMG1, whilst missing from fungi and *A. thaliana*, is ubiquitous in the plant kingdom, functions in the NMD pathway of moss and is needed for normal moss development. An RNA-seq analysis of transcripts regulated by SMG1 in moss revealed that NMD is important for regulating the unfolded protein response and is also involved in the DNA repair pathway. Taken together, SMG1 has been demonstrated to be an ancient kinase, which functions in the NMD pathway in moss. Furthermore, NMD is important for normal moss development.

Table of Contents

List of Tables

List of Illustrative Material

List of Abbreviations

~	Approximately
%	Percentage
^{o}C	Degree Celsius
Δ	Deletion
µ	Micro
3-AT	3-amino-1,2,4-triazole
A. lyrata	*Arabidopsis lyrata*
A. thaliana/At	*Arabidopsis thaliana*
AD	Activation domain
Agrobacterium	*Agrobacterium tumefaciens*
ApE	A plasmid editor
ARK?	*A. thaliana replacement kinase of SMG1?*
AS	Alternative splicing
AT	diammonium tartrate
ATM	ATAXIA TELANGIECTASIA MUTATED
ATR	ATAXIA TELANGIECTASIA AND RAD3-RELATED
AZC	L-azetidine- 2-carboxylic acid
BD	Binding domain
bp	Base pair
BR	Biological replicate
c	Centi
C. elegans	*Caenorhabditis elegans*
C. merolae	*Cyanidioschyzon merolae*
cDNA	complementary DNA
CHX	Cycloheximide
COG	Clusters of orthologous groups
CPuORF	Conserved peptide uORF
CT	C-termini
DGE	Differential gene expression
DMSO	Dimethyl sulfoxide
DNA	Deoxyribonucleic acid

DNA-PKcs	DNA-dependent protein kinase, catalytic subunit
dNTPs	Deoxyribonucleotides triphosphates
Drosophila	*Drosophila melanogaster*
DSE	Downstream sequence element
E. coli	*Escherichia coli*
E. siliculosus	*Ectocarpus siliculosus*
EDTA	Ethylenediaminetetraacetic acid
EF1α	elongation factor 1 α
eIF	eukaryotic initiation factor
EJC	Exon-junction complex
eRF	eukaryotic release factor
FATC	FRAP, ATM, TRRAP C-terminal
g	Gram
g	Relative centrifugal force
G. lamblia	*Giardia lamblia*
gbrowser	Genome browser
GC	Guanine-cytosine
GFP	Green fluorescence protein
GO	Gene ontology
GT	Gene targeting
H	Histidine
HSF	heat-shock transcription factor
HSP	heat-shock protein
ID	Identifier
KAN	Kanamycin resistance gene
Kb	Kilo base (pair)
KD	Kinase domain
KOG	Eukaryotic orthologous groups
l	Litre (decimetre)
L	Leucine
LB	Lysogeny Broth or Luria-Bertani broth
LECA	Last eukaryotic common ancestor
LUC	*Luciferase*
m	Milli (if prefix) or metre

M	Moles per decimetre (litre)
mRNA	Messenger RNA
MTA	mRNA adenosine methylase
n	Nano
NMD	Nonsense-mediated mRNA decay
nt	Nucleotide
NT	N-termini
NTC	No template control
ORF	Open reading frame
P. patens/Pp	*Physcomitrella patens*
PABP	polyA-binding protein
PCR	Polymerase chain reaction
PEG	Polyethylene glycol
PIKK	phosphatidylinositol 3-kinase-related kinase
PP2A	Protein phosphatase 2A
PTB	Polypyrimidine tract-binding protein
PTC	Premature termination (stop) codon
qRT-PCR	quantitative RT-PCR
rpm	Revolutions per minute
RNA	Ribonucleic acid
RNAi	RNA interference
RT-PCR	reverse transcriptase-PCR
S. cerevisiae	*Saccharomyces cerevisiae*
S. pombe	*Schizosaccharomyces pombe*
SA	Salicylic acid
SD	Synthetic defined
SMG	*suppressors with morphological defects on genitalia*
SMGL	*SMG-lethal*
SQ	Serine-glutamine
sqRT-PCR	semi-quantitative RT-PCR
SR	Serine-arginine
TAE	Tris- Acetate EDTA
Tm	Tunicamycin
TOR	TARGET OF RAPAMYCIN

TPL	TOPLESS
TPR	Tetratricopeptide repeat
TQ	Threonine-arginine
TR	Technical replicate
TRRAP	TRANSFORMATION/TRANSCRIPTION DOMAIN-ASSOCIATED PROTEIN
uORF	upstream ORF
mORF	main ORF
UPF	up-frameshift
UPR	Unfolded protein response
UTR	Untranslated region
UV	Ultraviolet
v/v	Volume per volume
VIGS	Virus-induced gene silencing
W	Tryptophan
w/v	Weight per volume
WT	Wild type
YFP	YELLOW FLUORESCENT PROTEIN
YPDA	Yeast extract - Peptone - Dextrose - Adenine

1 General introduction

1.1 Nonsense-mediated mRNA decay

Regulating gene expression allows organisms to respond to changes in their surroundings and is necessary for development in multicellular organisms. Gene expression can be controlled at multiple steps, including recruitment of RNA polymerase to the promoter of a gene or stability of a protein. Transcription factors bind to the promoters of target genes and are important regulators of gene expression by either helping to recruit or exclude RNA polymerase by acting as activating or repressive transcription factors, respectively (Krogan and Long, 2009). This can work through altering the chromatin state of the target genes promoter. Although attention has mainly been paid to transcriptional regulation, the importance of post-transcriptional gene regulation is being increasingly recognised. Alternative splicing of a primary transcript can lead to several different transcripts, differing in the protein sequence and/or stability of the transcript. The amount of a transcript present in a cell is the product of both the rate of transcription and the rate of decay. Several pathways have been identified that can target a selected transcript for decay, therefore altering the level of expression without necessarily affecting the rate of transcription. The RNA silencing pathway, for example, uses small RNAs, which have complementarity to specific mRNAs to target the specific mRNAs for destruction (Urlaub et al., 1989; Chen, 2009; Cheng and Maquat, 1993; Huntzinger and Izaurralde, 2011; Carter et al., 1996; Belostotsky and

Sieburth, 2009). Another RNA decay pathway is the nonsense-mediated mRNA decay (NMD) pathway. This pathway was first identified as an RNA quality control mechanism (see Section 1.1.1) but it has now become clear that NMD has a role in controlling the expression of physiologically important transcripts (see Section 1.3). The NMD pathway is initiated when a stop codon is recognised to be in an 'unusual' contexts by the terminating ribosome, although what constitutes an 'unusual' context varies between species and many exceptions to the proposed rules have been observed (see Section 1.2).

1.1.1 NMD as an mRNA surveillance pathway

Unexpected results from early mutant screens led to the identification of NMD. If a mutation introduced a nonsense codon, also known as a premature termination codon (PTC), not only would the level of the corresponding full-length protein be reduced but also the mRNA level would be lower than from the wild-type (WT) copy of the gene (Belgrader et al., 1994; Losson and Lacroute, 1979; Carter et al., 1996; Gozalbo and Hohmann, 1990; Maquat, 2004). These observations suggested a novel way to control gene expression, dependent on the presence of a PTC. The decrease in mRNA levels was shown to be the result of a translation-dependant decrease in stability of the transcript rather than a decrease in the rate of transcription (Singh et al., 2008; Losson and Lacroute, 1979; Gozalbo and Hohmann, 1990). However the molecular mechanism of this NMD pathway was unclear.

Mutant screens in budding yeast (*Saccharomyces cerevisiae*) and

Table 1.1: A table of NMD effectors

NI represents not identified.

	NMD effector	Budding yeast	*C. elegans*	*Drosophila*	*Vertebrates*	*A. thaliana*
'Core' NMD effectors	UPF1	UPF1	SMG2	UPF1	UPF1	AtUPF1
	UPF2	UPF2	SMG3	UPF2	UPF2	AtUPF2
	UPF3	UPF3	SMG4	UPF3	UPF3a	AtUPF3
					UPF3b	
Kinase	SMG1	NI	SMG1	SMG1	SMG1	NI
SMG5-7 family	SMG5	EBS1/ SMG7	SMG5	SMG5	SMG5	AtSMG7 and AtSMG7L
	SMG6		SMG6	SMG6	SMG6	
	SMG7		SMG7	NI	SMG7	
Additional regulators	SMG8	NI	SMG8	SMG8	SMG8	NI
	SMG9	NI	SMG9	SMG9	SMG9	NI
	SMG1L	NI	SMG1L	NI	SMG1L	AtSMG1L
	SMG2L	NI	SMG2L	SMG2L	SMG2L	NI
Core EJC	Mago	NI	Mag	Mago nashi	Magoh	AtMago
	Y14	NI	Y14	Y14	Y14	AtY14
	eIF4AIII	NI	eIF4AIII	eIF4AIII	eIF4AIII	AteIF4A3
	MLN51/ CASC3/ Barentsz	NI	Barentsz	Barentsz	Barentsz	AtBarentsz1 and AtBarentsz2

Caenorhabditis elegans identified components of the NMD pathway (Table 1.1). In budding yeast, the mutants were designated *up-frameshift 1*, *2* and *3* (*upf1*, *2* and *3*) because transcripts with nonsense codons, introduced from frameshifts, were over-expressed in the mutant backgrounds (Trcek et al., 2013; Losson and Lacroute, 1979; Leeds et al., 1992). In *C. elegans*, seven genes involved in the NMD pathway were identified through mutant screens originally looking to identify genes suppressing multiple mutant phenotypes in the nematode (Hwang et al., 2010; Hodgkin et al., 1989; Hosoda et al., 2005; Pulak and Anderson, 1993; Cali et al., 1999). These mutants all had abnormal genitals so were named *suppressor with morphological effect on genitalia* (*smg*) (Rufener and Mühlemann, 2013; Hodgkin et al., 1989; Durand and Lykke-Andersen, 2013; Pulak and Anderson, 1993; Cali et al., 1999). It soon became apparent that the NMD pathways of budding yeast and *C. elegans* were related, with *SMG2*, *SMG3* and *SMG4* from *C. elegans* encoding homologues of *UPF1*, *UPF2* and *UPF3* from budding yeast, respectively (Hwang et al., 2010; Serin et al., 2001; Hosoda et al., 2005; Aronoff et al., 2001; Page et al., 1999). From hereon in, NMD effectors will be named by the organisms they were first discovered in, for example, UPF1 for SMG2.

1.1.2 An overview of the molecular mechanism of the NMD pathway in animals

A model was developed in *C. elegans* centring around UPF1 (SMG2) and its phosphorylation (Rufener and Mühlemann, 2013; Page et al., 1999; Durand and Lykke-Andersen, 2013). This model proposed that UPF1 phosphorylation is

dependent on SMG1, UPF2 (SMG3) and UPF3 (SMG4) while dephosphorylation requires the three related proteins SMG5, SMG6 and SMG7 and this cycle of phosphorylation and dephosphorylation is needed for destruction of a PTC-containing transcript (Gatfield et al., 2003; Page et al., 1999; Rehwinkel et al., 2005; Metzstein and Krasnow, 2006). Subsequent studies in *C. elegans* and mammals have shown that SMG1 is the kinase that phosphorylates UPF1 (Gatfield et al., 2003; Grimson et al., 2004; Chen et al., 2005; Yamashita et al., 2001; Rehwinkel et al., 2005; Metzstein and Krasnow, 2006) and that UPF2 and UPF3 stimulate this by forming a complex with UPF1 (Kashima et al., 2006), SMG5, SMG6 and SMG7 are likely to be involved in recruiting a phosphatase to dephosphorylate UPF1 (Anders et al., 2003; Page et al., 1999; Chiu et al., 2003). This model is an over-simplistic view of NMD in animals but forms the foundations of the signalling pathway that occurs after PTC recognition (Figure 1.1). Once a PTC has been recognised by a ribosome, UPF1 is recruited to the ribosome through interaction with eukaryotic release factor 1 and 3 (eRF1 and 3), which in turn recruit SMG1 (Kashima et al., 2006). UPF1 interacts with UPF3 indirectly via UPF2 and formation of this complex induces the phosphorylation of UPF1 by SMG1 (Kashima et al., 2006). SMG5, SMG6 and SMG7 have 14-3-3-like domains at their N-termini (Fukuhara et al., 2005; Jonas et al., 2013), which then bind to phosphorylated UPF1 and recruit both the dephosphorylation machinery and the decay machinery (Unterholzner and Izaurralde, 2004; Fukuhara et al., 2005). SMG6 binds to the N-terminus of phosphorylated UPF1 (Okada-Katsuhata et al., 2011) and acts as an endonuclease cutting the transcript near the PTC (Huntzinger et al., 2008; Eberle et al., 2009) while SMG7 binds the C-terminus, in complex with SMG5,

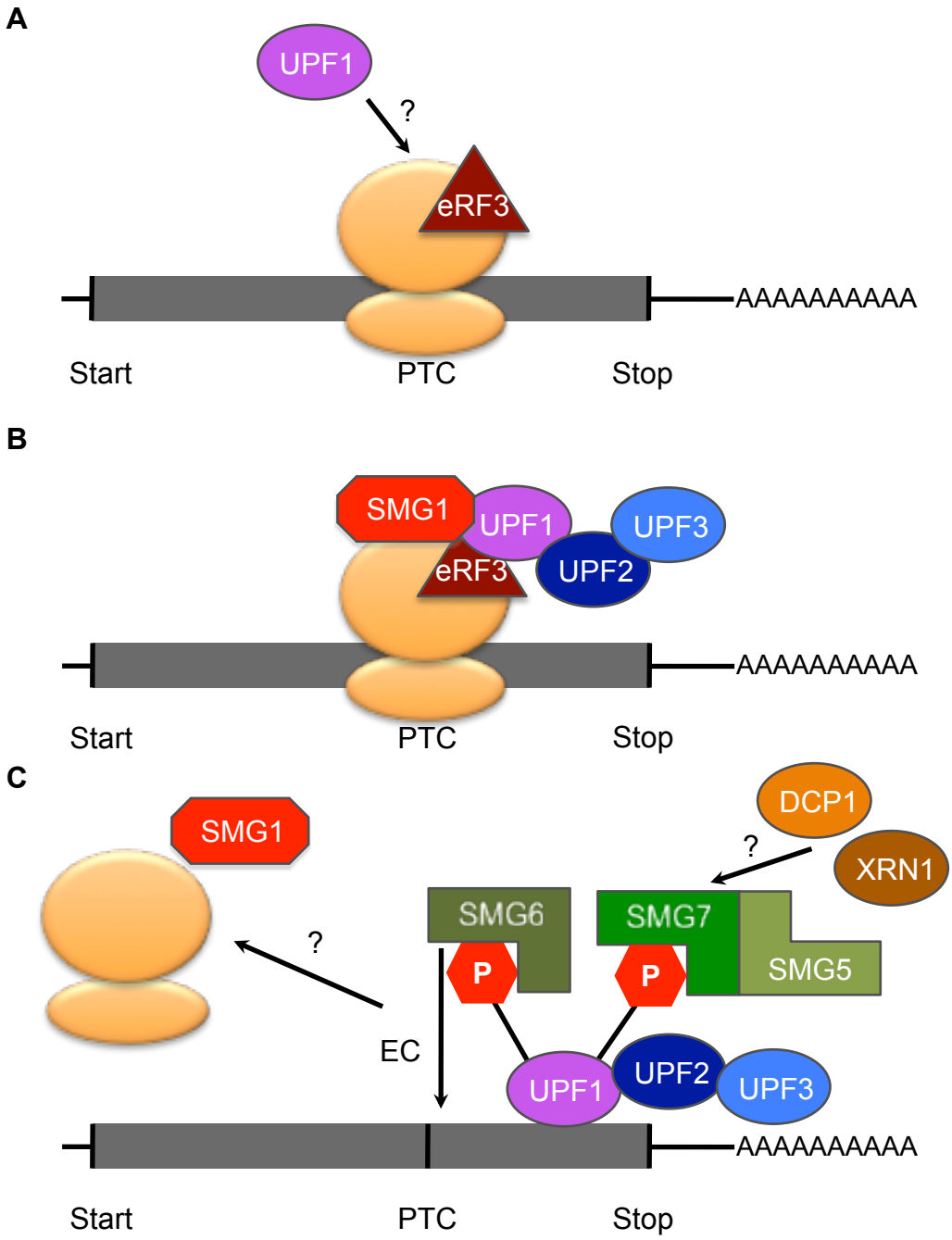

Figure 1.1: Model of NMD in animals

(A) Upon termination at the PTC, the ribosome stalls and UPF1 might be recruited at this stage by eRF3. **(B)** UPF1 forms a complex with SMG1, UPF2 and UPF3, which stimulates the kinase activity of SMG1. **(C)** Phosphorylated UPF1 recruits SMG6 to the N-terminus and SMG7:SMG5 is recruited to the C-terminus. *Legend continued on next page*

Figure 1.1: *Continued…*

The ribosome and SMG1 might have disassociated at this stage and UPF1 binds to the transcript. SMG6 causes an endonucleolytic cleavage (EC) slightly downstream of the PTC. SMG7:SMG5 may recruit the decapping complex and the endonuclease XRN1.

and appears to recruit the decapping complex (Nicholson and Mühlemann, 2010). This extended model of NMD depicted in Figure 1.1 appears to hold true in animals where UPF1-3, SMG1 and the SMG5-7 family are present (Table 1.1), but cannot be the case in organisms like budding yeast and *A. thaliana*, which lack some of the key components of the pathway, such as SMG1 (Table 1.1). UPF1-3 and SMG7 are important for NMD in many eukaryotes, however only UPF1 appears to be universal in eukaryotes and the simple pathway described in Figure 1.1 cannot be conserved across all organisms where NMD operates (Grimson et al., 2004; Chen et al., 2008; Delhi et al., 2011; Conti and Izaurralde, 2005).

Excavates such as the parasites *Giardia lamblia* and trypanosomes represent the most reduced NMD pathways yet described. It is not yet clear if the NMD pathway is functional in these organisms and if it does it would mean that these pathways have only a few recognisable components discovered in other organisms. For example, in trypanosomes, the NMD pathway appears to consist of UPF1 and UPF2 (Delhi et al., 2011), while the only recognisable NMD effector in the genome of *G. lamblia* is UPF1 (Chen et al., 2008). These organisms represent extremely reduced examples of the NMD pathway and have not been well studied. Differences between the better studied NMD pathways of model organisms will be examined further in Section 1.2.

A question central to understanding NMD is ‘How does the terminating ribosome differentiate between an authentic stop codon and a PTC?’ Early work lead to the proposal of two broad models trying to explain how a stop codon

could be identified as ‘aberrant’. These are 1) the long 3’ UTR model and 2) the downstream sequence element model. Proposed molecular mechanisms for each have now been described in different organisms and there appears to be some degree of evolutionary conservation, although early reports suggested that these were lineage specific mechanisms (Mühlemann et al., 2008).

1.1.3 The downstream sequence element model

The downstream sequence element (DSE) model proposes that a sequence element downstream of a PTC binds to a protein or protein complex and signals to the terminating ribosome that this stop codon is aberrant and that NMD should be activated. In budding yeast, a DSE was cloned and found to bind the RNA binding protein HRP1 (Zhang et al., 1995; Peltz et al., 1993; González et al., 2000), which interacts with the NMD effectors UPF1 and UPF2 to activate NMD (Figure 1.2A) (González et al., 2000; Wang et al., 2006). However, this model is specific to only a few transcripts in budding yeast and cannot account for a significant proportion of transcripts targeted to NMD. A variant of the DSE model involves exon-exon junctions acting as the DSE if the junction is greater than 50-55 nucleotides downstream of the stop codon in mammals (Mühlemann et al., 2008; Brogna and Wen, 2009). At many exon-exon junctions, a protein complex called the exon-junction complex (EJC) is deposited, consisting of four core proteins (Table 1.1), which are involved with activating NMD at the upstream terminating ribosome (Gehring et al., 2003; 2005; Ferraiuolo et al., 2004). In mammals, UPF3 is an auxiliary EJC component and acts to link the EJC to the NMD pathway (Figure 1.2B) (Buchwald et al., 2010; Gehring et al., .

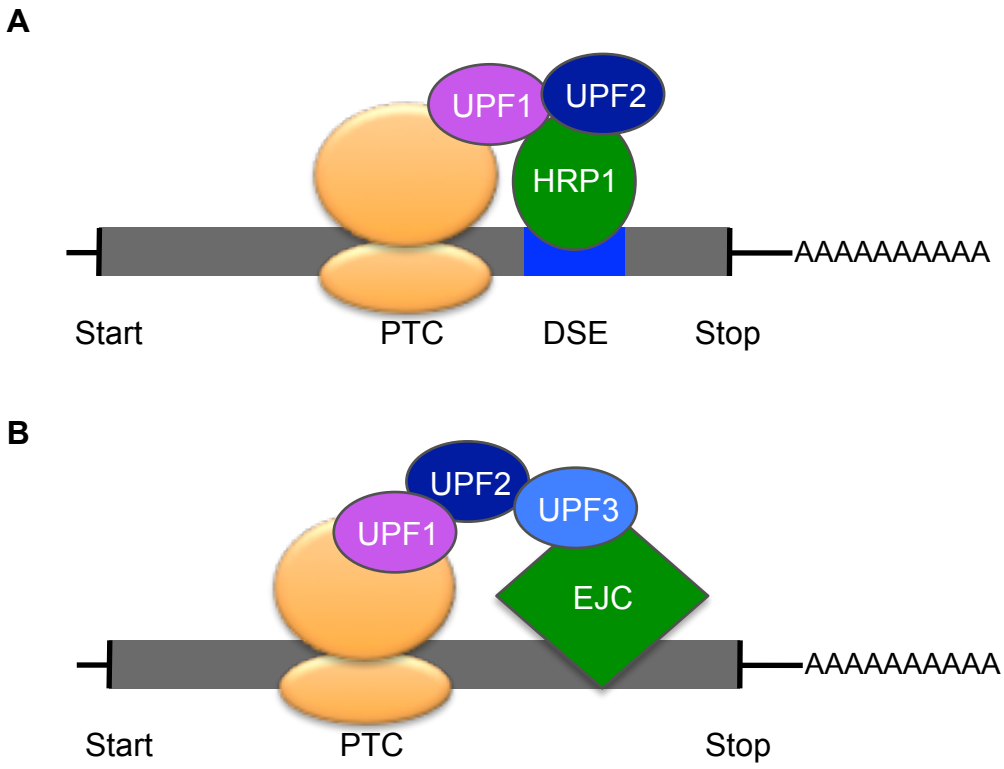

Figure 1.2: DSE and EJC models of NMD activation

(A) The DSE model of NMD activation in budding yeast. HRP1 binds to a DSE (blue) in budding yeast transcripts and binds UPF1 and UPF2. **(B)** The EJC model of NMD activation in mammals. The EJC binds close to an exon-exon junction with UPF3 as an auxiliary component and UPF2 and UPF1 link it to the terminating ribosome.

2003; Kashima et al., 2006). If a ribosome terminates upstream of an EJC then UPF3 and UPF1 interact via UPF2 and this activates the phosphorylation of UPF1 by SMG1 and recruitment of downstream decay machinery (Kashima et al., 2006). However, there are several examples where this does not act as a targeting feature (Brogna and Wen, 2009). It remains to be determined if UPF1 is recruited to every termination event or only at PTC termination events (Nicholson et al., 2010).

An exon-exon junction acting as a DSE is not universally conserved across eukaryotes. Budding yeast have few introns in their genome and have lost their homologues of the EJC. Early experiments studying model transcripts in *C. elegans* and *Drosophila* suggested that the EJC model did not apply in these organisms either (Longman et al., 2007; Gatfield et al., 2003), however, subsequent work has found that the EJC does have a role in NMD in *Drosophila* (Saulière et al., 2010). This suggests that the EJC might have a conserved function in the NMD pathway across animals. It also acts as a warning that it is unwise to use only a small number of model transcripts to make general conclusions about the mechanisms of NMD.

Even organisms as evolutionarily divergent from animals as flowering plants use the EJC in determining a PTC. NMD reporter genes consisting of the GFP coding sequence and an NMD targeting feature have been expressed in a tobacco leaf transient assay. Briefly, *Agrobacterium* carrying the NMD target construct is infiltrated into a tobacco leaf and expression of the NMD target is measured by GFP or transcript level. Different components of NMD can be

silenced used virus-induced gene silencing (VIGS) or expression of a dominant negative variant of UPF1 (Kertész et al., 2006; Kerényi et al., 2008). Using this system, it was shown that reporter constructs with introns downstream of the stop codon can trigger NMD in angiosperms and that this is dependent on the core components of the EJC (Kertész et al., 2006; Kerényi et al., 2008; Nyikó et al., 2013). Taken together, these data suggest that in both plants and animals the EJC downstream of a stop codon can trigger NMD and therefore that this mechanism could be evolutionarily ancient. However, more work is needed to establish whether the mechanisms linking the EJC and NMD are conserved between mammals, invertebrates and flowering plants.

1.1.4 The long 3' UTR model

In flowering plants, animals and budding yeast, a stop codon followed by a long 3' UTR can act as a trigger to activate NMD (Kertész et al., 2006; Kerényi et al., 2008; Losson and Lacroute, 1979; Meaux et al., 2008; Longman et al., 2007; Bühler et al., 2006). Several non-mutually exclusive models have suggested how the length of the 3' UTR is sensed. The faux 3' UTR model proposes that a 'normal' 3' UTR has features that allow normal translation termination, whereas a false (faux) 3' UTR does not terminate normally and triggers NMD (Behm-Ansmant et al., 2007b). It has been proposed a long 3' UTR increases the distance between the terminating ribosome and the polyA-binding protein (PABP), which leads to NMD being triggered (Behm-Ansmant et al., 2007b). During normal termination the PABP binds to eRF3 and it has been suggested that a long 3' UTR inhibits this interaction, allowing UPF1 to bind eRF3 and

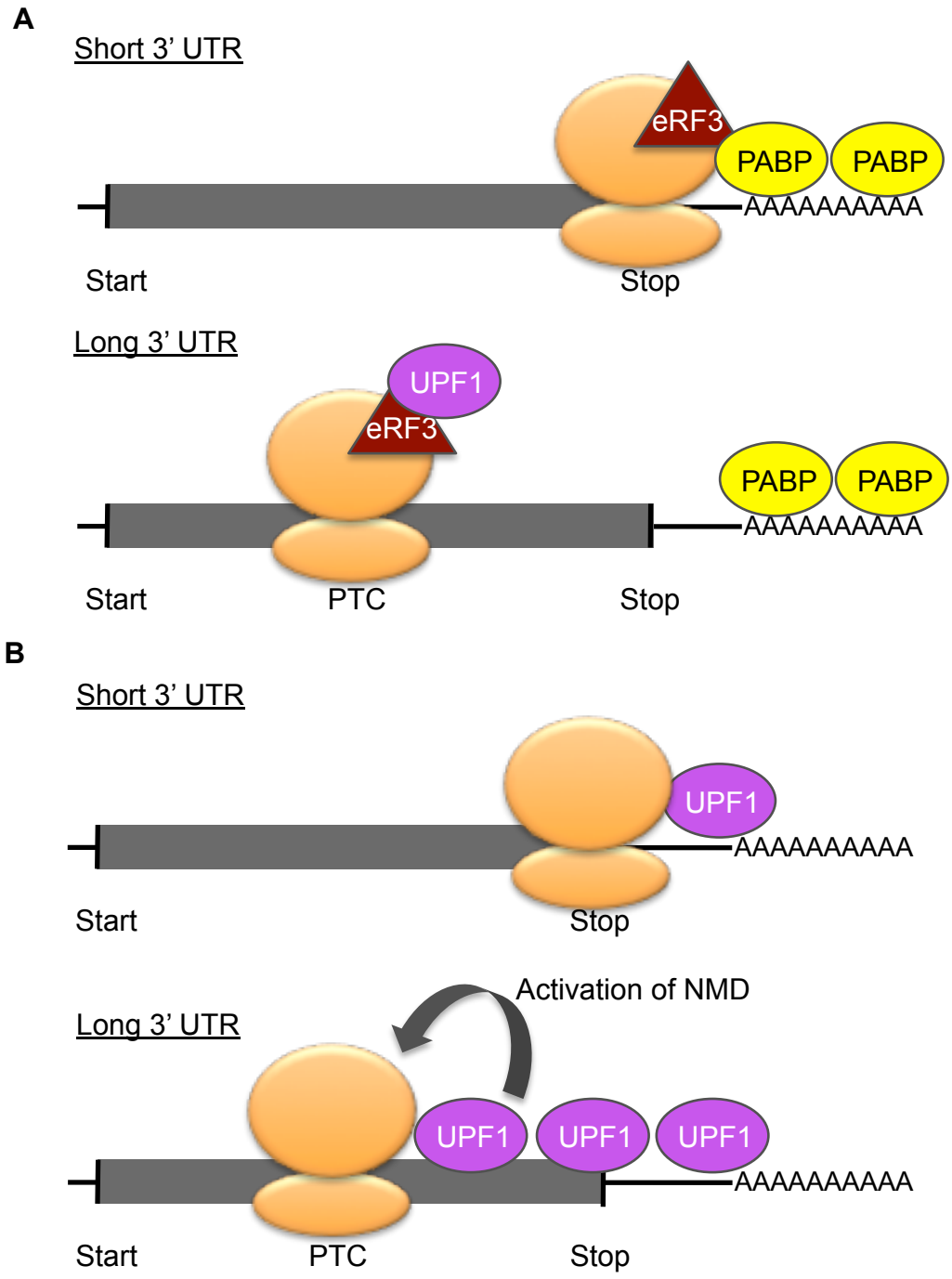

Figure 1.3: Models of 3' UTR length sensing in NMD

(A) When the ribosome terminates near the polyA tail (short 3' UTR), PABP and eRF3 interact and normal, non-NMD inducing termination occurs. Long 3' UTRs increase the distance between the terminating ribosome and PABP, which allows UPF1 to interact with eRF3 and activate NMD. *Legend continued on next page*

Figure 1.3: *Continued…*

(B) Few UPF1 molecules binds to short 3' UTRs and do not lead to NMD being triggered. Long 3' UTRs recruit more UPF1 molecules relative to shorter 3' UTR, which in turn triggers NMD.

leading to the activation of NMD (Figure 1.3) (Singh et al., 2008). Evidence has been gathered in budding yeast, *Drosophila*, angiosperms and mammals supporting this model that a PABP close to the PTC abolishes NMD (Behm-Ansmant et al., 2007a; Amrani et al., 2004; Kerényi et al., 2008; Singh et al., 2008). However, three lines of inquiry have questioned the importance of the PABP in sensing the 3' UTR length. Firstly, the PABP has been implicated in general mRNA decay and tethering PABP to an NMD target therefore might increase stability independently of competition with UPF1 (Mangus et al., 2003; Brogna and Wen, 2009). For example, a transcript can be stablised by tethering PABP to it independently of a PTC being present in budding yeast (Coller et al., 1998; Tsuboi and Inada, 2010). Secondly, a report by Meaux et al. (2008) demonstrated that NMD through 3' UTR length sensing in budding yeast did not require the either a polyA-tail or PABP (Meaux et al., 2008). Finally, deleting the region of PABP thought to be important for linking PABP to the terminating ribosome had no effect of NMD targets in budding yeast (Simón and Séraphin, 2007; Brogna and Wen, 2009), suggesting that the competition between PABP and UPF1 for eRF3 cannot account for coupling a long 3' UTR to NMD. Therefore, it appears that if PABP has a role in long 3' UTR recognition, it is redundant to other mechanisms.

Other more speculative models for how the length of the 3' UTR is measured have been suggested, although they currently lack experimental support (Niu and Cao, 2010; Brogna and Wen, 2009). A recently proposed model, based on experimental evidence, suggests that the length of a 3' UTR is sensed directly by UPF1 (Figure 1.3) (Hogg and Goff, 2010). UPF1 can bind to the 3' UTR of

any transcript (Zünd et al., 2013) in a length dependent manner (Figure 1.3) (Hogg and Goff, 2010). It has been shown that UPF1 binds along the whole length of the transcript and is displaced from the coding sequence by translation (Zünd et al., 2013). The amount of UPF1 bound along the 3' UTR can act as a molecular sensor of the length of the 3' UTR and targets the transcript to NMD (Hogg and Goff, 2010). Some transcripts with naturally long 3' UTRs have been reported to be resistant to NMD. However these do not accumulate UPF1, suggesting that they have evolved a mechanism to escape NMD, despite having a long 3' UTR, by modulating their ability to be bound by UPF1 (Hogg and Goff, 2010). The suggestion that UPF1 binds to many transcripts and senses the length of the 3' UTR is consistent with the finding in budding yeast that UPF1 is 10-fold more abundant than UPF2 (Maderazo et al., 2000).

It still remains to be resolved how UPF1 is recruited to the terminating ribosome in the long 3' UTR and the EJC models. It remains possible that it is due to an increase in the local concentration of UPF1, because of the long 3' UTR, recruitment via UPF2 and UPF3 bound to the EJC, or UPF1 binding to all terminating ribosomes but only being activated at a subset. Additionally, it remains unclear how UPF2 and UPF3 are recruited to UPF1 during long 3' UTR NMD; whether this is simply through UPF1 recruitment or whether there is another step of NMD that has not been characterised. Unlike in the EJC model where UPF3 is recruited via the EJC it is unclear what recruits UPF2 and UPF3 to the terminating ribosome in the long 3' UTR models.

1.2 The NMD pathway varies between species

1.2.1 The NMD pathway of mammals

The model of NMD in animals described in Section 1.1.2 and Figure 1.1 is highly reductionist and additional components of the NMD pathway have been identified in mammals. However, the mechanistic significance for many of these additional components has not been well characterised and could be specific to mammals (Longman et al., 2007; Yamashita et al., 2009). An extra complication is that some of the 'core' NMD effectors, such as UPF2 and UPF3, are not required for decay of all NMD targets, suggesting that additional 'branches' of the NMD pathway exist (Saltzman et al., 2011; Chan et al., 2007; Gehring et al., 2005).

An RNAi screen identified SMGL1 (NAG/NBAS) and SMGL2 (DHX34) as additional regulators of NMD. Unlike SMG1-7, these are essential for survival in *C. elegans* (Longman et al., 2007) and they also function in the NMD pathways of humans and zebrafish (Longman et al., 2007; Anastasaki et al., 2011), although their mechanistic role in NMD is not clear. Separate work has identified SMG8 and SMG9 and proteins that form a complex with SMG1 to suppress its kinase activity in mammalian NMD (Yamashita et al., 2009). Initial reports performing RNAi knockdown analysis suggested that *SMG8* also functions in the NMD pathway of *C. elegans* (Yamashita et al., 2009). However, recent analysis of *SMG8* mutants in *C. elegans* has shown no role for SMG8 in

NMD in this organism (Rosains and Mango, 2012), questioning the mechanistic conservation of SMG8 in NMD. Additional components of the SMG1 complex have been identified, revealing that the ATPases RUVBL1 (TIP49a or Pontin) and RUVBL2 (TIP49b or Reptin) interact with SMG1 and other related kinases to affect their function (Izumi et al., 2010). Knockdown of *RUVBL1* and *RUVBL2* reveals a weak loss of NMD in both humans and *C. elegans*, suggesting a minor role in NMD (Izumi et al., 2010). SMG10 and RPB5 were also identified as SMG1 complex interactors (Izumi et al., 2010). Knockdown of *RPB5* (a subunit of RNA polymerases), but not *SMG10*, had an effect on NMD in mammals (Izumi et al., 2010). However, it was suggested that *SMG10* has only a minor role in NMD in *C. elegans* (Izumi et al., 2010). Their mechanistic role in NMD stills needs to be resolved. Work has also identified that the vertebrate specific PNRC2 protein is a components of the NMD pathway (Cho et al., 2009; 2012; Lai et al., 2012). PNRC2 is believed to bind to phosphorylated UPF1 and recruit the decapping complex to help degrade the targeted transcript (Cho et al., 2009; Lai et al., 2012). PNRC2 forms a complex with SMG5 and has been suggested to be a more important interactor of SMG5 than SMG7 in NMD (Cho et al., 2012).

Some reports have suggested that different ‘branches’ of the NMD exist. This is where specific transcripts vary in their requirement of NMD effectors to be successfully degraded by NMD (Saltzman et al., 2011; Chan et al., 2007; Gehring et al., 2005). UPF1 appears to be essential for NMD but UPF2 and UPF3 are only required for a subset of NMD targets in mammals (Saltzman et al., 2011; Chan et al., 2007; Gehring et al., 2005). It is possible that BTZ from

the EJC can interact directly with UPF1 to activate NMD without the need for UPF2 and UPF3 (Gehring et al., 2009). Such ‘branches’ of the NMD pathway that do not require the full complement of NMD effectors have not been well studied in other model systems.

Much time has been spent debating 1) whether NMD occurs in the nucleus of mammals and 2) whether NMD is confined to the pioneer round of translation in mammals, which is usually associated with the nuclear cap-binding complex. 1) Early work suggested that targets of NMD in mammals were associated with the nucleus (Urlaub et al., 1989; Cheng and Maquat, 1993; Carter et al., 1996), however, this could be due to transcripts being degraded during export from the nucleus (Belgrader et al., 1994; Carter et al., 1996; Maquat, 2004). Expression of peptides that inhibit NMD specifically in the nucleus has been show to have no effect on NMD (Singh et al., 2008). More recent work used visualisation techniques to show that degradation of NMD targets occurs in the cytosol and not in the nucleus (Trcek et al., 2013). Taken together these studies appear to have resolved the debate of nuclear NMD, at least in the case of the tested transcripts. 2) Some work has suggested that UPF1 and the CBC (nuclear) cap-binding complex interact and that this increases the rate of NMD (Hwang et al., 2010; Hosoda et al., 2005). Soon after the export from the nucleus, the cap-binding complex (consisting of CBP80:CBP20) is replaced by the eIF4E cap (Rufener and Mühlemann, 2013; Durand and Lykke-Andersen, 2013) and it has been suggested that transcripts are then protected from NMD in mammals (Hwang et al., 2010; Hosoda et al., 2005). However, two independent studies have now shown that NMD can occur on eIF4E-bound transcripts in mammals

(Rufener and Mühlemann, 2013; Durand and Lykke-Andersen, 2013), therefore, while the bulk of NMD may occur during the pioneer round, this is not a pre-requisite for NMD in mammals as it can occur after the nuclear cap-binding complex has been removed.

A great deal of focus has been directed towards mammalian NMD, leading to a detailed understanding of both the central, possibly conserved, aspects of NMD as well as several apparently animal-specific aspects. Details of differences between the NMD pathway of mammals and other eukaryotes are discussed below. However, NMD has been less well studied outside the animal kingdom and therefore there is likely to be much left to be discovered.

1.2.2 The differences between the NMD pathways of model animals

The overview of the NMD pathway described in Section 1.1.2 is largely based on work on mammals and *C. elegans* and does not hold true for NMD in other animals. *Drosophila* has copies of many of the central NMD factors found in mammals (Table 1.1). Although it lacks SMG7, it retains SMG5 and SMG6 from the SMG5-7 family (Table 1.1). The functions of many of these components have been analysed in *Drosophila* and they have been implicated in NMD (Gatfield et al., 2003; Rehwinkel et al., 2005; Metzstein and Krasnow, 2006). However, the function of *SMG1* in the NMD pathway of *Drosophila* has been questioned (Gatfield et al., 2003; Chen et al., 2005; Rehwinkel et al., 2005; Metzstein and Krasnow, 2006). At least for a subset of transcripts, SMG1 does appear to operate in NMD, but loss of *SMG1* has a lesser effect in *Drosophila*

than loss of *UPF1*, *UPF2* or *SMG6* (Metzstein and Krasnow, 2006; Frizzell et al., 2012). It is possible that NMD in *Drosophila* does not completely rely on phosphorylation of UPF1, as it does in *C. elegans* and mammals. An alternative explanation is that there is a redundant kinase that acts to phosphorylate UPF1 and activate the NMD pathway. More work is needed to differentiate between the two possibilities.

Zebrafish (*Danio rerio*) have recently become established as a model organism to study NMD in vertebrates (Anastasaki et al., 2011; Wittkopp et al., 2009; Longman et al., 2013). Zebrafish are excellently placed as a model organism, as it is a basal vertebrate that can reveal insights true for NMD across vertebrates (Wittkopp et al., 2009). UPF1, SMG6, SMGL1 (NAG/NBAS) and SMGL2 (DHX34) have all had their functions in zebrafish NMD confirmed (Anastasaki et al., 2011; Wittkopp et al., 2009). Additionally, knockdown of *UPF2*, *SMG5* and *SMG6* have similar developmental phenotypes to knockdown of *UPF1*, suggesting a role for them in the same pathway (Wittkopp et al., 2009). *UPF3a* and *UPF3b* appear to have redundant roles in zebrafish NMD (Wittkopp et al., 2009). No developmental phenotype was observed in zebrafish depleted of *SMG1* (Wittkopp et al., 2009). This could be the result of inefficient knockdown, no role for SMG1 in NMD in zebrafish or a redundant kinase compensating for the loss of SMG1.

1.2.3 The NMD pathway of yeasts

While mutant screens revealed the three core NMD effectors UPF1, UPF2 and UPF3 in budding yeast, the NMD pathway is likely to be more complex than it first appears. SMG5-7 have important roles in the NMD pathway of animals (Pulak and Anderson, 1993; Cali et al., 1999) but these share similarities to a regulator of telomere length in budding yeast EST1 and SMG6 (EST1A) is also an important controller of telomere length in mammals (Redon et al., 2007). Functional analysis of the two homologues in budding yeast revealed that EBS1 but not EST1 has a role in the NMD pathway (Luke et al., 2007), suggesting subfunctionalization of the NMD and telomere regulatory roles in budding yeast between EBS1 and EST1, respectively, while at least SMG6 has retained a role in both (Redon et al., 2007; Eberle et al., 2009). It was proposed that EBS1 should be renamed SMG7 in budding yeast, because EBS1 is a functional homologue of the SMG5-7 family of animals and its domain structure best matches the domain structure of SMG7 rather than SMG5 or SMG6 (Luke et al., 2007). Deletion of *SMG7*/*EBS1* in budding yeast did not lead to as severe an NMD compromised phenotype as mutation of *UPF1*, suggesting that SMG7/EBS1 does not have as essential a role in NMD as UPF1, UPF2 and UPF3 (Luke et al., 2007). This might explain why it was not identified in the original mutant screens that identified *UPF1*, *UPF2* and *UPF3*. It could also be that the role of SMG7/EBS1 in NMD is not an essential step and loss of *SMG7*/*EBS1* only results in a minor loss of NMD efficiency. On the other hand, *SMG7*/*EBS1* could be partially redundant to another factor that has a similar role in budding yeast NMD. The exact nature of the role of SMG7/EBS1 in

budding yeast NMD is unknown. However, given the sequence homology to the SMG5-7 family and in particular to the 14-3-3-like domain, it is tempting to speculate that SMG7/EBS1 binds to phosphorylated UPF1 to recruit the decay machinery. It is also unknown whether UPF1 phosphorylation in budding yeast is essential for NMD. However, budding yeast UPF1 has been shown to be a phospho-protein (Wang et al., 2006). Yeasts have no orthologue of *SMG1* (Grimson et al., 2004) and no other NMD-associated kinase has been identified in fungi. More work is needed to establish a role for phosphorylation in yeast, to identify a kinase and to dissect out the exact roles of UPF1, UPF2, UPF3 and SMG7/EBS1.

Additional work is needed to understand how a transcript is targeted to NMD. As previously discussed (Section 1.1.3), budding yeast lacks an EJC complex (Table 1.1), therefore must rely on other signals to act to trigger NMD. A DSE, which binds HRP1 (Zhang et al., 1995; Peltz et al., 1993; González et al., 2000) has been characterised but is only applicable to a small subset of NMD targets. A long 3' UTR has been shown to be an effective trigger of NMD in budding yeast, possibly through the distance of the stop codon from the PABP, but this does not fully explain how a long 3' UTR can be sensed (Meaux et al., 2008) (see Section 1.1.4 for more details). It will be important to search for more DSEs and untangle how the cell senses 3' UTRs in budding yeast to gain a complete picture of NMD.

