The nonsense-mediated decay pathway of *Arabidopsis thaliana* and *Vitis vinifera* studied by a new gene depletion-complementation assay

Anna Hangyáné Benkovics

Budapest, 2011

PhD School
Name: Doctoral School of Horticultural Sciences

Field: Crop Sciences and Horticulture

Head of Ph.D. School: Prof. Dr. Magdolna Tóth
Doctor of the Hungarian Academy of Sciences
Head of Department of Fruit Sciences
CORVINUS UNIVERSITY OF BUDAPEST,
Faculty of Horticultural Sciences

Supervisors: **Dr György Dénes Bisztray**
professor, chair
Corvinus University of Budapest
Faculty of Horticultural Sciences
Institute of Viticulture and Oenology
Department of Viticulture

Dr Dániel Silhavy
senior scientist, group leader
Agricultural Biotechnology Center, Gödöllő
Plant RNA Biology Group

The applicant met the requirement of the PhD regulations of the Corvinus University of Budapest and the thesis is accepted for the defence process.

.....

....
Head of Ph.D. School Supervisors

Table of contents

I. Introduction

Grapevine is a plant of great economical importance grown on 7 337 364 hectars in 2008 according to FAO (Food and Agriculture Organization). Grapevine is one of the most important plants grown on mediterranean areas and the economically most important fruit plant of the world.

As a consequence of the global climate change, the weather of grape growing areas is becoming more and more extreme. For this reason, the usage of stress resistant grapevine varieties is becoming unavoidable. Weather of most grapevine growing areas has already been significantly changed during the past half century. However, currently water deficiency is the most important limiting factor of grapevine growth. Even in areas where the amount of the annual precipitation is high, the unequal distribution of rainfall throughout the year can results in water deficit in the growing period of grapevine. For this reason, more and more grapevine growing areas need irrigation. However, the growing population has an increasing water demand, which will results in decrease of grapevine growing areas which need to be irrigated. For this reason the need for the use of drought stress resistant grapevine varieties is increasing.

Drought resistance is a complex trait, requiring very efficient functioning of complicated regulation pathways. During drought periods, plants have to reprogram their gene expression, the gene products contributing to stress survival have to replace the previously expressed transcrips. In this process, not only the change of the transcriptional patterns is important, but also several posttranscriptional mechanisms. Changes of mRNA splicing, translational intensity and mRNA degradation pathways all have an important role in the reprogramming of gene expression. The different mRNA degradation pathways are especially important during the adaptation to changed environment. Stress responsive mRNA decay systems are able to quickly and selectively remove the transcipts coding for proteins unnecessary or harmful in the new environmental conditions, making it possible that the new, adaptive proteins quickly replace the previous factors. For this reason, knowledge of plant mRNA degradation pathways is a prerequisite of the understanding of the molecular biology bases of plant stress responses. However, only very little is known about plant mRNA decay pathways, despite their fundamental importance.

Hence the long-term aim of our program is to gain more insight into the mRNA degradation pathways playing a role in grapevine stress responses. As a first step, we started the investigation of one of the essential mRNA decay systems of grapevine, nonsense-

mediated decay (NMD). NMD plays an important role in reprogramming gene expression in response to different stress conditions in both animals and yeast, and presumably also in plants. During our program we have functionally identified the grapevine paralogs of the essential core NMD factor, SMG7, to gain a better knowledge of the molecular stress response pathways of grapevine.

I. 1. The nonsense-mediated decay pathway

Different mRNAs can have substantially different half lives, which is determined partially by specific *cis* elements in the mRNA sequence and *trans* factors recognizing these *cis* elements and attracting different mRNA degradation enzymes to the transcript (Narsai *et al.,* 2007). In this way special *cis* elements make mRNA degradation pathways target specific: stability of mRNAs having a given *cis* element can be increased or decreased through the trans factors by which it is recognized.

One of the target specific degradation pathways of plants is the nonsense-mediated mRNA decay (NMD) pathway, which recognizes and degrades mRNAs containing premature translation termination codons (PTCs), which could give rise to C-terminally truncated proteins. PTC containing mRNAs can originate from nonsense mutated genes, or wild-type genes as a result of aberrant transcription or alternative splicing. NMD is a conserved pathway working with different efficiency and somewhat different function, but with conserved core factors and similar mechanisms in all eukaryotic organisms: in unicellular organisms, budding and fission yeast, *Caenorhabditis elegans*, *Drosophila melanogaster*, fish, mammals and plants (Stalder and Muhlemann 2008).

Plant NMD recognizes mRNAs having either an unusually long 3' UTR (longer than ~300 nucleotides), or an intron at least ~50 nucleotides downstream of the stop codon as targets. This means that the *cis* elements marking the specific group of NMD target mRNAs are long 3' UTRs and 3' UTRs introns (Hori and Watanabe 2007, Kertesz *et al.,* 2006, Schwartz *et al.,* 2006). According to the recognized *cis* elements, NMD events can be sorted into two different categories: 'long 3' UTR-based' and 'intron-based' NMD. Long 3' UTR-based NMD is the dominant NMD mechanism in invertebrates and budding yeast, but it also exists in mammals. Intron-based NMD is the predominant NMD mechanism in mammals, but it also exists in *D. melanogaster* and fission yeast.

Interestingly, in addition to aberrant transcripts and mRNAs transcribed from nonsense mutated genes, wild-type mRNAs can also have NMD-inducing features. Several normal transcripts have an unusually long 3' UTR or a 3' UTR intron. For instance, almost 40% of grapevine genes have a 3' UTR longer than 300 nucleotides. In Arabidopsis, transcript

profiling experiments have shown that expression of ~0.5% of its protein coding genes is elevated in NMD deficient protoplasts (Kurihara *et al.,* 2009).

In conclusion, NMD plays a double function in plants: in addition to degrading aberrant transcripts, it also regulates the expression of several wild-type genes.

II. Aims

Knowledge of the selective mRNA degradation systems of grapevine and the function of genes playing a role in these pathways can enhance our understanding of the molecular biological basis of grapevine stress responses. However, grapevine is not a convenient target of functional genomic studies. For this reason, a prerequisite of our studies was the establishment of a transient *Nicotiana benthamiana* gene depletion-complementation system, which can be used to easily and quickly test the function of different grapevine genes.

Our further aim was to study the grapevine NMD pathway using the previously established transient *N. benthamiana* system.

Specific aims of my thesis:

- As a first step, we wanted to identify the potential grapevine orthologs of the known plant NMD and mRNA degradation factors using bioinformatic methods.
- Next we wanted to establish a transient gene depletion-complementation system in *N. benthamiana,* in which different endogeneous *N. benthamiana* genes can be depleted and then the compromised function can be complemented by the addition of a heterolog (for instance, grapevine) gene.
- Then – using this system - we wanted to test whether the bioinformatically identified grapevine ortohologs of the NMD core factors SMG7 are able to complement SMG7-deficiency in SMG7-silenced *N. benthamiana* leaves.
- We wanted to functionally map the protein domains of the grapevine SMG7 homologs.
- We wanted to test whether the SMG7 homolog SMG7L has a role in plant NMD.
- Finally, we wanted to test, whether the grapevine SMG7 homologs are theirselves NMD targets, similarly to the Arabidopsis SMG7.

III. Materials and methods

III. 1. *N. benthamiana* plants

For agroinfiltration, we have used ~3 weeks old *N. benthamiana* plants, grown under long-day conditions (16 hours day / 8 hours night) on 22 °C.

