Seed dormancy in Arabidopsis thaliana is controlled by alternative polyadenylation of DOG1


Published in:
Plant Physiology

Document Version:
Early version, also known as pre-print

Queen's University Belfast - Research Portal:
Link to publication record in Queen's University Belfast Research Portal

Publisher rights
© 2016 The Authors
This is a preprint version of an article accepted for publication by The Plant Cell.

General rights
Copyright for the publications made accessible via the Queen's University Belfast Research Portal is retained by the author(s) and / or other copyright owners and it is a condition of accessing these publications that users recognise and abide by the legal requirements associated with these rights.

Take down policy
The Research Portal is Queen's institutional repository that provides access to Queen's research output. Every effort has been made to ensure that content in the Research Portal does not infringe any person’s rights, or applicable UK laws. If you discover content in the Research Portal that you believe breaches copyright or violates any law, please contact openaccess@qub.ac.uk.
Alternative polyadenylation controls seed dormancy

Szymon Swiezewski

Institute of Biochemistry and Biophysics
Pawinskiego 5a, 02-106 Warsaw, Poland
48 22 5925725
sswiez@ibb.waw.pl

Research Area:
Genes, Development, and Evolution
Systems and Synthetic Biology
Seed dormancy in *Arabidopsis thaliana* is controlled by alternative polyadenylation of *DOG1*


*These authors contributed equally to the article.

2 Institute of Biochemistry and Biophysics, Pawinskiego 5a, 02-106 Warsaw, Poland

3 Queen's University Belfast, School of Biological Sciences, Belfast BT9 7BL

Northern Ireland

One-sentence summary:

*DOG1* is alternatively polyadenylated and proximally polyadenylated *DOG1* is required for seed dormancy.
This work was supported by MSHE grant number IdP2011-000461 and 2011/01/D/NZ8/03690 to SŚ. HF, MC, AŚ, ZP were supported by Foundation for Polish Science grant number: TEAM2010-5/9. LB was supported by IP2011 004671.

Yanwu Guo present address:
Friedrich Miescher Institute
4002 Basel, Switzerland

Corresponding author
Szymon Swiezewski
sswiez@ibb.waw.pl
Abstract

DOG1 is a key regulator of seed dormancy in Arabidopsis and other plants. Interestingly, the C-terminus of DOG1 is either absent or not conserved in many plant species. Here, we show that in Arabidopsis, DOG1 transcript is subject to alternative polyadenylation. In line with this, mutants in RNA 3’ processing complex display weakened seed dormancy in parallel with defects in DOG1 proximal polyadenylation site selection, suggesting that the short DOG1 transcript, is functional. This is corroborated by the finding that the proximally polyadenylated short DOG1 mRNA is translated in vivo and complements the dog1 mutation. In summary, our findings indicate that the short DOG1 protein isoform produced from the proximally polyadenylated DOG1 mRNA is a key player in the establishment of seed dormancy in Arabidopsis and characterize a set of mutants in RNA 3’ processing complex required for production of proximally polyadenylated functional DOG1 transcript.
INTRODUCTION

The ability to postpone germination, to align it with permissive environmental conditions, greatly enhances the chances of plants survival (Koornneef et al., 2002; Penfield and King, 2009). Therefore, seed dormancy is extensively controlled by many factors including: humidity, temperature and light, and is also subject to parental memory (Chiang et al., 2011; Finch-Savage and Leubner-Metzger, 2006; Graebner et al., 2012; Nonogaki, 2014). DOG1 (Delay of Germination 1) has been identified as a major QTL (Quantitative Trait Locus) for seed dormancy variability among natural Arabidopsis accessions and dog1 T-DNA insertional mutants exhibit reduced seed dormancy (Bentsink et al., 2006). The expression of DOG1 increases during seed maturation and the mRNA disappears quickly after imbibition, although the DOG1 protein is more stable (Nakabayashi et al., 2012). A recent report indicated that DOG1 enforces seed dormancy by strengthening the seed coat, which occurs via modulation of the expression of gibberellin metabolism genes (Graebner et al., 2014). The expression of DOG1 is extensively regulated, being strongly induced by ABA (Abscisic acid) and low temperature during seed maturation (Chiang et al., 2011; Kendall et al., 2011). Other factors required for DOG1 expression include histone-modifying enzymes such as the histone H2B ubiquitin transferase HUB (Liu et al., 2007) and H3 lysine 9 methyl transferase KYP (Zheng et al., 2012). DOG1 expression is also highly dependent on transcription elongation factor TFIIS (Grasser et al., 2009).