The fission yeast *Schizosaccharomyces pombe* has also been used to study NMD and has yielded some unexpected results (Wen and Brogna, 2010). Like

budding yeast, *S. pombe* lacks *SMG1* (Grimson et al., 2004) and relies on *UPF1* and *UPF2* for NMD (Wen and Brogna, 2010). Unlike budding yeast, *S. pombe* has an EJC, therefore, the role of introns in PTC recognition along with the EJC was explored (Wen and Brogna, 2010). Wen and Brogna (2010) found that the splicing of an intron did increase recognition of a PTC, but unexpectedly this occurred whether the PTC was upstream or downstream of the intron and did not require the EJC, suggesting a novel mechanism for PTC recognition (Wen and Brogna, 2010). It is still unclear whether the EJC has a role in NMD at all in *S. pombe*, but in these reporter transcripts splicing can increase perception of a PTC independently of an EJC (Wen and Brogna, 2010). They also found that there was no clear correlation between the length of a 3' UTR and targeting of the transcript to NMD in *S. pombe* (Wen and Brogna, 2010). However, caution should be used when trying to generalise work from a few model artificial transcripts to general rules for how NMD functions in an organism.

1.2.4 The NMD pathway in flowering plants

Work has begun to characterise NMD pathways in plants. This work has mostly focused on the flowering plant *Arabidopsis thaliana*, however, some work has been performed using a tobacco transient assay (described in Section 1.1.3). Using *A. thaliana* and the tobacco transient assay, the roles for many NMD effectors have been determined. *UPF1*, *UPF2*, *UPF3* and *SMG7* all have functional homologues in plants (Riehs et al., 2008; Arciga-Reyes et al., 2006; Kerényi et al., 2008). Examining the *A. thaliana* genome reveals that the NMD

pathway is similar to the budding yeast's NMD pathway (Table 1.1). Like budding yeast, *A. thaliana* also lacks *SMG1* (Grimson et al., 2004), or any known kinase involved with the NMD pathway. When *A. thaliana* UPF1 (AtUPF1) is expressed in tobacco cells, it becomes phosphorylated at the N- and C-termini and these regions are needed for AtUPF1 to function in tobacco (Mérai et al., 2012). Additionally, SMG7 has a role in NMD, in particular, residues important for binding phosphorylated UPF1 in the 14-3-3-like domain are essential for its role in NMD (Riehs et al., 2008; Benkovics et al., 2011; Kerényi et al., 2008; Mérai et al., 2012). These data suggest that NMD in plants involves UPF1 phosphorylation, although no kinase involved in NMD has been characterised outside the animal kingdom. This highlights the need to study NMD intensively in more non-animal models to better understand the NMD pathway.

As discussed previously (Sections 1.1.3 and 1.1.4), reporter genes that are targeted to NMD through both the long 3' UTR and EJC models have been described using the tobacco transient assay (Kertész et al., 2006; Kerényi et al., 2008), but, the exact mechanism for both is yet to be reported. For example, it would be interesting to identify whether UPF3 binds to the EJC, as has been reported in animals (Buchwald et al., 2010; Gehring et al., 2003; Kashima et al., 2006) and whether there are plant-specific mechanisms to target a transcript to NMD.

1.2.5 NMD targeting features in flowering plants

Work using the tobacco transient expression system has found that transcripts with either a long 3' UTR or an intron located in the 3' UTR are targeted to NMD in flowering plants (Kertész et al., 2006; Kerényi et al., 2008). In tobacco it was found that a 3' UTR of 600 nucleotides could efficiently target a transcript to NMD but a shorter 3' UTR of 300 nucleotides could also target a transcript to NMD, although less efficiently than the longer 3' UTRs (Kertész et al., 2006; Kerényi et al., 2008). The average length of a 3' UTR in *A. thaliana* (TAIR10) is 237 nucleotides long (Zimmer et al., 2013). Analysis of 270 genes in *A. thaliana* found that transcripts with a 3' UTR of >350 nucleotides were enriched among those with increased steady states in NMD mutants (Kalyna et al., 2012). A transcriptome-wide study of *A. thaliana* found that transcripts with lengths of over 347 nucleotides (representing the longest 10% of transcripts in TAIR10) were enriched in NMD mutants (Drechsel et al., 2013). This study also found that there is an enrichment of intron-containing 3' UTRs in *A. thaliana* NMD mutants, suggesting these are targeting features conserved across flowering plants (Drechsel et al., 2013). Taken together, this suggests that both of the well charactorised features that can target a transcript to NMD operate in plants and could work through similar mechaisms. For example, the EJC is important in the targeting of transcripts with introns downstream of the stop and addition of a series of adenosine residues to recruit PABP in a long 3' UTR can stabilise a transcript (Kerényi et al., 2008; Nyikó et al., 2013).

Upstream open reading frames (uORFs) are short ORFs located in the 5' UTR of a transcript. This places a start and stop codon very distant from the polyA tail, thus creating a long 3' UTR, often upstream of many introns. Therefore many uORFs can target a transcript to NMD (Nyikó et al., 2009; Saul et al., 2009). uORFs are very common in *A. thaliana* genes (approximately 20% contain them), but only a small fraction appear to be targeted to NMD (Nyikó et al., 2009). Therefore work has started to define why some uORFs can target a transcript to NMD. From modifying a small number of uORFs in front of a *GFP* reporter in the tobacco, it has been suggested that short uORFs (<36 amino acids) are less likely to trigger NMD and that longer uORFs (Nyikó et al., 2009). An analysis of 270 genes in *A. thaliana* found that uORFs that overlap with the main ORF (where the stop codon of the uORF is in the coding sequence of the main ORF) are highly enriched in transcripts where the steady state level was increased in NMD mutants (Kalyna et al., 2012). Finally, uORFs where the peptide sequence is conserved between transcripts between different species (called conserved peptide uORFs; CPuORFs) are greatly enriched among transcripts with increased steady state levels in *A. thaliana* NMD mutants and have been suggested to be direct targets of NMD (Rayson et al., 2012a; 2012b). These CPuORFs are enriched among NMD genes in A. thaliana NMD mutants with an increased steady state level independent of length, suggesting that they are targeted to NMD through a mechanism related to the conserved nature of their uORF, rather than simply being long uORFs (Rayson et al., 2012a; 2012b). CPuORFs have also been placed in front of a GFP reporter and found to be able to target a transcript to NMD (Nyikó et al., 2009) suggesting they directly target transcripts to NMD. Why long, overlapping and conserved

peptide uORFs appear to target transcripts to NMD while many other uORFs do not is unclear and further research is needed.

1.3 NMD is important for controlling gene expression

As NMD was originally characterised as a quality control mechanism, it was predicted that many targets would be the result of nonsense mutations, PTC-containing alternative splice variants and errors resulting from transcription or splicing (Mühlemann et al., 2008). However, transcriptomic analysis of NMD mutants in budding yeast, *Drosophila*, *C. elegans*, mammals and *A. thaliana* have all revealed that 1-10% of genes are influenced by NMD, and many of these genes do not have NMD targeting features (Guan et al., 2006; He et al., 2003; Mendell et al., 2004; Ramani et al., 2009; Rayson et al., 2012a; Rehwinkel et al., 2005). The NMD influenced genes will include both direct and indirect targets of the mRNA decay pathway. It is difficult to quantify the influence of changes to the transcriptome that are induced by truncated proteins, however, many changes in gene expression are the result of NMD targeting transcripts, which encode full-length and physiologically important proteins. These 'physiological' targets of NMD can be targeted to NMD by a number of different targeting features. One of the best characterised physiological targets of NMD is *AtSMG7* (Benkovics et al., 2011; Kerényi et al., 2008; Nyikó et al., 2013; Rayson et al., 2012a). *AtSMG7* has a long 3' UTR with two introns downstream of the 'natural' stop codon and through both the long 3' UTR-mediated recognition and EJC-mediated recognition, the 'natural' stop codon appears to the terminating ribosome to be a PTC and *AtSMG7* is

degraded by NMD (Benkovics et al., 2011; Kerényi et al., 2008; Nyikó et al., 2013; Rayson et al., 2012a). This appears to generate an autoregulatory loop to control the activity of the NMD pathway (Kerényi et al., 2008) and similar negative feedback loops have been reported to control the activity of NMD effector-encoding genes in animals (Huang et al., 2011; Yepiskoposyan et al., 2011). Therefore, naturally long 3' UTRs and/or introns downstream of the stop codon can target physiologically important transcripts to NMD (Sureau et al., 2001; Kertész et al., 2006; Nyikó et al., 2013; Yepiskoposyan et al., 2011).

As previously mentioned, uORFs can also target transcripts to NMD (Nyikó et al., 2009). If a uORF is translated, it will place a stop codon upstream of a long 3' UTR and/or introns downstream of the stop codon, to target the uORF-containing transcript to NMD. However, the transcriptomes of eukaryotes have many uORF-containing transcripts but only a subset of these are targets of NMD (Rayson et al., 2012a). One possibility is that stop codons early along a transcript are not targeted to NMD, because in the closed loop model of eukaryotic translation, termination occuring near the 5' end of the transcript is also near the PABP. The PABP can therefore bind to the 5' cap complex and trigger normal translation termination and therefore does not activate NMD, according to the faux 3' UTR model (Brogna and Wen, 2009). However, many stop codons early in the transcript do target the transcript to NMD (Brogna and Wen, 2009) and for those that cannot, this can sometimes be explained by the ribosome reinitiating at downstream start codons abolishing NMD (Zhang and Maquat, 1997). Two non-mutually exclusive models can be generated to explain why uORFs do not always trigger NMD; 1) if uORFs are translated by

the ribosome, the ribosomes are able to re-initiated at the start codon of the main ORF negating NMD or 2) uORFs are not always translated by the ribosome. Further work is needed to understand under which category uORF-containing transcripts fall.

It would be easy to view NMD as having a static role in suppressing gene expression, with transcripts targeted by NMD being degraded under all conditions, however, two mechanisms can actively modulate transcript levels via NMD:

1) Regulation of the activity of the NMD pathway
2) Altering the susceptibility of specific transcripts to NMD

Both mechanisms have been reported in mammals. During brain development in vertebrates a microRNA targeting UPF1 is up-regulated, decreasing the activity of the NMD pathway (Bruno et al., 2011). Some transcripts required for normal neural development are predicted to be NMD targets. Reducing the overall level of NMD at a critical developmental stage allows these transcripts to be up-regulated and facilities the differentiation of stem cells into neurons (Bruno et al., 2011). Phosphorylation of eIF2-α is known to inhibit translation on a global scale but it also inhibits NMD through an unknown mechanism (Gardner, 2008; Wang et al., 2011). eIF2-α is phosphorylated in response to stress conditions such as the unfolded protein response (UPR) or hypoxia. Transcription factors responsible for combating these stresses are targeted to NMD by uORFs present in their transcripts and repression of NMD allows these factors to be expressed under stress conditions (Gardner, 2008; Wang et al., 2011). In addition to global inhibition of NMD, there are also reports of individual

transcripts being able to evade NMD in a condition-specific manner. The uORF of the *CPA* transcript in budding yeast only targets the transcript to NMD when arginine is present in the growth medium (Gaba et al., 2005). Transcripts can also be alternatively spliced to target them to NMD, even in a manner that maintains the functionality of the protein. For example, *SC35* in mammals has a cryptic intron in its 3' UTR, which is spliced out if levels of the SC35 protein are high, creating a feedback loop (Sureau et al., 2001; Lareau et al., 2007). The splicing event deposits an EJC downstream of the 'normal' stop codon making it a target of NMD (Sureau et al., 2001). Therefore, there is a range of ways in which a transcript important for growth or stress response can be brought into or released from the sphere of influence of NMD.

1.3.1 AS-coupled NMD

AS-coupled NMD is an important mechanism for degrading transcripts to regulate gene expression. Splicing factors can cause changes in the splicing of specific primary transcripts to change splicing towards a PTC-containing variant. In the case of the mammalian *SC35*, the PTC is the authentic stop codon, which becomes recognised by NMD upon splicing of a cryptic intron and introduction of a EJC downstream of this stop (Sureau et al., 2001; Lareau et al., 2007). If a PTC is introduced in the middle of the coding sequence, then NMD removes this as waste to prevent production of truncated proteins. These truncated proteins might have dominant negative effects if they were not degraded by NMD. Many transcripts are regulated in the process, including splicing factors themselves to produce a autoregulatory feedback loop (Sureau

et al., 2001; Lareau et al., 2007; Stauffer et al., 2010; Palusa and Reddy, 2010). AS-coupled NMD is a highly regulated and is associated with highly conserved sequence element in mammals, suggesting a functional importance (Lareau et al., 2007). Approximately 20-30% of animal genes produce splice variants that are targets of NMD (Ramani et al., 2009; Weischenfeldt et al., 2012). In *A. thaliana*, a recent studies have found that around 13-17% of intron-containing genes produce a PTC-containing splice variant that has increased steady state levels in NMD mutants (Drechsel et al., 2013; Kalyna et al., 2012). It is worth noting that many of the genes associated with AS-coupled NMD in *A. thaliana* appear to have a role in salt stress (Drechsel et al., 2013), highlighting the potential importance of AS-coupled NMD in controlling properties of plants, in particular in processes important in agriculture. Interestingly, there is an emerging difference between plants and animals in regard to AS-coupled NMD, namely the types of AS events that occur in a transcript and whether this transcript is targeted to NMD (Drechsel et al., 2013; Kalyna et al., 2012). Transcripts containing intron retention events are targeted to NMD in a range of organisms including budding yeast and animals (Sayani et al., 2008; Ramani et al., 2009), however, it appears such transcripts are not targeted to NMD in *A. thaliana* (Drechsel et al., 2013; Kalyna et al., 2012), even when the PTC introduced by a retained intron is the same as those introduced by other splicing events and these transcripts are targeted to NMD (Kalyna et al., 2012). It is possible some of these transcripts accumulate in the nucleolus (Kim et al., 2009). One report has shown that the purpose of such intron retained transcripts is to act as stored transcripts, which are rapidly spliced at the correct developmental stage (Boothby et al., 2013). Although it is unclear how splicing

of such transcripts are regulated and how they evade NMD. The mostly likely scenario would be that they are not translated. However, there are examples of transcripts with intron retention events being translated and producing a protein in plants, for example in the *A. thaliana* HSFA2-encoding gene (Liu et al., 2013). These studies suggest that AS-coupled NMD is a highly regulated and far-reaching process for controlling gene expression in both animals and plants, however more research is needed to understand the significance of this in plants and how AS-coupled NMD differs between the two groups.

1.3.2 A biological function of NMD in plants

To understand the function of NMD in flowering plants, NMD mutants have been examined in *A. thaliana* (Yoine et al., 2006; Arciga-Reyes et al., 2006; Riehs et al., 2008). A full null mutant of *UPF1* (*upf1-3*) is seedling lethal (Yoine et al., 2006; Arciga-Reyes et al., 2006). *A. thaliana* NMD mutants have increased steady state levels of genes involved in pathogen response, and it was found that the seedling lethality of the *upf1-3* allele was due to increased production of the defense hormone salicylic acid (SA) (Rayson et al., 2012a; Riehs-Kearnan et al., 2012; Jeong et al., 2011; Rayson et al., 2012b). Therefore, NMD might repress the pathogen response of *A. thaliana* (Rayson et al., 2012a; Riehs-Kearnan et al., 2012; Jeong et al., 2011; Rayson et al., 2012b). It would be interesting to know if NMD is repressed during pathogen attack and this is what leads to increased expression of pathogen-regulated genes. It has been suggested that genes encoding NMD effectors are down-regulated during pathogen attack and this could repress NMD, which in turn

would lead to expression of pathogen related genes (Jeong et al., 2011). If NMD is repressed during pathogen, this would allow for the expression of novel truncated proteins, some of which might have a role in fighting the invading pathogen(s).

1.4 The origins of the NMD pathway

The NMD pathway has an important role in controlling gene expression as both a quality control mechanism (Mühlemann et al., 2008) and also as a regulator of physiological transcripts (see Section 1.3). However, the origins of the NMD pathway and its original function are unclear. Although there is some debate regarding the evolution of eukaryotes and which are the basal lineages, recent work has suggested the last eukaryotic common ancestor (LECA; stem eukaryote) diverged into two lineages, one that contains plants and another that contains animals and fungi (Richards and Cavalier-Smith, 2005; Derelle and Lang, 2012). In this situation, the NMD pathway would have been complex in the LECA, with UPF1, UPF2, UPF3, SMG7 and the EJC all likely to be functioning, as they function in both plants and animals (Section 1.2).

When focusing on diploids (and polyploids) such as animals and flowering plants, it is easy to imagine why NMD might have evolved and been maintained; to protect heterozygotes from damaging dominant negative truncated proteins, which could have deleterious effects on organismal survival. Many human genetic disorders have been described, for example β-Thalassemia, that are predicted to be suppressed in heterozygotes due to NMD

(Khajavi et al., 2006). However, there are many cases where NMD can make the symptoms of a disease worse by preventing functional, or partly functional truncated proteins from being generated, for example, in the case of some Duchenne muscular dystrophy patients, where some truncated protein would be better than no protein in the homozygotes (Bhuvanagiri et al., 2010). However, the NMD pathway probably did not evolve in a diploid eukaryote but in a haploid before LECA arose. It is unclear exactly when NMD evolved in relation to other innovations in eukaryotes but it has been suggested it was as a defence against group II introns.

Group II introns are selfish genetic elements and the likely ancestors of the modern day spliceosomal introns (Roy and Gilbert, 2006; Cavalier-Smith, 1991; Cech, 1986). It has been proposed that these underwent a massive expansion in early eukaryotes (Martin and Koonin, 2006). The splicing of these selfish genetic elements has been predicted to have been slower than translation (Poole, 2006) so it has been hypothesised that the nucleus evolved as a barrier to separate transcription and splicing from translation (Koonin, 2006; Martin and Koonin, 2006), therefore preventing truncated proteins arising from the translation of transcripts still undergoing splicing. The spliceosome itself is likely to be an outcome of the need to increase splicing efficiency (Koonin, 2006). It has been proposed that NMD may also have evolved as a quality control mechanism to prevent the translation of intron containing transcripts that were not correctly spliced in the early evolution of eukaryotes (Martin and Koonin, 2006; Koonin, 2006). Studies have shown that intron retention events are degraded by NMD in animals (Ramani et al., 2009), budding yeast (Sayani et

al., 2008) and the protist *Paramecium tetraurelia* (Jaillon et al., 2008). Therefore, NMD has allowed for the expansion of novel introns in eukaryotes and laid the ground for the increase in organismal complexity through alternative splicing. Farlow et al. (2010) have shown that in *Drosophila* species, newly arisen introns have weaker splice sites than older introns and that these new introns are hidden from selective pressures by NMD (Farlow et al., 2010). As mentioned previously, *A. thaliana* appears to be the exception and several examples of PTC-containing intron retention events do not target a transcript to NMD (Kalyna et al., 2012), suggesting a novel mechanism in plants to recognise intron retention events, although little is known about this.

This work demonstrates the importance of NMD not only to protect the organism from mutations and to regulate gene expression, but also in 'evolvability'. NMD is likely to have been a pre-requisite for the expansion of introns in eukaryotes and allowed for the evolution of alternative splicing. Today it has important additional roles in stress response (Wang et al., 2011; Rayson et al., 2012a; Riehs-Kearnan et al., 2012; Jeong et al., 2011).

1.5 Aims of this thesis

A great deal has been determined about the NMD pathway in animals by not just focusing on one model organism, but through complementary studies in multiple model systems, including, *C. elegans* and *Drosophila*, zebrafish, mice and human cell culture. These evolutionarily diverse organisms have allowed us not only to understand more about the central components of the pathway but

also to find organism/lineage specific mechanisms (Section 1.2.2). This has shown us that whilst SMG1 is important for NMD in *C. elegans* and mammals (Yamashita et al., 2001; Pulak and Anderson, 1993; McIlwain et al., 2010), this is not universal across all animals, since it does not appear to be essential for NMD in *Drosophila* or zebrafish (Wittkopp et al., 2009; Metzstein and Krasnow, 2006).

By studying evolutionarily diverse plants, a better understanding of the NMD pathways of plants can be deciphered, with both species-specific and conserved components being identified. In addition, the more we learn about NMD in plants, animals and fungi, the more we can predict about the NMD pathway of the LECA and the evolutionary origins of NMD.

The broad aim of this thesis is to understand the mechanism and importance of NMD in plants by primarily using the moss, *Physcomitrella patens* as a model organism. *P. patens* (also referred to simply as moss in this thesis) is an excellent model organism within the plant kingdom. It is currently the most evolutionarily basal land plant with a published genome sequence (Rensing et al., 2008), having diverged from flowering plants at least 450 million years ago (Kenrick and Crane, 1997). In addition to being useful for studying evolutionarily diverse land plants, *P. patens* is an excellent model for reverse genetics (Prigge and Bezanilla, 2010). *P. patens* has a higher rate of homologous recombination than flowering plants, allowing efficient gene targeting. By cloning approximately one kilobase upstream and downstream of a gene of interest around a selection cassette, the targeted gene can be effectively deleted

(Kamisugi et al., 2005). Since mosses spend much of their life cycle as haploids and can be sub-cultured vegetatively, knockout lines generated in this way can be studied directly, without the need for crossing (Prigge and Bezanilla, 2010).

Specific aims:

1 To identify and functionally characterise components of the NMD pathway of moss

2 To use knowledge from NMD in moss to understand better the NMD pathway in plants

3 To understand the physiological roles of NMD effectors in moss

2 Materials and Methods

2.1 Computer resources

2.1.1 Bioinformatics tools

Table 2.1: Computer resources.

Function	Tool name	Web address
Genome databases	COSMOSS TAIR NCBI Phytozome	http://cosmoss.org/ http://www.arabidopsis.org/ http://blast.ncbi.nlm.nih.gov/Blast.cgi http://www.phytozome.net/
Synteny analysis	Plant Genome Duplication Database	http://chibba.agtec.uga.edu/duplication/
Domain identification	SMART database	http://smart.embl-heidelberg.de/
Sequence alignment	CluastalX:	http://www.ebi.ac.uk/Tools/clustalw2/index.html
Tree viewer	FigTree	http://tree.bio.ed.ac.uk/software/figtree/
Primer design	Primer-BLAST Primer3	http://www.ncbi.nlm.nih.gov/tools/primer-blast/index.cgi?LINK_LOC=BlastHome http://frodo.wi.mit.edu/primer3/
Plasmid map generation	A plasmid editor (ApE):	http://www.biology.utah.edu/jorgensen/wayned/ape/
Image manipulation	ImageJ	http://rsbweb.nih.gov/ij/

2.1.2 Tree analysis

Conserved domains or whole protein sequences were aligned using the ClustalX program with default settings. SMART was used to identify a domain of interest. A neighbour joining tree was generated using ClustalX and visualised in FigTree. The tree was rooted using midpoint in FigTree. Bootstrapping was used to add significance to the nodes on the tree. A thousand replicates were performed and if a node had the support of ≥50% of the replicates, it was considered significant.

2.2 General growth media

LB medium for bacterial cultures

Tryptone	10 g/l
Yeast extract	5 g/l
NaCl	10 g/l
(Agar	15 g/l for solid media only)

YPDA medium for yeast cultures

Tryptone	20 g/l (times two for 2xYPDA)
Yeast extract	10 g/l (times two for 2xYPDA)
Glucose	20 g/l (times two for 2xYPDA)

Adenine 3 g/l (times two for 2xYPDA)

(Agar 20 g/l for solid media only)

SD medium for yeast cultures

Yeast nitrogen base 6.7 g/l

Glucose 20 g/l

Appropriate dropout supplement of amino acids (as specified by the manufacturer; Clonetech)

(Agar 20 g/l for solid media only)

2.3 Plant material and culture conditions

2.3.1 Moss growth conditions

Physcomitrella patens ssp. patens (Hedwig) ecotype 'Gransden 2004' (Rensing et al., 2008; Ashton and D Cove, 1977) was used in this study and is also referred to as *P. patens* or moss. Moss was cultured under sterile conditions in Petri dishes (Sterilin) on BCD or BCD supplemented with diammonium tartrate (BCDAT) medium, as described previously (Nishiyama et al., 2000), at 25 ^{o}C under continuous light. Moss was sub-cultured by transferring a small proportion of the plant with forceps to a new plate or by homogenisation of moss tissue in a previously autoclaved blender before growth in a Petri dish where the BCDAT had been overlaid with a sterile cellophane disc.

2.3.2 Treatment of moss with cellular process changing drugs

To phenotype moss exposed to DNA damaging conditions or induction of the unfolded protein response (UPR), moss tissue five- to six-days post-homogenisation, was inoculated as ‘spot inocula’ on BCDAT supplemented with drug or solvent control and grown for three weeks. To induce DNA damage moss was treated with 8 ng/ml bleomycin (Euro Nippon Kayaku GmbH), as described previously (Kamisugi et al., 2011). To induce the unfolded protein response of the endoplasmic reticulum BCDAT medium was supplemented with tunicamycin (Tm) 2.5 µg/ml or DMSO solvent control. To induce the unfolded protein response across the whole cell BCDAT medium was supplemented with 10 mM L-azetidine- 2-carboxylic acid (AZC), a proline analogue or 10 mM proline as a control for growth. To monitor the expression of unfolded protein responsive genes, moss tissue five- to six-days post-homogenisation was inoculated as ‘spot inocula’ on BCDAT supplemented with 1 µg/ml of tunicamycin or DMSO as solvent control and grown for two weeks before tissue was collected for RNA extraction. The expression of NMD targets was monitored after exposure to the inhibitor of translation and NMD cycloheximide (CHX; Sigma). For CHX treatment, 100 mg of six-day old homogenised lawns on BCDAT medium with cellophane was picked and incubated on a BCDAT medium plate supplemented with 20 µM CHX or DMSO solvent control plate for six hours.

2.3.3 Moss colony area size estimation

The image software ImageJ was used to convert the images into binary format and counting the number of pixels corresponding to a colony/plant measured moss colony or plant size on Petri dish. Moss colony size was normalised between plates and converted into mm^2 by estimating the area of the plate (Kamisugi et al., 2011).

2.4 General nucleic acid methods

2.4.1 Genomic DNA extraction

Moss genomic DNA was isolated from moss spot inoculates grown for 1-3 weeks using the ISOLATE Plant DNA Miniprep Kit (Bioline), according to manufacturer's instructions.

2.4.2 RNA extraction

Extraction was done using the RNeasy Plant Mini kit (Qiagen) and approximately 100 mg of moss tissue following manufacturer's instructions. To remove contaminating genomic DNA, the on-column DNase I digestion (Qiagen) was performed according to manufacturer's instructions.

2.4.3 cDNA synthesis

For cloning and semi-quantitative reverse transcriptase-polymerase chain reaction (sqRT-PCR), first-strand cDNA was synthesised from 0.3-0.7 µg of total RNA and oligo$(dT)_{12-18}$ using SuperScript II Reverse Transcriptase (Invitrogen) following manufacturer's instructions. For quantitative (q)RT-PCR, first strand cDNA was synthesised from 0.8 and 1.0 µg of total RNA using the iScript cDNA Synthesis kit (Biorad) following the manufacturer's instructions. This kit is designed to amplify short (<1 Kb) cDNA fragments for qRT-PCR analysis using both oligo(dT) and random hexamer primers in the reaction mix. *A. thaliana* RNA used for cDNA synthesis was kindly donated by Sandra Biewers.

2.4.4 Nucleic acid quantity and quality quantification

The quantity of RNA and DNA was measured using the NanoDrop spectrometer (Thermo Scientific), by measuring the absorbance at 260 nm. The quality of RNA was monitored using the Bioanalyser (Agilent) following the manufacturer's instructions. Briefly, RNA is denatured and then loaded into a chip and run through micro-channels containing gel and size is measured by run-time of the RNA fragments through the micro-channels and comparison to an RNA ladder. This gives a high-resolution output with the 18S and 28S ribosomal RNA ratio allowing for identification of degradation of the RNA.

2.4.5 Restriction enzyme digestions

Restriction reactions were set up in volumes between 10 and 20 µl with 0.5-1.0 units of appropriate FastDigest enzyme(s) (Fermentas) and universal buffer (Fermentas). Digestions were incubated at 37 ^{o}C for 30 minutes. The predicted size of bands were determined using the a plasmid editor (ApE) computer software.

2.4.6 Polymerase chain reaction

Polymerase chain reactions (PCR) were made up to contain dNTPs (0.2 mM each), forward and reverse primers (0.2 µM each) with DNA template, appropriate buffer and an appropriate DNA polymerase. Phusion (Finnzymes) was used for cloning, yeast colony PCR and genotyping while DreamTaq (Fermentas) to generate the linear fragments for moss transformations and sqRT-PCR analysis. The PCR additives 10 µg/ml ET-SSB (NEB) and 20 mM betaine were added to DreamTaq when amplifying PCR products for moss transformations.

2.4.7 Quantitative polymerase chain reaction

For qRT-PCR, a 20 µl reaction was made up of 10 µl SsoFast^{TM} Evagreen® supermix, primers to final concentration of 500 nM each, diluted cDNA template and water (to bring the volume to 20 µl). A qRT-PCR cycle consists of a single

initial denaturisation at 95 ^{o}C for 30 seconds followed by the combination of 95 ^{o}C for three seconds for denaturisation followed by 60 ^{o}C for five seconds for 40 cycles. The level of DNA is indirectly measured by the amount of fluorescence generated by the Evagreen® dye binding to the double-stranded DNA. To check that only one product was amplified, a melt curve was generated. The CPX96^{TM} Real Time system (Biorad) with the C1000^{TM} Thermo cycler (Biorad) was used in all qRT-PCR experiments. The newly synthesised double-stranded DNA was melted at different temperatures, starting at 65 ^{o}C and increasing in 0.2 ^{o}C increments to 95 ^{o}C. The temperature that the DNA melted at is plotted and gives an output to estimate the number of PCR products in a reaction through a reduction of fluorescence at a particular temperature. The relative abundances of transcripts were calculated using the moss reference genes *PpEF1α* (Pp1s7_457V6.1) (Khraiwesh et al., 2010) or Pp1s54_156V6.1 (Clathrin adapter complex subunit) (Kamisugi et al., 2011) (see figure legend of qRT-PCR data for reference gene used). The starting quantity calculation was done using the standard curve method. A dilution series was made from a cDNA reaction and the efficiency of the PCR reaction was estimated from this curve and used to accurately calculate the starting quantity of the target in each reaction. The data were analysed for anomalies, which were removed before the average of the technical replicates were calculated.

2.4.8 DNA agarose gel electrophoresis

Agarose gels for electrophoresis were made using 1% w/v agarose (Severn Biotech Ltd) for cloning perposes or 2% w/v agarose for analysis of sqRT-PCR

results in 1x TAE (Severn Biotech Ltd) buffer dissolved by heating in a microwave. Ethidium bromide was added to a final concentration of 0.5 µg/ml. Loading buffer was added to DNA samples prior to loading. DNA was run along side the GeneRuler^{TM} 1 Kb plus molecular weight ladder (Fermentas) for between 25 minutes and 1 hour at 50-100 volts using a gel tank (Biorad) and power pack (Biorad) and visualised using a UV transilluminator and photographed (Syngene).

2.4.9 Purification of DNA

To purify a single PCR product, the post-cycling reaction mixture was purified using the QIAquick PCR purification kit (Qiagen) following the manufacturer's instructions. Otherwise bands were extracted from agarose gels (Section 2.4.10). Ethanol precipitation was used to purify linear DNA for moss transformation and to keep it sterile. This was achieved by adding two to three volumes of 100% ethanol and a tenth of a volume of sodium acetate (3 M; pH 5.2) and incubating at -20 ^{o}C for 20-30 minutes before DNA was pelleted by centrifugation for 30 minutes between 18000-21000 relative centrifugal force (*g*) at 4 ^{o}C. The pellet was washed twice with 70% (v/v) ethanol before being re-suspended in sterile distilled water.

2.4.10 Extraction of DNA from agarose gels

Following electrophoresis (2.3.6) and viewing under a UV transilluminator in the dark room, the DNA band of interest was excised with a razor blade. The gel slice was trimmed to remove excess agarose and DNA was purified using the QIAquick PCR purification kit (Qiagen) using the Buffer QG (Qiagen) following the manufacturer's instructions.

2.4.11 Sanger sequencing

Sequencing of DNA was carried out by either MWG Biotech or Beckman Coulter Genomics using plasmid purified by miniprep or purified PCR products.

2.5 Cloning and sub-cloning

2.5.1 Primer design

Primers used for cloning, genotyping or sqRT-PCR of moss transformants were designed using Primer-BLAST, checking for predicted off-target amplification using the moss or *A. thaliana* databases. Primers used for qRT-PCR were designed using Primer3 using the following changes from default, product size between 90-150 bp, annealing temperature 60±1 ^{o}C and GC content between 40 and 60%.

2.5.2 Bacterial growth conditions

E. coli was grown at 37 ^{o}C overnight on LB agar or in LB liquid medium in a shaking incubator at 200 rpm. LB agar plates were inverted during incubation and were stored at 4 ^{o}C for up to 4 weeks before the cells were re-streaked on selective LB agar plates. Kanamycin (50 µg/ml), ampicillin (100 µg/ml) and gentamicin (10 µg/ml) were used as selective antibiotic in LB agar and liquid media. Molten LB agar was cooled to approximately 55 ^{o}C before the antibiotic was added and immediately poured into 9 cm diameter Petri dishes (Sterilin) under sterile conditions.

2.5.3 Transformation of bacteria

Escherichia coli α-select cells (Bioline) were transformed according to manufacturer's instructions by heat shock at 42 ^{o}C before being incubated in non-selective liquid LB at 37 ^{o}C for one to two hours. Cultures were then spread on LB agar plates with appropriate antibiotic selection.

2.5.4 Plasmid extractions

Minipreps were carried out on 1.5 ml of overnight culture using the Qiagen miniprep kit (Qiagen) following the manufacturer's instructions.

2.5.5 Gateway™ reactions

To generate entry clones for Gateway™ cloning (Invitrogen), the BP reaction was carried out by mixing 15-100 ng of purified PCR product with 100-300 ng of miniprep purified or manufacturer provided plasmid and 2 µl BP clonase II™, making the reaction up to 10 µl with sterile distilled water and incubating it for one hour or overnight at room temperature. To generate expression clones for Gateway™ cloning (Invitrogen), the LR reaction was carried out by mixing 100-300 ng of miniprep purified entry clone with 100-300 ng of miniprep purified destination vector and 2 µl LR clonase II™, making the reaction up to 10 µl with sterile distilled water and incubating it overnight at room temperature.

2.5.6 Bacterial glycerol stocks

Overnight bacterial culture (1.5 ml) was taken and pelleted by centrifugation and washed with LB liquid medium before being re-suspended in 300 µl of LB liquid, mixed with 300 µl of 80% glycerol (v/v) and stored at -70 ^{o}C.

2.6 Moss transformation and selection

Stable transformation of moss was performed using polyethylene glycol (PEG)-mediated delivery of DNA to protoplasts as previously described (Schaefer et al., 1991; Kamisugi et al., 2005). One to two plates of moss tissue homogenate was grown on cellophane on BCDAT agar medium for 5-6 days. Tissue was

harvested and digested with 10 mg/ml driselase (Sigma) to remove the cell walls and protoplasts were separated from undigested tissue by filtration through a 100 µm metal mesh (Sigma). Protoplasts were recovered by centrifugation (100 *x g*; 4 minutes) and washed in 8% (w/v) mannitol solution three times, before being finally resuspended at a density of 1.6 x 10^6 protoplasts/ml. Then 2.4 x 10^5 protoplasts were incubated with 10-15 µg of linear DNA and 20% (w/v) PEG-6000 and heat shocked for five minutes at 45 ^oC. The protoplasts were diluted with 8% mannitol before being recovered by centrifugation and resuspended in BCDAT containing 6% (w/v) mannitol and 10 mM $CaCl_2$ for incubation in darkness at 25 ^oC overnight. Protoplasts were embedded in BCDAT containing 6% (w/v) mannitol, 10 mM $CaCl_2$ and 0.4% agar on a cellophane overlaying the same medium containing 0.55% agar and allowed to regenerate for 5-6 days. The cellophane overlay was transferred onto BCDAT-agar medium containing 50 µg/l G418 for two weeks, to select transformed plants. Because the initial transformants comprise a mixture of stably transformed plants (transgene integrated in the genome) and unstable transformants (transgene maintained extrachromosomally), G418-resistant plants were then cultured for two weeks on medium lacking G418, followed by subculture for a further two weeks on medium containing G418 in order to recover only stably transformed plants.

2.7 Yeast two-hybrid

2.7.1 Strains, plasmids and growth conditions used in yeast two-hybrid

The budding yeast strains used in this work are AH109 (MATa) and Y187 (MATα) (Clontech), which are two compatible mating strains. These strains are auxotropic mutants unable to grow on medium lacking the amino acids tryptophan (W), leucine (L) and histidine (H). AH109 carries the *HIS3* reporter gene regulated by the GAL4-responsive *cis*-element. For untransformed yeast, growth was achieved on YPDA medium (Section 2.2; contains all essential amino acids for auxotropic mutants) but for yeast transformed with a plasmid, selection of transformants was achieved by growing on SD medium lacking the appropriate amino acid(s) (Section 2.2). The pGBKT7 plasmid (Clontech) modified for Gateway^{TM} was used as the destination vector to generate the yeast two-hybrid bait construct. This plasmid has a W biosynthesis gene, allowing selection of transformants with SD-W medium. The pGADT7 Rec plasmid (Clontech) modified for Gateway^{TM} was used as the destination vector to generate the yeast two-hybrid prey expression clones, which has L biosynthesis gene allowing for selection on SD-L medium. Diploid yeast containing both bait and prey plasmids were selected on SD media minus both W and L. To test for interactions, yeast was grown on SD media minus W, L and H with appropriate amounts of 3-amino-1,2,4-triazole (3-AT) to increase the stringency of the assay. To test for autoactivation of the *HIS3* reporter gene by the bait construct alone, the yeast strain AH109 was transformed with the bait

construct and yeast growth was monitored on SD-WH with varying concentrations of 3-AT (see figure legends for details). Yeast was grown at 29 ^{o}C on solid medium or in liquid medium in a shaking incubator at 200 rpm.

2.7.2 Yeast transformation

Small-scale yeast transformations were performed to introduce bait and prey yeast two-hybrid expression vectors into their appropriate host strains, using a previously published Li^{+}/PEG-mediated protocol (Causier and Davies, 2002), except that YPDA was used in place of YPD. Yeast transformations were plated on SD medium lacking appropriate amino acid to select for transformed colonies.

2.7.3 Yeast mating to test for individual interactions

To test for interactions between two proteins, mating of yeast strains AH109 and Y187 carrying vectors of interest was carried out as described previously (Causier and Davies, 2002), but using YPDA agar instead of YPD agar. Diploids were selected on SD lacking tryptophan and leucine.

2.7.4 Library screening

Bait vector transformed into the Y187 strain was used to screen previously generated libraries in AH109. These *A. thaliana* libraries are normalised to

ensure that both rare and abundant transcripts are represented to similar extents in the cDNA library rather than being proportional to expression level in *A. thaliana*. The libraries were generated using the Clontech SMART cDNA synthesis and recombination-mediated cloning strategies, in the pGADT7 Rec vector. One library was produced using oligo(dT) primers to generate of full-length or near full-length cDNAs from *A. thaliana* and the other was produced using random hexamers to generate cDNAs of random truncated fragments of *A. thaliana* genes and has previously been used in the Davies laboratory (Causier et al., 2012). Briefly, Y187 yeast containing the bait vector was grown for approximately 21 hours in 50 ml of liquid SD lacking W at 29 ^{o}C in a shaking incubator at 200 rpm, until the total cell count number reached approximately 2 x 10^{9}. These cells were pelleted and resuspended in fresh medium for mating (50 ml of 2xYPDA). For each library, 2 x 10^{7} cells were added to the bait culture. Mating proceeded for 24 hours at 29 ^{o}C in a shaking incubator at 50 rpm. Finally the mating solution was plated out across approximately 50 square Petri plates (120 mm x 120 mm; Gosselin) with SD-LWH supplemented with 3-AT (2.5 mM) and incubated at 29 ^{o}C for six days. To estimate the number of diploids, yeast culture was plated on SD-WL. Positive colonies that grew on SD-LWH supplemented with 3-AT (2.5 mM) after six days were picked and spread on a new SD-LWH supplemented with 3-AT (2.5 mM) to maintain the clone and used as template in yeast colony PCR (Section 2.4.6). Finally, PCR products were purified (Section 2.4.9) and sent for sequencing (Section 2.4.11).

2.8 Analysis of high-throughput RNA-sequencing

2.8.1 Sample preparation

Moss tissue grown for five-days post homogenisation on BCDAT medium overlaid with cellophane was collected for RNA-seq analysis. RNA was extracted using the RNeasy Plant Mini kit (Qiagen) following the manufacturer's instructions. To remove contaminating genomic DNA, the on-column DNase I digestion (Qiagen) was performed according to the manufacturer's instructions.

2.8.2 Library generation and sequencing

Library generation and RNA-sequencing (RNA-seq) were performed by GATC Biotech. Approximately 2 µg of RNA from each sample was used to generate a library after quality control check using the Bioanalyser (Agilent; Section 2.4.4) using the TruSeq sample prep kit (Illumina). Sequencing was performed using the Illumina HiSeq200 platform set to 100 bp single-end reads. In two channels, only three libraries were multiplexed to ensure a high number of reads for each library. In the third channel, five libraries were re-sequenced due to low number of reads generated in the first two and these were used as technical replicates.

2.8.3 Read mapping to the moss genome

Adapter sequences were identified and removed from reads generated from RNA-seq. Subsequently, bad quality reads were removed before alignment to the moss genome. Short reads (<30 nt) post-trimming were removed from analysis. The RNA-seq reads were aligned to the moss reference genome Version 1.6 (Zimmer et al., 2013). TopHat (tophat-1.3.1.Linux_x86_64 custom version supporting additional bowtie1 parameters) was used to map reads spanning exon-exon junctions (Trapnell et al., 2009). Normalisation of RNA-seq read count data was achieved by adjusting read counts to the total gene length and the number of total reads mapped in that sample. This is the reads per kilobase of exon model per million mapped reads (RPKM) method and was used to calculate the fold-changes between samples (Mortazavi et al., 2008). To test for differential gene expression (DGE), three different statistical tools were used; DESeq (Anders and Huber, 2010), edgeR (Robinson et al., 2010) and NOISeq (Tarazona et al., 2011). Both DESeq and edgeR models count data using a negative binomial distribution. On the other hand, NOISeq makes no assumptions regarding distribution and is a non-parametric test. For initial analysis of the data, differentially expressed genes were defined as those that were predicted to be significantly over- or under-expressed by two or three of these three statistical tools.

2.8.4 Functional clustering of genes whose expression is changed in *smg1Δ* lines

Gene ontology (GO) is a gene functional classification system that defines the predicted functions and properties of genes. Therefore, many unrelated genes whose gene products are predicted to function in the same pathway are grouped together within a single GO term. MapMan classifies genes in a similar manner to GO terms but was developed in a plant centric way, focusing on research in plants to aid classifications. For GO or MapMan analysis. Benjamini-Hochberg false discovery rate tests were performed to identify significantly enriched terms in either genes that were over- or under-expressed in one treatment compared to another. The corrected for multiple testing p-values were used to assign enrichment of a GO term with a cut off set to $p<0.05$.