III. 2. Molecular clonings

For the agroinfiltrations, genes were cloned into Bin61S binary vector or into the derivatives of Bin61S between the 35S promoter and terminator of *cauliflower mosaic virus*. The different *N. benthamiana, Arabidopsis thaliana* and *Vitis vinifera* sequences were amplified from cDNA synthesized from total RNA extract, using KOD Hot Start DNA Polymerase (EMD Chemicals).

III. 3. Agroinfiltration

Optical density (OD_{600}) of agrobacterium cultures was measured using spectrophotometer, then they were diluted by a MES buffer containing 0,01 M $MgCl_2$ and acetosyringon to the final concentration, which was always 0.2 for P14 and 0.4 for other constructs. GFP expression was studied using a 100 W hand-hold UV lamp (UV products, Upland, CA 91786, Black Ray model B 100AP) 3 days after agroinfiltration.

III. 4. Virus-induced gene silencing (VIGS)

To initiate VIGS, ~21 days old plants were agroinfiltrated with a mixture of three different agrobacterium cultures. One contained the P14 silencing suppressor construct, the second the TRV RNA1 sequence (BINTRA6 vector) (Ratcliff *et al.,* 2001), and the third the TRV-PDS vector containing the test gene to be silenced. 2-2 middle-aged leaves of plants were infiltrated in patches of ~1 cm^2 size. Reporter and test constructs, as well as P14 were agroinfiltrated into silenced leaves 10-14 days post infiltration (dpi).

III. 5. RNA extractions and Northern blots

RNA was extracted from ~1 cm^2 patches using the described method (Szittya *et al.,* 2002). Samples were run on a 1.5% agarose-MAE gel containing formaldehyde, then they were

transferred to nitrocellulose membrane (Hybond-N, Amersham Biosciences) with capillary blotting method. The blots were hybridized with P^{32} cytosin marked P14 and GFP probes. Intensity of bands was measured using PhosphorImager equipment. GFP band intensities were normalized to P14 band intensities.

III. 6. Tethering experiments

As a tethering reporter, we used the GFP-3'boxB (G-3'bB) construct, which contains 5 boxB sequences in its 3' UTR. We cloned the genes to be tested as λN-fusion constructs.

III. 7. Sequence searches

SMG7 homologs were searched in the Phytozome and the NCBI GenBank databases (www.phytozome.org; www.ncbi.nlm.nih.gov) using the tblastn program.

IV. Results

IV. 1. Grapevine homologs of plant mRNA degradation factors

As a first step towards the better understanding of the role mRNA degradation systems play in grapevine stress responses, we have bioinformatically identified the grapevine orthologs of the known plant general mRNA degradation and NMD factors.

The core NMD factors UPF1, UPF2 and UPF3 have one homolog in Arabidopsis (Arciga-Reyes *et al.,* 2006, Kerenyi *et al.,* 2008, Yoine *et al.,* 2006, Yoine *et al.,* 2006). Similarly, we have found one homolog of UPF1 and UPF2 in grapevine, however, we have found three grapevine paralogs of UPF3 (Table 1.). The SMG5-7 gene family, members of which play an essential role in NMD in both plants and animals, has two or three members in the animal model organisms (*C. elegans*, *Drosophila*, *Danio rerio*, human). Similarly, Arabidopsis has two SMG7 homologs, SMG7T and SMG7L. In grapevine, we have found two SMG7T paralogs and one SMG7L copy.

Several other NMD- and general mRNA degradation factors have more than one copies in grapevine (Table 1.). This fact has to be taken into consideration during the investigation of NMD and mRNA decay pathways in grapevine. During the functional testing of these factors it has to be tested which of the paralogs have retained and which have lost the original function of the gene.

The aim of our work was the functional identification of the three grapevine homologs of SMG7, because we have previously shown that this gene has an essential function in both long 3' UTR- and intron-based plant NMD, but its exact function is not known. In addition, the mammalisan SMG7 factors plays a role in both the PTC recognition and the mRNA degradation steps of the NMD pathway, and the three grapevine SMG7 paralogs may have divided this function between them. As a continuation of our work, our gene depletion-complementation system can be used for the identification of further NMD- and mRNA decay factors of grapevine.

Arabidopsis homolog	Gene name	Grapevine homolog
At5g47010	**UPF1**	XM_002279268.1
At2g39260	**UPF2**	XM_002275610.1
At1g33980	**UPF3**	XR_077501.1; XR_077493.1; XM_002264240.1
At5g19400; At1g28260	**SMG7**	XM_002276153.1; XM_002272651.1; XM_002274224.1
-	**SMG-1**	XM_002281236.1
At3g48190	**ATM**	XM_002279366.1
At5g40820	**ATR**	XM_002278373.1
At1g50030	**TOR**	XM_002275578.1
At1g51510	**Y14**	XM_002281192
At1g02140	**MAGO**	XM_002281258.1; XM_002272217.1
At3g19760	**eIF4AIII**	XM_002274975.1
At1g80000; At1g15280	**Barentz**	XM_02280036.1
At5g47880 / At1g12920 / At3g26618	**eRF1**	XM_002282991.1; XM_002271620.1; XM_002271421.1
At1g18070	**eRF3**	XM_002280946.1
At1g08370	**DCP1**	XM_002273389.1; XM_002281741.1
At5g13570	**DCP2**	XM_002274322.1
At3g13300	**VCS**	XM_002269539.1; XM_002269528.1
At1g77680	**SOV**	XM_002284383.1
At1g54490	**XRN4**	XM_002271951.1; XM_002275524.1
At3g44260; At5g22250; At1g27820; At1g27890; At1g61470; At3g44240; At1g06450; At1g15920; At5g10960; At1g80780; At2g32070	**CAF1**	XM_002271600.1; XM_002281102.1; XM_002271393.1; XM_002285266.1; XM_002285265.1; XM_002271432.1; XM_002272129.1; XM_002272173.1; XM_002279205.1
At3g58580; At3g58560; At3g18500; At1g02270	**CCR4**	XM_002280954.1; XM_002284124.1; XM_002275494.1
At1g55870	**PARN**	XM_002266033.1
At2g17510; At1g77680; At1g54440	**RRP6**	XM_002276903.1 / XM_002276873.1; XM_002284383.1
At3g61620	**RRP41**	XM_002280266.1
At2g17510	**RRP44**	XM_002276903.1 / XM_002276873.1

Table 1. Arabidopsis and grapevine homologs of plant NMD and mRNA decay factors. *Left column*: TAIR database ID of the experimentally or bioinformatically identified Arabidopsis orthologs of mRNA decay factors. *Middle column*: gene name of plant mRNA decay factors. *Right column*: RefSeq database ID of bioinformatically identified grapevine mRNA decay gene orthologs. Purple color indicates genes with two, blue with three, and green with 3< grapevine homologs. IDs separated by '/' are alternative transcripts of the same gene.

IV. 2 Experimental systems to study the plant NMD pathway

IV. 2. 1. *N. benthamiana* gene depletion-complementation system

As a first step towards the functional identification of grapevine NMD factors, we have established a *N. benthamiana* virus-induced gene silencing (VIGS)-based transient gene depletion-complementation system, which can be used for the quick functional testing of heterolog genes.