The DOG1 gene is comprised of 3 exons, and the exon 2-exon 3 junction is subject to alternative splicing, generating 4 different forms of mRNA (Bentsink et al., 2006). The function of these alternatively spliced transcript isoforms is unknown, and their relative ratio remains unchanged during seed development (Bentsink et al., 2006). Our recent analysis of the regulation of alternative splicing of DOG1 in a mutant defective in PolII elongation suggested that the rate of transcript elongation regulates alternative splice site selection in accordance with the kinetic coupling model and implies that DOG1 splicing is co-transcriptional (Dolata et al., 2015).

To our knowledge complementation of the dog1 mutation in Arabidopsis has not been achieved using a DOG1 cDNA, while the seed dormancy phenotype of this mutant was complemented using a genomic DOG1 clone from Lepidium sativum (L. sativum) (Graebner et al., 2014). Notably, the L. sativum DOG1 gene lacks the exon 3,
so can only encode a short two-exonic mRNA, with no alternatively spliced exon 2-
exon 3 isoforms.

Many alternatively spliced genes are also subject to alternative
polyadenylation (APA) (Di Giammartino et al., 2011). Alternative polyadenylation
leads to the generation of transcripts with different 3’ ends, through a series of steps
catalyzed by components of RNA 3’ processing complexes, like: the Cleavage and
Polyadenylation Specificity Factor (CPSF) and Cleavage Stimulation Factor (CstF)
complexes, and poly(A) polymerases (Gruber et al., 2014; Proudfoot, 2011; Mandel
et al., 2008). APA is common and widespread in animals, plants and other eukaryotic
organisms (Tian et al., 2005; Pickrell et al., 2010; Sun et al., 2012; Wu et al., 2015;
Shi, 2012). In animals APA is involved in a range of developmental processes
including cell differentiation and has been implicated in cancer (Danckwardt et al.,
2008; Mayr and Bartel, 2009; Lianoglou et al., 2013; Lin et al., 2012). Similarly, in
plants APA has been shown to control key developmental processes.

In one case, the nuclear RNA-binding protein FCA interacts with FY, a
component of CPSF complex, to promote the usage of a proximal polyadenylation
site in its own gene, leading to the production of a non-functional RNA isoform
(Simpson et al., 2003). In addition, FCA functions with CstF complex to promote
proximal polyadenylation of the non-coding antisense transcript of FLC (Flowering
Locus C), leading to suppression of FLC expression (Liu et al., 2010). Alternative
polyadenylation has also been implicated in the control of pathogen resistance in
Arabidopsis through the selection of proximal polyadenylation sites in the RPP7
(Recognition Of Peronospora Parasitica 7) gene (Tsuchiya and Eulgem, 2013).

Given the key function of DOG1 in Arabidopsis seed survival and the potential
role of DOG1 homologs in controlling seed dormancy in other species, it is important
to understand both the mechanisms of DOG1 locus regulation and the function of the
encoded protein. Here we describe the process of DOG1 alternative polyadenylation
that produce two alternatively polyadenylated isoforms of the DOG1 transcript. We
characterize mutants in proteins that control this mechanism, and show that the
proximally polyadenylated mRNA isoform is translated in vivo, and is functional as it
complements the dog1 mutant. Demonstrating that alternative polyadenylation of
DOG1 gene in Arabidopsis plays a fundamental role in regulation of seed dormancy.
RESULTS

Mutants of RNA 3’ processing factors display weak seed dormancy phenotype

The majority of Arabidopsis genes have multiple polyadenylation sites (Wu et al., 2011), but the biological consequences of the alternative polyadenylation are in most cases unknown. FY is the Arabidopsis homologue of the yeast RNA 3’ processing factor Pfs2p (Simpson et al., 2003). A previous study demonstrated that the fy-1 mutant in Landsberg erecta (Ler) background has weak seed dormancy (Jiang et al., 2012). To check if the function of FY in seed dormancy control is independent of genetic background, we assayed seed dormancy of fy-2 mutant in Col-0 background. Compared to the Col-0 wild type, fy-2 seeds showed very weak dormancy (Figure 1A) demonstrating that FY plays a significant role in controlling this process in Arabidopsis. Since FY functions with other RNA 3’ processing factors in flowering time control (Liu et al., 2010; Manzano et al., 2009), we analyzed seed dormancy of the mutants of RNA 3’ processing factors PCFS4 (PCF11P-Similar Protein 4) and ESP1 (Enhanced Gene Silence 1). Like fy mutants, the pcsf4-1 and esp1-2 mutants showed weakened seed dormancy (Figure 1B), suggesting that the observed seed dormancy defect is caused by misregulation of RNA 3’ processing of a gene or genes involved in seed dormancy control.