3 Chapter Three: Characterisation of the NMD Pathway of Moss

3.1 Introduction

3.1.1 The NMD pathway of plants is understudied relative to the NMD pathways of animals and fungi

As discussed in Chapter One, NMD plays an important role in regulating a range of transcripts, including those with uORFs and splice variants that contain PTCs (Gaba et al., 2005; Nyikó et al., 2009; Lareau et al., 2007). NMD regulates 1-10% of the transcriptomes of all examined eukaryotes (Guan et al., 2006; He et al., 2003; Mendell et al., 2004; Ramani et al., 2009; Rayson et al., 2012a; Rehwinkel et al., 2005) by degrading selected, non-aberrant transcripts (Mühlemann et al., 2008). Work on identifying and characterising components of the NMD pathway has focused on animal and yeast models. Only a few studies have been performed looking at NMD in more evolutionarily diverse organisms. Plants are distantly separate from animals/fungi and some evidence has suggested that the last common ancestor of plants and animals/fungi was the last eukaryotic common ancestor (LECA), also called the stem eukaryote (Richards and Cavalier-Smith, 2005; Derelle and Lang, 2012). Therefore, studying the NMD pathway in plants and comparing it to the NMD pathways of animals and fungi will reveal how the NMD pathway of the stem eukaryote

operated. Some work has been performed on plant NMD, however, all the studies have used angiosperms (flowering plants). Using *A. thaliana* or a tobacco transient expression system (expressing *A. thaliana* or grape genes in tobacco leaves), the predicted roles of UPF1, UPF2, UPF3 and SMG7 in the NMD pathway of flowering plants have been confirmed (Arciga-Reyes et al., 2006; Riehs et al., 2008; Kerényi et al., 2008; Benkovics et al., 2011).

In budding yeast, the UPF1-UPF2-UPF3 complex is essential for NMD (He et al., 1997) and this is conserved in animals (Chamieh et al., 2008). In animals but not budding yeast, UPF1 takes a central role in NMD by being phosphorylated by the SMG1 kinase, a step promoted by UPF2 and UPF3 (Kashima et al., 2006). Once UPF1 is phosphorylated in animals, SMG5-7 are recruited to specific phosphorylated residues and SMG6 acts as an endonuclease cutting the transcript near the PTC, while SMG7 and SMG5 form a complex and recruit both the XRN1 exoribonuclease and the exosome complex, to degrade the transcript in the 5'-3' and 3'-5' directions, respectively (Unterholzner and Izaurralde, 2004; Jonas et al., 2013). The details of the signaling pathway that leads to mRNA decay in plant NMD are still unclear, but it is known that at least the UPF1-UPF2-UPF3 complex is conserved in plants (Kerényi et al., 2008). The *A. thaliana* UPF1 is phosphorylated when expressed in tobacco cells and loss of the phosphorylated regions of UPF1 eliminates its ability to function in NMD (Mérai et al., 2012). In addition, SMG7, predicted to bind to phosphorylated UPF1 through its 14-3-3-like domain, is important for NMD in *A. thaliana* (Riehs et al., 2008; Rayson et al., 2012a) and this function is dependent on conserved sites predicted to bind to the phosphorylated residues

of UPF1 in tobacco cells (Mérai et al., 2012). These data suggest that plant NMD involves phosphorylation of UPF1 to recruit SMG7, which then recruits the RNA decay machinery. However, no kinase has yet been identified that functions in the NMD pathway of a plant. *SMG1* is absent from the genomes of *A. thaliana* and fungi (Grimson et al., 2004), leading to the dogma that SMG1 is an animal specific kinase (Izumi et al., 2010).

To gain a better understanding of the evolution of the NMD pathway across eukaryotes, the moss *P. patens* was chosen as the model plant. Moss has the advantages of a completely sequenced genome to facilitate identification of NMD effectors (Rensing et al., 2008), is distantly related to all other plants used to study NMD and has a high rate of homologous recombination, allowing for the generation of knockout moss lines to study gene function (Kamisugi et al., 2005).

3.1.2 The aims of studying the NMD pathway of model basal land plant *P. patens*

The aims of the experimental work of this chapter are:

1. To develop *P. patens* as a new model for NMD in the plant kingdom to determine how widespread the mechanisms of plant NMD are
2. To identify all the putative NMD effectors of *P. patens* using homology searches
3. To compromise the NMD pathway of *P. patens* using targeted gene replacement to studying the types of NMD targets in moss

3.2 Results

3.2.1 Identification of putative NMD effectors

To examine the level of conservation of NMD effectors across land plants, the recently published moss genome was utilised and reciprocal BLASTp searches were performed using *A. thaliana* or animal sequences as input. The moss genome contains homologues of many of the NMD effectors identified in other organisms (Table 3.1). Moss has genes encoding the 'core' NMD effectors UPF1-3 and also SMG7. Some of these NMD effectors are encoded by multigene families in moss. Moss has two copies of both *UPF1* and *UPF3* and three copies of *SMG7* (Table 3.1). In the case of *SMG7*, both *PpSMG7-1* and *PpSMG7-2* have been previously identified (Benkovics et al., 2011) and encode proteins much more similar to *A. thaliana* SMG7 (E-values of e-116 and e-117, respectively) than *PpSMG7-3* does (E-value of 9e-50). A putative third copy of *UPF3* was identified in the older gene models (Pp1s41_44V2.1) but this encodes a short truncated protein and has a very low expression level, measured using RNA-seq (Chapter 5) and is therefore likely to be a pseudogene. Unexpectedly, *SMG1*, the gene that encodes the kinase that activates NMD in animals, was identified in moss despite its reported absence in *A. thaliana* (Table 3.1) (Grimson et al., 2004). *SMG8* and *SMG9*, which encode interactors of SMG1 that are also absent from the *A. thaliana* genome, have been identified in the genomes of rice and grape (Yamashita, 2013) and are also present in the moss genome (Table 3.1).

Table 3.1: A list of putative NMD effectors identified by BLASTp searches

NI represents not identified

	NMD effector	Moss gene	Gene model	*A. thaliana* homologue
'Core' NMD effectors	UPF1	PpUPF1a	Pp1s44_135V6.1	AtUPF1 (AT5G47010)
		PpUPF1b	Pp1s10_103V6.1	
	UPF2	PpUPF2	Pp1s123_14V6.1	AtUPF2 (AT2G39260)
	UPF3	PpUPF3a	Pp1s246_94V6.1	AtUPF3 (AT1G33980)
		PpUPF3b	Pp1s13_385V6.1	
Kinase	SMG1	PpSMG1	Pp1s51_180U2__zimmer.1	NI
SMG5-7 family	SMG7	PpSMG7-1	Pp1s80_14V6.1	AtSMG7 (AT5G19400) and AtSMG7L (AT1G28260)
		PpSMG7-2	Pp1s311_73V6.1	
		PpSMG7-3	Pp1s28_218V6.1	
Additional regulators	SMG8	PpSMG8	Pp1s292_13V6.1	NI
	SMG9	PpSMG9	Pp1s83_237V6.1	NI
	SMG1L	NI	NI	AtSMG1L (AT5G24350)
	SMG2L	NI	NI	NI
Core EJC	Mago	PpMago1	Pp1s63_105V6.1	AtMago (AT1G02140)
		PpMago2	Pp1s125_34V6.1	
	Y14	PpY14a	Pp1s136_114V6.1	AtY14 (AT1G51510)
		PpY14b	Pp1s31_259V6.1	
	eIF4AIII	PpeIF4AIIIa	Pp1s275_62V6.1	AteIF4A3 (AT3G19760)
		PpeIF4AIIIb	Pp1s519_18V6.1	
	MLN51/ CASC3/ Barentsz	PpMLN51a	Pp1s193_55V6.1	AtBarentsz1 (AT1G80000) and AtBarentsz2 (AT1G15280)
		PpMLN51b	Pp1s193_54V6.1	

Additional NMD effectors have been identified but have been less well characterised. SMG Lethal-1 (SMGL1/NAG/NBAS) and SMG Lethal-2 (SMGL2/DHX34) were identified as essential in development of *C. elegans* and functions in NMD (Longman et al., 2007), however, the exact role in NMD is unclear. A protein with weak homology to SMGL1 has been identified in *A. thaliana* AT5G24350 (Table 3.1). BLASTp identified no protein with homology to SMGL1 in moss (Table 3.1). Interestingly, SMGL1 from *C. elegans* has no identifiable domains, while AT5G24350 has a Sec39 domain, suggesting a role in the secretary pathway rather than in NMD. When using *C. elegans* SMGL2 as input, BLASTp searches reveal that both *A. thaliana* and moss encode several putative homologues but these do not identify the *C. elegans* SMGL2 as the top BLASTp hit during the reciprocal BLASTp search.

In the case of the EJC, each of the core EJC members (Mago, Y14, BTZ and eIF4AIII) appear to be represented by two genes in moss. Mago has an additional third locus that resembles a pseudogene and was predicted by an old gene model (Pp1s63_110V2.1). This pseudogene contains a PTC and mutations at highly conserved positions and is expressed at a relatively low level (RNA-seq data; Chapter 5). It is perhaps not surprising that moss has duplicated copies of many NMD and EJC components, since it has undergone a recent genome duplication leading to a large number of duplicated genes (Rensing et al., 2007; 2008), although this also raises the possibility of EJCs comprised of different subunits with specialized properties.

3.2.2 Bioinformatic analysis of plant SMG1

The surprising presence of a putative SMG1 kinase in moss revealed by BLASTp searches presented the opportunity to better understand the role of a kinase in the NMD pathway of plants. *SMG1* is absent from the genomes of *A. thaliana* and the yeasts *S. cerevisiae* and *S. pombe* (Grimson et al., 2004) and was therefore proposed to be an animal specific component of the NMD pathway (Izumi et al., 2010), although some reviews have suggested the presence of *SMG1* in grape and/or rice based on unpublished BLAST searches (Templeton and Moorhead, 2005; Yamashita, 2013). SMG1 is a member of the phosphatidylinositol 3-kinase-related kinase (PIKK) family of kinases (Lempiäinen and Halazonetis, 2009). This family also contains TARGET OF RAPAMYCIN (TOR), an inactive kinase TRANSFORMATION/TRANSCRIPTION DOMAIN-ASSOCIATED PROTEIN (TRRAP), the DNA repair activating kinases ATAXIA TELANGIECTASIA MUTATED (ATM), ATAXIA TELANGIECTASIA AND RAD3-RELATED (ATR) and DNA-DEPENDENT PROTEIN KINASE, CATALYTIC SUBUNIT (DNA-PKcs) (Lempiäinen and Halazonetis, 2009). *A. thaliana* has four PIKKs: ATM, ATR, TRRAP and TOR (Templeton and Moorhead, 2005). To confirm the *PpSMG1* identified here is a true orthologue of animal *SMG1* rather than a different PIKK, a phylogenetic tree was constructed using ClustalX and visualised in FigTree using the midpoint to root the tree (Figure 3.1). The kinase domain was used in the alignments to minimize the introduction of artifacts caused by poor gene model prediction and variation of domain composition and spacing in the different PIKKs. All plant SMG1 sequences, including the moss

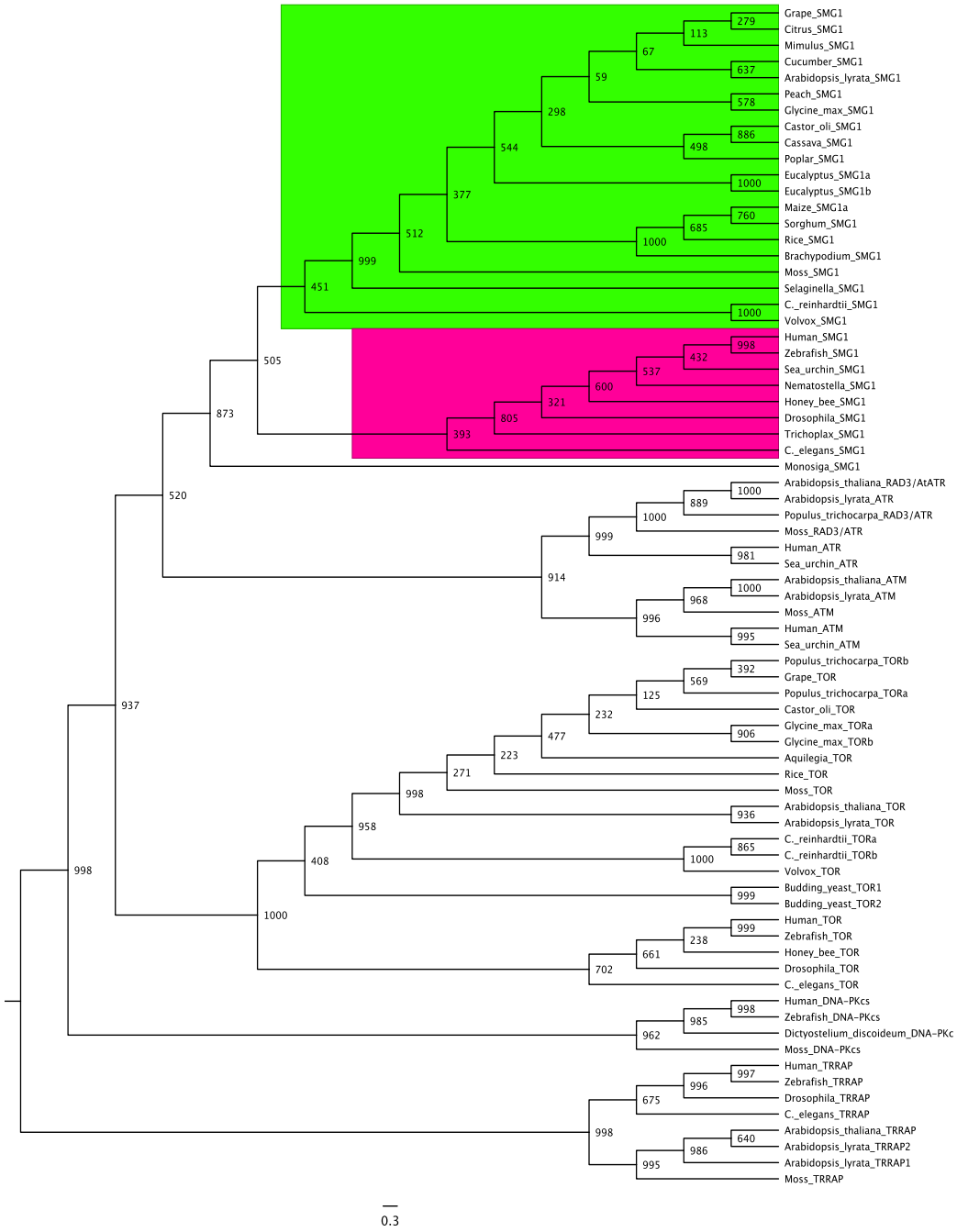

Figure 3.1: Neighbor joining phylogenetic tree of the KD of PIKK

Sequences of multiple PIKK proteins were collected and the protein kinase domains (KD) were identified using SMART and aligned using ClustalX. Bootstrap values (1000 replicates) are indicated at nodes. Midpoint rooting was performed. The plant SMG1 clade is shaded green and the animal SMG1 clade is shaded smoky red.

sequence, form a sister clade to the animal SMG1 clade (Figure 3.1), suggesting that plant *SMG1* sequences are true orthologues of animal *SMG1* sequences. Further confirmation of the phylogenetic analysis of PIKKs was obtained by examining the PIKK domain structure. All PIKKs, with the exception of SMG1, have a kinase domain (KD) adjacent to the FRAP, ATM, TRRAP C-terminal (FATC) domain in the C-terminal region of the protein (Lempiäinen and Halazonetis, 2009) (Figure 3.2A). All SMG1 proteins identified phylogenetically in this study, from both animals and plants, have a centrally located KD, separated from the FATC domain by a large middle region (Figure 3.2A). These data suggest SMG1 is not animal specific and is widespread in the plant kingdom.

The presence of SMG1 outside the animal kingdom suggests that the mechanism used to phosphorylate the NMD effector UPF1 is more widely conserved than previously thought. *A. thaliana* is the only green plant examined that does not contain an *SMG1* orthologue (Figure 3.1) (Grimson et al., 2004). To explore this further, the expected location of *SMG1* was identified in the *A. thaliana* genome using synteny with the genome of its close relative, *A. lyrata* which does contain *SMG1* (Figure 3.2B). Comparison of the gene content in this syntenic region confirms the absence of *SMG1* from the *A. thaliana* genome. At the expected position of *SMG1* the *A. thaliana* genome instead contains two unrelated transposable elements bracketing a short predicted ORF of 59 amino acids (*ORF59*, Figure 3.2B). This ORF displays weak homology to *SMG1* in its reverse strand, revealing the remains of the kinase-encoding locus that was lost within the last 5-10 million years (Hu et al., 2011). These results

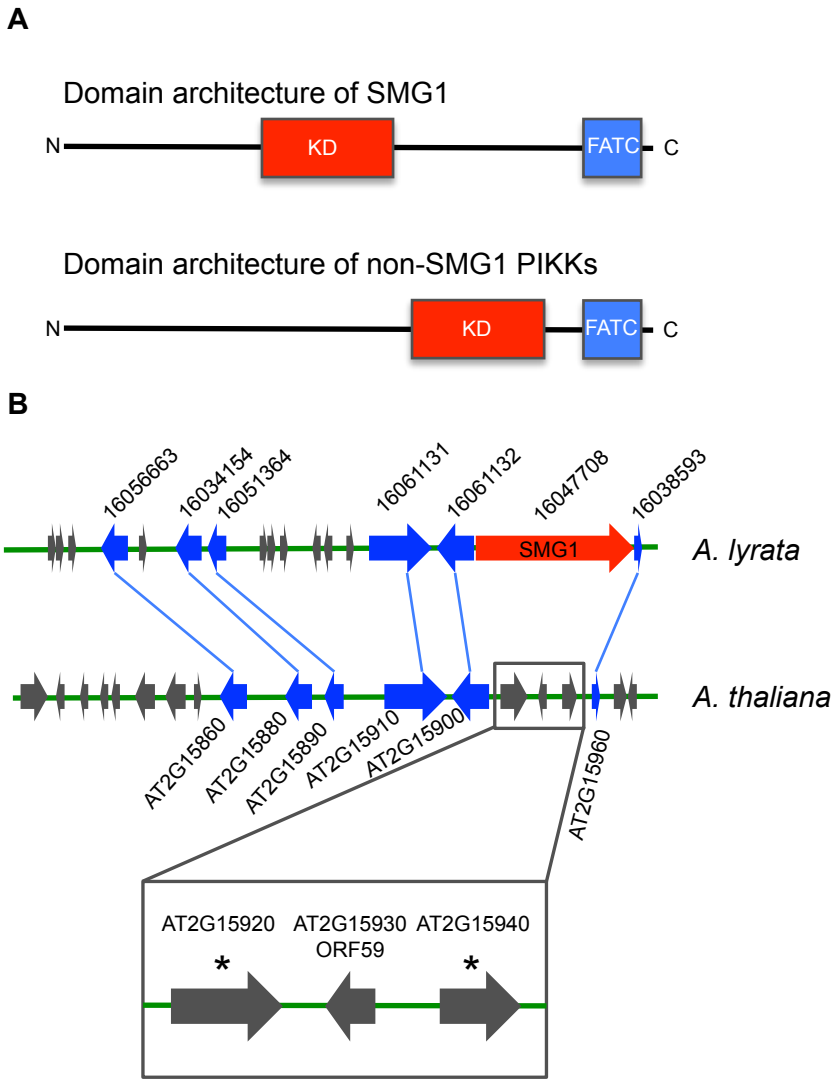

Figure 3.2: Conservation of SMG1 in some plants but loss in *A. thaliana*

(A) Domain structure of PIKKs**.** The kinase domain (KD) of SMG1 is centrally located, distant from the C-terminally located FATC domain in both animals and plants. The KD of other PIKKs (ATM, ATR, TOR, TRRAP and DNA-PKcs) is directly adjacent to the FATC domain at the C-terminal. **(B)** The *Arabidopsis* syntenic region that contains *SMG1* in *A. lyrata* has no SMG1 in *A. thaliana*. The expected position of *SMG1* contains different genes, including two unrelated transposable elements (denoted by asterisks). Synteny was examined using the Plant Genome Duplication Database (http://chibba.agtec.uga.edu/duplication/).

show that orthologues of *SMG1* are present in diverse eukaryotes, including green algae and land plants. However, they also confirm the absence of *SMG1* from fungi and demonstrate that the *SMG1* gene has been recently lost in the *A. thaliana* lineage, despite *A. thaliana* retaining a competent NMD pathway (Rayson et al., 2012a; Arciga-Reyes et al., 2006; Riehs et al., 2008; Kerényi et al., 2008).

These data suggest that SMG1 is more ancient than previously suggested. Recent work has suggested that the clade including plants/oomycetes and the one including animals/fungi are on opposing sides of the root of eukaryotes (Richards and Cavalier-Smith, 2005; Derelle and Lang, 2012). The presence of *SMG1* in both clades therefore suggests that *SMG1* was present in the LECA (Figure 3.3), dating its origin to over two billion years ago (Brocks et al., 1999).

3.2.3 Generating moss knockout lines lacking *SMG1*

As the function of the SMG1 kinase has never been studied outside the animal kingdom before, moss was chosen as an ideal model organism to understand the role of SMG1 in NMD in plants. Moss has a higher rate of homologous recombination than flowering plants, meaning targeted gene replacement can easily be achieved (Kamisugi et al., 2005). Therefore, a gene of interest can be effectively deleted by replacing it with an antibiotic selection cassette. Gene model prediction in the moss genome (Version 1.6) wrongly identified and split *PpSMG1* into two gene models (Pp1s51_180V6.1 and Pp1s51_182V6.1) (Zimmer et al., 2013), therefore these gene models were fused together to

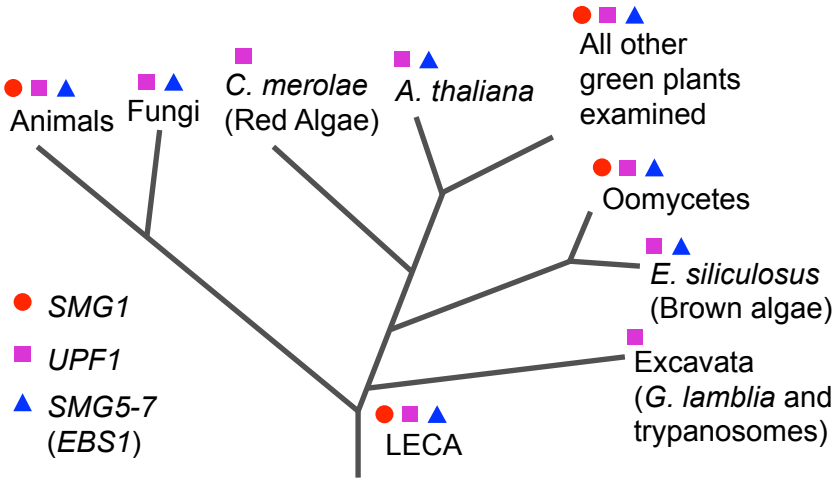

Figure 3.3: Conservation of NMD effectors across the eukaryotic domain

Tree of relationships between selected eukaryotes with root placed between plants and animals as suggested by Richards and Cavalier-Smith (2005) and Derelle and Lang (2012). Conservation of the NMD effectors UPF1, SMG5-7 (EBS1) and SMG1 (indicated by symbols), as assessed by homology searches and phylogenetic analysis. Note independent losses of the ancestral SMG1 (red circle), in fungi, *A. thaliana*, the red algae *C. merolae*, the brown algae *E. siliculosus* and excavata (*Trypanosoma brucei* and *Giardia lamblia*) and its predicted presence in the last eukaryotic common ancestor (LECA).

make a single accurate gene model named Pp1s51_180U2__zimmer.1 (D. Lang and A. Zimmer, personal communication). Approximately one Kb of sequence upstream of the start codon and approximately one Kb of sequence downstream of the stop codon was used to generate the knockout construct (Figure 3.4A). The knockout construct (designated pKO SMG1) was generated using three-site multisite Gateway™. The selection cassette consists of the *nptII* coding sequence (kanamycin resistance gene) driven by the 35S promoter with a 35S terminator sequence, flanked by loxP sites. The backbone used was the pDEST22 yeast two-hybrid construct because it had an appropriate bacterial selection, is high-copy number and Gateway™ compatible. To obtain the DNA required to transform and generate knockout lines of *PpSMG1*, the linear *smg1*KO fragment (Figure 3.4A) was amplified by PCR. However, the desired four Kb fragment was not amplified using the Phusion™ polymerase. A two Kb PCR product was instead observed (Figure 3.5), which sequencing revealed to be the expected PCR product but without the selection cassette. Given the high homology of the Gateway™ sites within the plasmid, it was possible that complex base-pairing/looping within the plasmid was causing the DNA polymerase in the PCR to strand-switch. Therefore, a single-stranded DNA binding protein (ET-SSB) sold as a PCR additive (NEB) along with betaine to improve processivity was added to the PCR reactions. It was found that the optimal PCR condition to bias the reaction towards making the full-length PCR product of 4 Kb using DreamTaq (Fermentas) with ET-SSB (NEB) and 20 mM betaine (Figure 3.5).

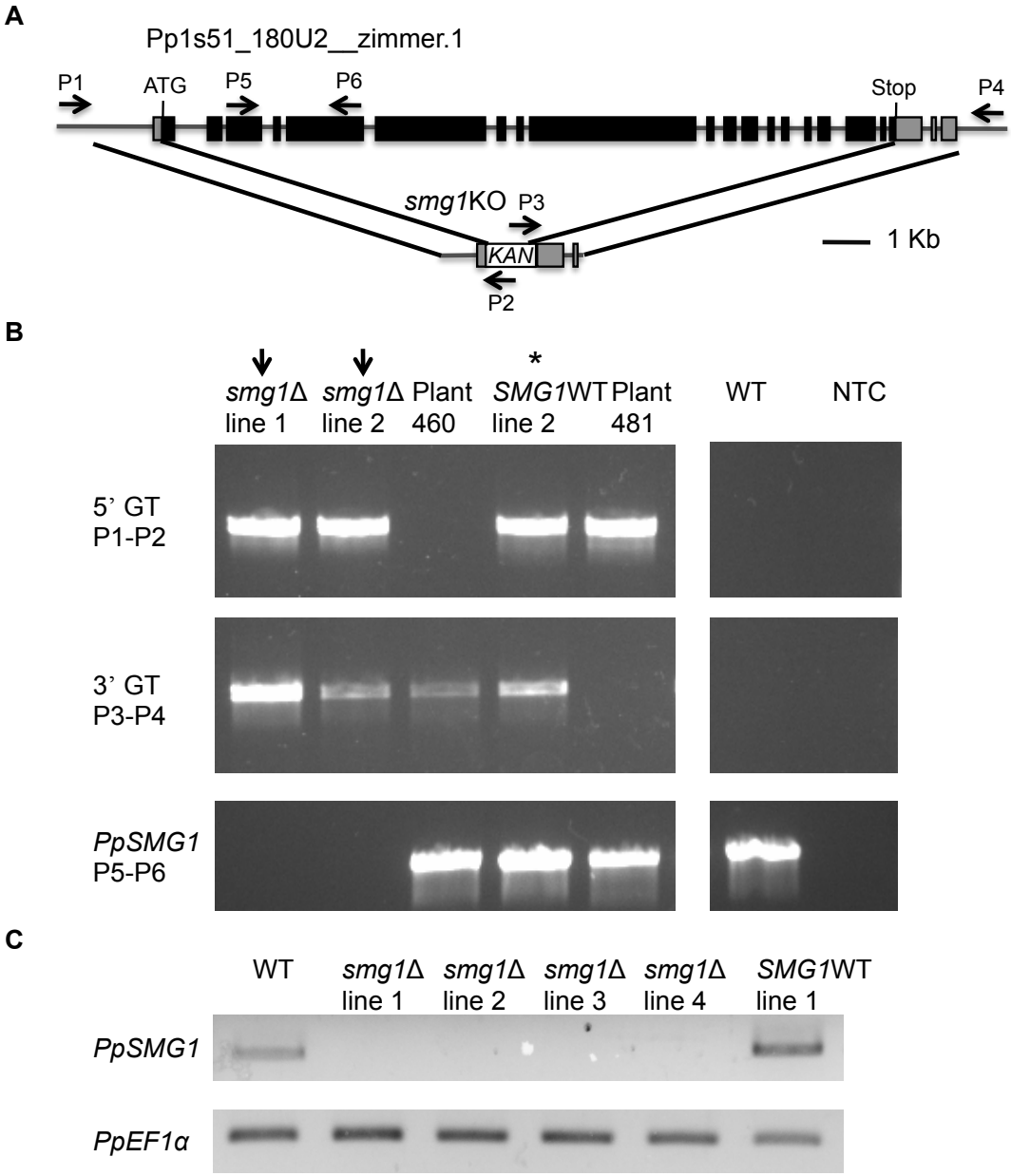

Figure 3.4: Targeted disruption of *SMG1* in moss

(A) The structure of the *PpSMG1* gene in moss with coding exons in black boxes, UTRs in grey boxes. The whole coding region (Pp1s51_180U2__zimmer.1) was replaced by the *P35S-nptII-g6term* selection cassette, shown as a white box (KAN; kanamycin resistance gene) with regions of homology used in gene targeting on either side. *Legend continued on next page*

Figure 3.4: *Continued…* The selection cassette and homology regions are cloned from pKO SMG1. Primers used for PCR genotyping are shown as arrows P1-P6. **(B)** Example genotyping. PCR determined whether successful gene targeting (GT) was achieved at the 5' (primers P1 and P2) and (primers P3 and P4) 3' ends of the *SMG1* locus and whether lines also carried an undisrupted copy of *PpSMG1* (primers P5 and P6). Wild type (WT) and no template controls (NTC) were set up. Some lines have GT events at both ends of *PpSMG1*, successfully deleting *PpSMG1* (arrows). Other lines have successful GT events at both ends but still retain a copy of *PpSMG1* (asterisk). Examples of unsuccessful 5' or 3' targeting are also shown (Plant 460 and Plant 481). **(C)** sqRT-PCR expression analysis in WT, four mutant lines with 5' and 3' gene targeting and no genomic *PpSMG1* (*smg1Δ* lines 1-4) and one mutant line with gene targeting and genomic *PpSMG1* (*SMG1WT* line 1).

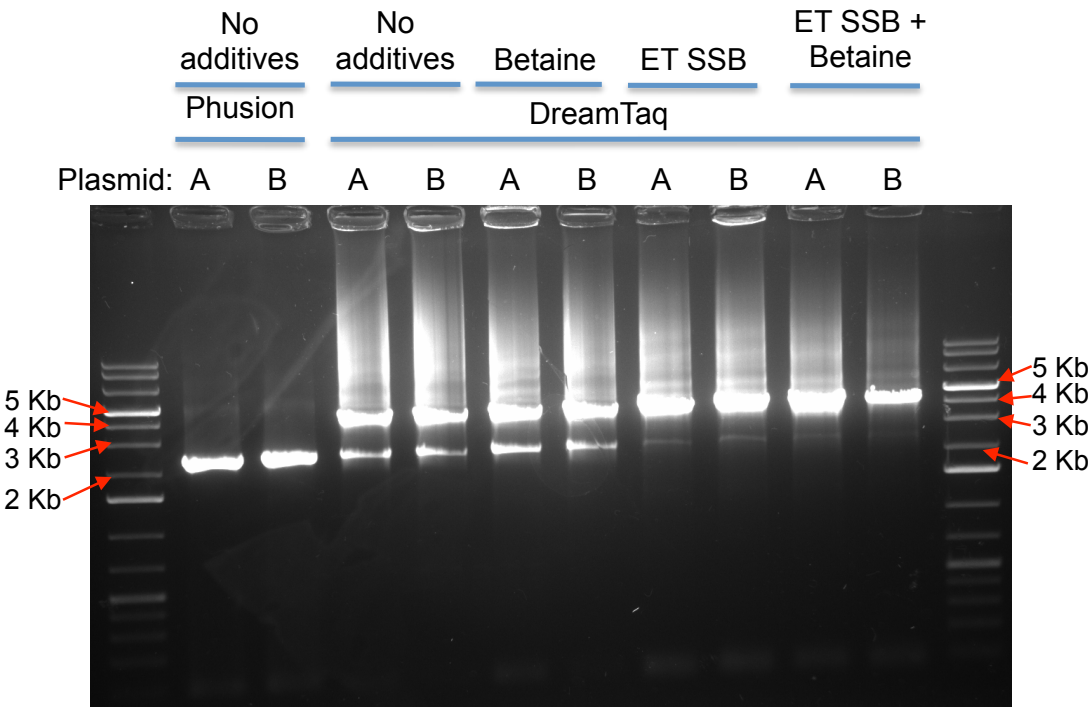

Figure 3.5: Optimisation of PCR conditions to generate the transforming DNA PCR product from pKO SMG1

The pKO SMG1 plasmid was used as template (A) or the transforming region was cloned into another plasmid backbone (pGAD424) by multisite Gateway™ (B).

The *PpSMG1* (Pp1s51_180U2__zimmer.1) locus was knocked out using homologous recombination, replacing *PpSMG1* with the kanamycin selection cassette. Successful gene targeting (GT) was measured by PCR amplification, testing successful integration at the upstream (5') and downstream (3') sites (Figure 3.4). Thirty-four transformants had GT events at both the upstream and downstream sites suggesting they were complete knockouts but only four lines also lacked the WT locus (Figure 3.4B). This suggests a duplication of genomic sequence that contains *PpSMG1* in many lines. Semi-quantitative (sq)RT-PCR was used to confirm that these four independent lines lacked *PpSMG1* expression (Figure 3.4C). It is unclear how it is possible for *PpSMG1* knockout lines to also contain a genomic copy of *PpSMG1* (*SMG1WT* lines). These lines could have resulted from protoplast fusion events where a diploid cell is formed or perhaps a localised recombination event could have occurred generating a WT site and a knockout site. These *SMG1WT* lines do however, provide a useful control line, in addition to WT, to analyse the role of loss of *PpSMG1* in NMD.

3.2.4 Plant SMG1 functions in the NMD pathway

To examine the role of SMG1 in the NMD pathway of moss, putative NMD targets were identified. If SMG1 functions in the moss NMD pathway then targets of NMD should be up-regulated in *smg1*Δ lines. The transcripts of splicing factors, like SR proteins or polypyrimidine binding (PTB) proteins, are common targets of NMD in *A. thaliana* and animals, creating a feedback loop (Stauffer et al., 2010; Wollerton et al., 2004; Boutz et al., 2007; Palusa and

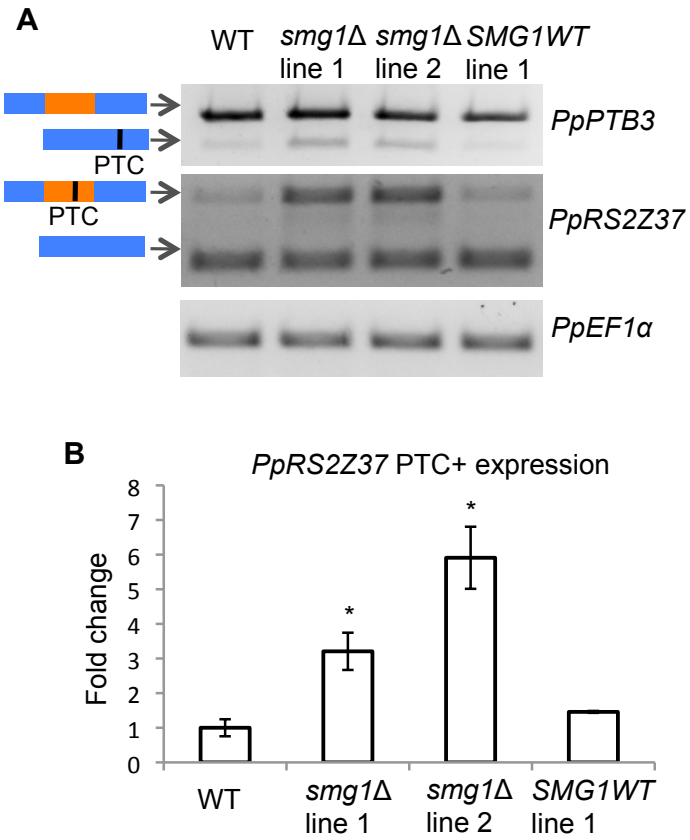

Figure 3.6: PTC+ splice variants are over-expressed in moss *smg1Δ* lines

(A) sqRT-PCR analysis of two alternatively spliced targets of NMD. *PpPTB3* produces two splice variants by exon skipping, the shorter PTC+ variant and the longer, PTC- variant. Alternative acceptor site selection of *PpRS2Z37* produces a longer, PTC+ variant compared to the shorter, PTC- variant. *PpEF1α* level is shown as a control for RNA loading. PTCs are indicated by a vertical black line. Constitutive exon sequences are blue and alternative exon sequences are in orange. **(B)** qRT-PCR analysis of *PpRS2Z37* PTC+ variant. 'Fold change' is the amount of target expression normalised to *PpEF1α* and relative to WT levels. Error bars represent the standard error of the mean from three biological replicates. Asterisks represent lines with a statistical difference from WT using an unpaired *t* test $p < 0.05$.

Reddy, 2010). The moss *PTB* homologue, *PpPTB3* (Pp1s48_128V6) was amplified by sqRT-PCR from WT and *smg1*Δ and two splice variants were identified (Figure 3.6A). Sequencing the PCR products revealed that the long variant encoded part of the full-length transcript (Pp1s48_128V6.1) and the short variant encoded a novel PTC containing (PTC+) variant, due to skipping of exon six. sqRT-PCR showed that the PTC+ variant was more highly expressed in the two mutant lines than in the WT or *SMG1*WT line 1 (Figure 3.6A). The SR protein-encoding transcript *PpRS2Z37* (Pp1s69_23V6) also undergoes AS (alternative acceptor site) to produce a PTC+ variant. sqRT-PCR and quantitative (q)RT-PCR show that this PTC+ splice variant is up-regulated three- to six-fold in *smg1*Δ lines, when compared to WT or control *SMG1WT* line 1 (Figure 3.6A and B).

SMG7, a conserved NMD effector-encoding gene, has also previously been shown to be a direct target of NMD in flowering plants due to the presence of a long 3' UTR and two introns located downstream of the stop codon (Benkovics et al., 2011; Kerényi et al., 2008; Nyikó et al., 2013; Rayson et al., 2012a). Regulation of *SMG7* by NMD creates an autoregulatory loop to control the level of NMD activity (Figure 3.7A), a common feature in the NMD pathway of animals (Rehwinkel et al., 2005; Huang et al., 2011; Yepiskoposyan et al., 2011). Therefore the expression of *PpSMG7-1* (Pp1s80_14V6), *PpSMG7-2* (Pp1s311_73V6) and *PpSMG7-3* (Pp1s28_218V6) was tested in WT/*SMG1*WT line 1 and *smg1*Δ lines (Figure 3.7B-D). The expression of both *PpSMG7-2* and *PpSMG7-3* was increased between 1.5 to 2 fold in the *smg1*Δ lines compared to WT (Figure 3.7D and E). This is in contrast to *PpSMG7-1*, which had

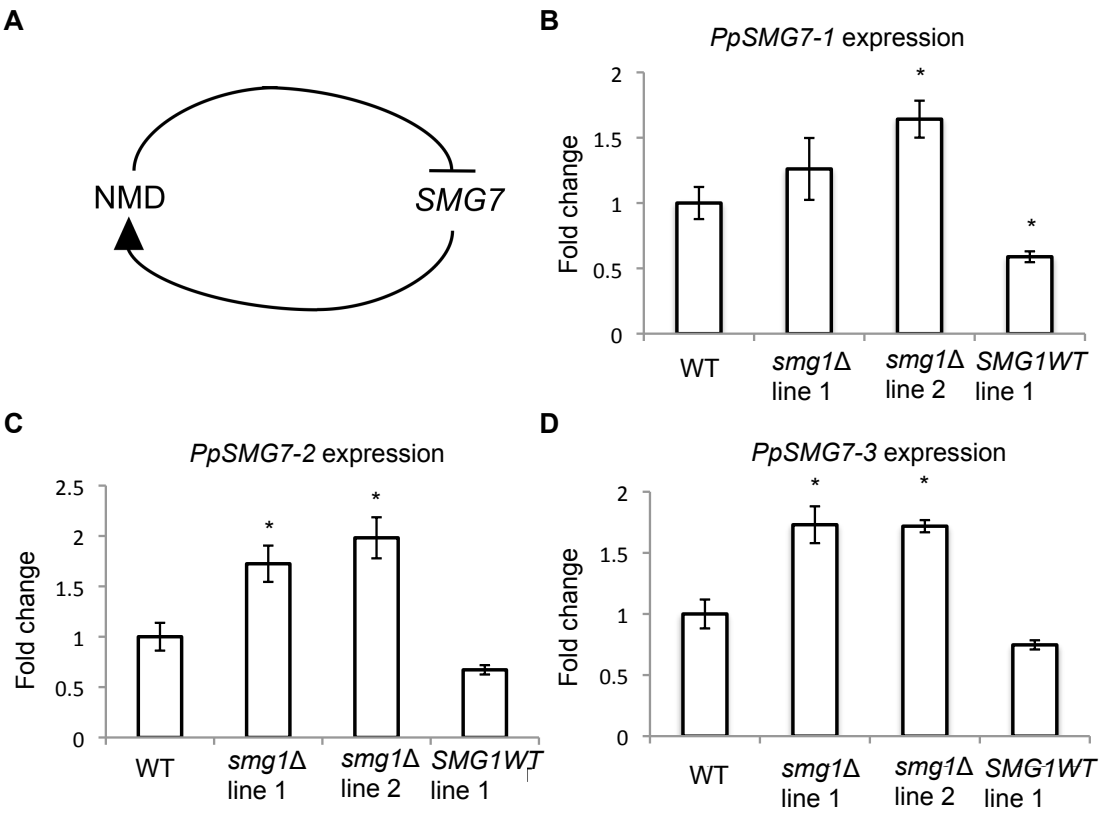

Figure 3.7: Moss *SMG7* homologues are over-expressed in moss *smg1Δ* lines

(A) A model of the autoregulatory feedback of NMD and *SMG7* in plants. **(B)** qRT-PCR analysis of *PpSMG7-1*. **(C)** qRT-PCR analysis of *PpSMG7-2*. **(D)** qRT-PCR analysis of *PpSMG7-3*. **(B-D)** 'Fold change' is the amount of target expression normalised to *PpEF1α* and relative to WT levels. Error bars represent the standard error of the mean from three biological replicates. Asterisks represent lines with a statistical difference from WT using an unpaired *t* test $p < 0.05$.

increased steady state level significantly in *smg1*Δ line 2 and non-significantly in *smg1*Δ line 1 (Figure 3.7B). It appears that *PpSMG7-1* can be up-regulated upon a reduction in NMD activity but this is less extreme and/or more variable than *PpSMG7-2* and *PpSMG7-3*. The 3' UTR structure of *SMG7*-encoding transcripts are conserved across flowering plants (Kerényi et al., 2008; Nyikó et al., 2013; Benkovics et al., 2011). The 3' UTRs of plant *SMG7* usually contain two introns and are abnormally long compared to other 3' UTRs in plants and these features target the transcript to NMD (Kerényi et al., 2008; Nyikó et al., 2013; Benkovics et al., 2011). In the currently released gene models of the moss genome (Version 1.6), the *SMG7* loci do not have predicted 3' UTRs due to poor EST coverage. By using our recent RNA-seq experiments performed on WT and *smg1*Δ lines 1 and 2 (Chapter 5), new gene models were predicted, which have 3' UTRs for all three *SMG7* loci. All three models have abnormally long 3' UTRs and contain two introns as do the flowering plant *SMG7*-encoding transcripts. These data suggest that the NMD autoregulatory loop involving *SMG7* through its abnormal 3' UTR structure is conserved across land plants.