In this experimental system, we first silence the *N. benthamiana* ortholog of the gene to be functionally identified in another species (for instance, SMG7 of grapevine). This is done by inserting a 600-700 nucleotides long sequence of the gene to be silenced into a vector containing the genome of *Tobacco rattle virus* (TRV), together with a sequence from the phytoene desaturase (PDS) gene (TRV-PDS-SMG7, Figure 1B). This vector is then conjugated into the C58C1 agrobacterium strain. In *N. benthamiana* leaves infiltrated with a culture of this agrobacterium the TRV virus replicates through a double-stranded RNA intermedier, which is recognized and cut by the silencing system of the plant. One strand of these short double-stranded RNAs is incorporated into RNA-induced silencing complexes (RISC) and inhibits expression of homolog transcripts in the cell. In this way the endogeneous SMG7 and PDS genes are silenced together with the TRV virus. The PDS gene plays a role in carotenoid biosynthesis and its silencing results in photobleaching (Figure 1.). The short double-stranded RNA sequences get into neighbouring cells through plasmodesmata, causing systemic silencing in leaves not infected by the virus. Photobleaching of systemic leaves in our system occurs 10-14 days after agroinfiltration. Then silenced leaves are agroinfiltrated by the Gc-I NMD target reporter construct, the P14 silencing suppressor and the test construct (for example, grapevine SMG7). Gc-I is a strong target of NMD, which contains a 200 nucleotides long stuffer sequence from the coding region of the PHA gene (‘c’) between the GFP coding region and the 35S terminator, which makes its 3’ UTR longer than 300 nucleotides, and an intron from the LS gene, which induces NMD (Figure 1.D). The P14 silencing suppressor is a double-stranded RNA binding protein, which makes the expression of co-infiltrated genes significantly stronger by inhibiting agroinfiltration-induced gene silencing. If co-infiltrated with only P14, Gc-I has strong green fluorescence and high mRNA level, as the expression of the *N. benthamiana* endogeneous SMG7 is low and NMD efficiency is strongly decreased. However, if Gc-I is co-infiltrated with both P14 and the Arabidopsis SMG7 gene, its

expression is weak, as the heterolog protein complements SMG7-deficiency and restores NMD (Figure 1.C). Using this system the ability of different grapevine SMG7 homologs to complement SMG7-deficiency and restore NMD in *N. benthamiana* can be tested, thus revealing the function of the individual grapevine SMG7 paralogs. (Figure 1.C).

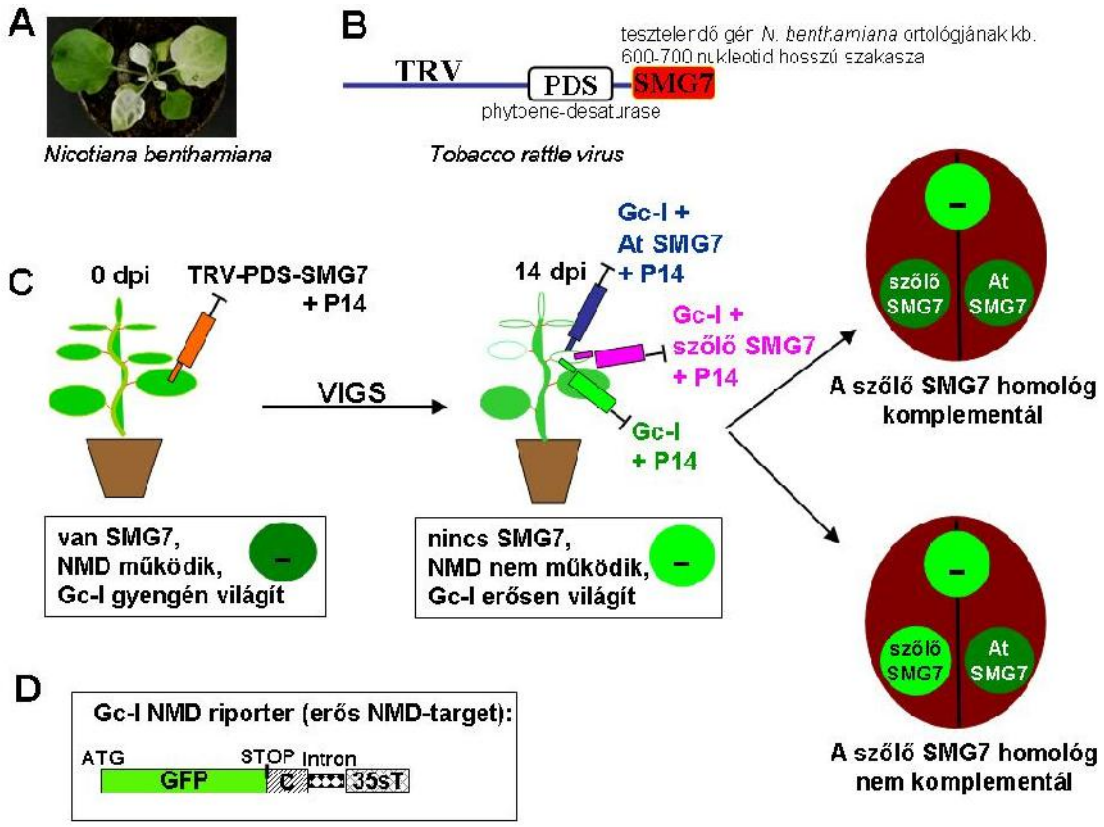

Figure 1. ***N. benthamiana*** **transient gene depletion-complementation system. A** PDS-silenced *N. benthamiana* plant. **B** The TRV-PDS-SMG7 construct. **C** Gene depletion and complementation in *N. benthamiana*. **D** Gc-I NMD target reporter construct. dpi: *days post infiltration,* 35sT: 35S terminator.

IV. 2. 2. *N. benthamiana* tethering system to test the function of plant NMD factors

SMG7 plays a double role in mammalisan NMD: first it recognizes and binds phosphorylated UPF1 and indirectly the NMD target transcript through its 14-3-3-like

phosphoserine-binding domain, then it induces mRNA decay through its C-terminal domain (Unterholzner and Izaurralde 2004).

For this reason, if an SMG7 homolog cannot complement SMG7-deficiency in *N. benthamiana*, it either has lost the whole NMD function, or one of its domains has lost its original NMD function, but the other has preserved it. For instance, if the N-terminal domain of an SMG7 homolog has lost its original function, the ability to bind the NMD target transcript, but its C-terminal domain has preserved the decay inducing function, then if this homolog is artificially bound (tethered) to the mRNA, it could be able to induce its degradation.