Next, we examined whether the weak seed dormancy phenotype of fy-2 may be due to misregulation of DOG1 transcription. RT-qPCR (reverse transcription and quantitative PCR) analysis revealed no significant change of DOG1 mRNA level in fy-2 mutant (Figure 1C). However, western blot analysis using a DOG1-specific antibody revealed a consistent reduction in the DOG1 protein level in fy-2 seeds when compared with Col-0 wild type (Wt) (Figure 1D and Supplemental Figure S1). Thus, we concluded that FY is required for proper DOG1 protein expression and hypothesized that the fy-2 mutant may be defective in DOG1 RNA processing, which leads to the suppression of translation. Therefore, we searched for potential DOG1 polyadenylation defects in fy-2. Surprisingly, 3’ RACE experiments initiated from exon 3 of DOG1 showed no obvious differences in either polyadenylation site selection or the length of the poly(A) tail, between Col-0 wild type and fy-2 seeds (Figure 1E).
Alternative polyadenylation of DOG1 gene results in production of two mRNA isoforms.

As we were unable to detect any defect in the use of the canonical polyadenylation site of DOG1 in the fy mutant, we searched for additional
Polyadenylation site prediction along the DOG1 gene performed with PASPA software. Diagram of DOG1 gene structure with alternatively spliced isoforms of intron 2 labeled as alpha, beta, gamma and delta, plus the position of the alternative polyadenylation site. Lines above represent the fragments amplified in RT-qPCR analysis using the numbered primer sets. Below shown in scale with the DOG1 gene structure is the diagram of the quality score representing the likelihood of a polyadenylation site predicted by PASPA software.

Schematic presentation of two alternatively polyadenylated DOG1 mRNA isoforms.

RT-qPCR along the DOG1 gene to examine the expression of the different mRNA isoforms in Col-0, fy-2 and dog1-3 seeds. Values were normalized to the level of UBC mRNA. Expression in mutants is normalized to the Col-0 value. Data represent the means of three biological replicates; error bars show SD.

Long DOG1/short DOG1 ratio in Col-0 and mutants of RNA 3' processing factors. RT-qPCR results were normalized to the level of UBC mRNA, and the expression in mutants is then normalized to the Col-0 value. Data represent the means of three biological replicates with error bars showing SD.

Alternative DOG1 polyadenylation site selection during seed development. RT-qPCR specific for shDOG1 and lgDOG1 was used to measure relative levels of each mRNA isoform. Manually pollinated flowers were collected at different days after pollination (dap). Values were normalized to the level of UBC mRNA. Data represent the means of three biological replicates with error bars showing SD.

The most common polyadenylation signal motif in plants is AAUAAA and UUGUUU positioned 19 and 7 nt respectively upstream of the cleavage site (Sherstnev et al., 2012). The identification of likely polyadenylation sites with PASPA (Ji et al., 2015), a software designed for mRNA
polyadenylation site prediction in plants, revealed three potential polyadenylation clusters in the \textit{DOG1} gene. One cluster corresponds to the predicted full length polyadenylation site of \textit{DOG1}, and the two others to additional internal polyadenylation sites: one in intron 1 and another in intron 2 (Figure 2A). To validate these predictions, we reanalyzed published Direct RNA Sequencing based mapping data to detect Arabidopsis polyadenylation sites in the \textit{DOG1} locus (Supplemental Figure S2) (Sherstnev et al., 2012). We identified two distinct polyadenylation clusters: a distal one used to produce the predicted full-length mRNA, which is named hereafter as the long \textit{DOG1} (\textit{lgDOG1}) form, and a proximal one matching one of the predicted internal polyadenylation sites located in intron 2 that would produce a truncated mRNA, which we name short \textit{DOG1} (\textit{shDOG1}). The second predicted internal polyadenylation site lies within intron 1, but analysis of RNA sequencing data revealed no indication of its usage (Supplemental Figure S2).