Recent work in *A. thaliana* suggested that CPuORFs can target transcripts to NMD in plants (Rayson et al., 2012b; 2012a; Nyikó et al., 2009). Searching the moss genome for predicted transcripts which show sequence homology to both a CPuORF and its associated downstream main ORFs (mORF) from the list of NMD targets in *A. thaliana*, identified two moss *eIF5* related genes, *PpeIF5-like1* (*PpeIF5L1*; Pp1s626_4V6) and *PpeIF5-like2* (*PpeIF5L2*; Pp1s93_126V6). The largest uORFs of these moss genes show homology to the CPuORF upstream of the *A. thaliana eIF5*-related gene AT1G36730, which shows

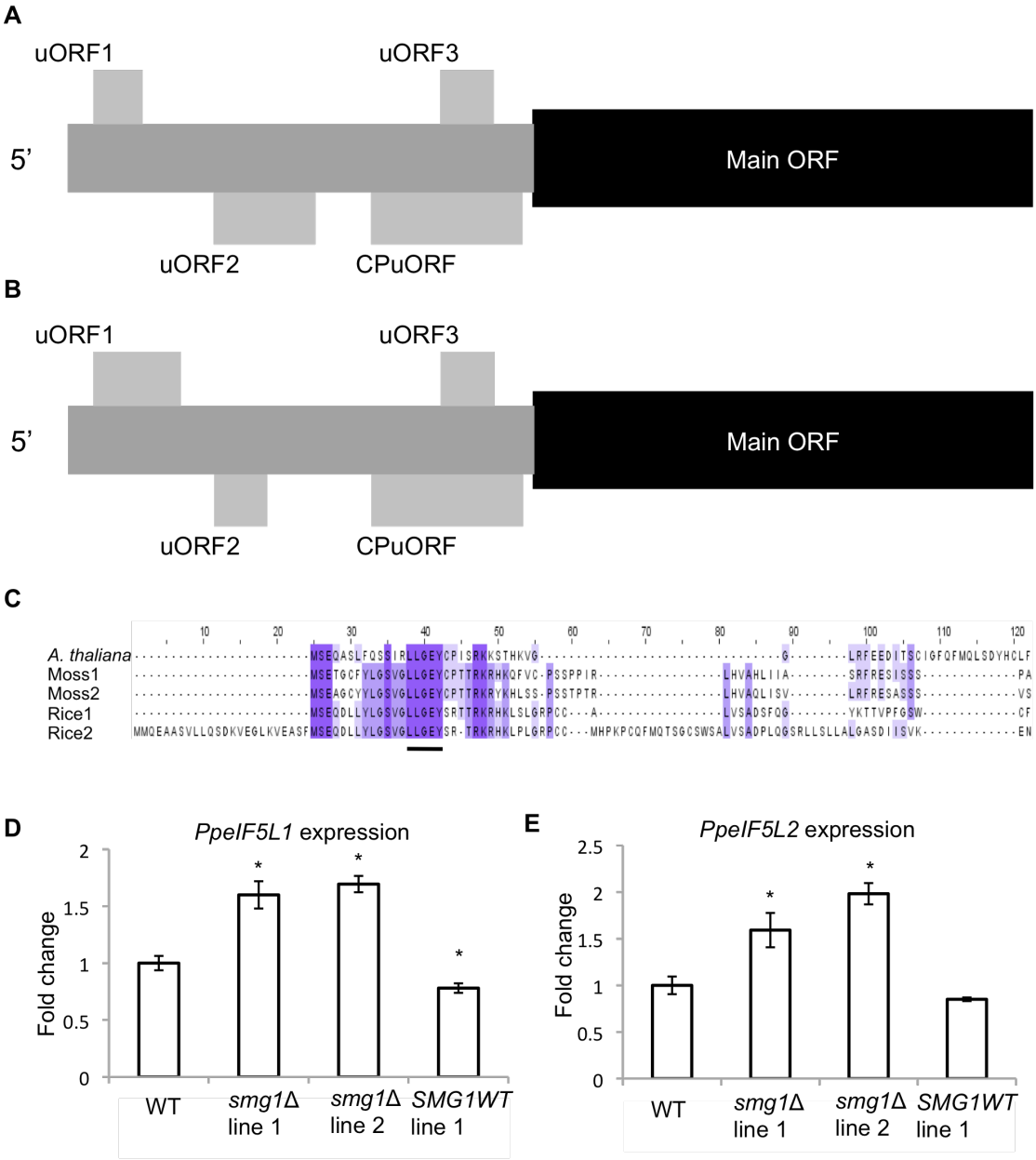

Figure 3.8: CPuORF-containing transcripts are over-expressed in moss *smg1*Δ lines

(A) The presence of uORFs including a CPuORF in *PpeIF5L1* (Pp1s626_4V6) relative to the main ORF. **(B)** The presence of uORFs including a CPuORF in *PpeIF5L2* (Pp1s93_126V6) relative to the main ORF. **(C)** Alignment of peptides encoded by the CPuORFs of *eIF5*-like genes in plants. *Legend continued on next page*

Figure 3.8: *Continued…* The deduced protein sequences of uORF sequences from putative *eIF5* encoding transcripts have been aligned. This includes the *A. thaliana* sequence (AT1G36730), two rice sequences, Rice1 (AK067685) and Rice2 (LOC_Os06g48350) and two moss sequences, Moss1 (Pp1s646_4V6) and Moss2 (Pp1s93_126V6). **(D)** qRT-PCR analysis of *PpeIF5L1*. **(E)** qRT-PCR analysis of *PpeIF5L2*. **(D-E)** 'Fold change' is the amount of target expression normalised to *PpEF1α* and relative to WT levels. Error bars represent the standard error of the mean from three biological replicates. Asterisks represent lines with a statistical difference from WT using an unpaired *t* test $p < 0.05$.

elevated expression in three NMD mutants *upf1-5*, *upf3-1* and *smg7-1* (Rayson et al., 2012a) (Figure 3.8A-C). The conserved association between this CPuORF and mORF in this *eIF5*-related transcript, from bryophyte to angiosperm, strongly implies a functional dependence between the CPuORF and mORF across all land plants. As found for the *A. thaliana eIF5*-related transcript, both have *PpeIF5L1* and *PpeIF5L2* increased steady state levels in *smg1* mutant lines (Figure 3.8D and E).

To demonstrate further that the observed increased steady state levels of these genes in *PpSMG1* knockout lines is attributable to a reduction in NMD, moss was exposed to cycloheximide (CHX). CHX is an inhibitor of translation, and since NMD is a translation-dependent decay mechanism, CHX has previously been used to identify targets of NMD (Kalyna et al., 2012). Three putative NMD targets were chosen to test this, each representative of a different class of NMD targeting feature: AS-coupled NMD (*PpRS2Z37*), uORF-containing transcript (*PpeIF5L1*) and long 3' UTR (*PpSMG7-2*). A six-hour incubation with CHX caused an increase in the steady state level of each of these transcripts, consistent with their status as targets of NMD (Figure 3.9).

Taken together, these data showing the increase in steady state levels of putative targets of NMD in *smg1*Δ relative to WT and *SMG1*WT lines plants, demonstrate that SMG1 is an NMD effector in moss and therefore likely to be important in other plants and possibly in the LECA.

3.2.5 SMG1 is important but not essential in moss development

A compromised NMD pathway has previously been shown to have an adverse effect on development in animals and plants (McIlwain et al., 2010; Arciga-Reyes et al., 2006; Riehs et al., 2008; Wittkopp et al., 2009; Hodgkin et al., 1989; Metzstein and Krasnow, 2006; Frizzell et al., 2012). Phenotypic analyses of the moss *smg1*Δ lines, showed that the mutants are viable, but produce fewer leafy structures (gametophores) than WT (Figure 3.10A-D). The few gametophores produced by *smg1*Δ lines predominantly grow downwards into the agar, rather than equally upwards out of the agar and downwards into the agar as in wild type plants (Figure 3.10E). Since the moss NMD mutants are *PpSMG1* complete knockout lines, this result shows that although *PpSMG1* is needed for normal moss growth it is not essential under standard laboratory conditions.

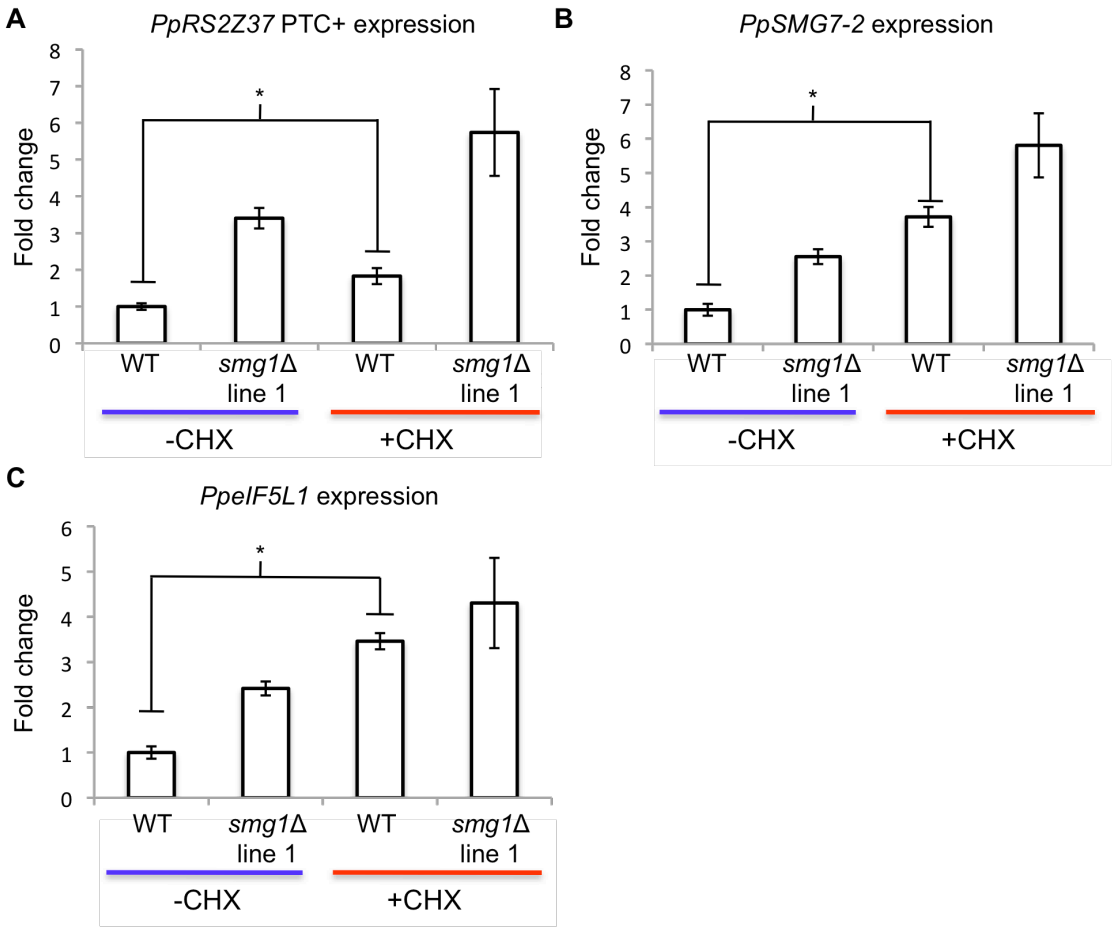

Figure 3.9: Targets of NMD are over-expressed when moss is exposed to CHX

(A) qRT-PCR analysis of *PpRS2Z37* PTC+ variant. **(B)** qRT-PCR analysis of *PpSMG7-2*. **(C)** qRT-PCR analysis of *PpeIF5L1*. **(A-C)** Fold change is the amount of target expression normalised to reference gene Pp1s54_156V6.1 and relative to WT DMSO levels. Error bars represent the standard error of the mean from three biological replicates. Asterisks represent lines with a statistical difference from WT DMSO to WT CHX using an unpaired *t* test $p < 0.05$.

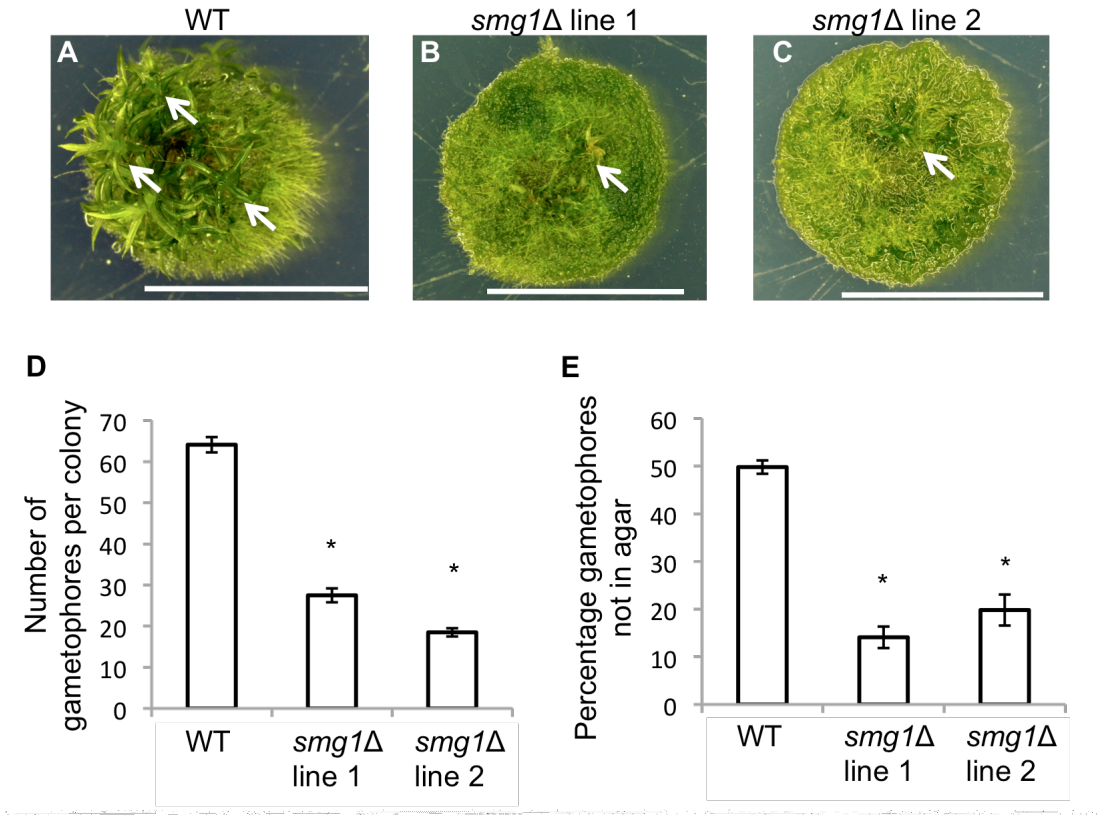

Figure 3.10: *smg1*Δ lines produce fewer leafy structures which mainly grow into the media

(A) Three-week-old wild-type colony (WT). **(B)** Three-week-old *smg1*Δ line 1. **(C)** Three-week-old *smg1*Δ line 2. White arrows point to example leafy structures (gametophores) growing from the filamentous tissue. **(D)** Total number of gametophores per colony after three weeks. **(E)** The percentage of gametophores not submerged in agar after three weeks. In **A-C** scale bars represent 1 cm. In **D** and **E** the n = 18. Error bars are SEM. Asterisks represent knockout line with a statistical difference from WT using an unpaired *t* test $p < 0.05$.

3.3 Discussion

3.3.1 An ancient role for SMG1 in the NMD pathway

The NMD pathway plays important roles, both in protecting organisms from the production of truncated proteins and in regulating gene expression (Mühlemann et al., 2008). In animals, the SMG1 kinase activates NMD by the phosphorylation of UPF1, following its recruitment to the terminating ribosome (Grimson et al., 2004; Yamashita et al., 2001; Kashima et al., 2006). Phosphorylation of UPF1 facilitates the subsequent binding of the SMG5-7 proteins (Okada-Katsuhata et al., 2011), allowing SMG6 to cut the transcript near the PTC (Huntzinger et al., 2008) and SMG5 and SMG7 to help recruit endonucleases (Unterholzner and Izaurralde, 2004; Jonas et al., 2013), leading to the destruction of the PTC-containing transcript. The work presented here has shown that the SMG1 kinase is not an animal-specific component of the NMD pathway and that SMG1 functions in NMD in a basal plant, a role that is likely to be conserved in all plants that retain *SMG1*. The involvement of SMG1 in the NMD pathways of both animals and plants is consistent with SMG1 functioning in NMD in the LECA (Figure 3.3). This suggests that the origins of SMG1 are very ancient; two to three billion years ago (Brocks et al., 1999).

SMG8 and SMG9 were reported to regulate the kinase activity of SMG1 and function in the NMD pathway of mammals (Yamashita et al., 2009), however, they may not have conserved roles in NMD (Rosains and Mango, 2012;

Yamashita et al., 2009). RNAi knockdown of *SMG9* did not lead to a reduction in NMD in *C. elegans* (Yamashita et al., 2009). RNAi knockdown of *SMG8* did reduce NMD slightly in *C. elegans* (Yamashita et al., 2009), however, another study mutated *SMG8* (deleted the first 65 amino acid residues) and found that these mutants had a normal level of NMD activity (Rosains and Mango, 2012). These data suggest SMG8 and SMG9 might not have an evolutionary conserved role in NMD. Interestingly, proteins with homology to SMG8 and SMG9 have been identified in moss (Table 3.1), rice and grape but not in *A. thaliana* (Yamashita, 2013). If they do play a role in plant NMD, they appear to only be conserved in species that also use SMG1 in the NMD pathway.

3.3.2 The evolutionary history of SMG1 suggests an alternative NMD-associated kinase

Despite the conserved role of SMG1 in the NMD pathway of plants and animals, *SMG1* has been independently lost in multiple eukaryotic lineages (Figure 3.3). In budding yeast, the lack of an NMD-associated kinase, through an early evolutionary loss of *SMG1* in the fungal lineage (Figure 3.1), led to the suggestion that yeast NMD does not rely on UPF1 phosphorylation (Figure 3.11D) (Gatfield et al., 2003). Nevertheless, budding yeast UPF1 is phosphorylated (Wang et al., 2006) and the loss of the yeast 14-3-3-like protein SMG7/EBS1, which is predicted to bind phosphorylated UPF1, results in a partially compromised NMD pathway (Luke et al., 2007). These data indicate that phosphorylation of UPF1 could be important for NMD in fungi and that, in the absence of an SMG1 kinase, an alternative kinase acts in NMD (Figure

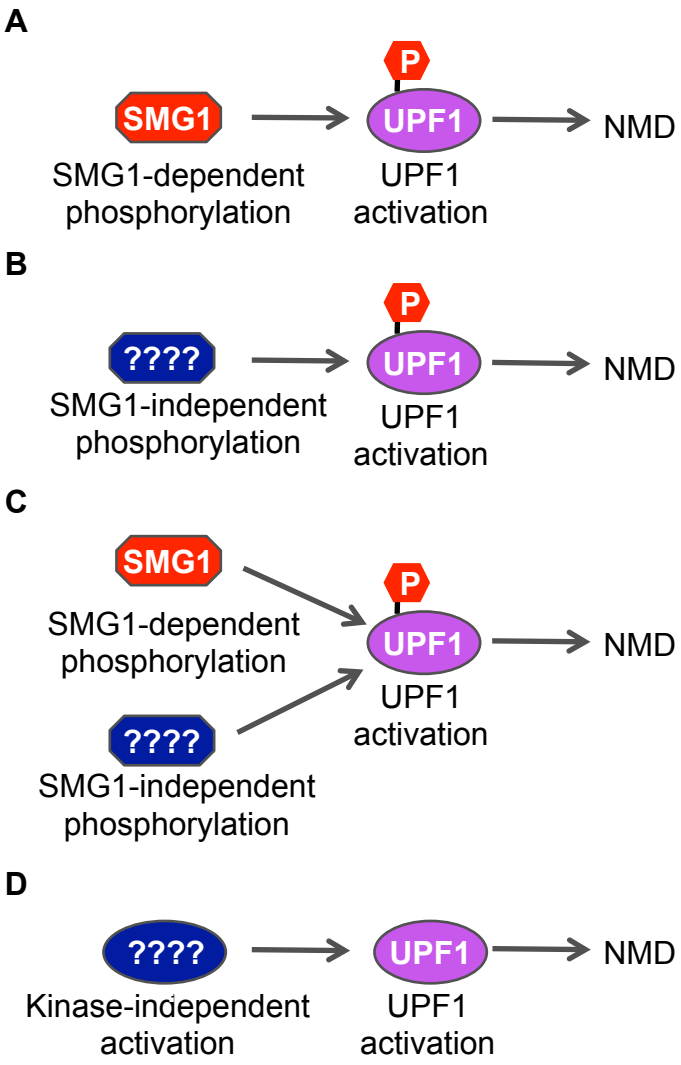

Figure 3.11: Proposed models for UPF1 activation in NMD

(A) The NMD pathway of organisms reliant on SMG1-dependant phosphorylation of UPF1 (e.g. *C. elegans*). **(B)** The NMD pathway of organisms that have replaced SMG1 with an undiscovered kinase, such as fungi and *A. thaliana*. **(C)** The proposed NMD pathway of organisms with two kinases. These may include the NMD pathways of *Drosophila*, zebrafish and the LECA. **(D)** The proposed NMD pathway of organisms that do not require UPF1 phosphorylation. These may include the NMD pathways of yeasts and excavates.

3.11B). A more recent loss of *SMG1* has occurred in *A thaliana* within the last 5-10 million years (Hu et al., 2011). Even though SMG1 is not present in the *A. thaliana* genome (Figure 3.1-3) (Grimson et al., 2004), the *A. thaliana* UPF1 is phosphorylated in tobacco cells (Mérai et al., 2012) and SMG7 is required for NMD in *A. thaliana* (Riehs et al., 2008). This suggests that phosphorylation of UPF1 remains necessary for NMD, again implicating an alternative kinase (Figure 3.11B).

Both SMG1 and SMG7 have been lost in the analysed members of the Excavate group, *Giardia lamblia* and *Trypanosoma brucei*, and in the red algae *Cyanidioschyzon merolae* (Figure 3.3). Despite the presence of UPF1 in these organisms, the absence of both its kinase (SMG1) and the protein that recognises its phosphorylated form (SMG7), could suggest that they no longer require the UPF1 phosphorylation/dephosphorylation cycle (Delhi et al., 2011; Chen et al., 2008) (Figure 3.11D). Alternatively, these eukaryotes could have independently acquired replacements for both SMG1 and SMG7 (Figure 3.11B). An investigation of the phosphorylation status of UPF1 in these species will help to discriminate between these possibilities. It is interesting to consider the diversity of NMD pathway arrangements that exist in different organisms (Figure 3.11). Evidence suggests that there are distinct ‘branches’ of the NMD pathway where different substrates have specific effector requirements (Chan et al., 2007; Gehring et al., 2005; Huang et al., 2011). For example, in mammals it has been reported there is a branch of the NMD pathway that does not require UPF3 (Chan et al., 2007). There is a possibility that currently unidentified NMD effectors could be involved in these branches of the NMD pathway in different

eukaryotes. PNRC2 has recently been identified as a vertebrate specific protein that interacts with SMG5 to link UPF1 phosphorylation status to the RNA decay machinery (Cho et al., 2012; Lai et al., 2012; Cho et al., 2009).

Multiple independent losses of *SMG1* in eukaryotes which have conserved SMG7 and phosphorylation of UPF1 (Luke et al., 2007; Wang et al., 2006; Mérai et al., 2012; Riehs et al., 2008), suggest that an alternative kinase is capable of activating NMD in many organisms (Figures 3.11B and C). One possibility is that at each loss event, SMG1 was replaced by an independent kinase (Figure 3.11B). A more appealing explanation is that redundancy between two or more kinases in a common ancestor allowed the independent losses of *SMG1* to be replaced by a pre-existing alternative kinase(s), which could be as ancient as SMG1. Loss of SMG1 in *Drosophila* and zebrafish has been shown to have little or no effect on NMD (Wittkopp et al., 2009; Metzstein and Krasnow, 2006), suggesting redundancy between SMG1 and another kinase(s) (Figure 3.11C). Redundancy would also explain the differential requirement for SMG1 for organismal survival. While UPF1 and UPF2 are required for *Drosophila* and zebrafish development, SMG1 is dispensable, possibly due to the presence of another NMD kinase (Figure 3.11C) (Wittkopp et al., 2009; Metzstein and Krasnow, 2006). Similarly, the plant NMD pathway could include an alternative kinase, providing an explanation for the relatively mild phenotype observed in moss *smg1*Δ lines compared to NMD-compromised phenotypes in *A. thaliana* and other non-plant species. Analysing the phenotypes of moss mutants that affect core NMD proteins would address this question.

If an alternative NMD-associated kinase is active, it will be interesting to look at the role of other PIKK family members in NMD. SMG1 is related to the *A. thaliana* DNA damage activated kinases ATM and ATR (Waterworth et al., 2011) and TOR, a regulator of translation (Deprost et al., 2007). The *A. thaliana* genomes also encodes a TRRAP protein, which is a kinase inactive member of the PIKK family (Templeton and Moorhead, 2005) and *A. thaliana* lacks DNA-PKcs (Figure 3.1). While a link to NMD has not been established, ATM and ATR can phosphorylate UPF1 in mammals. ATM phosphorylates UPF1 after DNA damage, but knockdown of ATM had no effect on NMD (Brumbaugh et al., 2004) suggesting UPF1 is involved in DNA damage repair, independently of its role in NMD (Brumbaugh et al., 2004). ATR also phosphorylates UPF1, regulating genome stability (Azzalin and Lingner, 2006). The lack of an NMD phenotype in PIKK mutants, other than *smg1*, could indicate that these kinases are unable to substitute for SMG1, a view supported by the difference in the domain structure of SMG1 proteins in animals and plants compared to other PIKKs (Figure 3.2A). Alternatively, it is possible that these PIKKs could act redundantly in some species and that this redundancy could mask their NMD effects.

3.3.3 Differential requirements of NMD and/or SMG1 for growth and development

Organisms vary in their requirement for NMD and/or SMG1 for development. *Drosophila*, vertebrates and *A. thaliana* all require NMD for organismal survival

(Arciga-Reyes et al., 2006; Metzstein and Krasnow, 2006; McIlwain et al., 2010; Wittkopp et al., 2009). In the case of *A. thaliana*, it was identified that seedling lethality was due to over-production of the defence hormone SA in the most NMD-compromised mutant lines (Riehs-Kearnan et al., 2012). This is in contrast to *C. elegans* and budding yeast where loss of NMD is not lethal (Hodgkin et al., 1989; Culbertson et al., 1980). The difference between organisms is likely due to different targets of NMD in different species as well as different sensitivity to altered expression of conserved targets, for example, splicing factors. In *C. elegans* and mice, loss of *SMG1* abolishes the NMD pathway and leads to the same phenotype as in other NMD mutants (Hodgkin et al., 1989; McIlwain et al., 2010) but in *Drosophila* and zebrafish, loss of *SMG1* has a mild or no phenotype, both at the level of NMD activity and at the level of whole organismal development (Wittkopp et al., 2009; Metzstein and Krasnow, 2006). As previously discussed (Section 3.3.2), this could be due to a non-essential role of SMG1 in the NMD pathway of these organisms because of redundancy with another NMD-associated kinase or a non-essential role of phosphorylation in the NMD pathway of these animals. In moss, loss of *PpSMG1* leads to a mild phenotype where only a few gametophores grow into the air rather than into the agar on plates, suggesting that 1) NMD is not essential for moss development, 2) PpSMG1 is redundant in the phosphorylation of UPF1 in moss or 3) phosphorylation is not essential for NMD in moss. Examination of other NMD mutants in moss will be helpful in elucidating the correct hypothesis.

3.3.4 Conservation of NMD target features

Transcripts encoding splicing factors are common targets of AS-coupled NMD, producing feedback loops between splicing factors and their transcripts (Lewis et al., 2003; Palusa and Reddy, 2010; Stauffer et al., 2010). AS-coupled NMD was identified in the moss SR protein-encoding transcript *PpRS2Z37* and polypyrimidine tract binding protein-encoding transcript *PpPTB3*. *PpRS2Z37* has an alternative acceptor site that introduces a PTC from intronic sequence (Figure 3.6) while *PpPTB3* undergoes exon skipping, resulting in a frameshift that introduces a PTC, targeting this splice variant to decay by NMD (Figure 3.6A). Exon skipping also targets mammalian PTB transcripts (Sawicka et al., 2008; Wollerton et al., 2004) and an *A. thaliana* homologue (Stauffer et al., 2010) to NMD. AS-coupled NMD is therefore likely to be as widespread in moss as it is in *A. thaliana* (Stauffer et al., 2010; Palusa et al., 2007; Kalyna et al., 2012) and animals (Ramani et al., 2009; McIlwain et al., 2010; Weischenfeldt et al., 2012).

Another group of NMD targeted transcripts are those that encode NMD effectors. In animals, several NMD effector-encoding transcripts are targeted by NMD to create an autoregulatory feedback loop (Rehwinkel et al., 2005; Huang et al., 2011; Yepiskoposyan et al., 2011). In flowering plants, *SMG7* is a direct target of NMD due to its long 3' UTR and two introns located downstream of the stop codon, a feature conserved between monocots and eudicots (Benkovics et al., 2011; Kerényi et al., 2008; Nyikó et al., 2013; Rayson et al., 2012a). *PpSMG7-2* and *PpSMG7-3* were found to have an increased steady state level

in *smg1*Δ lines indicating that this autoregulatory loop is also conserved across land plants. Surprisingly, *PpSMG7-1* was only significantly over-expressed in *smg1*Δ line 2 (Figure 3.7B). *PpSMG7-1* was under-expressed in *SMG1*WT line 1 suggesting that either 1) NMD is over-active in *SMG1*WT line 1 but this is unlikely as no significant reduction in other NMD targets was observed or 2) this line has a slight reduction in *PpSMG7-1* independent of NMD due to chance or line specific changes. *SMG1*WT line 1 could contain random insertion sites from the transforming DNA or could contain stochastic epigenetic changes resulting from the transformation process. An alternative hypothesis is that *SMG1*WT line 1 is a better control than WT because it has gone through the transformation process and better replicates *smg1*Δ line 1 than WT. It is impossible to conclude anything definitive at this stage, although WT has proven a reliable control for all other tested targets. Interestingly, plant *SMG7* genes have a conserved 3' UTR structure (but not sequence). This includes a longer than average 3' UTR and two introns, both features that are important for the targeting to NMD (Nyikó et al., 2013). The stop codon proximal intron is very close to the stop codon, too close to induce NMD under the >50 nucleotide downstream of a stop codon rule, first established in mammals (Nyikó et al., 2013). However, this stop codon proximal intron is important for splicing of the downstream stop codon that causes targeting to NMD in an EJC-dependant manner (Nyikó et al., 2013). Upon inspection of the newly predicted 3' UTRs of the three moss *SMG7* homologues, it was noticed all three have a longer than average 3' UTR and contain two introns. For *PpSMG7-1* and *PpSMG7-2*, the intron structure matches that of flowering plants, with the stop codon proximal intron being very close to the stop codon (<50 nucleotide downstream of a stop

codon) and the second being further downstream, suggesting that the first intron might be important for the splicing of the NMD-associated downstream intron. However, the stop codon proximal intron of *PpSMG7-3* is >50 nucleotides downstream of the stop codon so might function in NMD through association with an EJC, in addition to being involved with splicing of the downstream intron. These data suggest that NMD targeting of *SMG7* to create an autoregulatory feedback loop through targeting of a long 3' UTR containing introns might be ancient and conserved across land plants.

Previous work has suggested that CPuORFs can target transcripts to NMD (Rayson et al., 2012a; 2012b; Nyikó et al., 2009). Nearly half of all CPuORF-containing transcripts are up-regulated in at least one *A. thaliana* NMD mutant (Rayson et al., 2012a; 2012b) and the CPuORFs of AT1G70780 and AT3G18000 have previously been shown to be sufficient to target a reporter to NMD (Nyikó et al., 2009). The CPuORF of AT1G36730 is conserved in two moss genes (Figure 3.8). The mORFs of the two moss genes and the *A. thaliana* gene all encode eIF5-related proteins. Expression of AT1G36730 is elevated in *A. thaliana* NMD mutants (Rayson et al., 2012a) and is a direct target of NMD (S. Rayson, personal communication). In the moss *smg1*Δ lines, the expression of both *eIF5L* transcripts is elevated. CPuORFs fall into around 30 homology groups, where there is homology between uORFs within a homology group but not between them (Hayden and Jorgensen, 2007; Jorgensen and Dorantes-Acosta, 2012). Work showing that nearly half of all CPuORFs target a transcript to NMD under laboratory conditions (Rayson et al., 2012a; 2012b) suggests that despite the independent sequences and origins of

CPuORFs, many have a similar role in targeting transcripts to NMD. Our work suggests that CPuORFs may act to target transcripts to NMD across the plant kingdom. Interestingly, the *eIF5*-encoding transcripts of many animals contain uORFs and many of these are CPuORFs, however, the conservation is low between animal phyla (Loughran et al., 2012). It is likely that the peptide sequence of the uORF in front of the *eIF5* mORF is important, probably for regulatory reasons, but that several sequences have arisen independently during eukaryotic evolution. In plants, it appears that the regulatory purpose is to link the transcript to NMD. In mammals, the uORF is only translated when start codon selection efficiency is low (for example, when eIF5 levels are high) and translation of the uORF reduces translation of the mORF (Loughran et al., 2012). It was also suggested that translation of the uORF in animals could also target the transcript to NMD to reinforce the feedback loop (Loughran et al., 2012). Given that the start codon of the CPuORF of *A. thaliana* and moss *eIF5L* genes is in a poor context (S. Rayson, personal communication) it is likely such a feedback loop is conserved in plants, which therefore use NMD to reduce the levels of eIF5L.

These findings demonstrate the conservation of NMD mechanisms and targets across a large evolutionary span. The destruction of these conserved targets by NMD is likely to have important physiological outcomes, for example in regulating the activity of the NMD pathway itself or in the stringency of start codon selection.

4 Chapter Four: Using the Yeast two-hybrid Assay to Better Understand the NMD Pathway of Plants

4.1 Introduction

4.1.1 The putative role of an unidentified alternative NMD-associated kinase in *A. thaliana*

To gain a better insight into the NMD pathway of plants, the SMG1 kinase was identified in moss and the entire coding region of the *PpSMG1* gene was knocked out (Chapter 3). Upon deletion of *PpSMG1*, targets of the NMD pathway had an increased expression level (Chapter 3). It was subsequently speculated that phosphorylation of UPF1 by SMG1 could be an ancient mechanism, which operated in the stem eukaryote (Chapter 3). Although it has lost *SMG1*, *A. thaliana* has a competent NMD pathway, which is important for gene expression and development (Arciga-Reyes et al., 2006; Riehs et al., 2008; Rayson et al., 2012a). The absence of the SMG1 kinase raises the possibility that NMD in *A. thaliana* does not rely on phosphorylation of UPF1 or that another kinase has replaced SMG1 in phosphorylating UPF1. Literature searches have not yielded any reports that *A. thaliana* UPF1 (AtUPF1) is phosphorylated in *A. thaliana*, however, evidence has been reported that suggests AtUPF1 is phosphorylated and this is important in NMD in tobacco. The AtUPF1 protein is predicted to encode multiple SQ and TQ dipeptides,

mainly in the N- and C-termini (Mérai et al., 2012). These dipeptides are the sites of phosphorylation in animals (Okada-Katsuhata et al., 2011) and are common targets of PIKKs, with the exception of TOR (Izumi et al., 2012). These sites are phosphorylated when AtUPF1 is expressed in tobacco cells and these sites are essential for AtUPF1 to have a role in tobacco NMD (Mérai et al., 2012). It is likely tobacco has an orthologue of *SMG1*, although a published genome sequence is needed to confirm this. This suggests that SMG1 is the kinase responsible for the phosphorylation of AtUPF1 in tobacco cells. In addition to this, AtSMG7 requires its phospho-binding site to function in the NMD pathway of tobacco (Mérai et al., 2012). Work has shown AtSMG7 functions in the NMD pathway of *A. thaliana* (Riehs et al., 2008) suggesting a role for AtUPF1 phosphorylation in bridging a PTC to the decay machinery. These data suggest that AtUPF1 needs to be phosphorylated to function in the NMD pathway of plants despite the lack of an NMD-associated kinase in the *A. thaliana* genome. It was suggested that this alternative NMD-associated kinase might also be conserved (Figure 3.11). If an alternative kinase existed before the loss of SMG1 in different lineages, it is easy to imagine that changes in the expression or interaction strength of this alternative kinase would be a smaller evolutionary step than recruiting a new kinase to the NMD pathway. It was predicted that this hypothetical alternative NMD-associated kinase has been conserved because *SMG1* from both fungi and *A. thaliana* has been lost and because SMG1 has a non-essential role in the NMD pathway of some other organisms. Mutation of *SMG1* in *Drosophila* does not abolish NMD in all cases and knockdown of *SMG1* in zebrafish did not affect development, as it did with other NMD effectors (Metzstein and Krasnow, 2006; Wittkopp et al., 2009). It

has also been suggested that SMG1 in mammals is partially redundant to other wortmannin-sensitive kinases (Brumbaugh et al., 2004).

Therefore to try to identify a kinase that interacts with AtUPF1 the yeast two-hybrid system was used. The yeast two-hybrid assay was an attractive method given that it is a simple system to identify proteins that interact directly with a protein of interest.

4.1.2 Introduction to the yeast two-hybrid assay

The yeast two-hybrid assay can be used to identify proteins that directly interact with a protein of interest by expressing them in budding yeast and coupling this interaction to the expression of a reporter gene (Figure 4.1). In the assay, the first protein of interest is fused to a DNA-binding domain (BD) and is called the bait and the second protein is fused to the activation domain (AD) to produce a prey (Figure 4.1). The BD and AD are the two modular components of the yeast GAL4 transcription factor. When the BD and the AD are brought together by the interaction of two proteins of interest then they are able to induce the expression of reporter genes with the GAL4 binding sequence in their promoters (Figure 4.1). The *HIS3* reporter gene is used in this study and encodes an enzyme required for histidine biosynthesis in yeast. When no interaction is present, yeast cannot grow on media lacking histidine.

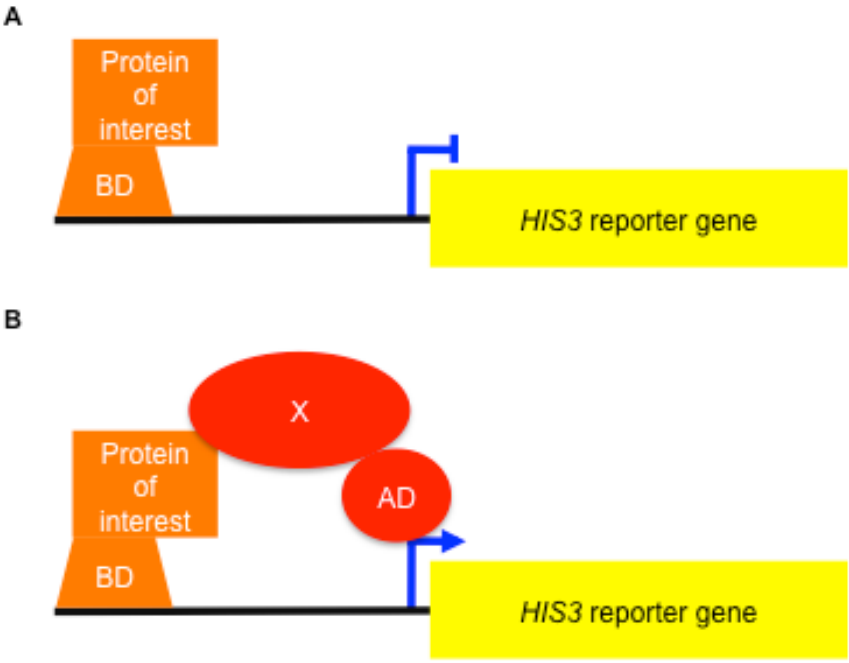

Figure 4.1: Model representing the yeast two-hybrid assay

(A) A fusion protein comprising of the GAL4 DNA-binding domain (BD) and the protein of interest. This should be unable to activate transcription of the *HIS3* reporter and the yeast should not grow on media lacking histidine. **(B)** Interaction between protein of interest and putative interactor (X), which is fused to the GAL4 activation domain (AD) recruits the AD to the promoter of the HIS3 reporter and induces transcription and allows for growth on media lacking histidine.

Proteins fused to the BD should not have independent transcription activation of the *HIS3* reporter, referred to as autoactivation. Therefore, test proteins should be screened for their autoactivation potential before a screen is attempted. Weak autoactivation can be overcome with the addition of 3-amino-1,2,4-triazole (3-AT) into the medium. 3-AT is a chemical inhibitor of the product of the *HIS3* reporter. This can be used to dampen autoactivation, allowing interactions to be detected.

The yeast two-hybrid system can be used to test pair-wise interaction between specific proteins or used to screen cDNA libraries cloned into AD-containing plasmids. This allows for the rapid identification of novel putative interacting proteins.

4.1.3 The aims of using the NMD effectors in the yeast two-hybrid assay

The aims of the experimental work of this chapter are:

1. To identify novel interactors of AtUPF1
2. To establish yeast two-hybrid assays, which can be used to identify a kinase phosphorylating AtUPF1
3. To explore the relationship between AtSMG7 and AtUPF1

The ease of use of the yeast two-hybrid assay means that it should be an excellent tool to detect an interacting kinase. Yeast two-hybrid has been used before to successfully identify kinase-substrate interaction (Tsai and Gazzarrini, 2012; Persak and Pitzschke, 2013). However, previous work has shown that

AtUPF1 is a powerful autoactivator (Kerényi et al., 2008). A library screen was performed using AtUPF1 as the bait, but since AtUPF1 is a powerful autoactivator, a very high concentration of 3-AT was used (100 mM) (Kerényi et al., 2008). This screen was able to pull out AtUPF2, a known interactor of AtUPF1 but no kinase was identified (Kerényi et al., 2008). It is possible that the high concentration of 3-AT meant that transient interactions, such as a kinase-substrate interaction, could not be identified. To allow a screen to be performed under conditions that are more favorable for the detection of a transient or weak interaction, AtUPF1 will be truncated to remove its autoactivation ability. Additionally, it has not yet been shown that AtSMG7 and AtUPF1 directly interact and whether this is dependent on AtUPF1 being phosphorylated, although mutational analysis of AtSMG7 showed that sites predicted to bind phosphorylated AtUPF1 are essential for NMD (Mérai et al., 2012).