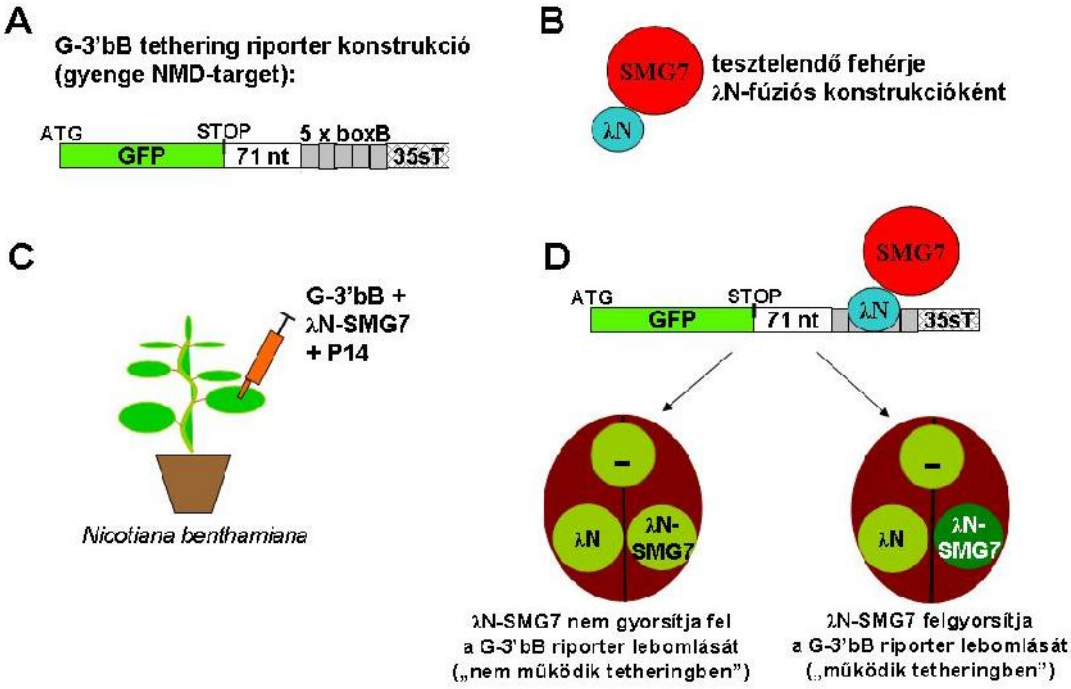

Figure 2. *N. benthamiana* tethering system. A The G-3'bB tethering reporter construct. **B** SMG7 construct fused to the λN peptid. **C** *N. benthamiana* agroinfiltration with the tethering reporter construct, the test construct and the P14 silencing suppressor. **D** A tethering experiment.

Our tethering experiments are based on the fact that the λN peptide binds strongly and specifically the secondary structure formed by the so called boxB RNA sequence. In this way, proteins fused to the λN peptide are artificially tethered to reporter constructs containing the boxB sequence (Figure 2.). In our experiments, we have used a GFP tethering reporter construct, which contains 5 boxB sequences 71 nucleotides downstream of its stop codon (G-

3'bB construct). The SMG7 gene to be tested is cloned immediately downstream of a DNA sequence coding for the λN peptide (λN-SMG7). The G-3'bB reporter construct, the λN-SMG7 test construct (or only λN as a negative control) and the P14 silencing suppressor are agroinfiltrated into *N. benthamiana* leaves. G-3'bB is a weak target of NMD, as the boxB sequences make its 3' UTR slightly longer than 300 nucleotides. For this reason, if co-infiltrated with only P14 and the negative control λN construct, its expression is of middle strenght, as it is weakly targeted by NMD. If the tested SMG7 homolog has conserved the mRNA decay inducing function, then it significantly decreases the expression level of the co-infiltrated G-3'bB construct, otherwise it does not affect G-3'bB level (Figure 2.D).

In this way, after the accomplishment of the gene depletion-complementation experiment, NMD function of the test gene can be further investigated in tethering experiment.

IV. 3. Functional analysis of grapevine SMG7 homologs

First we tested the function of the three grapevine SMG7 homologs by testing whether they can complement the NMD deficiency of SMG7-silenced *N. benthamiana* leaves.

IV. 3. 1. Gene depletion-complementation experiments

In gene depletion-complementation experiments, Gc-I was used as NMD reporter and Arabidopsis SMG7T (AtS7) as positive contol. We have found that if Gc-I was co-infiltrated with the grapevine SMG7-1 (VvS71) or the SMG7-2 (VvS72) constructs, its green fluorescence and mRNA level was significantly decreased, suggesting that both grapevine SMG7T paralogs have conserved the original NMD function. However, we have found that the SMG7L (VvS7L) construct could not complement NMD-deficiency, suggesting that it has lost the original NMD function (Figure 3.D).

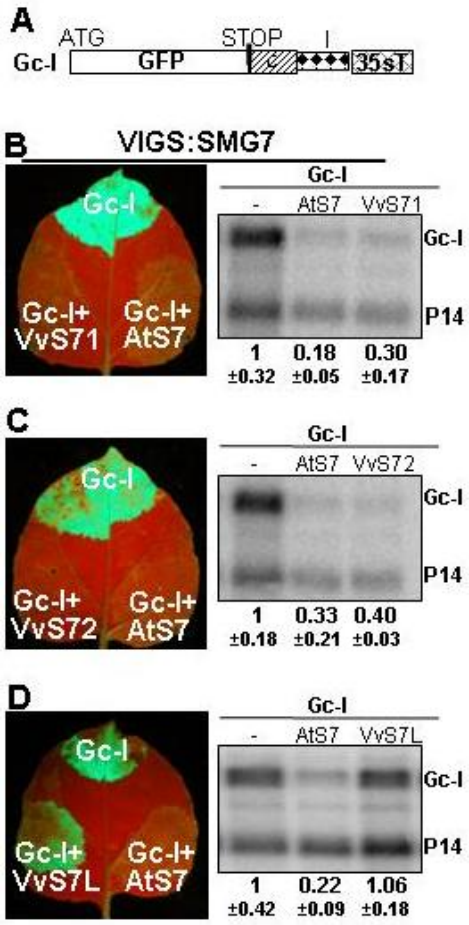

Figure 3. The grapevine SMG7-1 and SMG7-2 genes have conserved the NMD function, whilst SMG7L has lost it. A The Gc-I reporter construct. **B** SMG7-silenced *N. benthamiana* leaves were infiltrated either with Gc-I, or with Gc-I and the Arabidopsis SMG7 (AtS7) or the grapevine SMG7-1 construct (VvS71). Similarly to AtS7, VvS71 complements SMG7-deficiency, it decreases both the green fluorescence and mRNA expression of Gc-I significantly. **C** Similarly, grapevine SMG7-2 (VvS72) complements SMG7-deficiency. **D** Grapevine SMG7L (VvS7L) does not complement SMG7-deficiency.

IV. 3. 2. SMG7-1 has preserved, but SMG7L has lost the mRNA degradation function

Our gene depletion-complementation system showed that the SMG7L protein has lost the original NMD function of the SMG7 homolog proteins. However, it is possible, that its N- or C-terminal domain has preserved the original function. If the N-terminal domain has lost the original phosphoserine-binding function, bur the C-terminal domain has preserved the mRNA decay inducing function, then the SMG7L protein has lost the ability to bind the target mRNA through UPF1. However, if it is artificially bound to the mRNA in a tethering experiment, then

it should be able to induce mRNA destabilization. In the opposite case, if the N-terminal domain has preserved and the C-terminal domain has lost the original function, then the N-terminal domain of SMG7L fused to the C-terminal domain of an SMG7T homolog makes a functional SMG7 protein which is able to fulfill the NMD function.

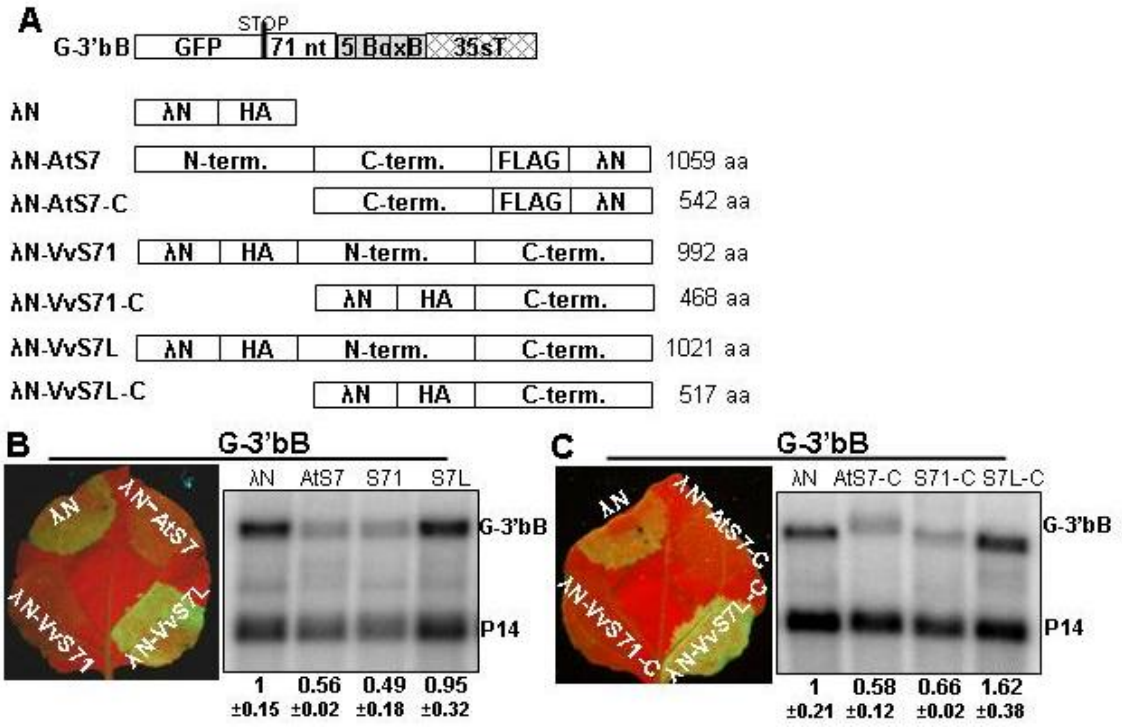

Figure 4. The grapevine SMG7-1 has preserved, whilst SMG7L has lost the mRNA decay inducing function. **A** Constructs used in tethering experiments. **B** Tethering in *N. benthamiana* plants using the full-length λN-Vv-SMG7-1 (λN-VvS71, S71) and λN-Vv-SMG7L (λN-VvS7L, S7L) constructs. The λN peptide was used as negative control, whilst the Arabidopsis SMG7-λN (λN-AtS7, AtS7) was used as positive control. The green fluorescence and mRNA level of the G-3'bB reporter was decreased by λN-VvS71 co-infiltration, similarly to λN-AtS7 co-infiltration, but it was not changed by λN-VvS7L co-infiltration. **C** Tethering experiment using the Arabidopsis SMG7-C-λN (λN-AtS7-C, AtS7-C) and the grapevine λN-Vv-SMG7-1-C (λN-VvS71-C, S71-C) and λN-Vv-SMG7L-C (λN-VvS7L-C, S7L-C) constructs. Green fluorescence and mRNA level of G-3'bB was significantly decreased by λN-AtS7-C- and λN-VvS71-C co-infiltration, but was not affected by λN-VvS7L-C.

First we wanted to test in tethering experiment whether the SMG7L protein is able to induce mRNA decay if artificially tethered to a transcript. We have found that whilst the λN-Vv-SMG7-1 construct (S71) induced decreased mRNA level and green fluorescence of G-3'bB like the Arabidopsis λN-SMG7 (AtS7), λN-Vv-SMG7L tethering did not affect G-3'bB levels (Figure 4.A,B).

Next we wanted to test whether the C-terminal domain of the SMG7 paralogs alone is able to induce mRNA decay in tethering experiment. We have found that whilst the C-terminal domain of Arabidopsis SMG7 (AtS7-C) and the grapevine SMG7-1 (S71-C) induced mRNA decay efficiently, the C-terminal domain of SMG7L did not induce destabilization of the reporter construct (Figure 4.C).

We concluded that the C-terminal domain of the plant SMG7T proteins is responsible for the induction of mRNA degradation, whilst the C-terminal domain of SMG7L has lost this function.

IV. 3. 3. The N-terminal domain of SMG7L has preserved the original function played in NMD

In the previous experiment we have found that the grapevine SMG7L gene has lost the ability of its C-terminal domain to induce mRNA decay. However, the N-terminal domain of the SMG7 homologs is more conserved, than the C-terminal domain. It is possible, that N-terminal domain of SMG7L has preserved the original NMD function, which is presumably phosphor-UPF1 binding.

To test this hypothesis, we cloned a fusion construct coding for the N-terminal domain of the grapevine SMG7L protein and the C-terminal domain of the grapevine SMG7-1 protein (S7LN-1C) and tested whether this construct is able to complement NMD deficiency in SMG7-silenced *N. benthamiana* plants (Figure 5.A). We used the grapevine SMG7-1 construct as positive control (S71). Interestingly, we have found that the VvS7LN-1C construct significantly decreased Gc-I reporter expression both on the mRNA- and protein level, although less efficiently than the SMG7-1 construct. We concluded that the N-terminal domain of SMG7L is able to fulfill the NMD function of the SMG7-1 N-terminal domain, suggesting that it may have preserved the original function, presumably phosphoserine-binding.

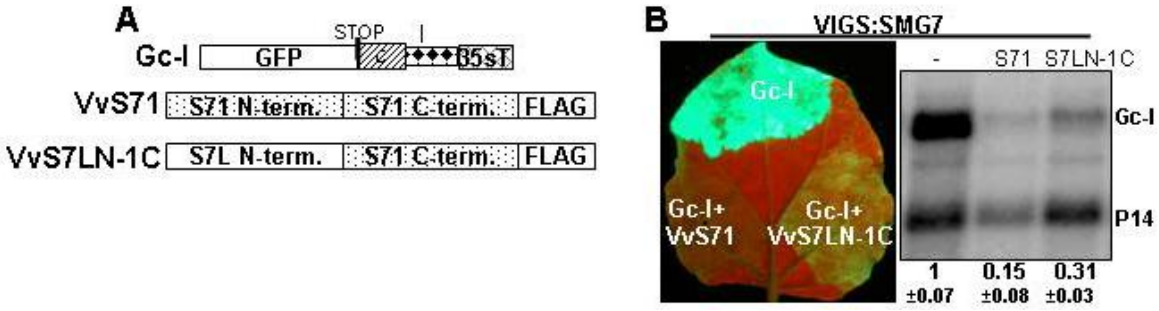

Figure 5. The N-terminal domain of SMG7L has preserved the original NMD function. A The used constructs. VvS7LN-1C consists of the N-terminal domain of grapevine SMG7L and the C-terminal domain of grapevine SMG7-1. **B** In SMG7-silenced *N. benthamiana* plants, VvS7LN-1C (S7LN-1C) co-infiltration decreases Gc-I expression significantly, although less efficiently, than that of VvS71 (S71 on the blot).

IV. 3. 4. NMD is efficient in SMG7L-silenced *N. benthamiana* plants

Although SMG7L could not complement SMG7-deficiency in *N. benthamiana* and its C-terminal domain has lost the mRNA decay inducing function of the original SMG7 factors, our results also suggest that its N-terminal domain has preserved the original phosphoserine-binding function. For this reason, it is possible that SMG7L plays a role in plant NMD.

To test this possibility, we wanted to test whether NMD efficiency is changed in SMG7L-silenced *N. benthamiana* plants compared to control PDS-silenced plants.
NMD efficiency was monitored by comparing the expression of the following GFP reporter constructs:

1.) GFP: wild-type GFP with ’normal’ 3’ UTR;
2.) Gc: weak target of NMD harbouring a 200 nt long stuffer sequence originating from the coding region of the PHA gene (’c’) in its 3’ UTR;
3.) G-L: strong target of NMD harbouring a 600 nt long stuffer sequence originating from the coding region of the PHA gene (’abc’) in its 3’ UTR;
4.) Gc-I: very strong target of NMD, which has an identical sequence to Gc, except that Gc-I contains an intron of the LS gene in the ‘c’ region (Figure 6.A).