4.2 Results

4.2.1 Autoactivation testing of truncated AtUPF1

To identify potential interacting partners of AtUPF1, truncations of AtUPF1 were designed for use in the yeast two-hybrid assay. As full-length AtUPF1 is a powerful autoactivator and a previous screen did not find an interacting kinase (Kerényi et al., 2008), two truncations of AtUPF1 were generated to try to eliminate the autoactivating region of AtUPF1 when it is used as a bait (Figure 4.2A). The first truncation, AtUPF1 NT, stretches from residue 2 to 808, which includes the UPF2 binding domain, RNA helicase domain and the putative phosphorylation sites at the N-terminal (Figure 4.2A). The second truncation, AtUPF1 CT, includes residues 618 to 1254, which includes half the RNA helicase domain and the putative phosphorylation sites at the C-terminal (Figure 4.2A). To test if these proteins autoactivate, as full-length AtUPF1 does, yeast was grown on medium lacking tryptophan (W) to select for the bait plasmid and histidine (H) to test for autoactivation. Yeast containing AtUPF1 NT grows strongly on SD–WH media suggesting that it can still autoactivate (Figure 4.2B). However, AtUPF1 CT does not grow well on SD–WH media, suggesting that it has little autoactivation ability. To dampen the residual autoactivation, a range of concentrations of 3-AT were tested. Yeast expressing AtUPF1 NT were still able to grow at concentration of 20 mM 3-AT (Figure 4.2B). No observable growth of yeast expressing AtUPF1 CT was seen at any concentration of 3-AT (Figure 4.2B). Therefore, AtUPF1 NT appears to

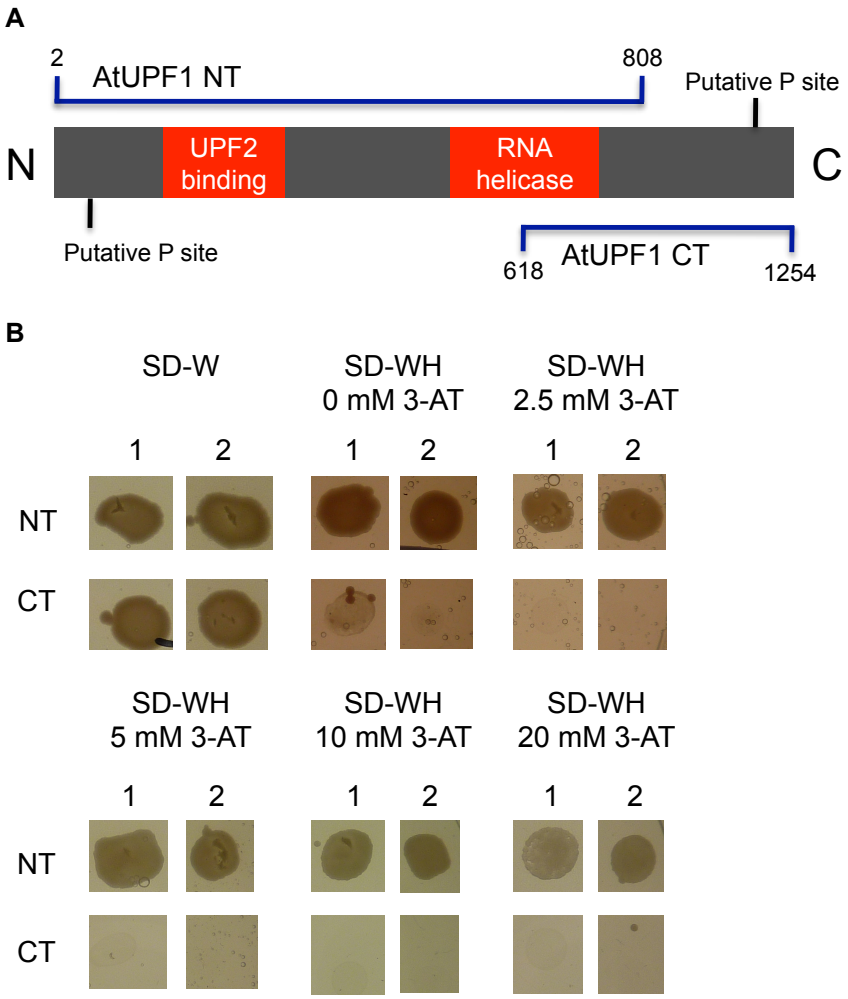

Figure 4.2: Autoactivation testing for BD vectors AtUPF1 NT and AtUPF1 CT

(A) Domain structure of AtUPF1 and fragments cloned for yeast two-hybrid, AtUPF1 (N-terminal) NT and AtUPF1 (C-terminal) CT. AtUPF1 has two conserved domains 1) UPF2 binding site and 2) RNA helicase domain. **(B)** Yeast was grown on SD-W media as control of growth and on SD-WH plates to test for autoactivation of the *HIS3* reporter. Growth of yeast on SD-WH plates shows that the bait protein can activate transcription alone. 3-AT was used to dampen activation. Two independent clones were tested for each construct. The yeast was photographed after five days of growth.

offer no advantage over the full-length protein as bait in the yeast two-hybrid assay due to high level of autoactivation, but AtUPF1 CT appears not to autoactivate and could be used in a library screen to identify novel interacting partners of AtUPF1.

4.2.2 The pAD AtUPF1 CT screen did not yield a kinase

To identify potential interacting proteins of the C-terminal of AtUPF1, a yeast two-hybrid screen was performed using the pBD AtUPF1 CT plasmid as bait. Two libraries were screened together, one made using random hexamer primed cDNA from *A. thaliana* and the other using oligo(dT) primed cDNA. After mating the yeast carrying AtUPF1 CT with the yeast containing the libraries, the diploids were plated out onto selection plates (SD-WLH) supplemented with 3-AT (2.5 mM). A total of 10.25 x 10^6 diploids were screened. Results from the screen are listed in Table 4.1 after filtering non protein-encoding sequences. No kinase was identified in the screen. The screen did not pull out any proteins with obvious links to the functions of UPF1, for example, NMD or genome stability (Table 4.1). The most frequently isolated protein from the screen was AT5G24620, which encodes a thaumatin domain containing protein. Clones corresponding to AT5G24620 accounted for 34% of the total clones sequenced that encoded protein-encoding genes from *A. thaliana*. Little is known about this gene in *A. thaliana* or the thaumatin domain in plants but it has a role in pathogen defense (Liu et al., 2010). Since none of the putative interactors identified in this screen corresponded to gene ontologies that have correctly

Table 4.1: A list of interactors of pAD AtUPF1CT.

Table 4.1 *continued on next page*

Locus	TAIR description	Number of clones
AT5G24620	Pathogenesis-related thaumatin superfamily protein	18
AT1G52510	alpha/beta-Hydrolases superfamily protein;	4
AT2G17900	Homology Subgroup S-ET - Protein containing an interrupted SET domain.	2
AT2G22360	DnaJ chaperone	2
AT3G04710	Encodes one of the 36 carboxylate clamp (CC)-tetratricopeptide repeat (TPR) proteins	2
AT5G08330	TCP11	2
AT1G04190	Encodes one of the 36 carboxylate clamp (CC)-tetratricopeptide repeat (TPR) proteins	1
AT1G07890	APX1, ASCORBATE PEROXIDASE 1, ATAPX01	1
AT1G13930	Involved in response to salt stress.	1
AT1G72160	Sec14p-like phosphatidylinositol transfer family protein	1
AT2G21270	ubiquitin fusion degradation 1 (UFD1)	1
AT2G42810	Encodes a phytochrome-specific type 5 serine/ threonine protein phosphatase	1
AT2G46240	A member of Arabidopsis BAG (Bcl-2-associated athanogene) proteins,	1
AT3G07090	PPPDE putative thiol peptidase family protein	1
AT3G12120	Major enzyme responsible for the synthesis of 18:2 fatty acids in the endoplasmic reticulum.	1
AT3G14210	Represses nitrile formation and favors isothiocyanate production during glucosinolate hydrolysis	1
AT3G63200	PATATIN-like protein 9 (PLP9)	1
AT3G63460	ransducin family protein / WD-40 repeat family	1
AT4G02290	glycosyl hydrolase 9B13	1
AT4G10760	mRNAadenosine methylase (MTA)	1
AT4G12400	Encodes one of the 36 carboxylate clamp (CC)-tetratricopeptide repeat (TPR) proteins	1
AT4G15500	Encodes a protein that might have sinapic acid:UDP-glucose glucosyltransferase activity.	1

Table 4.1 *continued*

Locus	TAIR description	Number of clones
AT4G33010	glycine decarboxylase P-protein 1 (GLDP1)	1
AT4G35090	Encodes a peroxisomal catalase	1
AT4G38770	Encodes one of four proline-rich proteins in Arabidopsis which are predicted to localize to the cell wal	1
AT5G38410	Ribulose bisphosphate carboxylase (small chain)	1
AT5G38430	Ribulose bisphosphate carboxylase (small chain) family protein	1
AT5G42050	DCD (Development and Cell Death) domain protein	1
AT5G43330	C-NAD-MDH2, CYTOSOLIC-NAD-DEPENDENT MALATE DEHYDROGENASE 2	1

been linked to NMD this screen was not pursued further.

4.2.3 Autoactivation testing of further AtUPF1 truncations

Truncations of AtUPF1 that consisted of about half of AtUPF1 from either the N- or C-terminus were unsuccessful in identifying a kinase that interacts with AtUPF1. The ‘tails’ at the N- and C-terminals of AtUPF1 contain no identified protein domains, but are enriched in S/TQ dipeptides, predicted to be the phosphorylation sites important in NMD (Mérai et al., 2012). In humans, two functional phosphorylation sites have been mapped, one at the N-terminal (T28) and one at the C-terminal (S1096) (Okada-Katsuhata et al., 2011) and an alignment of UPF1 protein sequences shows that these sites have homologues sites in *A. thaliana* (T29 and S1193, respectively) (Okada-Katsuhata et al., 2011) (Figure 4.3A). Therefore, new truncations of AtUPF1 were generated consisting of these ‘tails’ (Figure 4.3A). AtUPF1 NT2 includes resides 2 to 137 and contains putative phospho-site T29 (Figure 4.3A). AtUPF1 CT2 includes residues 1082 to 1254 and contains putative phospho-site S1193. Yeast containing AtUPF1 NT2 or AtUPF1 CT2 show strong growth on SD–WH media (Figure 4.3B). Increasing the concentration of 3-AT reduces but does not eliminate growth of AtUPF1 NT2, even at 40 mM 3-AT (Figure 4.3A). Surprisingly, AtUPF1 CT2 is a strong autoactivator, despite being a further truncated form of AtUPF1 CT, which did not autoactivate. Growth was reduced, but not eliminated, at the higher concentrations of 3-AT tested (20 and 40 mM). Taken together, these data suggest using the ‘tails’ of AtUPF1 cannot be helpful in the search for interacting partners in the yeast two-hybrid assay.

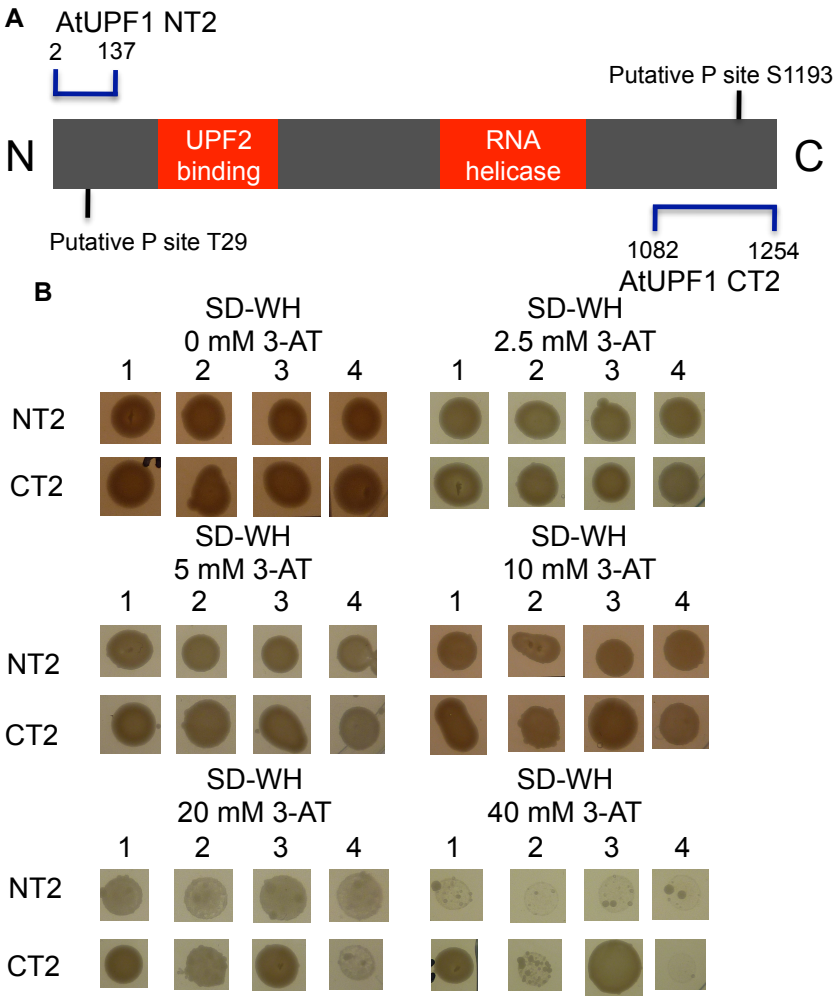

Figure 4.3: Autoactivation testing for BD vectors AtUPF1 NT2 and AtUPF1 CT2.

(A) Domain structure of AtUPF1 and fragments cloned for yeast two-hybrid, AtUPF1 (N-terminal) NT2 and AtUPF1 (C-terminal) CT2. AtUPF1 has two conserved domains 1) UPF2 binding site and 2) RNA helicase domain. **(B)** Yeast was grown on SD-WH plates to test for autoactivation of the *HIS3* reporter. Growth of yeast on SD-WH plates shows the bait protein can activate transcription of the reporter gene without an interacting protein with an AD. 3-AT was used to dampen activation. Four independent clones were tested for each construct. The yeast was photographed after five days of growth.

4.2.4 Phosphomimetic AtUPF1 interacts with AtSMG7 NT

In animals, it is essential that UPF1 is phosphorylated and recruits members of the SMG5-7 family for NMD (Okada-Katsuhata et al., 2011), but this has not been demonstrated in plants. It is, however, known that mutations of AtSMG7 at sites predicted to be involved in the recognition of phosphorylated AtUPF1, inhibit its role in NMD in tobacco cells (Mérai et al., 2012). We do not know if phosphorylated AtUPF1 is directly bound by AtSMG7, and if so, which residues are important for this interaction. A yeast two-hybrid assay where AtSMG7 is the bait and full-length AtUPF1 is the prey was designed. Whether AtUPF1 and AtSMG7 directly interact and whether this interaction is dependent on the phosphorylation status of particular residues of AtUPF1 was tested using this system. To mimic the phosphorylation of AtUPF1, site-directed mutagenesis was used to change residues predicted to be important for NMD in plants and changed these phosphorylatable residues (T or S) to glutamic acid (E), which is known to act a phosphomimetic (Wagner et al., 2004), and would therefore mimic the appearance and properties of a phosphorylated T or S residue.

First, a clone of AtSMG7 was needed and tested for autoactivation to see if it can be used as the bait in yeast two-hybrid. Attempts to clone full-length AtSMG7 into the entry vector pDONR207 failed so an N-terminal truncated AtSMG7 including residues 1 to 591 was constructed (AtSMG7 NT; Figure 4.4A). This region also includes the EST1 domain and the EST1 DNA/RNA binding domain, both predicted to function in telomere regulation (Figure 4.4A).

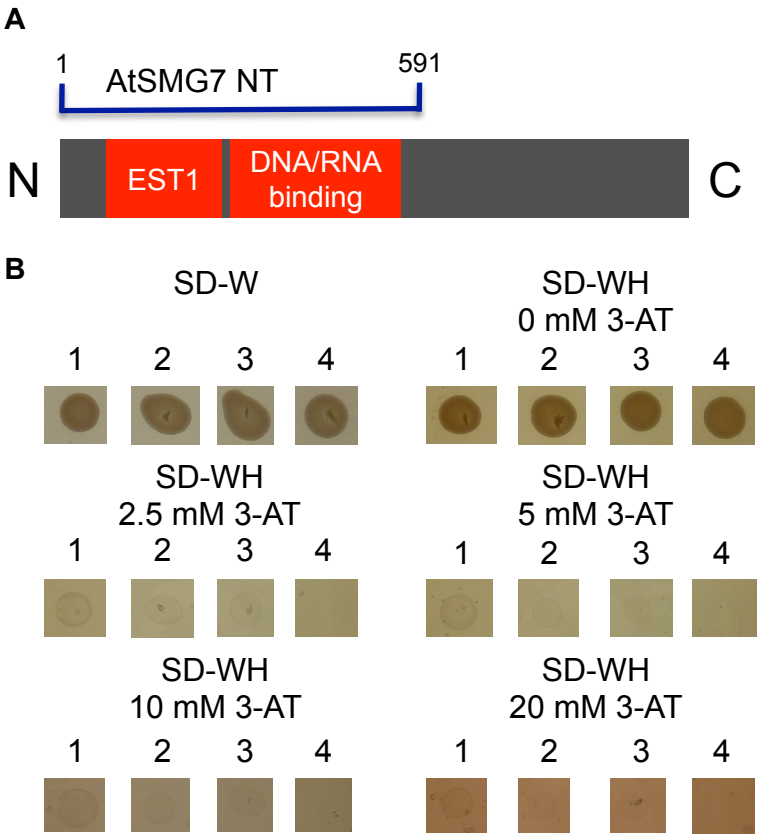

Figure 4.4: Autoactivation test of pBD SMG7 NT fragment

(A) The domain structure and fragments cloned for yeast two-hybrid, cloned fragment AtSMG7 (N-terminal) NT. AtSMG7 has two conserved domains at the N-terminal 1) EST1 and 2) EST1 DNA/RNA binding domain. The N-terminal region contains the 14-3-3-like domain (Mérai et al., 2012). **(B)** Yeast was grown on SD-WH plates to test for autoactivation of the *HIS3* reporter. Growth of yeast on SD-WH plates shows the bait protein can activate transcription of the reporter gene without an interacting protein with an AD. 3-AT was used to dampen activation. Four independent clones were tested for each construct. The yeast was photographed after three days of growth.

This region encodes the 14-3-3-like domain predicted to bind to phospho-UPF1 (Mérai et al., 2012). AtSMG7 NT grew on SD-WH plates without 3-AT but addition of 2.5 mM 3-AT inhibited growth and removed autoactivation. Therefore, AtSMG7 NT could be used as bait at the low concentration of 2.5 mM 3-AT.

To test whether a phosphomimetic residue at sites of AtUPF1 predicted to be important for NMD can stablise the interaction between AtUPF1 and AtSMG7, yeast containing the AtSMG7 NT and AtUPF1 WT (no site-directed mutation), AtUPF1 PM-CT (phosphomimetic S1193Q at the C-terminal) or AtUPF1 PM-NT (phosphomimetic T29Q at the N-terminal) was mated. The diploids were grown on medium selective for both plasmids (SD-WL) or interaction-selective media (SD-WLH 2.5 mM 3-AT) (Figure 4.5). Very little growth is observed for AtSMG7 NT and AtUPF1 WT (Figure 4.5), suggesting a very weak or no interaction is taking place. However, strong growth of yeast on interaction selective plates is observed when S1193 is mutated to a phosphomimetic (Figure 4.5). In contrast to this, no growth of yeast on interaction selective plates was observed when T29 is mutated to a phosphomimetic (Figure 4.5). These data suggest that AtSMG7 binds to phosphorylated AtUPF1 S1193 but not to phosphorylated T29, or that this site is not accurately represented by the phosphomimetic in a yeast system.

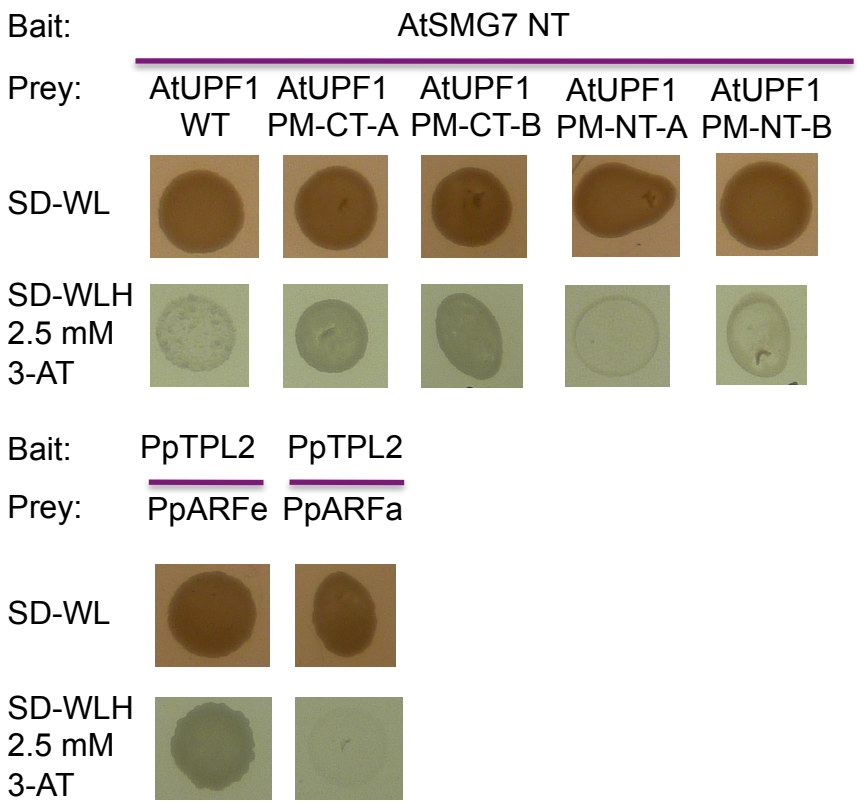

Figure 4.5: Phosphomimic AtUPF1 interacts with AtSMG7 NT

Bait (BD) and prey (AD) fused to proteins of interest are shown above the colony which is grown on SD-WL as a control of yeast growth or SD-WLH 2.5 mM 3-AT to measure interaction. PM-NT is phosphomimic AtUPF1 at T29 at the N-terminal (independent replicates A and B). PM-CT is phosphomimic AtUPF1 at S1193 at the C-terminal (independent replicates A and B). Moss TOPLESS2 (PpTPL2) and moss ARFe was used as positive control and PpTPL2 and moss ARFa was used as a negative control, for growth on selection plates. The yeast was photographed after three days of growth.

4.3 Discussion

The aim of this chapter was to discover proteins that directly interact with AtUPF1, with particular focus on identifying a kinase that could have replaced SMG1 in *A. thaliana* NMD, using the yeast two-hybrid assay. As AtUPF1 is a powerful autoactivator, several truncations were generated to try and remove the autoactivation ability of AtUPF1, however, only one of these did not autoactivate (AtUPF1 CT). AtUPF1 CT was used to screen *A. thaliana* yeast two-hybrid libraries, however, it did not identify a kinase as a putative interactor. Subsequently, we (me, Dr Barry Causier and Prof Brendan Davies) started to develop a novel yeast-three hybrid assay to identify proteins that can bridge the interaction between AtSMG7 and AtUPF1 such as a kinase.

4.3.1 Identification of putative novel interactors of AtUPF1 CT

After performing a library screen with AtUPF1 CT as bait, several putative interactors of the C-terminal of AtUPF1 have been reported (Table 4.1) but, without further validation of their interaction, it is unclear if any of these yeast two-hybrid interactors are genuine *in planta* interactors. It is possible that the truncation in the middle of the RNA helicase domain might have caused misfolding of AtUPF1, resulting in few meaningful interactors being identified. This is supported by the fact that two clones of putative interactors encode a chaperone protein DnaJ (AT2G22360). However, some of the identified interactors might have a role in NMD, discussed below. None of the putative

interactors identified in the previously published yeast two-hybrid screen using full-length AtUPF1, using a very high concentration (100 mM) of 3-AT (Kerényi et al., 2008), overlap with our list of interactors (Table 4.1).

Of the 53 yeast clones identified in the screen, eighteen contained a gene that best matched AT5G24620. AT5G24620 encodes a thaumatin domain containing protein. The thaumatin domain is found in many pathogen-related proteins, although the exact nature of the protein is not known. Thaumatin-like proteins have been shown to have anti-fungal properties but also have a range of ligands, which are endogenous plant components (Liu et al., 2010). It is possible that AT5G24620 encodes a genuine interactor of AtUPF1 but given the lack of evidence linking the thaumatin domain to RNA, translation or nuclear functions, investigating AT5G24620 further is not prioritised. AT1G52510 was represented by two clones and encodes an alpha/beta-hydrolase. This is unlikely to be a genuine interactor of AtUPF1 as the protein encoded by AT1G52510 is predicted to be located in the chloroplast and AtUPF1 has a cytosolic role in RNA decay and possibly a nuclear role in genome stability. SDG37 (AT2G17900) was also identified in two clones and is likely to have a role in histone methylation. This might be involved with nuclear functions of AtUPF1 but is unlikely to have a role in the cytosolic NMD pathway. Two clones encoding TPR10 (AT3G04710) and one clone encoding each of TPR3 (AT1G04190), HOP3 (AT4G12400) and PP5 (AT2G42810) were isolated in the screen and all encode tetratricopeptide repeat (TPR) domain-containing proteins. A bioinformatic analysis of *A. thaliana* TPR proteins identified all of these as putative interactors of heat-shock proteins (HSP) 70 and/or 90 (Prasad

et al., 2010). PP5 encodes a protein phosphatase and has been shown to dephosphorylate phytochrome A (Ryu et al., 2005). Plant homologues of PP5 and HOP have been shown to interact with HSP90 (van Bentem et al., 2005; Zhang et al., 2003). PP5 has also been shown to function as a chaperone independently of HSP90 in *A. thaliana* (Park et al., 2011). This is not the case for the PP5 from yeast (Park et al., 2011) and might not be representative of other TPR proteins. However, it remains possible that TPR proteins bind to AtUPF1 CT due to misfolding, likely due to its truncation. The TPR protein from animals P58(IPK) was shown to bind to unfolded proteins through the TPR domain and to recruit the BiP chaperone to re-fold the protein (Tao and Sha, 2011). This suggests a mechanism linking unfolded proteins, potentially truncated AtUPF1, to TPR proteins in yeast two-hybrid. On the other hand, these TPR proteins might have a functional role in stabilising the NMD complex, perhaps by recruiting chaperones. mRNA adenosine methylase (MTA; AT4G10760) was identified by a single clone (Table 4.1) and might be of interest as an interactor of AtUPF1. MTA covalently marks adenosine residues in mRNA, but the function of this mark is unclear (Bodi et al., 2012). This mark is enriched at the 3' end of a transcript and it has been suggested it might play a role in NMD (Bodi et al., 2012). It is unclear why UPF1 and MTA might interact, even if both function in NMD, as MTA is the enzyme leaving the methylation mark in the nucleus and UPF1 has a role in recognising stop codons in an unusual context in the cytosol. Therefore, if UPF1 has a role in directing the methylation of adenosine residues, it is unclear how this could work from a mechanistic point of view, given the cytosolic recognition of a PTC and a nuclear role for MTA. However, it is known that UPF1 has a role in

altering the release of a PTC+ transcript from the site of transcription by an unknown mechanism in mammals (De Turris et al., 2011), showing that UPF1 influences PTC-containing transcripts in the nucleus despite the evidence against NMD functioning in the nucleus (Trcek et al., 2013; Singh et al., 2008).

More work will be needed to confirm these potential interactors. The next step would be to clone the potential interactors into a bait vector and test if they interact with full-length AtUPF1 in a prey vector, which will overcome the problem of autoactivation from AtUPF1. If the potential interactors pass this stage of selection, then *in planta* interaction could be test for, for example, using the split YFP assay, also known as bimolecular fluorescence complementation. Following such confirmation of a physical interaction, it would be interesting to identify a role for the interactor in NMD through genetic approaches.

4.3.2 The role of phosphorylation of AtUPF1 in recruiting AtSMG7

It is currently unclear if AtUPF1 and AtSMG7 interact directly in plants and if this is dependent on phosphorylation. A yeast two-hybrid assay was developed to test if AtUPF1 and AtSMG7 NT interacted without phosphorylation of AtUPF1, which it does not (Figure 4.5), and then to determine which phosphorylated residues were important for this interaction by site-directed mutagenesis of S and T residues to a phosphomimetic amino acid (E).

It has been proposed that AtSMG7 binds to the phosphorylated N- and C-terminals of AtUPF1 in NMD (Mérai et al., 2012). In humans, the functionally

important phospho-sites have been mapped and sites homologous to sites in humans were identified in AtUPF1 (Okada-Katsuhata et al., 2011). These homologous sites (Figure 4.3A) were mutated to the phosphomimetic glutamic acid (E) in the hope that this would accurately mimic a phosphorylated S or T residue, as has been shown previously (Wagner et al., 2004). Indeed it was found that the S1193E mutation created a stronger interaction between AtSMG7 NT and AtUPF1 than between AtSMG7 NT and wild-type AtUPF1 (Figure 4.5). Surprisingly, the T29E mutation did not stabilised the interaction between AtSMG7 NT and AtUPF1 (Figure 4.5). This could be due to the E residue failing to accurately mimic a phosphorylated T in this situation. On the other hand, it might be that AtSMG7 does not bind phosphorylated T29. Interestingly, in humans, SMG7 only binds to the site at the N-terminus (S1096) rather than both sites and SMG6 is responsible for binding to the functional site at the N-terminus (T28) so it might not be surprising that SMG7 does not bind both of these sites. This model would suggest another unidentified factor binds to phosphorylated T29 in *A. thaliana* NMD (Figure 4.6) rather than the model where SMG7 binds to phosphorylated residues at both ends of the protein (Mérai et al., 2012). It is not unprecedented that a novel protein outside the SMG5-7 family is involved in recruiting the decay machinery to phosphorylated UPF1. PNRC2 is a vertebrate specific protein that, as a heterodimer with SMG5, functions in binding to phosphorylating UPF1 and recruiting the decapping complex (Cho et al., 2012; Lai et al., 2012; Cho et al., 2009). However, there is a high level of redundancy seen between phosphorylated

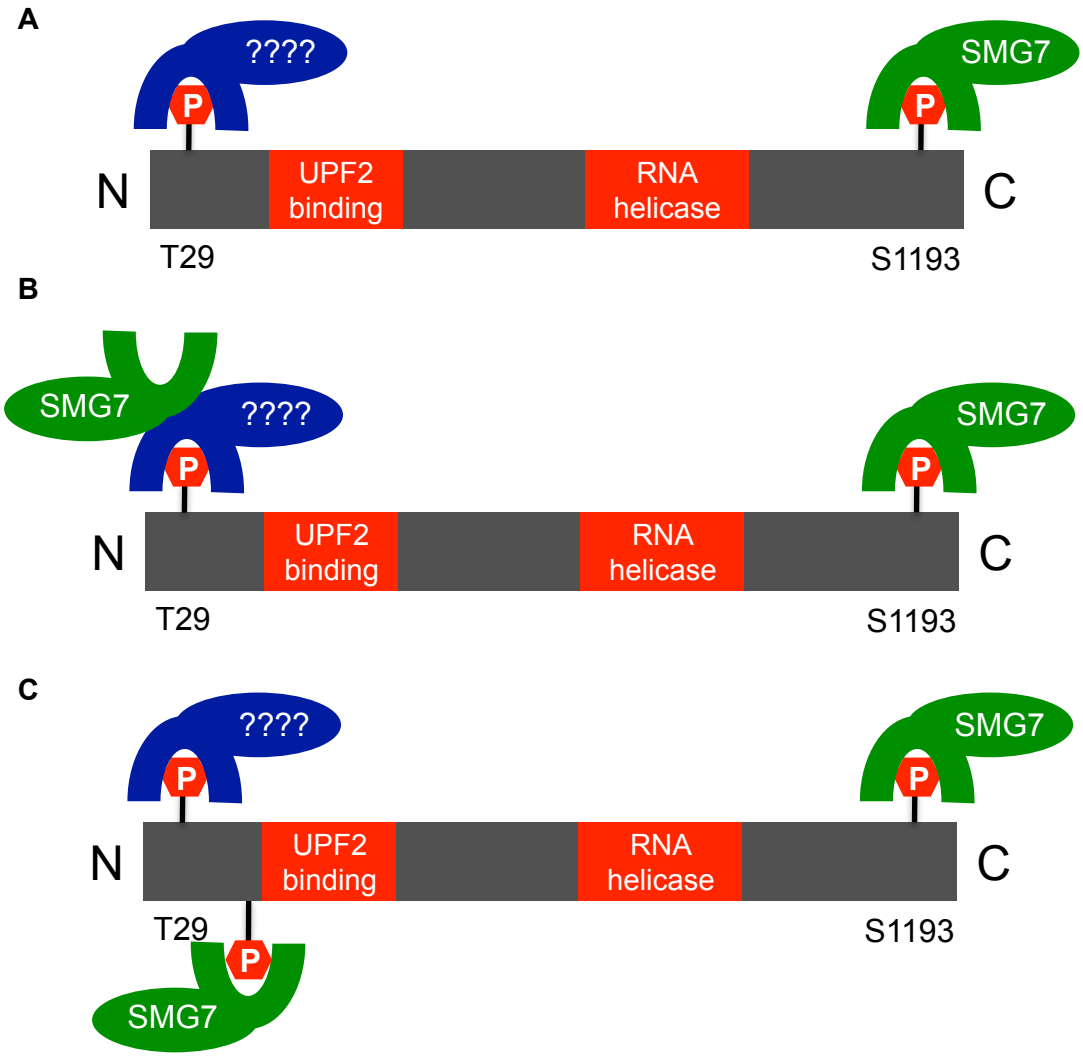

Figure 4.6: Model of AtUPF1 phosphorylation and protein recruitment

(A) Phosphorylated AtUPF1 recruits AtSMG7 to the C-terminal S1193 but not to T29 (phosphomimetic work, Figure 4.5) suggesting an alternative, unknown protein is recruited. **(B)** To explain the importance of AtSMG7 in NMD, AtSMG7 might dimerise with the unknown protein, which binds to phosphorylated T29. **(C)** To explain the importance of AtSMG7 in NMD, AtSMG7 might bind to another phosphorylated site at the N-terminus.

sites in AtUPF1, not seen in the human UPF1 where both the N- and C-terminuses have to be deleted to inhibit NMD (Mérai et al., 2012). While AtSMG7 might not bind phosphorylated T29, it might dimerise with the unidentified protein to aid function as SMG5 does with SMG7 and PNRC2 in humans or SMG7 might bind to another site in the N-terminus other than T29 (Figure 4.6B and C). It has been proposed that the heterodimer of SMG7:SMG5 might bind UPF1 through both SMG7 and SMG5 and the heterodimer helps stablise this interaction (Jonas et al., 2013), which is a fusion of the latter two models (Figure 4.6B and C).

4.3.3 The interaction of AtSMG7 NT with phosphomimetic AtUPF1 reveals a novel yeast-three hybrid system that could identify a kinase

Our findings that AtSMG7 NT interacts with phosphomimetic AtUPF1 reveals not only that it is likely that AtSMG7 needs AtUPF1 to be phosphorylated at specific residues but that this system could be used in a novel yeast-three hybrid assay to identify a kinase phosphorylating AtUPF1. Yeast-three hybrid works on the same principle as yeast two-hybrid, with the GAL4 BD and AD split between two proteins; however, these two proteins do not interact directly and need a third protein to bridge their interaction. This third protein is expressed from an expression vector, such as pTFT, and is not a fusion protein. In our proposed assay, the bridging protein would not actually take part in the complex *per se* but would phosphorylate AtUPF1 thus allowing AtSMG7 NT to interact with it without the need for a phosphomimetic mutation. This assay would require the generation of a pTFT library using *A. thaliana* cDNAs.

5 Chapter Five: RNA-seq Analysis Reveals the Importance of SMG1 in Moss Stress Responses

5.1 Introduction

5.1.1 The targets of NMD across eukaryotes

The NMD pathway has an important role in suppressing gene expression in many eukaryotes. As previously stated, NMD influences 1-10% of the transcriptomes of all examined eukaryotes, as revealed through analyzing NMD mutants and knockdowns (Guan et al., 2006; He et al., 2003; Mendell et al., 2004; Ramani et al., 2009; Rayson et al., 2012a; Rehwinkel et al., 2005). These changes in gene expression will be a mix of direct targets of NMD having an increased steady state level in mutants and knockdowns and up- and down-regulation of indirect targets of NMD. While NMD influences many physiologically important transcripts in all examined organisms, the overlap of targets is small (Rayson et al., 2012a), suggesting that NMD has been co-opted into regulating multiple pathways at separate times. One consistent role for NMD is in AS-coupled NMD. AS can generate splice variants with a PTC, for example by introduction of a cassette exon known as a ‘poison’ cassette exon. This type of transcript is then targeted to NMD in budding yeast, invertebrates, mammals and flowering plants (Sayani et al., 2008; Barberan-Soler et al., 2009; Palusa and Reddy, 2010; Lareau et al., 2007; Kalyna et al., 2012). In animals,

any sort of AS event appears to be able to introduce a PTC that is recognised by NMD, however, in *A. thaliana*, multiple transcripts with intron retention (IR) events were shown to evade NMD, making *A. thaliana* unusual among eukaryotes in this respect (Kalyna et al., 2012). It will be interesting to find out if this is the case in other plants or if it is an *A. thaliana* specific feature. AS-coupled NMD is particularly important for regulating the expression of splicing factors. Members of the GRP, PTB and SR families of splicing factors have been shown to regulate the splicing of their own primary transcripts; so that when the protein level is increased splicing of the PTC-containing variant is increased and this variant is subsequently degraded by NMD (Palusa and Reddy, 2010; Stauffer et al., 2010; Schoning et al., 2008). Over-expression of PTB in animals causes exon skipping which introduces a PTC due to a frame shift and becomes a target of NMD (Wollerton et al., 2004; Boutz et al., 2007). In *A. thaliana*, auto- and cross-regulation of *PTB* encoding transcripts has also been observed with exon skipping in *AtPTB3* leading to NMD (Stauffer et al., 2010). Exon skipping leading to NMD has also been observed in the moss homologue *PpPTB3* (Chapter 3). Twelve of the eighteen SR protein-encoding genes are alternatively spliced to produce targets of NMD in *A. thaliana*, thus suggesting that it is a major mechanism to control correct gene expression of splicing regulators in flowering plants (Palusa and Reddy, 2010) and AS-coupled NMD of one SR protein-encoding gene in moss has been identified (*PpRS2Z37*; Chapter 3).

While studies of the transcriptomes of NMD mutants and knockdowns of diverse organisms have not revealed many conserved targets, they have identified

altered pathways with important physiological relevance to the organism of study (Mendell et al., 2004; Rehwinkel et al., 2005; Rayson et al., 2012a). In humans, transcriptomic analysis revealed that many genes involved with amino acid starvation and the unfolded protein response were up-regulated upon loss of UPF1 (Mendell et al., 2004; Gardner, 2008; Wang et al., 2011). Follow-up work revealed that NMD represses the expression of genes important for recovery from hypoxia and accumulation of unfolded proteins through recognition of their uORF as a NMD targeting feature (Gardner, 2008; Wang et al., 2011). Hypoxia or induction of the unfolded protein response (UPR) triggers phosphorylation of eIF2-α in animals, which inhibits NMD and leads to increased expression of genes involved with protecting the cell from these stress conditions (Gardner, 2008; Wang et al., 2011). For example, in mammals the transcription factor ATF4 activates the UPR of the endoplasmic reticulum (ER) and is repressed by NMD through its uORF. Additionally, NMD represses the UPR of *C. elegans*. *HSP4* has an increased steady state level in NMD mutants of *C. elegans* (Sakaki et al., 2012).

In *A. thaliana*, many of the phenotypes associated with mutations in NMD effector-encoding genes are due to over-production of the plant defense hormone salicylic acid (SA) (Rayson et al., 2012a; Riehs-Kearnan et al., 2012; Jeong et al., 2011; Rayson et al., 2012b). The exact molecular link between NMD and the pathogen response has not been described. It is possible that NMD directly targets transcripts that activate SA production. It has been suggested that the activity of the NMD pathway is reduced during pathogen attack through a down-regulation of NMD effectors (Jeong et al., 2011). Recent

work has shown that eIF2-α is phosphorylated during pathogen attack (Pajerowska-Mukhtar et al., 2012). Given the NMD inhibitory effect of eIF2-α phosphorylation in animals during the UPR and hypoxia (Gardner, 2008; Wang et al., 2011), it is easy to speculate that eIF2-α phosphorylation inhibits NMD upon pathogen assault, allowing for direct targets of NMD that regulate SA biosynthesis to be up-regulated.

5.1.2 The aims of studying the transcriptomic differences between WT and *smg1*Δ lines

Now that *SMG1* has been identified and knocked out in moss, this gives us not only a model system in which to study the function of SMG1 in plant NMD (Chapter 3) but also an opportunity to study the biological function of NMD and/or PpSMG1 in moss growth.

The specific aims of the experimental work of this chapter are:

1. To identify differentially expressed genes and splice variants between WT and *smg1*Δ lines
2. To identify dysregulated pathways upon loss of *PpSMG1* to gain an insight into its biological role in moss
3. To confirm that the dysregulated pathways identified have a physiological role in moss.

5.2 Results

5.2.1 Transcriptomic analysis of moss compromised in NMD

To investigate changes in the transcriptome of moss upon loss of the NMD effector *PpSMG1*, short-read high-throughput RNA-sequencing (RNA-seq) on the Illumina HiSeq2000 platform set to 100 bp read length (single-end) was performed. Libraries were generated from pools of poly(A)-purified RNA. A total of nine libraries were sequenced across three channels of the HiSeq2000 (performed by GATC Biotech): four WT libraries, four *smg1*Δ line 1 libraries and three *smg1*Δ line 2 libraries. Each biological replicate (BR) was prepared by pooling RNA from two moss plants together. To account for the variation between individual sequencing reactions, some preparations were sequenced twice, which are the technical replicates (TR). Each sequenced library is named accordingly, for example, WT BR1.TR2 for the first biological replicate (BR) from WT but from the second sequencing reaction (TR). The total sequenced and subsequently mapped reads for each library are listed in Table 5.1. Mapping was performed by D. Lang and A. Zimmer of the University of Freiburg, Germany. For every library, except *smg1*Δ line 2 BR1.TR1, over 90% of the reads mapped to the moss genome, indicating a low level of reads resulting from contamination, reads mapping to un-sequenced parts of the moss genome and reads with sequencing errors. Of these reads, over 99% of all mapped reads for each library were mapped to the nuclear chromosomes

Table 5.1: Read statistics of RNA-seq data

Number of sequences represents the number of individual reads generated by the sequencing of each library. Sequences after clipping shows how many reads survived quality control after having adapter sequences removed. Mapped reads is the number of reads after clipping that were mapped to a location in the moss genome.

Library	Number of sequences	Sequences after clipping	% after clipping	Mapped reads	% reads mapped
WT BR2.TR2	31,500,663	31,141,725	98.86%	28,385,622	91.15%
WT BR2.TR1	53,969,218	53,405,472	98.96%	48,595,835	90.99%
WT BR1.TR2	29,880,653	29,651,357	99.23%	27,432,512	92.52%
WT BR1.TR1	39,860,598	39,600,315	99.35%	36,583,250	92.38%
smg1Δ line 2 BR2.TR2	36,609,523	36,334,460	99.25%	33,171,126	91.29%
smg1Δ line 2 BR2.TR1	41,789,739	41,495,868	99.30%	37,825,695	91.16%
smg1Δ line 2 BR1.TR1	75,157,724	74,654,346	99.33%	47,480,269	63.60%
smg1Δ line 1 BR2.TR2	32,559,328	32,268,264	99.11%	29,286,648	90.76%
smg1Δ line 1 BR2.TR1	46,679,512	46,294,924	99.18%	41,959,573	90.64%
smg1Δ line 1 BR1.TR2	35,174,129	34,900,457	99.22%	31,987,893	91.65%
smg1Δ line 1 BR1.TR1	45,812,224	45,466,354	99.25%	41,665,162	91.64%

rather than the plastid or mitochondrial genomes, as expected from poly-A enriched-coupled sequencing. Once the reads were mapped to the moss genome, the mapped reads were viewable on a personal track on the COMOSS.org genome browser (gbrowser). An example output from the gbrowser with mapped reads from example WT and *smg1*Δ lines is shown in Figure 5.1A. Figure 5.1A shows the *PpSMG1* locus (gene model: Pp1s51_180U2__zimmer.1), which was replaced by the kanamycin selection cassette in *smg1*Δ lines (Chapter 3). Figure 5.1A confirms that *PpSMG1* was deleted from *smg1*Δ lines as no reads from libraries generated from *smg1*Δ lines map to the coding sequence. Interestingly, the 3' UTR region of SMG1 that was not knocked out is over-expressed in both lines (Figure 5.1A). One possible explanation is that these reads are generated from the transcript produced by the kanamycin selection cassette. The selection cassette has the 35S terminator of transcription, however, this suggests that it is not very effective in moss and there is a great deal of read-through to the native *PpSMG1* terminator. This should serve as a warning to researchers who only knockout a proportion of a gene that they might be producing unexpected bicistronic transcripts including the selection cassette and the downstream proportion of the target gene. Surprisingly, this over-expression was much more pronounced in *smg1*Δ line 1 than in *smg1*Δ line 2. Both lines contain the same targeted gene replacement event so the changes could only be the result of stochastic epigenetic effects.

Once reads had been mapped to the genome, differential gene expression (DGE) analysis was performed (D. Lang and A. Zimmer, personal

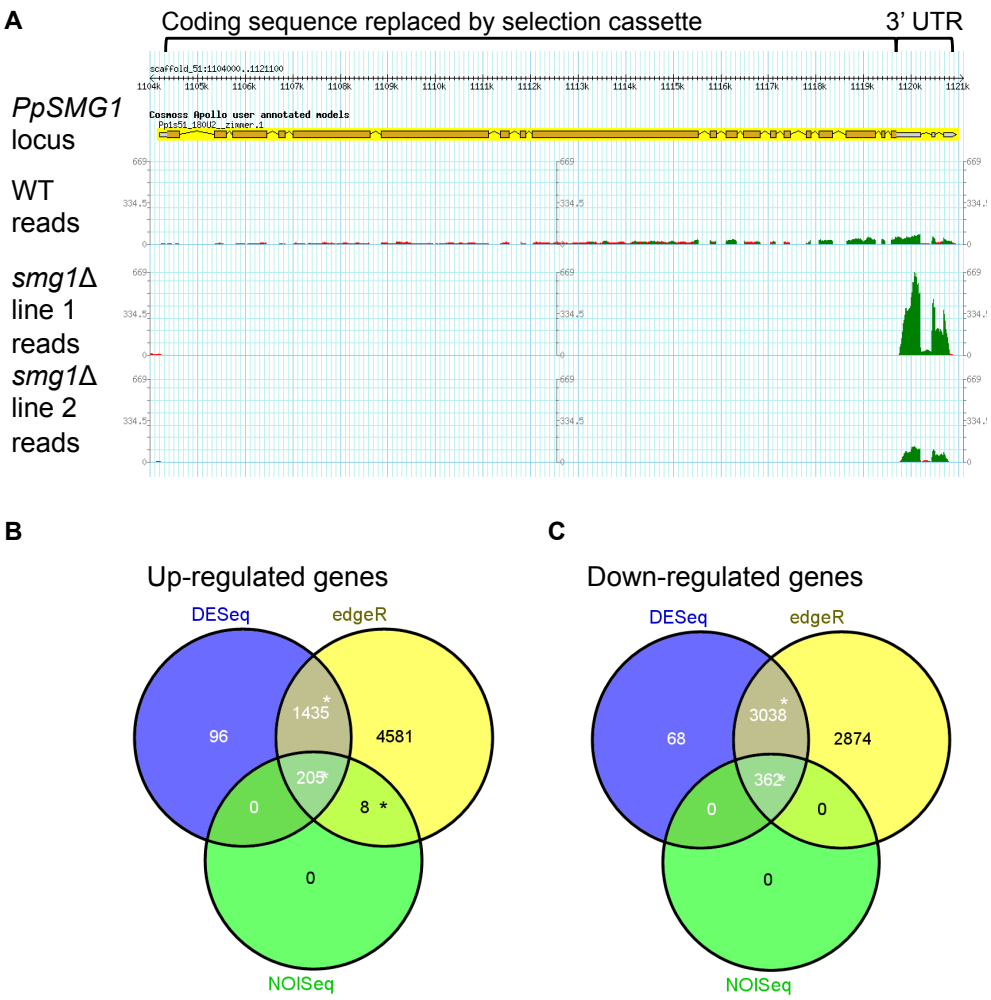

Figure 5.1: Downstream analysis of RNA-seq mapped reads

(A) Example of the output from the gbrowser of COSMOSS.org with mapped RNA-reads. The *PpSMG1* locus (Pp1s51_180U2__zimmer.1) with reads supporting the exon-intron organisation. The lack of reads in the *smg1*Δ lines confirms that *PpSMG1* has be successfully replaced in these lines. (B and C) Genes that are up-regulated and down-regulated, respectively, in *smg1*Δ lines when compared to WT ($p<0.05$). Three different tools were used to assess if a transcript was up- or down-regulated (DESeq, edgeR and NOISeq). During the first round of selection, only genes that were differentially regulated in at least two tools were taken forward (overlap is indicated with an *).

communication). Reads that fell into a gene model predicted by the version 1.6 of the moss genome (COSMOSS.org) were be counted to estimate fold change and its statistical significance. Statistical significance of DGE was used to generate a list of up- or down-regulated genes rather than setting a fold-change threshold for genes to be considered differentially expressed. Therefore, small but significant changes in gene expression were analysed. Three statistical packages/tools were used to assess if a gene was up- or down-regulated between WT and the *smg1*Δ lines; DESeq, edgeR and NOISeq. For the initial downstream analysis, such as GO term and MapMan analysis, genes identified as differentially expressed by two of these statistical tools were considered, giving a total of 1648 up-regulated genes and 3400 down-regulated genes (Figure 5.1).

5.2.2 GO/MapMan terms are enriched in differentially expressed genes

With thousands of up- and down-regulated genes revealed by RNA-seq in *smg1*Δ lines, gene ontology (GO) and MapMan terms were used to give an insight into the differentially regulated pathways and functions in these moss mutants (D. Lang and A. Zimmer, unpublished data). GO and MapMan identifiers (IDs) were assigned to moss genes by comparison to *A. thaliana* homologous genes and their IDs (Zimmer et al., 2013).

For up-regulated genes in *smg1*Δ lines (pathways normally repressed by NMD), 20 GO biological process terms were significantly enriched. A single gene can

be assigned to multiple GO terms, for example, there is overlap between ‘Auxin efflux’, ‘Positive gravitropism’, ‘Basipetal auxin transport’, and ‘Indoleacetic acid biosynthetic process’ (Table 5.2). These genes include homologues of ABC transporters involved in auxin transport in *A. thaliana* and enzymes involved in the biosynthesis of the biologically active auxin, IAA. The term ‘DNA recombination’ was the most significantly over-represented GO term from the up-regulated gene list (Table 5.2) and the ‘DNA recombination’ up-regulated genes included several DNA helicases predicted to function in DNA repair and DNA-PKcs, a PIKK family member involved in DNA repair in mammals (Table 5.2). ‘mRNA processing’ related genes were also identified as those normally repressed by *PpSMG1* (Table 5.2). This could be because many splicing factors generate splice variants targeted to NMD and loss of NMD means these variants are up-regulated.

Among GO terms enriched among down-regulated genes are those relating to ‘Translation’ and ‘Photosynthesis’ (Table 5.3), suggesting normal cellular homeostasis and energy acquisition is reduced. In contrast to *A. thaliana*, where defense against pathogen genes are up-regulated in NMD mutants, moss *smg1*Δ lines have GO terms such as ‘Defense response to fungus’, ‘Response to salicylic acid stimulus’ and ‘Systemic acquired resistance’ enriched among down-regulated genes (Table 5.3). This suggests that repressing a defense response is not conserved across plants.

There are five MapMan terms associated with genes with an increased steady state in *smg1*Δ lines (Table 5.4). These include ‘stress.abiotic.heat’ and

Table 5.2: Enriched GO terms relating to biological processes of up-regulated genes in the *smg1*Δ lines relative to the WT

GOid	Annotated genes in GOid	Number of differential genes in GOid	Expected number of genes for significance	p value for enrichment (Benjamini-Hochberg corrections)	GO term description
GO:0006310	43	10	2.56	0.000674531	DNA recombination
GO:0008361	66	12	3.93	0.000781396	Regulation of cell size
GO:0006139	2436	168	144.92	0.000885892	Nucleobase-containing compound metabolic process
GO:0006424	5	3	0.3	0.001914356	Glutamyl-tRNA aminoacylation
GO:0009958	10	4	0.59	0.001952205	Positive gravitropism
GO:0048767	10	4	0.59	0.001952205	Root hair elongation
GO:0010315	2	2	0.12	0.003533978	Auxin efflux
GO:0010540	2	2	0.12	0.003533978	Basipetal auxin transport
GO:0030655	2	2	0.12	0.003533978	Beta-lactam antibiotic catabolic process
GO:0042538	6	3	0.36	0.003659675	Hyperosmotic salinity response
GO:0006754	163	14	9.7	0.004852099	ATP biosynthetic process
GO:0030036	41	7	2.44	0.005355531	Actin cytoskeleton organization
GO:0006397	59	11	3.51	0.006187483	mRNA processing
GO:0009664	39	7	2.32	0.007408526	Plant-type cell wall organization
GO:0001522	22	5	1.31	0.008285205	Pseudouridine synthesis
GO:0005992	15	4	0.89	0.009999632	Trehalose biosynthetic process
GO:0051260	4	3	0.24	0.010153103	Protein homooligomerization
GO:0009684	3	2	0.18	0.010182698	Indoleacetic acid biosynthetic process
GO:0043481	3	2	0.18	0.010182698	Anthocyanin accumulation in tissues in response to UV light

‘chaperones and cochaperones.HSP70s.chaperones’, which are overlapping. All of the genes up-regulated in ‘stress.abiotic.heat’ are included in the ‘chaperones and cochaperones.HSP70s.chaperones’ and related to the classical heat shock response/UPR. This suggests that perhaps the NMD pathway represses the UPR in plants as it does in animals (Sakaki et al., 2012; Gardner, 2008). In agreement with what we (me, Dr Daniel Lang and Dr Andreas Zimmer) see in GO terms associated with up-regulated genes, we see ‘RNA.processing.splicing’, suggesting NMD normally represses splicing factors, possibility through AS-couple NMD.

As for down-regulated genes significantly enriched for MapMan terms, several terms relating to photosynthesis are present, agreeing with what was present in GO terms for down-regulated genes (Table 5.3 and 5.5). Additionally, several terms relate to the cell wall (Table 5.5), suggesting that the cell wall structure of *smg1*Δ lines might be altered.

Taken together, enrichment of GO and MapMan terms among the up- and down-regulated genes in *smg1*Δ lines suggests that NMD and/or *PpSMG1* have important roles in moss growth and development, as has been observed in various eukaryotes. NMD and/or *PpSMG1* appear to repress the expression of genes relating to the UPR and DNA repair but a physiologically relevant role still needs to be established.

Table 5.3: Enriched GO terms relating to biological processes of down-regulated genes in the *smg1*Δ lines relative to the WT

Table 5.3 *continued on next page*

GOid	Annotated genes in GOid	Number of differential genes in GOid	Expected number of genes for significance	p value for enrichment (Benjamini-Hochberg corrections)	GO term description
GO:0006412	753	136	67.76	1.47E-19	Translation
GO:0009813	16	14	1.44	2.11E-13	Flavonoid biosynthetic process
GO:0015979	186	55	16.74	6.05E-11	Photosynthesis
GO:0009765	46	18	4.14	1.39E-07	Photosynthesis, light harvesting
GO:0009698	36	25	3.24	3.15E-07	Phenylpropanoid metabolic process
GO:0006559	14	9	1.26	4.93E-07	L-phenylalanine catabolic process
GO:0006073	105	24	9.45	6.22E-07	Cellular glucan metabolic process
GO:0006979	127	27	11.43	9.25E-07	Response to oxidative stress
GO:0042545	63	21	5.67	8.30E-06	Cell wall modification
GO:0006629	445	67	40.04	3.13E-05	Lipid metabolic process
GO:0009753	37	12	3.33	5.84E-05	Response to jasmonic acid stimulus
GO:0030001	278	38	25.02	0.000424563	Metal ion transport
GO:0015717	3	3	0.27	0.00072663	Triose phosphate transport
GO:0000902	78	9	7.02	0.00129628	Cell morphogenesis
GO:0015995	21	7	1.89	0.001759703	Chlorophyll biosynthetic process
GO:0055114	874	96	78.65	0.002935677	Oxidation reduction
GO:0030244	30	8	2.7	0.004060251	Cellulose biosynthetic process
GO:0006032	19	6	1.71	0.005091134	Chitin catabolic process
GO:0006869	32	8	2.88	0.006212429	Lipid transport
GO:0010248	2	2	0.18	0.00809012	Establishment or maintenance of transmembrane electrochemical gradient
GO:0046686	85	15	7.65	0.008308782	Response to cadmium ion
GO:0009644	10	4	0.9	0.008793812	Response to high light intensity
GO:0006633	104	19	9.36	0.009949742	Fatty acid biosynthetic process

Table 5.3 *Continued*

GOid	Annotated genes in GOid	Number of differential genes in GOid	Expected number of genes for significance	p value for enrichment (Benjamini-Hochberg corrections)	GO term description
GO:0050832	22	7	1.98	0.012753792	Defense response to fungus
GO:0006541	24	8	2.16	0.014294344	Glutamine metabolic process
GO:0006542	7	3	0.63	0.01928642	Glutamine biosynthetic process
GO:0008272	7	3	0.63	0.01928642	Sulfate transport
GO:0009751	25	6	2.25	0.020886292	Response to salicylic acid stimulus
GO:0009827	6	3	0.54	0.022783447	Plant-type cell wall modification
GO:0046931	3	2	0.27	0.022817098	Pore complex assembly
GO:0031348	13	4	1.17	0.024078686	Negative regulation of defense response
GO:0009664	39	10	3.51	0.024736564	Plant-type cell wall organization
GO:0009651	79	14	7.11	0.028503606	Response to salt stress
GO:0009423	8	3	0.72	0.028821843	Chorismate biosynthetic process
GO:0033587	8	3	0.72	0.028821843	Shikimate biosynthetic process
GO:0009733	59	10	5.31	0.029793522	Response to auxin stimulus
GO:0009627	21	5	1.89	0.035192884	Systemic acquired resistance
GO:0006955	92	12	8.28	0.040226306	Immune response
GO:0006081	17	5	1.53	0.040255213	Cellular aldehyde metabolic process
GO:0042344	9	3	0.81	0.040394366	Indole glucosinolate catabolic process
GO:0006414	22	5	1.98	0.04228173	Translational elongation
GO:0019685	16	4	1.44	0.04284599	Photosynthesis, dark reaction
GO:0052331	4	2	0.36	0.042923281	Hemolysis of cells in other organism during symbiotic interaction
GO:0016998	45	8	4.05	0.045305498	Cell wall macromolecule catabolic process
GO:0016114	39	8	3.51	0.049103417	Terpenoid biosynthetic process

Table 5.4: Enriched MapMan terms of up-regulated genes in the *smg1*Δ lines relative to the WT

MapMan term	p value for enrichment (Benjamini-Hochberg corrections)
DNA.synthesis/chromatin structure	0.004577991
stress.abiotic.heat	0.004655027
transport.ABC transporters and multidrug resistance systems	0.014763367
chaperones and co-chaperones.HSP70s.chaperones	0.023272632
RNA.processing.splicing	0.026012463

Table 5.5: Enriched MapMan terms of down-regulated genes in the *smg1*Δ lines relative to the WT

Table 5.5 *continued on next page*

MapMan term	p value for enrichment (Benjamini-Hochberg corrections)
PS.lightreaction.photosystem II.PSII polypeptide subunits	0.000204908
transport.Major Intrinsic Proteins.PIP	0.000238753
signalling.receptor kinases.leucine rich repeat X	0.000854828
PS.lightreaction.photosystem II.LHC-II	0.001100448
secondary metabolism.phenylpropanoids.lignin biosynthesis.PAL	0.001100448
signalling.receptor kinases.leucine rich repeat XIV	0.001100448
signalling.receptor kinases.leucine rich repeat XII	0.001260012
misc.protease inhibitor/seed storage/ lipid transfer protein (LTP) family protein	0.002000159
signalling.receptor kinases.leucine rich repeat XI	0.002000159
cell wall.degradation.pectate lyases and polygalacturonases	0.002181506
PS.lightreaction.photosystem I.LHC-I	0.00658954
cell wall.cellulose synthesis.cellulose synthase	0.006656297
development.late embryogenesis abundant	0.007125125
misc.plastocyanin-like	0.007673018
misc.GDSL-motif lipase	0.007673018
misc.beta 1,3 glucan hydrolases.glucan endo-1,3-beta-glucosidase	0.015170978
RNA.regulation of transcription.zf-HD	0.015630842
cell wall.modification	0.016788378
cell wall.cell wall proteins.LRR	0.020612336

Table 5.5 *continued*

MapMan term	p value for enrichment (Benjamini-Hochberg corrections)
stress.abiotic.drought/salt	0.020612336
not assigned.unknown	0.020612336
secondary metabolism.isoprenoids	0.021616183
transport.amino acids	0.021616183
hormone metabolism.jasmonate.synthesis-degradation.lipoxygenase	0.022115347
lipid metabolism.FA synthesis and FA elongation.beta ketoacyl CoA synthase	0.02828771
lipid metabolism.FA synthesis and FA elongation.fatty acid elongase	0.02828771
secondary metabolism.wax	0.030075735
protein.synthesis.ribosomal protein.eukaryotic.40S subunit.S27	0.034296729
stress.biotic	0.041991118
signalling.receptor kinases.leucine rich repeat III	0.041991118
RNA.processing	0.042294296
transport.H+ transporting pyrophosphatase	0.047336748

5.2.3 NMD represses the unfolded protein response

As loss of *PpSMG1* lead to the over-expression of several genes relating to the UPR, identified through MapMan analysis (Table 5.4), the changes in this pathway resulting from the loss of *PpSMG1* were further characterised. The UPR-related transcripts identified encode a range of proteins, including seven HSP70-like proteins and small HSPs (two HSP18 proteins and five HSP17) predicted to be present in both the cytosol and the chloroplast, suggesting that the whole cell is responding to unfolded proteins rather than a specific sub-cellular compartment. The most well studied UPR pathway in plants and animals is that of the ER. Manual searches of genes in *smg1*Δ lines identified genes encoding proteins predicted to be responsible for ER homeostasis to be up-regulated. For example, BiP proteins are the major chaperones in the ER of many eukaryotes (Gupta and Tuteja, 2011) and *PpBiP1* (Pp1s181_3V6.1) and *PpBiP2* (Pp1s288_23V6.1) both have an increased steady state level in *smg1*Δ lines. However, *PpBiP2* was only identified by a single statistical tool (edgeR). In addition, *PpDerlin-1a* (Pp1s213_66V6.1), a protein that exports unfolded proteins from the ER for degradation, *PpERdj3A* (Pp1s368_19V6.1), a HSP70 co-chaperone and *PpIRE1b* (Pp1s34_189V6.1), a receptor kinase that regulates the UPR of the ER were all up-regulated in *smg1*Δ lines compared to WT, suggesting UPR of the ER was activated. The expression of these genes was confirmed in WT and *smg1*Δ line 1 by qRT-PCR (Figure 5.2). The expression of all five genes was significantly increased in *smg1*Δ line 1 compared to WT. The expression of a heat-shock transcription factor-encoding

transcript (*HSF*) was also tested and found to be significantly increased (Figure 5.2).

In addition to testing the expression of UPR of the ER related genes in WT and *smg1*Δ line 1, their expression when moss was exposed to 1 µg/ml of tunicamycin (Tm) dissolved in DMSO was tested. Tm inhibits N-site glycosolation, inducing unfolded proteins specifically in the ER and has been used to study the UPR of the ER in plants and animals (Sugio et al., 2009; Sakaki et al., 2012). The expression of *PpBiP1*, *PpBiP2* and *PpERjd3A* were found to be significantly up-regulated in WT treated with Tm compared to WT treated with DMSO solvent control (Figure 5.2). In the case of most transcripts, the effect of Tm was not additive to the loss of *PpSMG1* (Figure 5.2A, C-E). However, in the case of *PpBiP2* an even higher level of up-regulation after loss of *PpSMG1* than in WT is observed (Figure 5.2B). *PpBiP2* is a highly regulated gene that under normal conditions is expressed at a very low level (RNA-seq; D. Lang and A. Zimmer, personal communication) but is up-regulated 52-fold after a two-week exposure to Tm (Figure 5.2B). The expression of a heat shock transcription factor-encoding transcript (*HSF*; Pp1s31_388V6.1) was also examined and confirmed that it was up-regulated in *smg1*Δ line 1 on DMSO control but not in WT or *smg1*Δ line 1 exposed to Tm (Figure 5.2). This suggests that the HSF (Pp1s31_388V6.1) is not involved in the UPR, or at least the ER branch of the UPR, and is up-regulated through other changes induced by the loss of *PpSMG1*. In mammals, unfolded proteins and related stresses inhibit NMD through eIF2-α phosphorylation, which in turn de-represses transcription factors needed for a normal UPR that are targeted by NMD

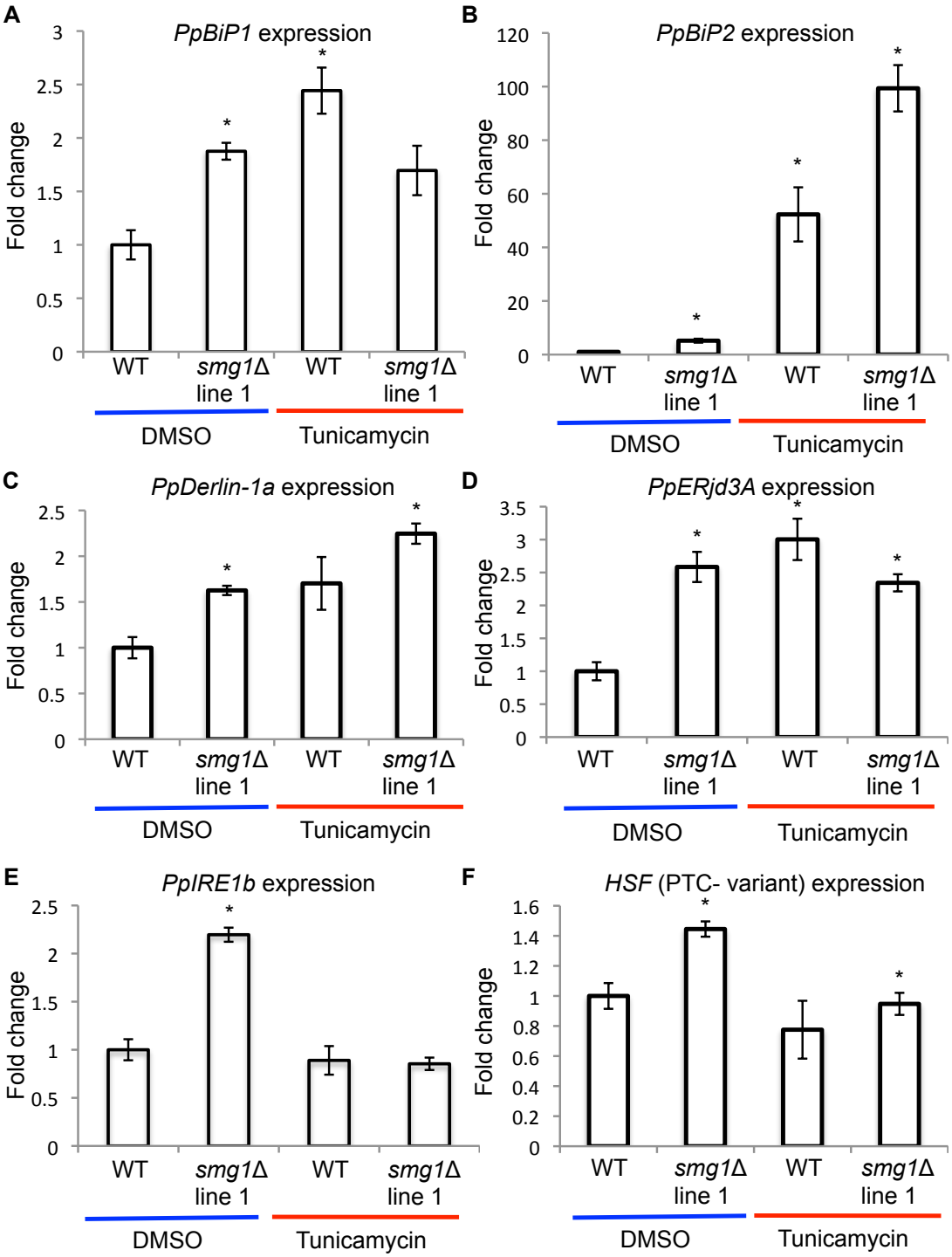

Figure 5.2: Expression of UPR-related genes in *smg1Δ* lines

(A) Expression *PpBiP1*. **(B)** Expression of *PpBiP2*. **(C)** Expression of *PpDerlin-1a*. **(D)** Expression of *PpERjd3A*. **(E)** Expression of *PpIRE1b*. **(F)** Expression of

Legend continued on next page

Figure 5.2: *Continued…*

HSF (PTC- variant). (**A**-**F**) Moss was grown for two-weeks on Tm (1 µg/ml) or solvent control (DMSO). 'Fold change' is the amount of target expression normalised to *PpEF1α* and relative to WT levels. Error bars represent the standard error of the mean from three biological replicates. Asterisks represent lines with a statistical difference from WT DMSO using an unpaired *t* test $p < 0.05$.

(Gardner, 2008; Wang et al., 2011). Therefore, whether NMD was inhibited by exposure to Tm (1 µg/ml) for two-weeks was tested by measuring the level of two transcripts with PTCs that are normally targeted by NMD (Figure 5.3). It was found that the PTC+ variants of *PpRS2Z37* and an *HSF*-encoding transcript (Pp1s31_388V6) were up- regulated in *smg1*Δ lines with or without Tm but not in WT exposed to Tm. These data suggest that NMD is not repressed in response to unfolded proteins as it is in mammals through eIF2-α phosphorylation.

To identify whether over-expression of UPR-related transcripts altered *smg1*Δ lines response to inducers of unfolded proteins, moss medium was initially supplemented with 2.5 µg/ml Tm or DMSO solvent control (Figure 5.4). Tm inhibited moss growth in both WT and *smg1*Δ lines, however, WT was more susceptible to Tm than the *smg1*Δ lines (Figure 5.4). The *smg1*Δ lines appear greener and more photosynthetically active (Figure 5.4A) and produce colonies with a larger area (Figure 5.4B). To confirm this effect was not drug specific and was indeed related to the UPR, moss was then exposed to 10 mM L-azetidine-2-carboxylic acid (AZC), a proline analogue known to induce unfolded proteins in plants and *A. thaliana* (Sugio et al., 2009) or 10 mM proline as a control (Figure 5.5). The *smg1*Δ lines appear greener and more photosynthetically active and produce colonies with a larger area (Figure 5.5), as was the case for exposure to Tm (Figure 5.4). Therefore, *smg1*Δ lines are less susceptible than WT to inducers of the unfolded protein response, possibly due to the increased expression of UPR-related genes like *PpBiP2*.

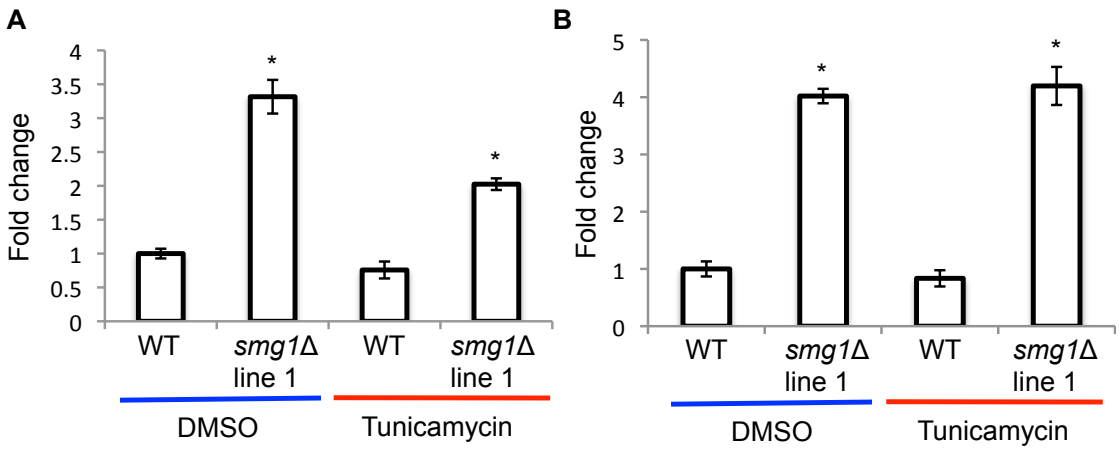

Figure 5.3: Tm exposure does not inhibit NMD

(A) Expression of *PpRS2Z37*. **(B)** Expression of *HSF* (PTC+ variant). (**A**-**B**) Moss was grown for two-weeks on Tm (1 µg/ml) or solvent control (DMSO). Fold change is the amount of target expression normalised to *PpEF1α* and relative to WT levels. Error bars represent the standard error of the mean from three biological replicates. Asterisks represent lines with a statistical difference from WT DMSO using an unpaired *t* test $p < 0.05$.

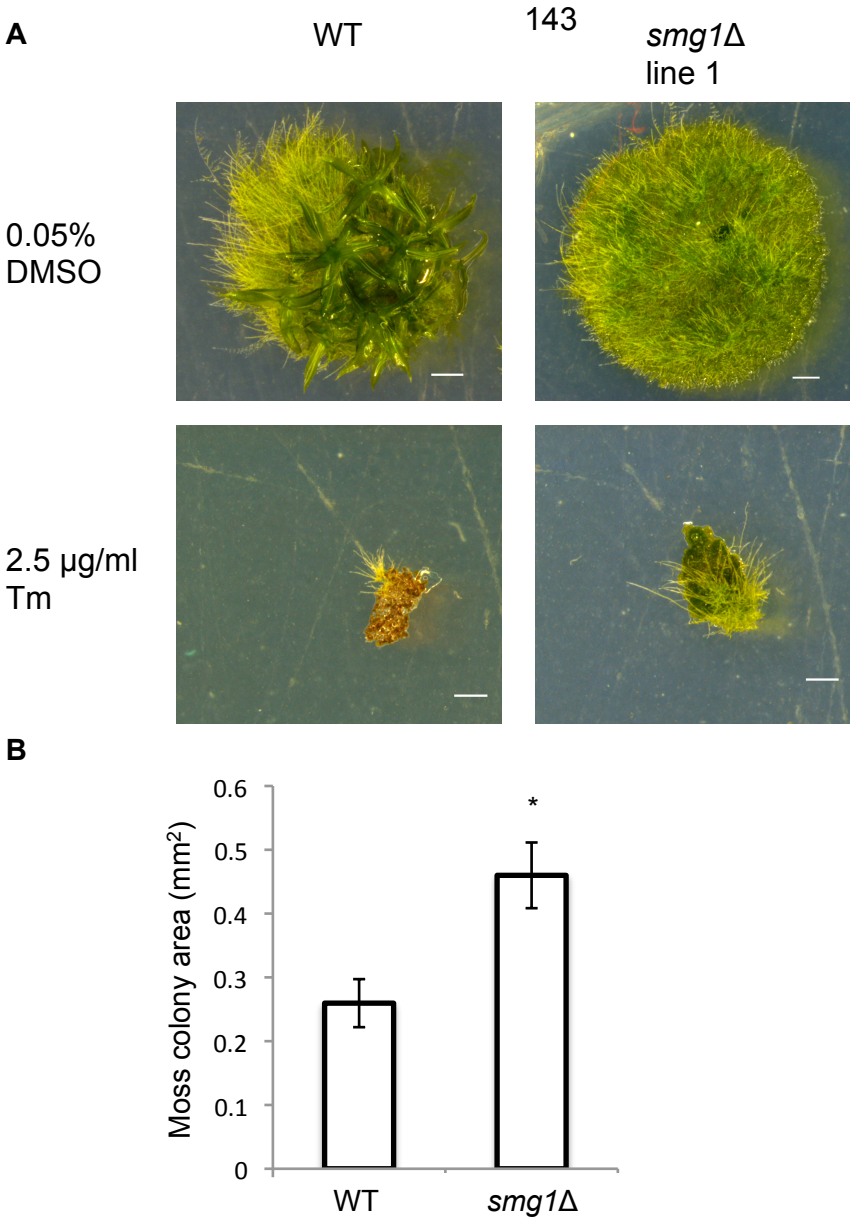

Figure 5.4: *smg1*Δ lines are partially resistant to Tm

(A) Three week old plants grown on Tm or solvent control (DMSO). Scale bar is 1 mm. **(B)** Moss colony size on Tm (2.5 µg/ml). n = 5-12. For *smg1*Δ, colonies of lines 1 and 2 were pooled. Asterisks represent lines with a statistically significant difference from WT using an unpaired *t* test $p < 0.05$.

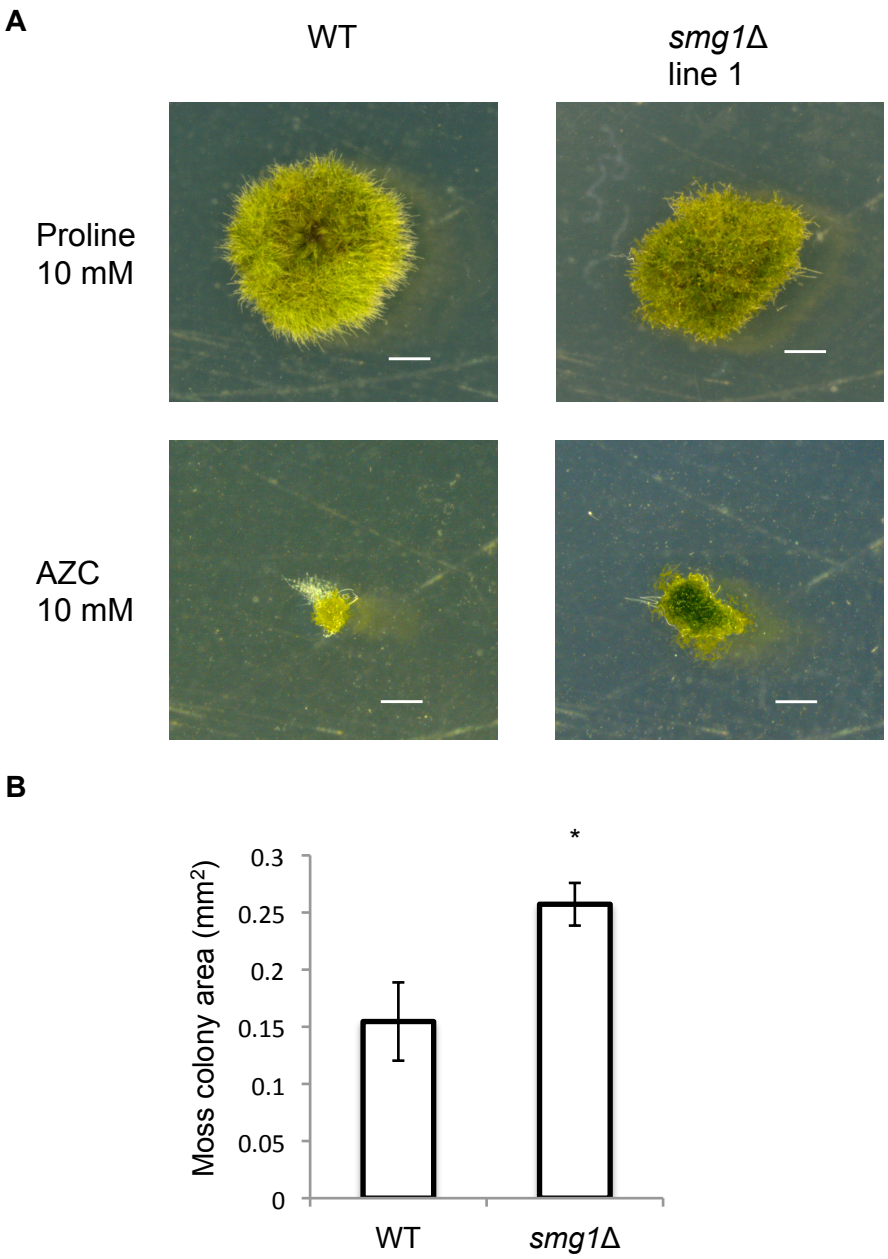

Figure 5.5: *smg1*Δ lines are partially resistant to AZC

(A) Three week old plants grown on AZC or proline control. Scale bar is 1 mm. **(B)** Moss colony size on AZC (10 mM). n = 6-12. For *smg1*Δ, colonies of lines 1 and 2 were pooled. Asterisks represent lines with a statistically significance difference from WT using an unpaired *t* test $p < 0.05$.

Taken together these data show that a compromised NMD pathway leads to up-regulation of the UPR-related transcripts and to partial resistances of *smg1*Δ lines to unfolded protein inducers. The mechanism linking NMD/PpSMG1 and the UPR is unclear but it is unlikely that unfolded proteins inhibit the NMD pathway in moss, as is the case in mammals (Wang et al., 2011).

5.2.4 PpSMG1 is involved in genome stability

Genes up-regulated in *smg1*Δ lines were enriched with those associated with the GO term 'DNA recombination' (Table 5.2) suggesting that DNA recombination or repair might be affected in *smg1*Δ lines. In mammals, SMG1 functions in the DNA repair pathway independently of the NMD pathway by phosphorylating p53 (Brumbaugh et al., 2004). In addition to the genes identified in the GO enrichment, manual searches revealed that other genes related to DNA recombination and repair were up-regulated. These included *PpATM* (Pp1s135_65V6.1), a PIKK family member that activates DNA repair, *PpRAD54b* (Pp1s236_78V6.1), a helicase predicted to be involved in DNA recombination, and *PpSOG1* (Pp1s251_11V6.1), a NAC family transcription family predicted to be involved with DNA repair (Yoshiyama et al., 2009; 2013). The activation of *PpATM* as opposed to *PpATR* (which is slightly down-regulated according to one statistical tool), suggests that double-strand-breaks are accumulating or that the signaling pathway responding to them has been de-repressed (Waterworth *et al.*, 2011). Interestingly, AtSOG1 acts downstream of AtATM in the DNA repair of double-strand break pathway and is needed for

the AtATM-dependent changes in gene expression (Waterworth et al., 2011; Yoshiyama et al., 2013; 2009).

The up-regulation of genes relating to DNA repair in the *smg1*Δ lines could be due to:

1. PpSMG1 repressing DNA repair, as it does the UPR, so that loss of *PpSMG1* leads to over-expression of DNA repair factors and potentially to an efficient or improved response to DNA damage.
2. PpSMG1 being involved in the DNA repair pathway, as it is in mammals (Gehen et al., 2008; Brumbaugh et al., 2004). In this situation the loss of *PpSMG1* might lead to an inefficient DNA repair pathway and other repair genes might be up-regulated to either compensate for the loss of *PpSMG1* or to help repair damage that accumulates in the absence of *PpSMG1*.

To assess the physiological impact of the loss of *PpSMG1* on the DNA repair pathway, *smg1*Δ lines were exposed to bleomycin (zeocin), a potent inducer of double-strand DNA damage (Chankova *et al.*, 2007; Kamisugi *et al.*, 2011). *smg1*Δ lines are more susceptible to bleomycin induced growth inhibition than WT, producing significantly smaller colonies after three weeks of exposure (Figure 5.6). These data suggest that SMG1 is needed for an efficient DNA repair pathway in moss as in animals (Gehen et al., 2008; Brumbaugh et al., 2004).

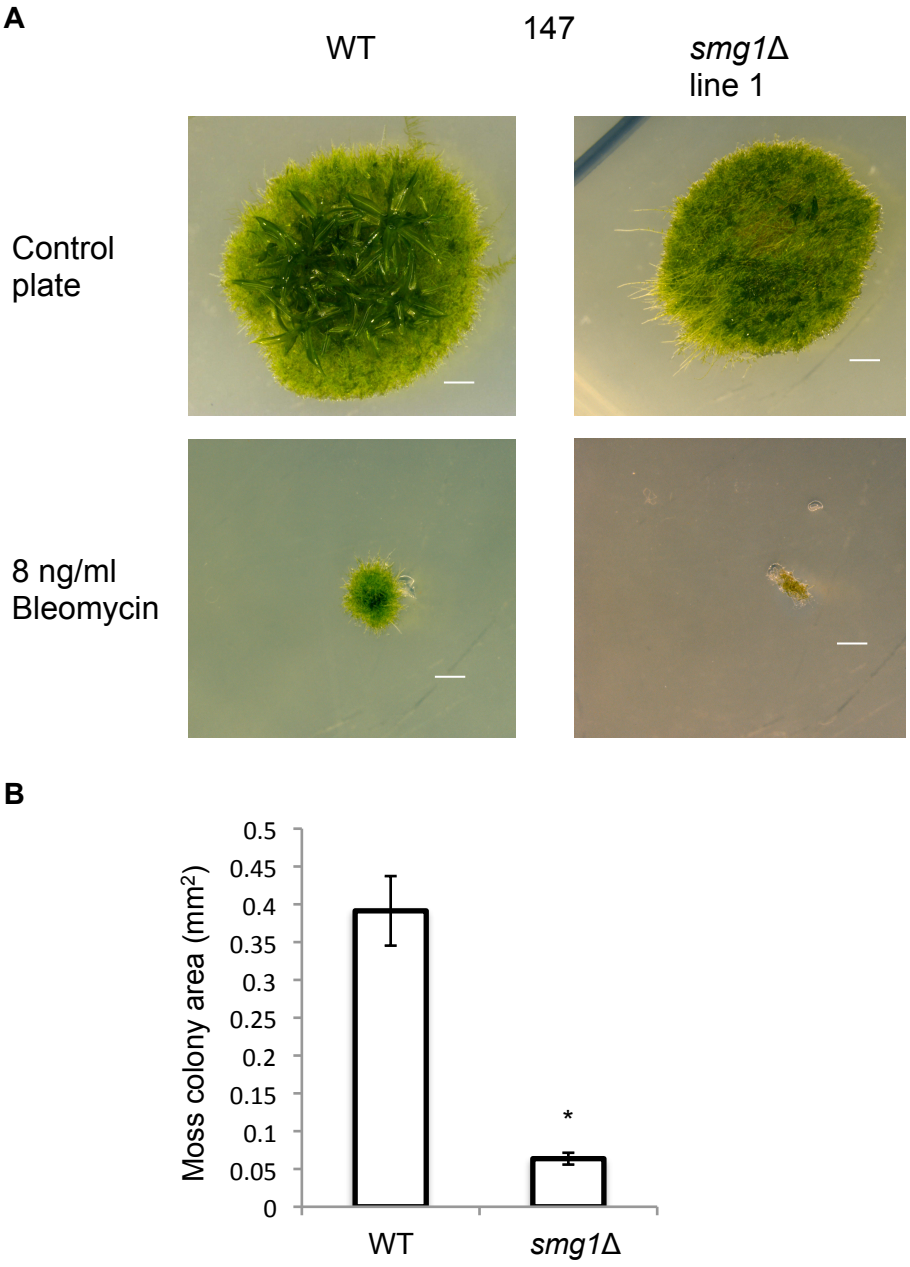

Figure 5.6: *smg1*Δ lines are more susceptible to bleomycin than WT

(A) Three week old plants grown on 8 ng/ml bleomycin or control plate. Scale bar is 1 mm. **(B)** Moss colony size on bleomycin (8 ng/ml). n = 6-12. For *smg1*Δ, colonies of lines 1 and 2 were pooled together. Asterisks represent lines with a statistically significant difference from WT using an unpaired *t* test $p < 0.05$.