In PDS-silenced control plants the expression of Gc is only the half, whilst that of G-L and Gc-I is only the fifth of the expression of wild-type GFP, because of efficient NMD (Figure 6.B). If SMG7L plays a role in NMD, NMD efficiency should be changed in PDS-

SMG7L-silenced plants. However, we have found that NMD efficiency in PDS-SMG7L-silenced plants was similar to PDS-silenced plants, as the ratio of NMD target and control constructs was very similar between these two types of plants (Figure 6.C). We concluded that SMG7L either does not play a role in NMD, or its function is not always essential for efficient NMD.

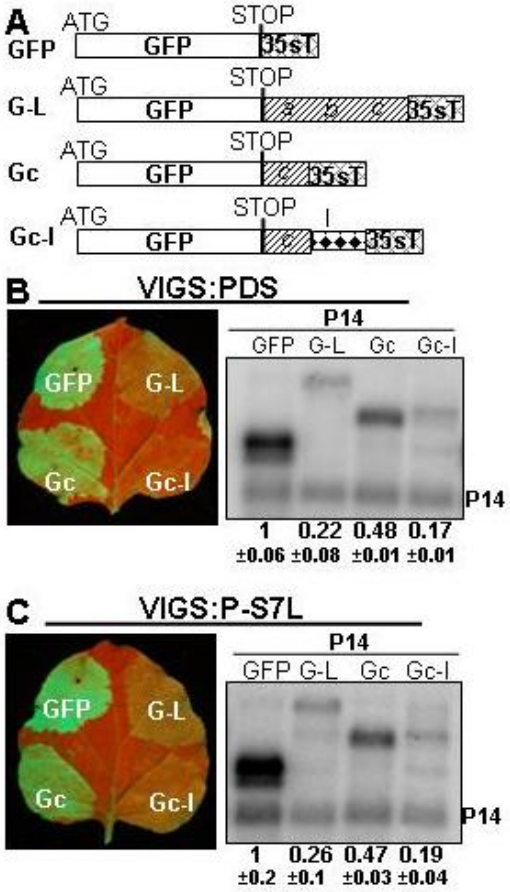

Figure 6. SMG7L does not play a role in NMD. A The used GFP reporter constructs. **B** NMD works efficiently in PDS-silenced *N. benthamiana* plants: expression of the NMD target G-L and Gc-I constructs is significantly lower than that of the control (GFP and Gc) constructs. **C** NMD works with similar efficiency in PDS-SMG7L-silenced *N. benthamiana* plants.

IV. 4. Regulation of grapevine SMG7 homologs

We have previously found that the Arabidopsis and *N. benthamiana* SMG7 genes are interestingly NMD targets themselves and both have 500-600 nucleotides long 3' UTRs and 3' UTR introns, which are rare among plant transcripts (Kerenyi *et al.,* 2008). Presumably this feature of SMG7 genes has been selevted for during evolution. For this reason, we wanted to test whether the three grapevine SMG7 homologs have preserved NMD regulation.

First we have collected the EST sequences available in the NCBI EST database for the three paralogs. We have found that the SMG7-1 and SMG7-2 genes have unusually long 3' UTRs (more than 750 and 550 nucleotides, respectively) and contain two 3' UTR introns in similar positions as the Arabidopsis SMG7 gene. In contrast, the 3' UTR of the SMG7L gene does not contain introns and it has two alternative polyadenylation sites, one 121 and the other 400 nucleotides downstream from the stop codon (Figure 7.). The terminators of SMG7-1 and SMG7-2 and the longer 3' UTR of SMG7L transcribed with the further polyadenylation site potentially trigger NMD.

To test this hypothesis experimentally, we have cloned the terminator regions of the three SMG7 paralogs downstream of the GFP coding region and agroinfiltrated these constructs with only P14 or with P14 and the UPF1DN (U1DN) construct into *N. benthamiana* leaves (UPF1DN is a dominant-negative point mutated form of the UPF1 NMD factor, which inhibits NMD if infiltrated into *N. benthamiana* leaves) (Figure 7.A). In this way we have shown that the terminator of SMG7-1 and SMG7-2 indeed trigger NMD, as if the G-VvS71-T or the G-VvS72-T constructs were infiltrated alone, their mRNA level and green fluorescence were low, while if they were infiltrated together with U1DN, their expression was significantly elevated (Figure 7.B). These results suggest that the SMG7-1 and SMG7-2 genes are directly regulated by NMD. In contrast, the G-VvS7L-T construct, which has the SMG7L terminator region cloned after the GFP coding sequence, showed strong green fluorescence and high mRNA levels if it was infiltrated alone, and U1DN co-infiltration did not affect its expression (Figure 7.B). We concluded that the terminator of SMG7L does not trigger NMD and this paralog has lost the NMD regulation together with the original NMD function.

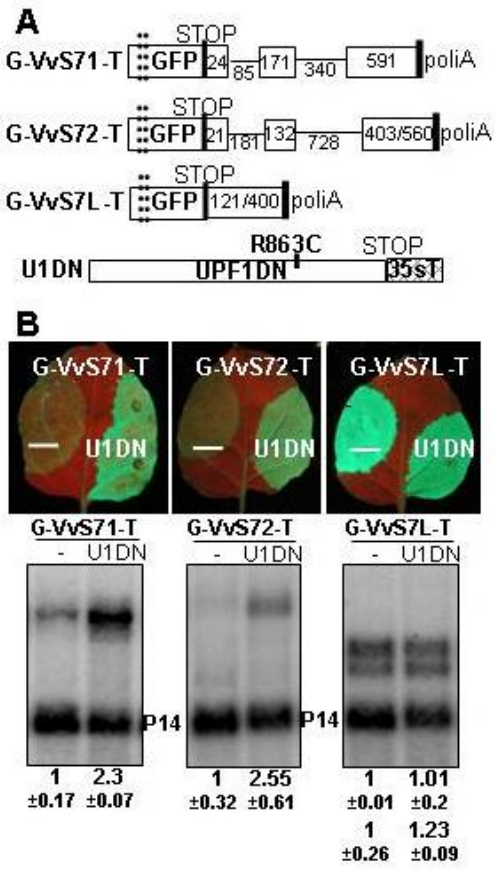

Figure 7. SMG7-1 and SMG7-2 are NMD-regulated, whilst SMG7L has lost the NMD-regulation of the original SMG7 genes. A The used constructs. Rectangles represent exons, lines represent introns. Nucleotide length of the individual sections is indicated by numbers below the lines or in the rectangles. In the case of the VvS72 and VvS7L terminators, we identified alternative polyadenylation sites. U1DN is a dominant-negative point mutant UPF1 construct, which inhibits NMD if infiltrated into *N. benthamiana* leaves. **B** *Left panels*: Fluorescence and mRNA level of G-VvS71-T is strongly enhanced by U1DN co-infiltration. *Middle panels*: Fluorescence and mRNA level of G-VvS72-T is strongly enhanced by U1DN co-infiltration. *Right panels*: fluorescence and mRNA level of G-VvS7L-T is high without U1DN and is not influenced by U1DN co-infiltration.

IV. 5. NMD targets in grapevine

IV. 5. 1. Grapevine orthologs of Arabidopsis NMD target genes

According to previous results, NMD regulates different transcripts in different organisms. However, it is possible that its targets are more overlapping between more closely

related species, such as Arabidopsis and grapevine, which both belong to the rosid clad of dicot plants.