5.2.5 NMD effectors and feedback loops

In *A. thaliana* and mammals, NMD effector encoding transcripts have been identified as targets of NMD creating an evolutionarily conserved negative feedback loop (Yepiskoposyan et al., 2011; Huang et al., 2011). The NMD pathway therefore regulates its own activity. The expression of all the core NMD and EJC components was analysed in the RNA-seq data (Table 5.6). The expression of EJC component encoding transcripts was not altered in *smg1*Δ lines (Table 5.6). Of *PpSMG7* encoding transcripts, only *PpSMG7-2* was designated as up-regulated by two statistical tools, while each of *PpSMG7-1* and *PpSMG7-3* were up in one tool. The over-expression of *PpSMG7-2* and *PpSMG7-3* have been confirmed by qRT-PCR (Figure 3.7), confirming the analysis of the RNA-seq analysis, even when only one statistical tool indicated differential expression in the case of *PpSMG7-3*. However, the over-expression of *PpSMG7-1* was not statistically significant in both *smg1*Δ lines when analysed by qRT-PCR, suggesting a small and border-line increase in expression (Figure 3.7). As previously stated, each of these transcripts have abnormally long 3' UTRs and two introns downstream of the stop codon, which can act as targeting features to NMD in flowering plants (Benkovics et al., 2011; Nyikó et al., 2013; Kerényi et al., 2008). The expression of *PpUPF1a* and *PpUPF1b* were slightly up-regulated according to only one tool but might represent a genuine up-regulation, like *PpSMG7-3* and *PpBiP2*, further work is needed to address this. The 3' UTRs of both *UPF1* homologues were abnormally long (both over 1200 nucleotides long), which could act as an NMD targeting feature. *PpUPF3a* was identified as up-regulated by two tools and has

Table 5.6: The expression of transcripts encoding NMD effectors in *smg1Δ*

NMD effector	Gene Model	Up-regulated in RNA-seq	Fold change in RNA-seq	Confirmed by qRT-PCR
PpUPF1a	Pp1s44_135V6.1	Yes[1]	1.32	ND
PpUPF1b	Pp1s10_103V6.1	Yes[1]	1.50	ND
PpUPF2	Pp1s123_14V6.1	No	1.30	ND
PpUPF3a	Pp1s246_94V6.1	Yes[2]	2.07	ND
PpUPF3b	Pp1s13_385V6.1	Yes[1]	1.50	ND
PpSMG7-1	Pp1s80_14V6.1	Yes[1]	1.63	Yes[ab]
PpSMG7-2	Pp1s311_73V6.1	Yes[2]	1.89	Yes[a]
PpSMG7-3	Pp1s28_218V6.1	Yes[1]	1.55	Yes[a]
PpMago1	Pp1s63_105V6.1	No	-1.08	ND
PpMago2	Pp1s125_34V6.1	No	-1.02	ND
PpY14a	Pp1s136_114V6.1	No	-1.03	ND
PpY14b	Pp1s31_259V6.1	No	-1.07	ND
PpeIF4AIIIa	Pp1s275_62V6.1	No	1.03	ND
PpeIF4AIIIb	Pp1s519_18V6.1	No	1.01	ND
PpMLN51a	Pp1s193_55V6.1	No	1.23	ND
PpMLN51b	Pp1s193_54V6.1	No	1.23	ND

[a] – The qRT-PCR results are shown in Figure 3.7. [b] – *PpSMG7-1* was only up-regulated in line 2. ND – The qRT-PCR was not performed for transcripts that were not up-regulated in the RNA-seq experiment. Superscripted number – Represents the number of statistical packages (DESeq, edgeR and NOISseq) that showed statistical significance in DGE analysis.

a long 3’ UTR (*ca*. 850 nucleotides long), which could target the transcript to NMD. *PpUPF3b* was predicted to be up-regulated by one statistical tool and has both a long 3’ UTR (850 nucleotides) and a uORF, which could target it to NMD. Interestingly, none of the EJC-component encoding genes were differentially expressed (Table 5.6). This is in contrast to the recent report that *AtBarentsz1* and *AtBarentsz2* are up-regulated in NMD mutants through their 3’ UTR located intron to form a feedback loop for the intron-based NMD pathway (Nyikó et al., 2013). In contrast to this, neither of the moss homologues contain a 3’ UTR located intron, nor are they over-expressed in the NMD-comprised lines, suggesting that this is not an evolutionarily conserved feedback loop in plants (Table 5.6). Taken together, these data suggest that multiple negative feedback loops could regulate the NMD pathway of moss, as is the case in flowering plants and mammals, but does not include the core EJC-components like in *A. thaliana*.

5.3 Discussion

The aim of the work in this chapter was to identify the differentially expressed genes and pathways upon the loss of *PpSMG1*. Using conservative criteria we (me, Dr Daniel Lang and Dr Andreas Zimmer) have identified 1648 up-regulated genes and 3400 down-regulated genes in *smg1*Δ lines compared to WT (Figure 5.1B) with several biological pathways being affected (Tables 5.2-5). Then the biological significance of these changes in gene expression was tested by analysing the ability of the *smg1* mutant moss plants to unfolded protein stress and DNA damage stress. Mutant moss was better able to survive exposure to inducers of unfolded proteins (Figure 5.4 and 5.5) while mutant moss was more susceptible to DNA damage induced inhibition of growth (Figure 5.6). These data suggest that NMD and/or PpSMG1 plays an important role in the normal physiology of moss.

5.3.1 NMD might regulate the UPR in moss

We (me, Dr Daniel Lang and Dr Andreas Zimmer) have shown that NMD and/or PpSMG1 normally keeps the UPR turned off. It is unclear whether PpSMG1 functions to directly repress the UPR as a regulator of gene expression or whether the UPR is turned on indirectly in *smg1*Δ lines, for example through an increase in the accumulation of unfolded proteins caused by the failure of NMD to prevent the production of truncated proteins. Work in *C. elegans* found that enhancers of UPR activation included factors involved with mRNA decay

(Sakaki et al., 2012). The researchers then showed that mutations in NMD effector genes caused up-regulation of an UPR reporter, leading the authors to speculate that the truncated proteins produced in the absence of NMD induce the UPR (Sakaki et al., 2012). Subsequently, they found HeLa cells where *SMG6* had been knocked down were more susceptible to Tm as an inducer of the UPR than cells where *SMG6* had not been knocked down. They concluded that loss of NMD leads to more truncated proteins and that these cells are already under UPR stress, so they are more susceptible to Tm as an inducer of unfolded proteins (Sakaki et al., 2012). This work is in contrast to our findings in moss and the findings of another group working on mammalian cells (Wang et al., 2011). Wang et al. (2011) found that knockdown of *UPF1* or *UPF2* leads to a mild increase in mouse embryonic fibroblasts surviving exposure to Tm. This group has demonstrated that inhibition of NMD is needed for a full UPR, as under normal conditions NMD represses transcription factors through their encoding transcripts' uORFs (Gardner, 2008; Wang et al., 2011). NMD targets were not up-regulated after a two-week exposure of moss to Tm (Figure 5.3), suggesting that unfolded proteins do not inhibit NMD in plants as they do in mammals. Inhibition of NMD in mammals is dependent on eIF2-α phosphorylation after exposure to unfolded proteins (Wang et al., 2011). However, the UPR of plants might not involve eIF2-α phosphorylation. It has been shown that the level of phosphorylation of eIF2-α is not increased after heat shock or exposure to Tm in *A. thaliana* or wheat (Gallie et al., 1997; Lageix et al., 2008; Kamauchi et al., 2005). It is unclear why two reports using mammalian systems have found differing responses in the survival rate of mammalian cells in response to loss of NMD and exposure to Tm (Sakaki et al.,

2012; Wang et al., 2011). The reasons could include 1) differences between species that these cells arose from, 2) differences in cell type or cell survival assays and 3) differences in the NMD effectors knocked down and their role in NMD/UPR. However, it appears that both plants and animals use NMD to repress the UPR, however, further work is needed to understand the mechanism in plants.

It is currently unclear whether NMD directly represses the UPR and, if it does, at what point the pathway is affected (Figure 5.7). An alternative explanation is the one proposed in *C. elegans* (Sakaki et al., 2012), where the UPR is activated due to the production of unfoldable truncated proteins increasing in the absence of NMD (Figure 5.7). It has not been established why this increase in UPR-related gene expression in the mutant lines leads to an increase in resistance to unfolded protein inducing drugs. One possibility is that the up-regulation of UPR genes in unexposed *smg1*Δ lines has primed the moss so that it responds more quickly to misfolded proteins, aiding the plants ability to grow on Tm or AZC. However, another possibility is that *smg1*Δ lines can launch a stronger response to unfolded proteins than WT because there is reduced inhibition of the UPR pathway. The expression of *PpBiP2* is much larger in *smg1*Δ line 1 when exposed to Tm than WT exposed to Tm (Figure 5.2B). The latter situation would be similar to what is seen in the SA response to pathogens in *A. thaliana* NMD mutants. *A. thaliana* NMD mutants produce more SA compared to WT under normal growth conditions and when plants are attacked by a pathogen, leading to a large response to pathogens and a partial resistance to pathogens (Rayson et al., 2012a).

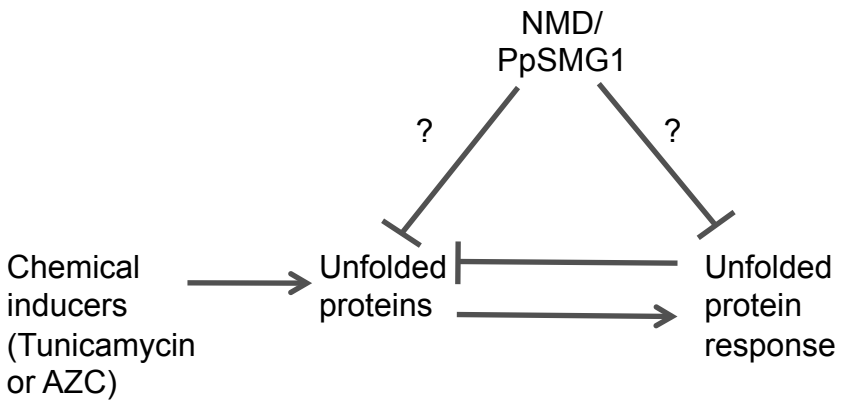

Figure 5.7: A model of the potential interaction between NMD/PpSMG1 and the UPR

NMD/PpSMG1 might directly repress the unfolded protein response or NMD might inhibit misfolded proteins accumulating by degrading transcripts encoding truncated proteins and thus indirectly repress the UPR.

It is interesting to ask whether eIF2-α phosphorylation can inhibit NMD in plants and if so, what the physiological significance of this may be. Pathogen attack and the plant defence hormone SA can both induce eIF2-α phosphorylation in *A. thaliana* (Lageix et al., 2008; Pajerowska-Mukhtar et al., 2012). NMD mutants have both an increased resistance to pathogens and elevated levels of SA (Rayson et al., 2012a; Riehs-Kearnan et al., 2012; Jeong et al., 2011) so it is possible that eIF2-α phosphorylation inhibits NMD in plants to help de-repress the pathogen response pathway. NMD in *A. thaliana* also targets multiple genes with CPuORFs (Rayson et al., 2012a; Nyikó et al., 2009). Interestingly, a transcription factor (AtTBF1), which is needed for *A. thaliana* to respond normally to pathogens is under the control of a CPuORF (Pajerowska-Mukhtar et al., 2012). The authors did not investigate whether the uORFs of this transcript control expression at the level of translation or decay, but suggested that they could repress translation of the main ORF unless eIF2-α was phosphorylated upon pathogen attack (Pajerowska-Mukhtar et al., 2012). It is possible that this transcript and others are normally repressed by NMD until eIF2-α is phosphorylated, releasing the repression and allowing the plants to launch a full pathogen response.

5.3.2 PpSMG1 is needed for genome stability

It was demonstrated that PpSMG1 is needed for an efficient DNA repair pathway in moss as loss of *PpSMG1* leads to plants not being able to survive exposure to the DNA damage inducing agent bleomycin (Figure 5.6). Interestingly, in mammals, loss of *SMG1* leads to spontaneous DNA damage

and increased susceptibility to ionizing radiation (Brumbaugh et al., 2004). The authors noted that knockdown of *SMG1* induced markers of DNA damage, such as H2AX phosphorylation. This might be analogous to the up-regulation of several DNA-repair related genes in *smg1*Δ lines (Table 5.2). It is possible that PpSMG1 regulates DNA repair through the NMD pathway (Figure 5.8A). However, in mammals SMG1 phosphorylates p53, the guardian of the genome, to activate the response to DNA damage, independently of NMD (Gehen et al., 2008; Brumbaugh et al., 2004). This phosphorylation of the S15 residue precedes the phosphorylation of this residue by ATM, which is an important regulator of p53 (Brumbaugh et al., 2004). Plants lack a homologue of p53, therefore, PpSMG1 might phosphorylate an unidentified moss equivalent of p53 to aid the DNA repair pathway (Figure 5.8B). It has been speculated that the transcription factor SOG1 is functionally analogous to p53 in plants (Yoshiyama et al., 2009; Waterworth et al., 2011). AtSOG1 is important in activating cell cycle arrest and cell death in response to DNA damage in *A. thaliana*, as p53 in does mammals (Yoshiyama et al., 2009; Waterworth et al., 2011). Recently it has been shown that AtATM-dependant phosphorylation of AtSOG1 at SQ residues is important for DNA repair (Yoshiyama et al., 2013). Therefore, it is tempting to suggest that SOG1 could be a target of SMG1 in plants and could be an important step in regulating the DNA repair pathway. Further work is needed to determine the exact role in the DNA repair pathway. For example, does plant SMG1 function in the repair of double-strand breaks through the homologous recombination pathway or through the non-homologous end-joining pathway.

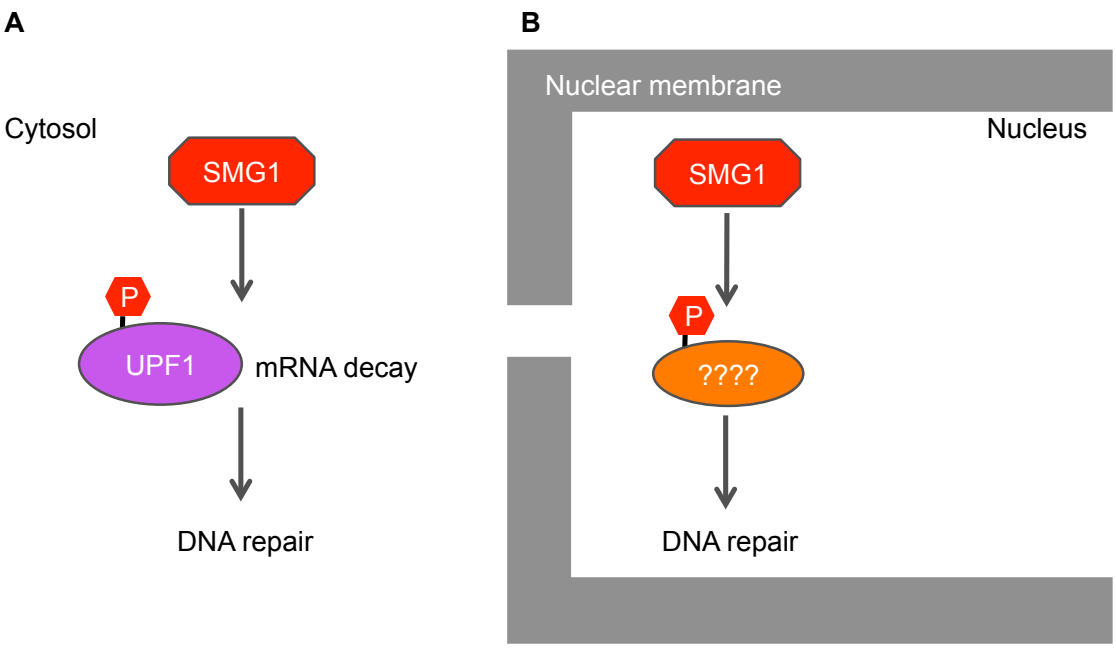

Figure 5.8: Models describing how PpSMG1 might function in the DNA repair pathway

(A) PpSMG1 could function in DNA repair through the NMD pathway in the cytosol. NMD could indirectly regulate the expression of DNA repair genes. **(B)** PpSMG1 could function in the nucleus by phosphorylating components of the repair machinery to activate normal repair as it does in mammals.

5.3.3 Multiple feedback loops regulate the NMD pathway in moss

In mammals and flowering plants, NMD has been shown to create multiple negative feedback loops to regulate the activity of the NMD pathway. Multiple genes encoding NMD effectors are targeted to NMD by uORFs, long 3' UTRs and/or by introns located in the 3' UTR. For example, in *A. thaliana*, *AtSMG7* is targeted to NMD by its long 3' UTRs and the two introns located downstream of the stop codon (Rayson et al., 2012a; Kerényi et al., 2008; Nyikó et al., 2013). The EJC component encoding *AtMLN51* is also targeted to NMD through 3' UTRs introns (Nyikó et al., 2013). It has also been suggested that *AtUPF3* is targeted to NMD through a uORF (Saul et al., 2009). In moss, multiple feedback loops are likely to exist as well. Whether *PpSMG1* is targeted to NMD could not be tested using our RNA-seq data, as this NMD effector was knocked out to collect the data. Both *UPF1* homologues were slightly up-regulated *smg1*Δ lines (Table 5.6). Both of these transcripts have long 3' UTRs, which is likely to be the NMD targeting feature. Both *UPF3* homologues were also found to be up-regulated in *smg1*Δ lines (Table 5.6). *PpUPF3a* has a long 3' UTR that could target it to NMD, while *PpUPF3b* also has a long 3' UTR but also a uORF, either or both of which could target it to NMD. Finally, all three homologues of *SMG7* appear to be targeted to NMD, to some extent (Table 5.6), which is likely to be through the long 3' UTRs and two 3' UTR located introns each of them contains. This suggests that multiple feedback loops control the activity of the NMD pathway in moss. In *A. thaliana*, *AtSMG7* is targeted to NMD through its long 3' UTR containing two introns and it has been suggested that *AtUPF3* is targeted to NMD through a uORF (Rayson et al., 2012a; Kerényi et al., 2008;

Nyikó et al., 2013; Saul et al., 2009). In mammals, multiple NMD effector-encoding transcripts are targeted to NMD to regulate the activity of the NMD pathway (Yepiskoposyan et al., 2011; Huang et al., 2011), for example, the long 3' UTRs of *UPF1*, *SMG5* and *SMG7* have been identified as the targeting features responsible (Yepiskoposyan et al., 2011).

In summary, we (me, Dr Daniel Lang and Dr Andreas Zimmer) have been able to use RNA-seq to identify transcripts deregulated by the loss of *PpSMG1* on a transcriptomic wide scale. The loss of PpSMG1 leads to a compromised NMD pathway (Chapter 3). Therefore, large fraction of differentially expressed genes are expected to be directly or indirectly regulated by NMD; however, a fraction of them could be regulated by NMD-independent roles of PpSMG1. SMG1 in mammals is known to function in the DNA repair pathway independently of NMD (Brumbaugh et al., 2004). A role of PpSMG1 in the DNA repair pathway was observed (Figure 5.6), but further work is needed to establish whether this is due to the loss of NMD or a role of PpSMG1 independent of NMD. Our work has also suggested a link between NMD and/or PpSMG1 and the UPR but more work will be needed to determine the exact nature of the relationship between PpSMG1 and the UPR and if this is conserved between species. This work has established that a loss of *PpSMG1* results in widespread changes in gene expression that have physiological significance, such as altering the ability of a plant to withstand changes to the amount of DNA damage or unfolded proteins it was exposed to. This work highlights that NMD is important in controlling gene expression across species.

6 General discussion

The work in this thesis has demonstrated that the NMD-associated kinase SMG1, which has previously been considered to be animal specific (Izumi et al., 2010), is both present in plants and functions in plant NMD (Chapter 3) as well as in animals (Grimson et al., 2004). Furthermore, development and evaluation of yeast-two and -three hybrid methods for identifying novel components of the NMD pathway in *A. thaliana*, such as a kinase that might have replaced SMG1 in *A. thaliana* has been performed (Chapter 4). Finally, we identified widespread changes in the transcriptome of moss upon loss of *PpSMG1*, indicating that NMD and/or PpSMG1 have important roles in moss growth and development, including influencing the UPR and DNA repair (Chapter 5).

6.1 An alternative kinase in the NMD pathway?

The model flowering plant *A. thaliana* lacks the SMG1 kinase, despite it being present in all other examined plants including the close relative of *A. thaliana*, *A. lyrata* (Chapter 3). Now that the role of SMG1 in the NMD pathway in plants has been demonstrated, the most likely scenario is that SMG1 phosphorylates UPF1 in plants, as it does in animals (Grimson et al., 2004). Others have previously demonstrated that AtUPF1 is phosphorylated in plants, using a expression tobacco system (which may or may not have a *SMG1* homologue), and that this is essential for NMD (Mérai et al., 2012), while the work presented here has shown that AtUPF1 needs to be modified, through phosphomimetic

mutation to interact with the important NMD effector AtSMG7 (Chapter 4). These data suggest that NMD in *A. thaliana* relies on an unidentified alternative kinase, which we (me and Prof Davies) have dubbed *A. thaliana* replacement kinase of SMG1 (ARK) until the true identity of the kinase has been identified. It is easy to predict that this kinase phosphorylates AtUPF1 at the same residues as SMG1 does in other plants (S/TQ dipeptides).

One possibility is that the related kinases ATM and ATR from the PIKK family might have replaced SMG1 in *A. thaliana*. NMD mutants in A. thaliana have an overlapping phenotype including curling of leaves, which is a result of an over production of the plant defence hormone SA (Rayson et al., 2012a; 2012b; Riehs-Kearnan et al., 2012). However, the phenotypes of the single and double mutants of *ATM* and *ATR* (Culligan et al., 2004) do not match the known NMD phenotypes. In animals, *ATM* was knocked down, but no NMD phenotype was observed suggesting that *ATM* does not have a role in NMD (Brumbaugh et al., 2004). However, this study also found that addition of wortmannin, an inhibitor of PIKKs, lead to a larger up-regulation of an NMD target than was observed in an *SMG1* knockdown. This indicates that either wortmannin is more effective at inactivating SMG1 than the *SMG1* knockdown lines, or that a second wortmannin sensitive kinase is involved in NMD in animals (Brumbaugh et al., 2004). More research is needed to confirm any role for these kinases in the NMD pathway of plants and animals.

If the ATM and ATR kinases are not involved in NMD, we have developed a yeast-three hybrid screen to identify the ARK kinase responsible for UPF1

phsophorylation in *A. thaliana* (Chapter 4). An alternative experimental approach would be to tag AtUPF1 (for example, with HA or GFP) and pull-down the AtUPF1 complex with anti-tag antibodies, identifying interacting proteins using mass spectrometry. Another option would be to adopt a genetic approach, using a mutant screen to identify novel components of the NMD pathway in *A. thaliana*. Since the physiological NMD mutant phenotype is caused by elevated SA (Rayson et al., 2012a; 2012b; Riehs-Kearnan et al., 2012), this phenotype cannot be used in the screen. To couple the loss of NMD to a readily identifiable phenotype, NMD could be made to target the activity of the fluorescence generating enzyme luciferase (LUC). The *LUC* encoding transcript could be engineered to contain an NMD targeting feature, for example a CPuORF or an alternative exon cassette with a PTC, so that the level of LUC activity is low in WT plants but high in NMD mutant plants. Our group is currently generating 35S::CPuORF::*LUC* plants and a *LUC* ORF with a cassette exon carrying a PTC, has already been shown to be targeted to NMD in *A. thaliana* (Hickey et al., 2012). As null mutations in NMD effector-encoding genes have a seedling lethal phenotype due to the overproduction of SA, the mutant screen could be performed in mutants that are undable to produce SA and have been shown to rescue SA induced phenotypes in NMD mutants (such as *pad4* or *sid2*) (Rayson et al., 2012b; Riehs-Kearnan et al., 2012). High-throughput sequencing has been used in the past to identify causative mutations much more quickly than conventional map-based cloning and is considered to be fast forward screens (Mokry et al., 2011; Schneeberger and Weigel, 2011). Such techniques could be employed in such screens and could

rapidly yield the identity of mutated genes in the newly identified NMD-comprised lines.

As previously discussed, we have hypothesised that the ARK kinase could be evolutionarily ancient, as has been demonstrated for SMG1 (Chapter 3). This is because *SMG1* has been independently lost multiple times in evolution, and work in *Drosophila* and zebrafish has suggested to us that there is redundancy between SMG1 and another, unidentified kinase in the NMD pathway (Chapter 3). Therefore, it is possible that in the LECA and many extant eukaryotes, including many plants, both SMG1 and ARK kinases phosphorylate UPF1 in NMD (Figure 6.1). Both kinases may have been conserved in the majority of plants and animals because each has non-NMD functions. For example, SMG1 has an independent role in DNA repair (Brumbaugh et al., 2004) (Chapter 5). If this is the case it will have widespread impact on NMD research in animals, fungi and plants and could be useful for directing research of medical and agricultural importance, given the roles of NMD in human and plant diseases.

6.2 The need for plant models alternative to *A. thaliana*

A. thaliana is an extremely useful model organism, with a relativity small genome, short generation time and large community resources such as a well curated genome database and insertion mutant collections. It is clear that this model has facilitated great advances in our basic understanding of plant biology (Koornneef and Meinke, 2010). However, *A. thaliana* is not representative of all

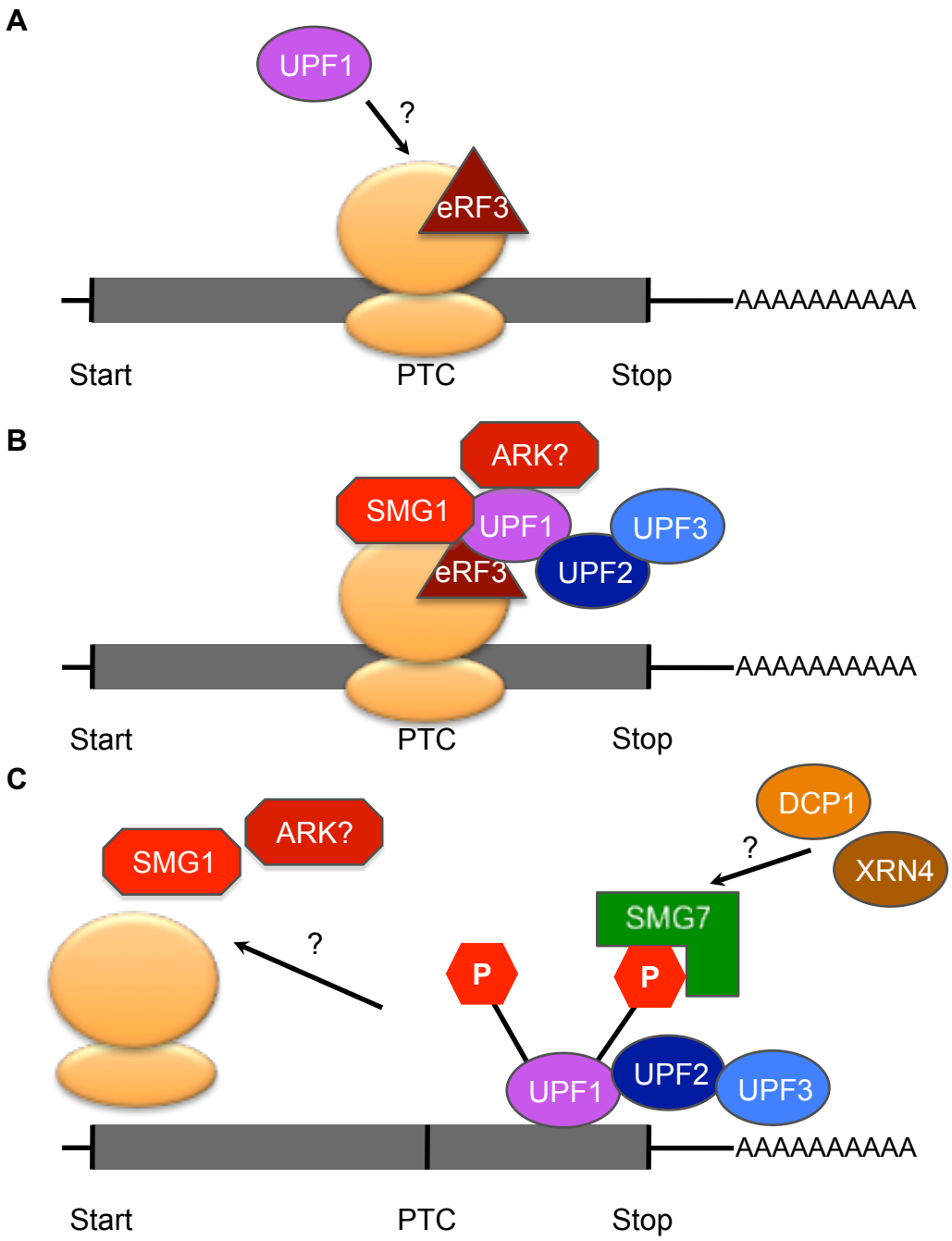

Figure 6.1: Model of NMD in organisms with two NMD-associated kinases

(A) Upon termination at the PTC, the ribosome stalls and UPF1 might be recruited at this stage by eRF3. **(B)** UPF1 forms a complex with SMG1/ARK, UPF2 and UPF3, which stimulates the kinase activity of SMG1/ARK. **(C)** SMG7 is recruited to at least the C-terminus phospho-site of UPF1. *Legend continued on next page*

Figure 6.1: *Continued…*

The ribosome and SMG1/ARK might have disassociated at this stage and UPF1 binds to the transcript. SMG7:SMG5 may recruit the decapping complex and the endonuclease XRN4.

other plants; not even flowering plants or other species in the *Arabidopsis* genus. For example, *A. thaliana* does not take part a symbiotic relationship with arbuscular mycorrhizal fungus (Kohlen et al., 2011), while 80% of all vascular plants do (Opik et al., 2010). Mutant analysis in *A. thaliana* has been important for the establishment of the ABCE model of flower development (Causier et al., 2010). However, work on other species are making it clear that the A-function is a largely *A. thaliana* specific feature, with mutants of A-class genes in other species not resembling their *A. thaliana* equivalents (Causier et al., 2010). This and other work has lead to a new model called the (A)BC model, which introduces multiple other genes that establish the flower into the new (A)-function (Causier et al., 2010).

Here, it has been demonstrated that unlike all other plants examined, including important crops like rice, maize, cassava and grape, *A. thaliana* lacks the SMG1 kinase from the NMD pathway (Chapter 3) (Grimson et al., 2004). Work in this thesis used moss to demonstrate that the SMG1 kinase functions in the NMD pathway of plants (Chapter 3), something that was not possible by using *A. thaliana* as a model organism. Additionally, a role for plant SMG1 in DNA repair has been identified, as is the case for animal SMG1 (Brumbaugh et al., 2004). More research is needed to establish the exact link between SMG1 and DNA repair. In humans, SMG1 directly phosphorylates p53 at serine 15, which precedes phosphorylation of this residue by fellow PIKK family member ATM, an important activator of the cell's response to double-strand breaks. An analogous situation might exist in plants. One possibility is that SOG1, a non-homologous plant equivalent of p53 (Yoshiyama et al., 2009) is a target of

SMG1. It has already been demonstrated that AtATM directly phosphorylates SQ dipeptides on AtSOG1 in order for normal DNA repair (Yoshiyama et al., 2013). It is therefore easy to speculate that SOG1 is a direct target of plant SMG1. However, *A. thaliana* cannot be used to establish a direct link between plant SMG1 and DNA repair and whether this involves SOG1 or another component of the DNA repair machinery.

The work presented in this thesis and highlighted from the literature reveals the need for more comparative studies and additional models in plant sciences. *A. thaliana* will no doubt continue to be the major model in plant sciences due to ease of use, wealth of background knowledge and strong community resources. However, animal research has been aided with the use of multiple model organisms, such as *C. elegans*, *Drosophila*, Zebrafish, mice, human cell culture and other less well used model organisms. In contrast, the plant community largely has focused on *A. thaliana* and no other secondary model has become established. Strong cases for organisms like moss, Petunia and *Brachypodium* can be made for being a useful model for comparative study (Gerats and Vandenbussche, 2005; Brkljacic et al., 2011). Moss with powerful reverse genetics as used in this study (Chapter 3), Petunia is another eudicot with a large insertion mutant collection and *Brachypodium* is a monocot and is becoming a model for the grasses. Additionally, high throughput genome sequencing means that any organism can be treated more like a model organism and we can determine what models to choose from post-genome sequence completion to develop resources for it, rather than generating vast resources before looking at the genome. However, plant scientists have not

rallied behind another model plant in the same way as with *A. thaliana*. Many other plants that have been used for research such as the grass rice have been chosen not for ease of use but because they are important crops and the hope is to apply the research directly to improving crop yields. While this has led to useful advances in crop sciences, such approaches are slow to yield novel advances in our understanding of basic biology, which is likely to lead to unexpected ways to manipulate crops in the long term. The outcome of the work presented in this thesis not only highlights the need for more plant models but has shown that the moss, *P. patens*, is a useful model. However, more flowering plant models are also required and a consensus in the plant sciences community is needed.

6.3 NMD as an important controller of gene expression

Loss of NMD in flowering plants, animals and fungi leads to widespread changes in gene expression (Guan et al., 2006; He et al., 2003; Mendell et al., 2004; Ramani et al., 2009; Rayson et al., 2012a; Rehwinkel et al., 2005). However, few targets are conserved among species (Rayson et al., 2012a). Only two clusters of orthologous groups/eukaryotic orthologous groups (COG/KOG) were up-regulated in *A. thaliana*, *Drosophila*, human and budding yeast NMD mutants or knockdowns (Rayson et al., 2012a). These were serine/threonine protein kinase and nonsense-mediated mRNA decay protein (where the SMG5-7 family was the conserved target of NMD) (Rayson et al., 2012a). Direct comparison with this dataset is not possible as moss genes have not yet been assigned to COG/KOG terms (D. Lang, personal communication).

However, here it has been shown that *SMG7* encoding transcripts are targets of NMD in moss (Chapters 3 and 5), suggesting conservation of an NMD feedback loop involving the SMG5-7 family across eukaryotes. COG/KOG terms associated with ABC transporters (also known as multidrug transporters) have been identified as up-regulated in *A. thaliana*, budding yeast and *Drosophila* NMD mutants/knock downs and we (me, Dr Daniel Lang and Dr Andreas Zimmer) have seen that a MapMan term was associated with ABC transporters among the over-expressed genes in *smg1*Δ lines (Chapter 5). It is unclear why ABC transporters would be conserved targets of NMD, whether they are direct or indirect targets and what the targeting feature is. Further study could reveal a novel mechanism to regulate these important membrane proteins. Analysis of GO and MapMan terms lead us to suspect *smg1*Δ lines had altered tolerance to cellular stresses such as DNA damage and unfolded proteins (Chapter 5). This hypothesis was tested by treating mutant and WT moss with inducers of DNA damage or the unfolded protein response and confirmed that the mutants had altered tolerances to these stresses (Chapter 5), demonstrating the usefulness of such ‘blunt’ instruments as GO and MapMan analysis.

Comparisons between diverse organisms at the individual gene level is difficult due to the difficulty in assigning orthology between such diverse genes *en masse*, therefore COG/KOG has been a good but limited proxy. However, this approach misses many important and conserved targets of NMD that are co-regulated with other proteins in a related biological process. For example, *eIF5* was identified as a conserved target of NMD across land plants (Chapter 3), which was not apparent from GO or MapMan analysis. *AteIF5L1* was identified

as a CPuORF containing gene that was over-expressed in NMD mutants (Rayson et al., 2012a) and it was demonstrated that the two moss homologues (*PpeIF5L1* and *PpeIF5L2*) are also CPuORF containing transcripts that are over-expressed in an NMD mutant (Chapter 3). It has been found that the animal homologue has an (unrelated) uORF and is part of an autoregulatory feedback loop of *eIF5* (Loughran et al., 2012). eIF5 regulates the translation of start codons in poor contexts, with more eIF5 leading to an increase in translation of start codons in a poor context (Figure 6.2) (Loughran et al., 2012). The uORFs of eIF5 in animals are in poor contexts and an increase in eIF5 balances the expression of eIF5 by promoting translation at the uORF start codon rather than the main ORF, possibly coupling it to NMD (Loughran et al., 2012). Such a feedback loop could exist in plants as the start codons of the moss and *A. thaliana eIF5*-encoding genes are also in poor contexts (S. Rayson, personal communication). An interesting consequence of this finding is that changes in the level of eIF5 and therefore the stringency of start codon selection, could alter the translation of many uORFs (Figure 6.2). If the stringency is low and more uORFs are consequently translated, this would bring more transcripts under the influence of NMD (Figure 6.2). It will be interesting to determine whether eIF5 levels control start codon selection in plants as it does in animals and whether this alters transcript susceptibility to NMD.

We have identified more than 1600 over-expressed genes and 3400 under-expressed genes in *smg1*Δ lines (Chapter 5). Subsequently it was shown that these changes in gene expression have a physiological effect on the moss in respect to the UPR and DNA damage (Chapter 5). However, the function of

many moss genes are unclear and the moss *smg1*Δ lines and our RNA-seq data provides a useful resource for studying the role of NMD and PpSMG1 in moss growth and development and transcritpomics.

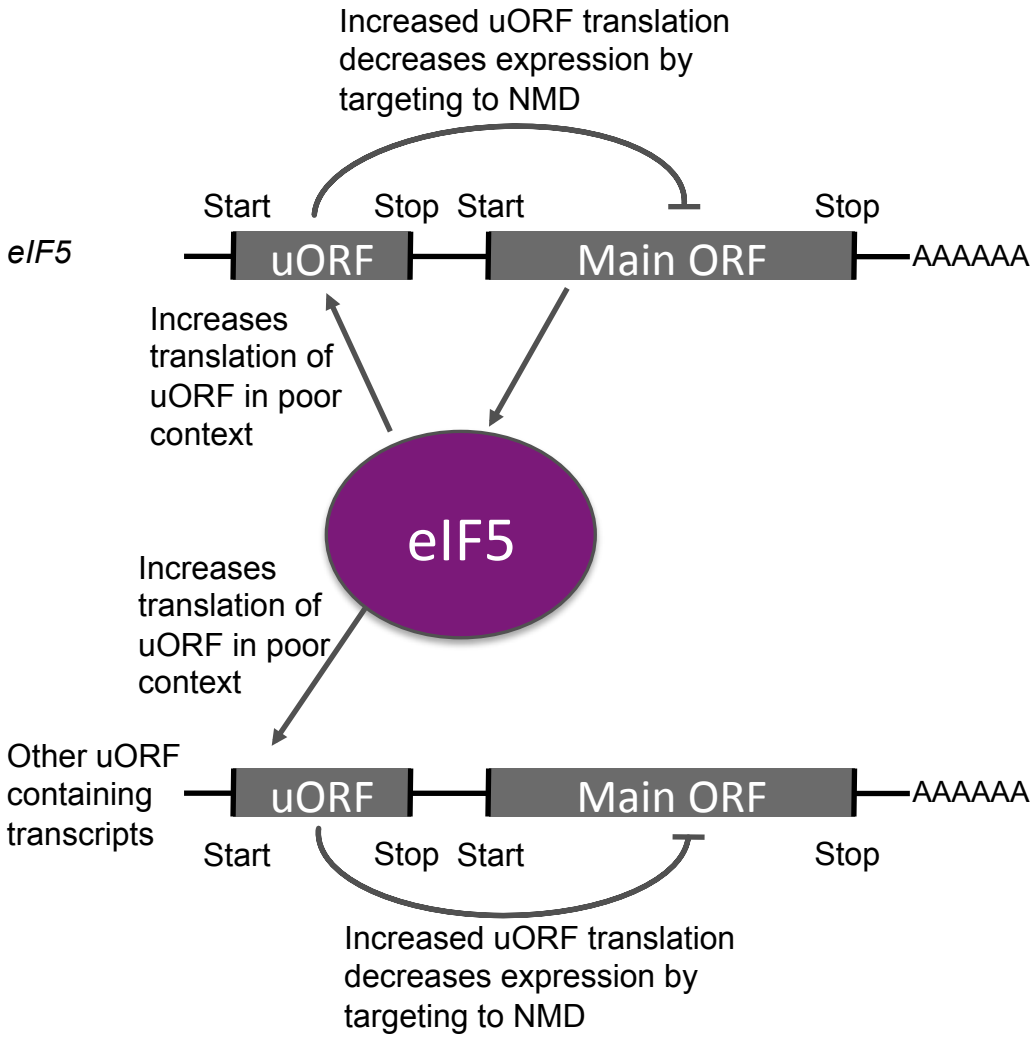

Figure 6.2: Model of eIF5-mediated regulation of NMD of transcripts containing uORFs in poor contexts

The eIF5 protein can increase the translation of uORFs in poor contexts and high levels of eIF5 protein could regulate its own transcript through NMD in an autoregulatory feedback loop and the NMD susceptibility of other transcripts containing uORFs in poor contexts.

References

Amrani, N., Ganesan, R., Kervestin, S., Mangus, D.A., Ghosh, S., and Jacobson, A. (2004). A faux 3'-UTR promotes aberrant termination and triggers nonsense-mediated mRNA decay. Nature **432**: 112–118.

Anastasaki, C., Longman, D., Capper, A., Patton, E.E., and Cáceres, J.F. (2011). Dhx34 and Nbas function in the NMD pathway and are required for embryonic development in zebrafish. Nucleic Acids Res **39**: 3686–3694.

Anders, K.R., Grimson, A., and Anderson, P. (2003). SMG-5, required for C.elegans nonsense-mediated mRNA decay, associates with SMG-2 and protein phosphatase 2A. EMBO J **22**: 641–650.

Anders, S. and Huber, W. (2010). Differential expression analysis for sequence count data. Genome Biol **11**: R106.

Arciga-Reyes, L., Wootton, L., Kieffer, M., and Davies, B. (2006). UPF1 is required for nonsense-mediated mRNA decay (NMD) and RNAi in Arabidopsis. The Plant Journal **47**: 480–489.

Aronoff, R., Baran, R., and Hodgkin, J. (2001). Molecular identification of smg-4, required for mRNA surveillance in C. elegans. Gene **268**: 153–164.

Ashton, N. and D Cove (1977). The isolation and preliminary characterisation of auxotrophic and analogue resistant mutants of the moss, Physcomitrella patens. Mol Gen Genet **154**: 87–95.

Azzalin, C.M. and Lingner, J. (2006). The human RNA surveillance factor UPF1 is required for S phase progression and genome stability. Curr Biol **16**: 433–439.

Barberan-Soler, S., Lambert, N.J., and Zahler, A.M. (2009). Global analysis of alternative splicing uncovers developmental regulation of nonsense-mediated decay in C. elegans. RNA **15**: 1652–1660.

Behm-Ansmant, I., Gatfield, D., Rehwinkel, J., Hilgers, V., and Izaurralde, E. (2007a). A conserved role for cytoplasmic poly(A)-binding protein 1 (PABPC1) in nonsense-mediated mRNA decay. EMBO J **26**: 1591–1601.

Behm-Ansmant, I., Kashima, I., Rehwinkel, J., Saulière, J., Wittkopp, N., and Izaurralde, E. (2007b). mRNA quality control: an ancient machinery recognizes and degrades mRNAs with nonsense codons. FEBS Lett **581**: 2845–2853.

Belgrader, P., Cheng, J., Zhou, X., Stephenson, L.S., and Maquat, L.E. (1994). Mammalian nonsense codons can be cis effectors of nuclear mRNA half-life. Mol Cell Biol **14**: 8219–8228.

Belostotsky, D.A. and Sieburth, L.E. (2009). Kill the messenger: mRNA decay and plant development. Curr Opin Plant Biol **12**: 96–102.

Benkovics, A.H., Nyikó, T., Mérai, Z., Silhavy, D., and Bisztray, G.D. (2011). Functional analysis of the grapevine paralogs of the SMG7 NMD factor using a heterolog VIGS-based gene depletion-complementation system. Plant Mol Biol **75**: 277–290.

Bhuvanagiri, M., Schlitter, A.M., Hentze, M.W., and Kulozik, A.E. (2010). NMD: RNA biology meets human genetic medicine. Biochem J **430**: 365–377.

Bodi, Z., Zhong, S., Mehra, S., Song, J., Graham, N., Li, H., May, S., and Fray, R.G. (2012). Adenosine Methylation in Arabidopsis mRNA is Associated with the 3' End and Reduced Levels Cause Developmental Defects. Front Plant Sci **3**: 48.

Boothby, T.C., Zipper, R.S., van der Weele, C.M., and Wolniak, S.M. (2013). Removal of retained introns regulates translation in the rapidly developing gametophyte of Marsilea vestita. Dev Cell **24**: 517–529.