Kurihara and coworkers have published an Arabidopsis gene list which have an increased expression in both *upf1* and *upf3* mutant plants. These genes are either direct targets of NMD or they are regulated by a direct NMD target, for example a transcription factor (Kurihara *et al.,* 2009).

We have screened the 5'- and 3' UTR regions of these genes looking for NMD *cis* elements: 3' UTRs longer than 300 nukleotides, introns at least 50 nucleotides downstream of the stop codon or uORFs. Then we selected the genes with NMD *cis* elements which have a grapevine ortholog in the Phytozome database (we have found 18 genes with these features). Then we screened the grapevine orthologs for NMD *cis* elements and have found that out of the 12 genes with available UTR data, 10 have NMD *cis* elements (Table 2.). These grapevine genes are presumably direct targets of NMD.

	Phytozome description	Arabidopsis gén potenciálisan NMD-indukáló tulajdonsága	szőlő homológ Phytozome azonosító	szőlő homológ RefSeq azonosító	Arabidopsis-szőlő hasonlóság	szőlő gén legközelebbi Arabidopsis homológja	szőlő gén potenciálisan NMD-indukáló tulajdonsága
At1g01060	LHY (LATE ELONGATED HYPOCOTYL); DNA binding / transcription factor	300 nt < 3' UTR	GSVIVT00026185001	XM_002267684.1	59,10%	At1g01060	NMD-gyanús uORF
At1g29950	transcription factor/ transcription regulator	300 nt < 3' UTR; NMD-gyanús uORF	GSVIVT00003416001	XM_002275329.1 XM_002275357.1	57,80%	At1g29950	300 nt < 3' UTR; NMD-gyanús uORF
At1g36730	eukaryotic translation initiation factor 5, putative / eIF-5, putative	400 nt < 3' UTR; NMD-gyanús uORF	GSVIVT00015309001	XM_002285789.1	69,20%	At1g36730	300 nt < 3' UTR; NMD-gyanús uORF
At1g65840	ATPAO4 (ARABIDOPSIS THALIANA POLYAMINE OXIDASE 4); amine oxidase/ polyamine oxidase	NMD-gyanús uORF	GSVIVT00026668001	XM_002264856.1	77,50%	At1g65840	NMD-gyanús uORF
At2g47490	mitochondrial substrate carrier family protein	300 nt < 3' UTR; NMD-gyanús uORF	GSVIVT00028055001	XM_002273538.1	90,10%	At2g47490	NMD-gyanús uORF
At4g19110	protein kinase, putative	NMD-gyanús uORF	GSVIVT00003204001	XM_002265323.1	79,30%	At4g19110	500 nt < 3' UTR; NMD-gyanús uORF
At4g30960	SIP3 (SOS3-INTERACTING PROTEIN 3); ATP binding / kinase/ protein kinase/ protein serine/threonine kinase	NMD-gyanús uORF	GSVIVT00013357001	XM_002262631.1	85,30%	At4g30960	400 nt < 3' UTR; NMD-gyanús uORF
At5g19400	SMG7	500 nt < 3' UTR; 50 nt < intron	GSVIVT00022273001 GSVIVT00000654001	XM_002276153.1 XM_002272651.1	74,0% 65,7%	At5g19400	50 nt < intron; 700 nt < 3' UTR
At5g47790	forkhead-associated domain-containing protein / FHA domain-containing protein	800 nt < 3' UTR	GSVIVT00004318001	XM_002265233	49,90%	At5g47790	300 nt < 3' UTR
At5g47880	ERF1-1 (EUKARYOTIC RELEASE FACTOR 1-1); translation release factor	300< 3'UTR; 50 nt < intron	GSVIVT00001852001 GSVIVT00014642001 GSVIVT00031808001	XM_002271620.1 XM_002282991.1 XM_002271421.1	94,3% 94,3% 93,1%	At5g47880	nincs adat 400 nt < 3' UTR nincs adat

Table 2. Potentially NMD target grapevine orthologs of probably NMD-regulated Arabidopsis genes. *First column*: TAIR ID of the Arabidopsis gene. *Second column*: description in the Phytozome database. *Third column*: NMD-inducing feature of the

Arabidopsis genes. *Fourth column*: Phytozome ID of the grapevine ortholog. *Fifth column*: RefSeq ID of the grapevine ortholog. *Sixth column*: similarity between the Arabidopsis and grapevine orthologs. *Seventh column*: closest Arabidopsis ortholog of the grapevine gene according to BLAST statistics. *Eights column*: potential NMD-inducing feature of the grapevine gene.

V. Discussion

We have established a *N. benthamiana* transient gene depletion-complementation system and have shown using this system that one of the core factors of plant NMD has three homologs in grapevine, two of which have retained the original NMD function and NMD regulation, and one has lost both and gained a new function during evolution, in which its N-terminal domain plays a similar function as the N-terminal domain of SMG7T in NMD.

V. 1. The *N. benthamiana* VIGS-based transient gene depletion-complementation system

The uniqueness of our experimental system is that we complement the deficiency of the gene silenced in *N. benthamiana* with an ortholog gene cloned from a different species. In this way, we combine the advantages of the *N. benthamiana* VIGS system with the advantage of being able to perform functional genomic studies of any species, as far as we are working on a conserved pathway. After the establishment of an appropriate reporter system, a broad array of grapevine genes can be functionally tested in this system, even genes essential for growth and development. The usage our system can speed up the identification pf grapevine genes playing a role in gene expression regulation pathways, which can help in the understanding of grapevine resistance to different stress conditions and the selection of stress resistant varieties.

This system can also be used for the functional mapping of plant genes: the ability of different truncated or point mutated constructs to complement their *N. benthamiana* ortholog's deficiency shows whether a given domain or aminos acid is essential to perform the function of the gene product. As far as we know, a similar transient system for the functional mapping of plant genes has not been known before.

V. 3. The role of SMG7 duplications in NMD

In plants, the ancient SMG7 gene has presumably duplicated in the common ancestor of dicot plants resulting in the ancient SMG7T and SMG7L genes, which were both stabilized

in the genome, possibly because their function diverged quickly and the new, probably NMD-independent function of SMG7L became important for the plant.

But why did selection preserve the grapevine SMG7-1 and SMG7-2 paralogs, which – according to our results – perform the same function in the plant? The ploidy level of plants changes often during evolution, resulting in several copy number changes of individual genes. However, most often the resulting copies are quickly lost during the diploidization process following the genome duplication event. However, in some cases duplication of a gene can provide a selective advantage for the plant, for example when the cells need a great amount of a given gene product, such as in the case of ribosomal or histone proteins. This is probably not the case for SMG7 proteins, as it was shown in different species and our unpublished results confirm that SMG7 overexpression inhibits NMD, probably because SMG7 relocalizes UPF1 into P-bodies (Luke *et al.,* 2007, Unterholzner and Izaurralde 2004). Indeed, duplication of proteins acting in a complex with other proteins is often harmful, as it changes the fine-tuned proportions of the interacting partners (Papp *et al.,* 2003).

Then what could be the selective advantage which preserved the SMG7-1 and SMG7-2 duplicates in the grapevine genome? According to Raser and O'Shea, there should be a selection pressure acting to decrease the noise of gene expression (Raser and O'Shea 2005). One way to decrease noise is gene duplication, as the noise in the expression of the paralogs is averaged. Another way to decrease noise suggested by both theoretical models and experiments performed with *Bacillus subtilis* is to translate relatively few proteins from a transcript (Ozbudak *et al.,* 2002). This could be the case for SMG7T gene, which are translated inefficiently and degraded after a few rounds of translation by NMD.

As SMG7T is presumably the limiting factor of NMD efficiency, it is possible that the balanced functioning of NMD requires very strict regulation of SMG7T expression to reach a balanced and stable SMG7T level in the cell. This could have lead to the conserved autoregulatory circle of SMG7T NMD-regulation and the maintenance of two SMG7T duplicates of grapevine, which help to keep noise level in SMG7T expression low.

Another possibility why NMD-regulation of the SMG7T genes could be important is that if the efficiency of plant NMD, like that of mammalian NMD, is regulated and SMG7 level can become limiting for it, then NMD regulation of SMG7 can prevent NMD from becoming hyperactive under special conditions. If NMD becomes hyperactive, it decreases the level of SMG7, which then becomes limiting for NMD and inhibits the further increase in its efficiency (Kerenyi *et al.,* 2008).

However, it is also possible that promoters of the SMG7-1 and -2 genes were subfunctionalized, the paralogs shared the spatial or temporal expression of the original SMG7

gene or they became responsive for different stress conditions, which makes both paralogs essential for grapevine. These possibilities should be tested in further experiments.

VI. New scientific results

1. We have bioinformatically identified the potential grapevine orthologs of Arabidopsis mRNA degradation factors, especially the factors of nonsense-mediated decay.

2. We have established a transient *N. benthamiana* gene depletion-complementation system, which can be used for the quick and easy functional identification of grapevine genes.

3. Using this system we have experimentally identified the grapevine orthologs of one of the core factors of plant NMD, SMG7.

4. We have shown that the SMG7L SMG7 homolog does not play a role in NMD.

5. We have functionally mapped the SMG7T and SMG7L genes and have shown that the function of their N-terminal domains is conserved, whilst that of the C-terminal domain is different, in the case of SMG7T it is the induction of NMD target mRNA decay, while the C-terminal domain of SMG7L has lost this function.

6. We have shown that the two grapevine SMG7T homologs are NMD-regulated, whilst SMG7L has lost the NMD regulation.

VII. Publications

PAPERS:

The paper forming the base of this thesis:

1. **Benkovics, A. H.** - Nyiko, T. - Merai, Z. - Silhavy, D. - 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, 277-90 **(IF. 3.978, H=0)**

Other papers in the subject of this thesis:

2. Kerenyi, Z. - Merai, Z. - Hiripi, L. - **Benkovics, A. -** Gyula, P. - Lacomme, C. - Barta, E. - Nagy, F. - Silhavy, D. (2008) Inter-kingdom conservation of mechanism of nonsense-mediated mRNA decay. *EMBO J,* 27, 1585-95 **(IF. 8.662, H=34)**
3. Nyiko, T. - Sonkoly, B. - Merai, Z. - **Benkovics, A. H. -** Silhavy, D. (2009) Plant upstream ORFs can trigger nonsense-mediated mRNA decay in a size-dependent manner. *Plant Mol Biol,* 71, 367-78 **(IF. 3.978, H=0)**

CONFERENCE ABSTRACTS in the subject of this thesis:

Hungarian conference abstracts:

1. **Benkovics, A.H. -** Silhavy, D. - Bisztray, Gy.D. (2009): Szőlőgének funkcionális vizsgálata heterológ in vivo rendszerben. 'Lippay János – Ormos Imre – Vas Károly' Tudományos Ülésszak, Budapest, Magyarország. Összefoglalók, 274. o.
2. Nyikó, T. – Sonkoly, B. – **Benkovics, A.** – Silhavy, D. (2008): 5' nem-transzlálódó régióban található nyílt leolvasási keretek (uORF) szerepe a génszabályozásban. XX. MBK Napok, Gödöllő, Magyarország, Összefoglalók, 15. o.
3. Mérai, Z.- **Benkovics, A.H.** – Nyikó, T. – Silhavy, D. (2009): A növényi nonsense-mediated decay útvonal kései lépései. XI. MBK Napok, Gödöllő, Magyarország, Összefoglalók, 13. o.

International conference abstracts:

1. Merai, Z. - Kerenyi, Z. - **Benkovics, A. -** Hiripi, L. - Silhavy, D. (2008): Characterization of Plant Nonsense-Mediated mRNA Decay (NMD) Machinery Revealed an Unexpected Conservation Within Eukaryotic NMD Systems. Thirteenth Annual Meeting of the RNA Society, Berlin, Germany. Abstracts, p. 509.
2. **Benkovics, A.H. -** Nyiko, T. - Silhavy, D. (2009): Regulation of the Plant Nonsense-Mediated Decay Pathway. FEBS Practical Course on Protein interaction modules, Split, Croatia. Abstracts, p. 2.
3. **Benkovics, A.H.** – Nyiko, T. – Mérai, Zs. – Bisztray, G.D. – Silhavy, D. (2010): Premature termination codon containing transcripts can be degraded by alternative pathways in plants. EMBO/ESF RNA Quality control workshop, Vienna, Austria. Abstracts, p. 44.

References

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

Hori K. and Watanabe Y. (2005) UPF3 suppresses aberrant spliced mRNA in Arabidopsis *Plant J,* 43, 530-540.

Kerenyi Z., Merai 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.

Kurihara Y., Matsui A., Hanada K., Kawashima M., Ishida J., Morosawa T., Tanaka M., Kaminuma E., Mochizuki Y., Matsushima A., Toyoda T., Shinozaki K., and Seki M. (2009) Genome-wide suppression of aberrant mRNA-like noncoding RNAs by NMD in Arabidopsis *Proc Natl Acad Sci U S A,* 106, 2453-2458.

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.

Narsai R., Howell K.A., Millar A.H., O'Toole N., Small I., and Whelan J. (2007) Genome-wide analysis of mRNA decay rates and their determinants in Arabidopsis thaliana *Plant Cell,* 19, 3418-3436.

Ozbudak E.M., Thattai M., Kurtser I., Grossman A.D., and van Oudenaarden A. (2002) Regulation of noise in the expression of a single gene *Nat Genet,* 31, 69-73.

Papp B., Pal C., and Hurst L.D. (2003) Dosage sensitivity and the evolution of gene families in yeast *Nature,* 424, 194-197.

Raser J.M. and O'Shea E.K. (2005) Noise in gene expression: origins, consequences, and control *Science,* 309, 2010-2013.

Ratcliff F., Martin-Hernandez A.M., and Baulcombe D.C. (2001) Technical Advance. Tobacco rattle virus as a vector for analysis of gene function by silencing *Plant J,* 25, 237-245.

Schwartz A.M., Komarova T.V., Skulachev M.V., Zvereva A.S., Dorokhov Iu L., and Atabekov J.G. (2006) Stability of plant mRNAs depends on the length of the 3'-untranslated region *Biochemistry (Mosc),* 71, 1377-1384.

Stalder L. and Muhlemann O. (2008) The meaning of nonsense *Trends Cell Biol,* 18, 315-321.

Szittya G., Molnar A., Silhavy D., Hornyik C., and Burgyan J. (2002) Short defective interfering RNAs of tombusviruses are not targeted but trigger post-transcriptional gene silencing against their helper virus *Plant Cell,* 14, 359-372.

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

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.

Yoine M., Ohto M.A., Onai K., Mita S., and Nakamura K. (2006) The lba1 mutation of UPF1 RNA helicase involved in nonsense-mediated mRNA decay causes pleiotropic phenotypic changes and altered sugar signalling in Arabidopsis *Plant J,* 47, 49-62.