Boutz, P.L., Stoilov, P., Li, Q., Lin, C.-H., Chawla, G., Ostrow, K., Shiue, L., Ares, M., and Black, D.L. (2007). A post-transcriptional regulatory switch in polypyrimidine tract-binding proteins reprograms alternative splicing in developing neurons. Genes Dev **21**: 1636–1652.

Brkljacic, J. et al. (2011). Brachypodium as a model for the grasses: today and the future. Plant Physiol **157**: 3–13.

Brocks, J.J., Logan, G.A., Buick, R., and Summons, R.E. (1999). Archean molecular fossils and the early rise of eukaryotes. Science **285**: 1033–1036.

Brogna, S. and Wen, J. (2009). Nonsense-mediated mRNA decay (NMD) mechanisms. Nat Struct Mol Biol **16**: 107–113.

Brumbaugh, K.M., Otterness, D.M., Geisen, C., Oliveira, V., Brognard, J., Li, X., Lejeune, F., Tibbetts, R.S., Maquat, L.E., and Abraham, R.T. (2004). The mRNA surveillance protein hSMG-1 functions in genotoxic stress response pathways in mammalian cells. Mol Cell **14**: 585–598.

Bruno, I.G., Karam, R., Huang, L., Bhardwaj, A., Lou, C.H., Shum, E.Y., Song, H.-W., Corbett, M.A., Gifford, W.D., Gecz, J., Pfaff, S.L., and Wilkinson, M.F. (2011). Identification of a microRNA that activates gene expression by repressing nonsense-mediated RNA decay. Mol Cell **42**: 500–510.

Buchwald, G., Ebert, J., Basquin, C., Sauliere, J., Jayachandran, U., Bono, F., Le Hir, H., and Conti, E. (2010). Insights into the recruitment of the NMD machinery from the crystal structure of a core EJC-UPF3b complex. Proc Natl Acad Sci USA.

Bühler, M., Steiner, S., Mohn, F., Paillusson, A., and Mühlemann, O. (2006).

EJC-independent degradation of nonsense immunoglobulin-mu mRNA depends on 3' UTR length. Nat Struct Mol Biol **13**: 462–464.

Cali, B.M., Kuchma, S.L., Latham, J., and Anderson, P. (1999). smg-7 is required for mRNA surveillance in Caenorhabditis elegans. Genetics **151**: 605–616.

Carter, M.S., Li, S., and Wilkinson, M.F. (1996). A splicing-dependent regulatory mechanism that detects translation signals. EMBO J **15**: 5965–5975.

Causier, B. and Davies, B. (2002). Analysing protein-protein interactions with the yeast two-hybrid system. Plant Mol Biol **50**: 855–870.

Causier, B., Ashworth, M., Guo, W., and Davies, B. (2012). The TOPLESS interactome: a framework for gene repression in Arabidopsis. Plant Physiol **158**: 423–438.

Causier, B., Schwarz-Sommer, Z., and Davies, B. (2010). Floral organ identity: 20 years of ABCs. Semin Cell Dev Biol **21**: 73–79.

Cavalier-Smith, T. (1991). Intron phylogeny: a new hypothesis. Trends Genet **7**: 145–148.

Cech, T.R. (1986). The generality of self-splicing RNA: relationship to nuclear mRNA splicing. Cell **44**: 207–210.

Chamieh, H., Ballut, L., Bonneau, F., and Le Hir, H. (2008). NMD factors UPF2 and UPF3 bridge UPF1 to the exon junction complex and stimulate its RNA helicase activity. Nat Struct Mol Biol **15**: 85–93.

Chan, W.-K., Huang, L., Gudikote, J.P., Chang, Y.-F., Imam, J.S., MacLean, J.A., and Wilkinson, M.F. (2007). An alternative branch of the nonsense-mediated decay pathway. EMBO J **26**: 1820–1830.

Chankova, S.G., Dimova, E., Dimitrova, M., and Bryant, P.E. (2007). Induction of DNA double-strand breaks by zeocin in Chlamydomonas reinhardtii and the role of increased DNA double-strand breaks rejoining in the formation of an adaptive response. Radiat Environ Biophys **46**: 409–416.

Chen, X. (2009). Small RNAs and their roles in plant development. Annu. Rev. Cell Dev. Biol. **25**: 21–44.

Chen, Y.-H., Su, L.-H., and Sun, C.-H. (2008). Incomplete nonsense-mediated mRNA decay in Giardia lamblia. Int J Parasitol **38**: 1305–1317.

Chen, Z., Smith, K.R., Batterham, P., and Robin, C. (2005). Smg1 nonsense mutations do not abolish nonsense-mediated mRNA decay in Drosophila melanogaster. Genetics **171**: 403–406.

Cheng, J. and Maquat, L.E. (1993). Nonsense codons can reduce the

abundance of nuclear mRNA without affecting the abundance of pre-mRNA or the half-life of cytoplasmic mRNA. Mol Cell Biol **13**: 1892–1902.

Chiu, S.-Y., Serin, G., Ohara, O., and Maquat, L.E. (2003). Characterization of human Smg5/7a: a protein with similarities to Caenorhabditis elegans SMG5 and SMG7 that functions in the dephosphorylation of Upf1. RNA **9**: 77–87.

Cho, H., Han, S., Choe, J., Park, S.G., Choi, S.S., and Kim, Y.K. (2012). SMG5-PNRC2 is functionally dominant compared with SMG5-SMG7 in mammalian nonsense-mediated mRNA decay. Nucleic Acids Res.

Cho, H., Kim, K.M., and Kim, Y.K. (2009). Human proline-rich nuclear receptor coregulatory protein 2 mediates an interaction between mRNA surveillance machinery and decapping complex. Mol Cell **33**: 75–86.

Coller, J.M., Gray, N.K., and Wickens, M.P. (1998). mRNA stabilization by poly(A) binding protein is independent of poly(A) and requires translation. Genes Dev **12**: 3226–3235.

Conti, E. and Izaurralde, E. (2005). Nonsense-mediated mRNA decay: molecular insights and mechanistic variations across species. Curr Opin Cell Biol **17**: 316–325.

Culbertson, M.R., Underbrink, K.M., and Fink, G.R. (1980). Frameshift suppression Saccharomyces cerevisiae. II. Genetic properties of group II suppressors. Genetics **95**: 833–853.

Culligan, K., Tissier, A., and Britt, A. (2004). ATR regulates a G2-phase cell-cycle checkpoint in Arabidopsis thaliana. Plant Cell **16**: 1091–1104.

De Turris, V., Nicholson, P., Orozco, R.Z., Singer, R.H., and Muhlemann, O. (2011). Cotranscriptional effect of a premature termination codon revealed by live-cell imaging. RNA: 1–14.

Delhi, P., Queiroz, R., Inchaustegui, D., Carrington, M., and Clayton, C. (2011). Is there a classical nonsense-mediated decay pathway in trypanosomes? PLoS ONE **6**: e25112.

Deprost, D., Yao, L., Sormani, R., Moreau, M., Leterreux, G., Nicolaï, M., Bedu, M., Robaglia, C., and Meyer, C. (2007). The Arabidopsis TOR kinase links plant growth, yield, stress resistance and mRNA translation. EMBO Rep **8**: 864–870.

Derelle, R. and Lang, B.F. (2012). Rooting the Eukaryotic Tree with Mitochondrial and Bacterial Proteins. Mol Biol Evol **29**: 1277–1289.

Drechsel, G., Kahles, A., Kesarwani, A.K., Stauffer, E., Behr, J., Drewe, P., Rätsch, G., and Wachter, A. (2013). Nonsense-mediated decay of alternative precursor mRNA splicing variants is a major determinant of the Arabidopsis steady state transcriptome. THE PLANT CELL ONLINE **25**: 3726–3742.

Durand, S. and Lykke-Andersen, J. (2013). Nonsense-mediated mRNA decay occurs during eIF4F-dependent translation in human cells. Nat Struct Mol Biol.

Eberle, A.B., Lykke-Andersen, S., Mühlemann, O., and Jensen, T.H. (2009). SMG6 promotes endonucleolytic cleavage of nonsense mRNA in human cells. Nat Struct Mol Biol **16**: 49–55.

Farlow, A., Meduri, E., Dolezal, M., Hua, L., and Schlötterer, C. (2010). Nonsense-mediated decay enables intron gain in Drosophila. PLoS Genet **6**: e1000819.

Ferraiuolo, M.A., Lee, C.-S., Ler, L.W., Hsu, J.L., Costa-Mattioli, M., Luo, M.-J., Reed, R., and Sonenberg, N. (2004). A nuclear translation-like factor eIF4AIII is recruited to the mRNA during splicing and functions in nonsense-mediated decay. Proc Natl Acad Sci USA **101**: 4118–4123.

Frizzell, K.A., Rynearson, S.G., and Metzstein, M.M. (2012). Drosophila mutants show NMD pathway activity is reduced, but not eliminated, in the absence of Smg6. RNA **18**: 1475–1486.

Fukuhara, N., Ebert, J., Unterholzner, L., Lindner, D., Izaurralde, E., and Conti, E. (2005). SMG7 is a 14-3-3-like adaptor in the nonsense-mediated mRNA decay pathway. Mol Cell **17**: 537–547.

Gaba, A., Jacobson, A., and Sachs, M.S. (2005). Ribosome occupancy of the yeast CPA1 upstream open reading frame termination codon modulates nonsense-mediated mRNA decay. Mol Cell **20**: 449–460.

Gallie, D.R., Le, H., Caldwell, C., Tanguay, R.L., Hoang, N.X., and Browning, K.S. (1997). The phosphorylation state of translation initiation factors is regulated developmentally and following heat shock in wheat. J Biol Chem **272**: 1046–1053.

Gardner, L.B. (2008). Hypoxic inhibition of nonsense-mediated RNA decay regulates gene expression and the integrated stress response. Mol Cell Biol **28**: 3729–3741.

Gatfield, D., Unterholzner, L., Ciccarelli, F.D., Bork, P., and Izaurralde, E. (2003). Nonsense-mediated mRNA decay in Drosophila: at the intersection of the yeast and mammalian pathways. EMBO J **22**: 3960–3970.

Gehen, S.C., Staversky, R.J., Bambara, R.A., Keng, P.C., and O'Reilly, M.A. (2008). hSMG-1 and ATM sequentially and independently regulate the G1 checkpoint during oxidative stress. Oncogene **27**: 4065–4074.

Gehring, N.H., Kunz, J.B., Neu-Yilik, G., Breit, S., Viegas, M.H., Hentze, M.W., and Kulozik, A.E. (2005). Exon-junction complex components specify distinct routes of nonsense-mediated mRNA decay with differential cofactor requirements. Mol Cell **20**: 65–75.

Gehring, N.H., Lamprinaki, S., Hentze, M.W., and Kulozik, A.E. (2009). The

hierarchy of exon-junction complex assembly by the spliceosome explains key features of mammalian nonsense-mediated mRNA decay. PLoS Biol **7**: e1000120.

Gehring, N.H., Neu-Yilik, G., Schell, T., Hentze, M.W., and Kulozik, A.E. (2003). Y14 and hUpf3b form an NMD-activating complex. Mol Cell **11**: 939–949.

Gerats, T. and Vandenbussche, M. (2005). A model system for comparative research: Petunia. Trends Plant Sci **10**: 251–256.

González, C.I., Ruiz-Echevarría, M.J., Vasudevan, S., Henry, M.F., and Peltz, S.W. (2000). The yeast hnRNP-like protein Hrp1/Nab4 marks a transcript for nonsense-mediated mRNA decay. Mol Cell **5**: 489–499.

Gozalbo, D. and Hohmann, S. (1990). Nonsense suppressors partially revert the decrease of the mRNA level of a nonsense mutant allele in yeast. Curr Genet **17**: 77–79.

Grimson, A., O'Connor, S., Newman, C.L., and Anderson, P. (2004). SMG-1 is a phosphatidylinositol kinase-related protein kinase required for nonsense-mediated mRNA Decay in Caenorhabditis elegans. Mol Cell Biol **24**: 7483–7490.

Guan, Q., Zheng, W., Tang, S., Liu, X., Zinkel, R.A., Tsui, K.-W., Yandell, B.S., and Culbertson, M.R. (2006). Impact of nonsense-mediated mRNA decay on the global expression profile of budding yeast. PLoS Genet **2**: e203.

Gupta, D. and Tuteja, N. (2011). Chaperones and foldases in endoplasmic reticulum stress signaling in plants. Plant Signal Behav **6**: 232–236.

Hayden, C.A. and Jorgensen, R.A. (2007). Identification of novel conserved peptide uORF homology groups in Arabidopsis and rice reveals ancient eukaryotic origin of select groups and preferential association with transcription factor-encoding genes. BMC Biol **5**: 32.

He, F., Brown, A.H., and Jacobson, A. (1997). Upf1p, Nmd2p, and Upf3p are interacting components of the yeast nonsense-mediated mRNA decay pathway. Mol Cell Biol **17**: 1580–1594.

He, F., Li, X., Spatrick, P., Casillo, R., Dong, S., and Jacobson, A. (2003). Genome-wide analysis of mRNAs regulated by the nonsense-mediated and 5“ to 3” mRNA decay pathways in yeast. Mol Cell **12**: 1439–1452.

Hickey, S.F., Sridhar, M., Westermann, A.J., Qin, Q., Vijayendra, P., Liou, G., and Hammond, M.C. (2012). Transgene regulation in plants by alternative splicing of a suicide exon. Nucleic Acids Res **40**: 4701–4710.

Hodgkin, J., Papp, A., Pulak, R., Ambros, V., and Anderson, P. (1989). A new kind of informational suppression in the nematode Caenorhabditis elegans. Genetics **123**: 301–313.

Hogg, J.R. and Goff, S.P. (2010). Upf1 Senses 3'UTR Length to Potentiate mRNA Decay. Cell **143**: 379–389.

Hosoda, N., Kim, Y.K., Lejeune, F., and Maquat, L.E. (2005). CBP80 promotes interaction of Upf1 with Upf2 during nonsense-mediated mRNA decay in mammalian cells. Nat Struct Mol Biol **12**: 893–901.

Hu, T.T. et al. (2011). The Arabidopsis lyrata genome sequence and the basis of rapid genome size change. Nat Genet **43**: 476–481.

Huang, L., Lou, C.-H., Chan, W., Shum, E.Y., Shao, A., Stone, E., Karam, R., Song, H.-W., and Wilkinson, M.F. (2011). RNA homeostasis governed by cell type-specific and branched feedback loops acting on NMD. Mol Cell **43**: 950–961.

Huntzinger, E. and Izaurralde, E. (2011). Gene silencing by microRNAs: contributions of translational repression and mRNA decay. Nat Rev Genet **12**: 99–110.

Huntzinger, E., Kashima, I., Fauser, M., Saulière, J., and Izaurralde, E. (2008). SMG6 is the catalytic endonuclease that cleaves mRNAs containing nonsense codons in metazoan. RNA **14**: 2609–2617.

Hwang, J., Sato, H., Tang, Y., Matsuda, D., and Maquat, L.E. (2010). UPF1 Association with the Cap-Binding Protein, CBP80, Promotes Nonsense-Mediated mRNA Decay at Two Distinct Steps. Mol Cell.

Izumi, N., Yamashita, A., and Ohno, S. (2012). Integrated regulation of PIKK-mediated stress responses by AAA+ proteins RUVBL1 and RUVBL2. Nucleus **3**: 29–43.

Izumi, N., Yamashita, A., Iwamatsu, A., Kurata, R., Nakamura, H., Saari, B., Hirano, H., Anderson, P., and Ohno, S. (2010). AAA+ proteins RUVBL1 and RUVBL2 coordinate PIKK activity and function in nonsense-mediated mRNA decay. Science Signaling **3**: ra27.

Jaillon, O. et al. (2008). Translational control of intron splicing in eukaryotes. Nature **451**: 359–362.

Jeong, H.-J., Kim, Y.J., Kim, S.H., Kim, Y.-H., Lee, I.-J., Kim, Y.K., and Shin, J.S. (2011). Nonsense-Mediated mRNA Decay Factors, UPF1 and UPF3, Contribute to Plant Defense. Plant Cell Physiol **52**: 2147–2156.

Jonas, S., Weichenrieder, O., and Izaurralde, E. (2013). An unusual arrangement of two 14-3-3-like domains in the SMG5-SMG7 heterodimer is required for efficient nonsense-mediated mRNA decay. Genes Dev **27**: 211–225.

Jorgensen, R.A. and Dorantes-Acosta, A.E. (2012). Conserved Peptide Upstream Open Reading Frames are Associated with Regulatory Genes in Angiosperms. Front Plant Sci **3**: 191.

Kalyna, M. et al. (2012). Alternative splicing and nonsense-mediated decay modulate expression of important regulatory genes in Arabidopsis. Nucleic Acids Res **40**: 2454–2469.

Kamauchi, S., Nakatani, H., Nakano, C., and Urade, R. (2005). Gene expression in response to endoplasmic reticulum stress in Arabidopsis thaliana. FEBS J **272**: 3461–3476.

Kamisugi, Y., Cuming, A.C., and Cove, D.J. (2005). Parameters determining the efficiency of gene targeting in the moss Physcomitrella patens. Nucleic Acids Res **33**: e173.

Kamisugi, Y., Schaefer, D.G., Kozak, J., Charlot, F., Vrielynck, N., Holá, M., Angelis, K.J., Cuming, A.C., and Nogué, F. (2011). MRE11 and RAD50, but not NBS1, are essential for gene targeting in the moss Physcomitrella patens. Nucleic Acids Res.

Kashima, I., Yamashita, A., Izumi, N., Kataoka, N., Morishita, R., Hoshino, S., Ohno, M., Dreyfuss, G., and Ohno, S. (2006). Binding of a novel SMG-1-Upf1-eRF1-eRF3 complex (SURF) to the exon junction complex triggers Upf1 phosphorylation and nonsense-mediated mRNA decay. Genes Dev **20**: 355–367.

Kenrick, P. and Crane, P. (1997). The origin and early evolution of plants on land. Nature: 1–7.

Kerényi, Z., Mérai, Z., Hiripi, L., Benkovics, A., Gyula, P., Lacomme, C., Barta, E., Nagy, F., and Silhavy, D. (2008). Inter-kingdom conservation of mechanism of nonsense-mediated mRNA decay. EMBO J **27**: 1585–1595.

Kertész, S., Kerényi, Z., Mérai, Z., Bartos, I., Pálfy, T., Barta, E., and Silhavy, D. (2006). Both introns and long 3'-UTRs operate as cis-acting elements to trigger nonsense-mediated decay in plants. Nucleic Acids Res **34**: 6147–6157.

Khajavi, M., Inoue, K., and Lupski, J.R. (2006). Nonsense-mediated mRNA decay modulates clinical outcome of genetic disease. Eur J Hum Genet **14**: 1074–1081.

Khraiwesh, B., Arif, M.A., Seumel, G.I., Ossowski, S., Weigel, D., Reski, R., and Frank, W. (2010). Transcriptional control of gene expression by microRNAs. Cell **140**: 111–122.

Kim, S.H., Koroleva, O.A., Lewandowska, D., Pendle, A.F., Clark, G.P., Simpson, C.G., Shaw, P.J., and Brown, J.W.S. (2009). Aberrant mRNA transcripts and the nonsense-mediated decay proteins UPF2 and UPF3 are enriched in the Arabidopsis nucleolus. Plant Cell **21**: 2045–2057.

Kohlen, W., Charnikhova, T., Liu, Q., Bours, R., Domagalska, M.A., Beguerie, S., Verstappen, F., Leyser, O., Bouwmeester, H., and Ruyter-Spira, C. (2011). Strigolactones are transported through the xylem and play a key role in shoot architectural response to phosphate deficiency in

nonarbuscular mycorrhizal host Arabidopsis. Plant Physiol **155**: 974–987.

Koonin, E.V. (2006). The origin of introns and their role in eukaryogenesis: a compromise solution to the introns-early versus introns-late debate? Biol Direct **1**: 22.

Koornneef, M. and Meinke, D. (2010). The development of Arabidopsis as a model plant. The Plant Journal **61**: 909–921.

Krogan, N.T. and Long, J.A. (2009). Why so repressed? Turning off transcription during plant growth and development. Curr Opin Plant Biol **12**: 628–636.

Lageix, S., Lanet, E., Pouch-Pélissier, M.-N., Espagnol, M.-C., Robaglia, C., Deragon, J.-M., and Pélissier, T. (2008). Arabidopsis eIF2alpha kinase GCN2 is essential for growth in stress conditions and is activated by wounding. BMC Plant Biol **8**: 134.

Lai, T., Cho, H., Liu, Z., Bowler, M.W., Piao, S., Parker, R., Kim, Y.K., and Song, H. (2012). Structural Basis of the PNRC2-Mediated Link between mRNA Surveillance and Decapping. Structure **20**: 2025–2037.

Lareau, L.F., Inada, M., Green, R.E., Wengrod, J.C., and Brenner, S.E. (2007). Unproductive splicing of SR genes associated with highly conserved and ultraconserved DNA elements. Nature **446**: 926–929.

Leeds, P., Wood, J.M., Lee, B.S., and Culbertson, M.R. (1992). Gene products that promote mRNA turnover in Saccharomyces cerevisiae. Mol Cell Biol **12**: 2165–2177.

Lempiäinen, H. and Halazonetis, T.D. (2009). Emerging common themes in regulation of PIKKs and PI3Ks. EMBO J **28**: 3067–3073.

Lewis, B.P., Green, R.E., and Brenner, S.E. (2003). Evidence for the widespread coupling of alternative splicing and nonsense-mediated mRNA decay in humans. Proc Natl Acad Sci USA **100**: 189.

Liu, J., Sun, N., Liu, M., Liu, J., Du, B., Wang, X., and Qi, X. (2013). An autoregulatory loop controlling Arabidopsis HsfA2 expression: role of heat shock-induced alternative splicing. Plant Physiol **162**: 512–521.

Liu, J.-J., Sturrock, R., and Ekramoddoullah, A.K.M. (2010). The superfamily of thaumatin-like proteins: its origin, evolution, and expression towards biological function. Plant Cell Rep **29**: 419–436.

Longman, D., Hug, N., Keith, M., Anastasaki, C., Patton, E.E., Grimes, G., and Cáceres, J.F. (2013). DHX34 and NBAS form part of an autoregulatory NMD circuit that regulates endogenous RNA targets in human cells, zebrafish and Caenorhabditis elegans. Nucleic Acids Res.

Longman, D., Plasterk, R.H.A., Johnstone, I.L., and Cáceres, J.F. (2007). Mechanistic insights and identification of two novel factors in the C. elegans

NMD pathway. Genes Dev **21**: 1075–1085.

Losson, R. and Lacroute, F. (1979). Interference of nonsense mutations with eukaryotic messenger RNA stability. Proc Natl Acad Sci USA **76**: 5134–5137.

Loughran, G., Sachs, M.S., Atkins, J.F., and Ivanov, I.P. (2012). Stringency of start codon selection modulates autoregulation of translation initiation factor eIF5. Nucleic Acids Res **40**: 2898–2906.

Luke, B., Azzalin, C.M., Hug, N., Deplazes, A., Peter, M., and Lingner, J. (2007). Saccharomyces cerevisiae Ebs1p is a putative ortholog of human Smg7 and promotes nonsense-mediated mRNA decay. Nucleic Acids Res **35**: 7688–7697.

Maderazo, A.B., He, F., Mangus, D.A., and Jacobson, A. (2000). Upf1p control of nonsense mRNA translation is regulated by Nmd2p and Upf3p. Mol Cell Biol **20**: 4591–4603.

Mangus, D.A., Evans, M.C., and Jacobson, A. (2003). Poly(A)-binding proteins: multifunctional scaffolds for the post-transcriptional control of gene expression. Genome Biol **4**: 223.

Maquat, L.E. (2004). Nonsense-mediated mRNA decay: splicing, translation and mRNP dynamics. Nat Rev Mol Cell Biol **5**: 89–99.

Martin, W. and Koonin, E.V. (2006). Introns and the origin of nucleus-cytosol compartmentalization. Nature **440**: 41–45.

McIlwain, D.R., Pan, Q., Reilly, P.T., Elia, A.J., McCracken, S., Wakeham, A.C., Itie-Youten, A., Blencowe, B.J., and Mak, T.W. (2010). Smg1 is required for embryogenesis and regulates diverse genes via alternative splicing coupled to nonsense-mediated mRNA decay. Proc Natl Acad Sci USA.

Meaux, S., van Hoof, A., and Baker, K.E. (2008). Nonsense-mediated mRNA decay in yeast does not require PAB1 or a poly(A) tail. Mol Cell **29**: 134–140.

Mendell, J.T., Sharifi, N.A., Meyers, J.L., Martinez-Murillo, F., and Dietz, H.C. (2004). Nonsense surveillance regulates expression of diverse classes of mammalian transcripts and mutes genomic noise. Nat Genet **36**: 1073–1078.

Metzstein, M.M. and Krasnow, M.A. (2006). Functions of the nonsense-mediated mRNA decay pathway in Drosophila development. PLoS Genet **2**: e180.

Mérai, Z., Benkovics, A.H., Nyikó, T., Debreczeny, M., Hiripi, L., Kerényi, Z., Kondorosi, E., and Silhavy, D. (2012). The late steps of plant nonsense-mediated mRNA decay. Plant J.

Mokry, M., Nijman, I.J., van Dijken, A., Benjamins, R., Heidstra, R., Scheres, B., and Cuppen, E. (2011). Identification of factors required for meristem function in Arabidopsis using a novel next generation sequencing fast forward genetics approach. BMC Genomics **12**: 256.

Mortazavi, A., Williams, B.A., McCue, K., Schaeffer, L., and Wold, B. (2008). Mapping and quantifying mammalian transcriptomes by RNA-Seq. Nat Methods **5**: 621–628.

Mühlemann, O., Eberle, A.B., Stalder, L., and Zamudio Orozco, R. (2008). Recognition and elimination of nonsense mRNA. Biochim Biophys Acta **1779**: 538–549.

Nicholson, P. and Mühlemann, O. (2010). Cutting the nonsense: the degradation of PTC-containing mRNAs. Biochem Soc Trans **38**: 1615–1620.

Nicholson, P., Yepiskoposyan, H., Metze, S., Zamudio Orozco, R., Kleinschmidt, N., and Mühlemann, O. (2010). Nonsense-mediated mRNA decay in human cells: mechanistic insights, functions beyond quality control and the double-life of NMD factors. Cell Mol Life Sci **67**: 677–700.

Nishiyama, T., Hiwatashi, Y., Sakakibara, I., Kato, M., and Hasebe, M. (2000). Tagged mutagenesis and gene-trap in the moss, Physcomitrella patens by shuttle mutagenesis. DNA Res **7**: 9–17.

Niu, D.-K. and Cao, J.-L. (2010). Nucleosome deposition and DNA methylation may participate in the recognition of premature termination codon in nonsense-mediated mRNA decay. FEBS Lett.

Nyikó, T., Kerényi, F., Szabadkai, L., Benkovics, A.H., Major, P., Sonkoly, B., Mérai, Z., Barta, E., Niemiec, E., Kufel, J., and Silhavy, D. (2013). Plant nonsense-mediated mRNA decay is controlled by different autoregulatory circuits and can be induced by an EJC-like complex. Nucleic Acids Res.

Nyikó, T., Sonkoly, B., Mérai, Z., Benkovics, A.H., and Silhavy, D. (2009). Plant upstream ORFs can trigger nonsense-mediated mRNA decay in a size-dependent manner. Plant Mol Biol **71**: 367–378.

Okada-Katsuhata, Y., Yamashita, A., Kutsuzawa, K., Izumi, N., Hirahara, F., and Ohno, S. (2011). N- and C-terminal Upf1 phosphorylations create binding platforms for SMG-6 and SMG-5:SMG-7 during NMD. Nucleic Acids Res.

Opik, M., Vanatoa, A., Vanatoa, E., Moora, M., Davison, J., Kalwij, J.M., Reier, U., and Zobel, M. (2010). The online database MaarjAM reveals global and ecosystemic distribution patterns in arbuscular mycorrhizal fungi (Glomeromycota). New Phytol **188**: 223–241.

Page, M.F., Carr, B., Anders, K.R., Grimson, A., and Anderson, P. (1999). SMG-2 is a phosphorylated protein required for mRNA surveillance in

Caenorhabditis elegans and related to Upf1p of yeast. Mol Cell Biol **19**: 5943–5951.

Pajerowska-Mukhtar, K.M., Wang, W., Tada, Y., Oka, N., Tucker, C.L., Fonseca, J.P., and Dong, X. (2012). The HSF-like transcription factor TBF1 is a major molecular switch for plant growth-to-defense transition. Curr Biol **22**: 103–112.

Palusa, S.G. and Reddy, A.S.N. (2010). Extensive coupling of alternative splicing of pre-mRNAs of serine/arginine (SR) genes with nonsense-mediated decay. New Phytologist **185**: 83–89.

Palusa, S.G., Ali, G.S., and Reddy, A.S. (2007). Alternative splicing of pre-mRNAs of Arabidopsis serine/arginine-rich proteins: regulation by hormones and stresses. The Plant Journal **49**: 1091–1107.

Park, J.H. et al. (2011). Heat-induced chaperone activity of serine/threonine protein phosphatase 5 enhances thermotolerance in Arabidopsis thaliana. New Phytol **191**: 692–705.

Peltz, S.W., Brown, A.H., and Jacobson, A. (1993). mRNA destabilization triggered by premature translational termination depends on at least three cis-acting sequence elements and one trans-acting factor. Genes Dev **7**: 1737–1754.

Persak, H. and Pitzschke, A. (2013). Tight interconnection and multi-level control of Arabidopsis MYB44 in MAPK cascade signalling. PLoS ONE **8**: e57547.

Poole, A.M. (2006). Did group II intron proliferation in an endosymbiont-bearing archaeon create eukaryotes? Biol Direct **1**: 36.

Prasad, B.D., Goel, S., and Krishna, P. (2010). In silico identification of carboxylate clamp type tetratricopeptide repeat proteins in Arabidopsis and rice as putative co-chaperones of Hsp90/Hsp70. PLoS ONE **5**: e12761.

Prigge, M.J. and Bezanilla, M. (2010). Evolutionary crossroads in developmental biology: Physcomitrella patens. Development **137**: 3535–3543.

Pulak, R. and Anderson, P. (1993). mRNA surveillance by the Caenorhabditis elegans smg genes. Genes Dev **7**: 1885–1897.

Ramani, A.K., Nelson, A.C., Kapranov, P., Bell, I., Gingeras, T.R., and Fraser, A.G. (2009). High resolution transcriptome maps for wild-type and nonsense-mediated decay-defective Caenorhabditis elegans. Genome Biol **10**: R101.

Rayson, S., Arciga-Reyes, L., Wootton, L., De Torres Zabala, M., Truman, W., Graham, N., Grant, M., and Davies, B. (2012a). A role for nonsense-mediated mRNA decay in plants: pathogen responses are induced in Arabidopsis thaliana NMD mutants. PLoS ONE **7**: e31917.

Rayson, S., Ashworth, M., De Torres Zabala, M., Grant, M., and Davies, B. (2012b). The salicylic acid dependent and independent effects of NMD in plants. Plant Signal Behav **7**.

Redon, S., Reichenbach, P., and Lingner, J. (2007). Protein RNA and protein protein interactions mediate association of human EST1A/SMG6 with telomerase. Nucleic Acids Res **35**: 7011–7022.

Rehwinkel, J., Letunic, I., Raes, J., Bork, P., and Izaurralde, E. (2005). Nonsense-mediated mRNA decay factors act in concert to regulate common mRNA targets. RNA **11**: 1530–1544.

Rensing, S.A. et al. (2008). The Physcomitrella genome reveals evolutionary insights into the conquest of land by plants. Science **319**: 64–69.

Rensing, S.A., Ick, J., Fawcett, J.A., Lang, D., Zimmer, A., Van de Peer, Y., and Reski, R. (2007). An ancient genome duplication contributed to the abundance of metabolic genes in the moss Physcomitrella patens. BMC Evol Biol **7**: 130.

Richards, T.A. and Cavalier-Smith, T. (2005). Myosin domain evolution and the primary divergence of eukaryotes. Nature **436**: 1113–1118.

Riehs, N., Akimcheva, S., Puizina, J., Bulankova, P., Idol, R.A., Siroky, J., Schleiffer, A., Schweizer, D., Shippen, D.E., and Riha, K. (2008). Arabidopsis SMG7 protein is required for exit from meiosis. J Cell Sci **121**: 2208–2216.

Riehs-Kearnan, N., Gloggnitzer, J., Dekrout, B., Jonak, C., and Riha, K. (2012). Aberrant growth and lethality of Arabidopsis deficient in nonsense-mediated RNA decay factors is caused by autoimmune-like response. Nucleic Acids Res.

Robinson, M.D., McCarthy, D.J., and Smyth, G.K. (2010). edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. Bioinformatics **26**: 139–140.

Rosains, J. and Mango, S.E. (2012). Genetic characterization of smg-8 mutants reveals no role in C. elegans nonsense mediated decay. PLoS ONE **7**: e49490.

Roy, S.W. and Gilbert, W. (2006). The evolution of spliceosomal introns: patterns, puzzles and progress. Nat Rev Genet **7**: 211–221.

Rufener, S.C. and Mühlemann, O. (2013). eIF4E-bound mRNPs are substrates for nonsense-mediated mRNA decay in mammalian cells. Nat Struct Mol Biol.

Ryu, J.S. et al. (2005). Phytochrome-specific type 5 phosphatase controls light signal flux by enhancing phytochrome stability and affinity for a signal transducer. Cell **120**: 395–406.

Sakaki, K., Yoshina, S., Shen, X., Han, J., Desantis, M.R., Xiong, M., Mitani, S., and Kaufman, R.J. (2012). RNA surveillance is required for endoplasmic reticulum homeostasis. Proc Natl Acad Sci USA.

Saltzman, A.L., Pan, Q., and Blencowe, B.J. (2011). Regulation of alternative splicing by the core spliceosomal machinery. Genes Dev **25**: 373–384.

Saul, H. et al. (2009). The upstream open reading frame of the Arabidopsis AtMHX gene has a strong impact on transcript accumulation through the nonsense-mediated mRNA decay pathway. Plant J **60**: 1031–1042.

Saulière, J., Haque, N., Harms, S., Barbosa, I., Blanchette, M., and Le Hir, H. (2010). The exon junction complex differentially marks spliced junctions. Nat Struct Mol Biol.

Sawicka, K., Bushell, M., Spriggs, K.A., and Willis, A.E. (2008). Polypyrimidine-tract-binding protein: a multifunctional RNA-binding protein. Biochem Soc Trans **36**: 641–647.

Sayani, S., Janis, M., Lee, C.Y., Toesca, I., and Chanfreau, G.F. (2008). Widespread impact of nonsense-mediated mRNA decay on the yeast intronome. Mol Cell **31**: 360–370.

Schaefer, D., Zryd, J.P., Knight, C.D., and Cove, D.J. (1991). Stable transformation of the moss Physcomitrella patens. Mol Gen Genet **226**: 418–424.

Schneeberger, K. and Weigel, D. (2011). Fast-forward genetics enabled by new sequencing technologies. Trends Plant Sci **16**: 282–288.

Schoning, J.C., Streitner, C., Meyer, I.M., Gao, Y., and Staiger, D. (2008). Reciprocal regulation of glycine-rich RNA-binding proteins via an interlocked feedback loop coupling alternative splicing to nonsense-mediated decay in Arabidopsis. Nucleic Acids Res **36**: 6977–6987.

Serin, G., Gersappe, A., Black, J.D., Aronoff, R., and Maquat, L.E. (2001). Identification and characterization of human orthologues to Saccharomyces cerevisiae Upf2 protein and Upf3 protein (Caenorhabditis elegans SMG-4). Mol Cell Biol **21**: 209–223.

Simón, E. and Séraphin, B. (2007). A specific role for the C-terminal region of the Poly(A)-binding protein in mRNA decay. Nucleic Acids Res **35**: 6017–6028.

Singh, G., Rebbapragada, I., and Lykke-Andersen, J. (2008). A competition between stimulators and antagonists of Upf complex recruitment governs human nonsense-mediated mRNA decay. PLoS Biol **6**: e111.

Stauffer, E., Westermann, A., Wagner, G., and Wachter, A. (2010). Polypyrimidine tract-binding protein homologues from Arabidopsis underlie regulatory circuits based on alternative splicing and downstream control. Plant J **64**: 243–255.

Sugio, A., Dreos, R., Aparicio, F., and Maule, A.J. (2009). The cytosolic protein response as a subcomponent of the wider heat shock response in Arabidopsis. Plant Cell **21**: 642–654.

Sureau, A., Gattoni, R., Dooghe, Y., Stévenin, J., and Soret, J. (2001). SC35 autoregulates its expression by promoting splicing events that destabilize its mRNAs. EMBO J **20**: 1785.

Tao, J. and Sha, B. (2011). Structural insight into the protective role of P58(IPK) during unfolded protein response. Methods in Enzymology **490**: 259–270.

Tarazona, S., García-Alcalde, F., Dopazo, J., Ferrer, A., and Conesa, A. (2011). Differential expression in RNA-seq: a matter of depth. Genome Res **21**: 2213–2223.

Templeton, G.W. and Moorhead, G.B.G. (2005). The phosphoinositide-3-OH-kinase-related kinases of Arabidopsis thaliana. EMBO Rep **6**: 723–728.

Trapnell, C., Pachter, L., and Salzberg, S.L. (2009). TopHat: discovering splice junctions with RNA-Seq. Bioinformatics **25**: 1105–1111.

Trcek, T., Sato, H., Singer, R.H., and Maquat, L.E. (2013). Temporal and spatial characterization of nonsense-mediated mRNA decay. Genes Dev.

Tsai, A.Y.-L. and Gazzarrini, S. (2012). AKIN10 and FUSCA3 interact to control lateral organ development and phase transitions in Arabidopsis. Plant J **69**: 809–821.

Tsuboi, T. and Inada, T. (2010). Tethering of poly(A)-binding protein interferes with non-translated mRNA decay from the 5' end in yeast. J Biol Chem **285**: 33589–33601.

Unterholzner, L. and Izaurralde, E. (2004). SMG7 acts as a molecular link between mRNA surveillance and mRNA decay. Mol Cell **16**: 587–596.

Urlaub, G., Mitchell, P.J., Ciudad, C.J., and Chasin, L.A. (1989). Nonsense mutations in the dihydrofolate reductase gene affect RNA processing. Mol Cell Biol **9**: 2868–2880.

van Bentem, S., Vossen, J.H., de Vries, K.J., van Wees, S., Tameling, W., Dekker, H.L., de Koster, C.G., Haring, M.A., Takken, F., and Cornelissen, B. (2005). Heat shock protein 90 and its co-chaperone protein phosphatase 5 interact with distinct regions of the tomato I-2 disease resistance protein. Plant J **43**: 284–298.

Wagner, L.E., Li, W.-H., Joseph, S.K., and Yule, D.I. (2004). Functional consequences of phosphomimetic mutations at key cAMP-dependent protein kinase phosphorylation sites in the type 1 inositol 1,4,5-trisphosphate receptor. J Biol Chem **279**: 46242–46252.

Wang, D., Zavadil, J., Martin, L., Parisi, F., Friedman, E., Levy, D., Harding,

H., Ron, D., and Gardner, L.B. (2011). Inhibition of nonsense-mediated RNA decay by the tumor microenvironment promotes tumorigenesis. Mol Cell Biol **31**: 3670–3680.

Wang, W., Cajigas, I.J., Peltz, S.W., Wilkinson, M.F., and González, C.I. (2006). Role for Upf2p phosphorylation in Saccharomyces cerevisiae nonsense-mediated mRNA decay. Mol Cell Biol **26**: 3390–3400.

Waterworth, W.M., Drury, G.E., Bray, C.M., and West, C.E. (2011). Repairing breaks in the plant genome: the importance of keeping it together. New Phytol **192**: 805–822.

Weischenfeldt, J., Waage, J., Tian, G., Zhao, J., Damgaard, I., Jakobsen, J.S., Kristiansen, K., Krogh, A., Wang, J., and Porse, B.T. (2012). Mammalian tissues defective in nonsense-mediated mRNA decay display highly aberrant splicing patterns. Genome Biol **13**: R35.

Wen, J. and Brogna, S. (2010). Splicing-dependent NMD does not require the EJC in Schizosaccharomyces pombe. EMBO J.

Wittkopp, N., Huntzinger, E., Weiler, C., Saulière, J., Schmidt, S., Sonawane, M., and Izaurralde, E. (2009). Nonsense-mediated mRNA decay effectors are essential for zebrafish embryonic development and survival. Mol Cell Biol **29**: 3517–3528.

Wollerton, M.C., Gooding, C., Wagner, E.J., Garcia-Blanco, M.A., and Smith, C.W.J. (2004). Autoregulation of polypyrimidine tract binding protein by alternative splicing leading to nonsense-mediated decay. Mol Cell **13**: 91–100.

Yamashita, A. (2013). Role of SMG-1-mediated Upf1 phosphorylation in mammalian nonsense-mediated mRNA decay. Genes Cells **18**: 161–175.

Yamashita, A. et al. (2009). SMG-8 and SMG-9, two novel subunits of the SMG-1 complex, regulate remodeling of the mRNA surveillance complex during nonsense-mediated mRNA decay. Genes Dev **23**: 1091–1105.

Yamashita, A., Ohnishi, T., Kashima, I., Taya, Y., and Ohno, S. (2001). Human SMG-1, a novel phosphatidylinositol 3-kinase-related protein kinase, associates with components of the mRNA surveillance complex and is involved in the regulation of nonsense-mediated mRNA decay. Genes Dev **15**: 2215–2228.

Yepiskoposyan, H., Aeschimann, F., Nilsson, D., Okoniewski, M., and Mühlemann, O. (2011). Autoregulation of the nonsense-mediated mRNA decay pathway in human cells. RNA **17**: 2108–2118.

Yoine, M., Nishii, T., and Nakamura, K. (2006). Arabidopsis UPF1 RNA helicase for nonsense-mediated mRNA decay is involved in seed size control and is essential for growth. Plant Cell Physiol **47**: 572–580.

Yoshiyama, K., Conklin, P.A., Huefner, N.D., and Britt, A.B. (2009).

Suppressor of gamma response 1 (SOG1) encodes a putative transcription factor governing multiple responses to DNA damage. Proc Natl Acad Sci USA **106**: 12843–12848.

Yoshiyama, K.O., Kobayashi, J., Ogita, N., Ueda, M., Kimura, S., Maki, H., and Umeda, M. (2013). ATM-mediated phosphorylation of SOG1 is essential for the DNA damage response in Arabidopsis. EMBO Rep **14**: 817–822.

Zhang, J. and Maquat, L.E. (1997). Evidence that translation reinitiation abrogates nonsense-mediated mRNA decay in mammalian cells. EMBO J **16**: 826–833.

Zhang, S., Ruiz-Echevarria, M.J., Quan, Y., and Peltz, S.W. (1995). Identification and characterization of a sequence motif involved in nonsense-mediated mRNA decay. Mol Cell Biol **15**: 2231–2244.

Zhang, Z.M., Quick, M.K., Kanelakis, K.C., Gijzen, M., and Krishna, P. (2003). Characterization of a plant homolog of Hop, a cochaperone of hsp90. Plant Physiol **131**: 525–535.

Zimmer, A.D., Lang, D., Buchta, K., Rombauts, S., Nishiyama, T., Hasebe, M., Van de Peer, Y., Rensing, S.A., and Reski, R. (2013). Reannotation and extended community resources for the genome of the non-seed plant Physcomitrella patens provide insights into the evolution of plant gene structures and functions. BMC Genomics **14**: 498.

Zünd, D., Gruber, A.R., Zavolan, M., and Mühlemann, O. (2013). Translation-dependent displacement of UPF1 from coding sequences causes its enrichment in 3' UTRs. Nat Struct Mol Biol.