In summary, the \textit{lgDOG1} transcript is comprised of three exons and corresponds to the previously described Arabidopsis \textit{DOG1} mRNA (Bentsink et al., 2006), whereas the newly identified \textit{shDOG1} transcript comprises only exons 1 and 2, and therefore lacks the alternative splicing isoforms described (Figure 2B).

Next, to quantify the different \textit{DOG1} mRNA isoforms in Col-0 wild type and \textit{fy-2} mutant seeds we designed sets of primers within the \textit{DOG1} locus for use in RT-qPCR analysis. Amplification with primer set 4 showed no difference in the level of \textit{lgDOG1} mRNA between \textit{fy-2} and wild type seeds (Figure 2A and Figure 2C). In contrast, use of primer set 3 demonstrated a clear reduction of the \textit{DOG1} transcript in \textit{fy-2} seeds. The primers of set 3 span the exon 2-intron 2 border that is specific for the short alternatively polyadenylated \textit{shDOG1} mRNA isoform. Therefore, this result suggested a reduction in usage of the proximal polyadenylation site in the \textit{fy-2} mutant. To confirm this, we designed primers to amplify the full-length short and long \textit{DOG1} transcript isoforms and obtained similar results (Supplemental Figure S3).

Different mRNA isoforms generated through APA often differ in their stability (Krol et al., 2015). However, a cordycepin-dependent RNA stability assay (Golisz et al., 2013) showed that both the short and long \textit{DOG1} mRNA isoforms have a similar half-life of about 1 h (Supplemental Figure S4). This shows that usage of \textit{DOG1} proximal polyadenylation site leads to the production of a stable transcript.

A change in the long/short \textit{Dog1} transcript ratio, similar to that observed in \textit{fy-2}, was also detected in mutants of other factors required for RNA 3' processing,
which showed weak seed dormancy (Figure 2D and Figure 1B). This finding led us to speculate that the reduction in the \textit{shDOG1} mRNA level in these mutants, may be the underlying cause of their reduced seed dormancy.

Having established that the \textit{DOG1} gene is subjected to alternative polyadenylation, we questioned whether this process is developmentally regulated. \textit{DOG1} expression is tightly controlled during seed development and has previously been shown to peak between 9-16 days after pollination and then slowly decay (Nakabayashi et al., 2012; Zhao et al., 2015). Alternative splicing of the \textit{lgDOG1} mRNA isoform has been reported to be unaffected during seed development (Bentsink et al., 2006). Our analysis of short and long \textit{DOG1} mRNA levels showed that both forms are induced during seed development and show a similar expression profile (Figure 2E).

\textbf{Short \textit{DOG1} mRNA is translated \textit{in vivo} and produce a conserved protein}

The previously recognized long \textit{DOG1} mRNA transcript (\textit{lgDOG1}) encodes a protein of \(~32\text{ kD}. The proximally polyadenylated \textit{shDOG1} transcript described in this study codes for a slightly smaller \(~30\text{ kD} protein. These short and long \textit{DOG1} proteins only differ by several amino acids at their C-terminal ends, encoded by exon 3 in \textit{lgDOG1} and a sequence from intron 2 in \textit{shDOG1} (Figure 3A). Interestingly, a multiple sequence alignment with \textit{DOG1} proteins from other species showed that the sequence unique to shDOG1 is conserved at a similar level compared to that encoded by exon 2, while the amino acids unique to lgDOG1 (exon 3) are either absent due to stop codons or are weakly conserved (Figure 3A). Evolutionary conservation is often considered an indication of functionality, suggesting that the shDOG1 protein is functional.

Next, we determined which of the \textit{DOG1} mRNA isoforms is translated. \textit{DOG1} antibodies used previously and that described here do not distinguish the long and short \textit{DOG1} proteins because they were raised to peptides shared by the two isoforms (Nakabayashi et al., 2012). In addition, the small difference in protein mass does not allow distinction between the shDOG1 and lgDOG1 proteins on a western blot: only one \textit{DOG1} band is observed in samples from freshly harvested seeds (Figure 1D). Therefore, we created a \textit{GFP::DOG1} genomic fusion driven by the \textit{DOG1} promoter. Samples from transgenic plants expressing this fusion protein were used in a GFP pull-down assay (Figure 3B) with subsequent analysis by mass
spectrometry. Notably, no long DOG1-specific peptides were identified, but we could clearly detect peptides unique to the shDOG1 protein (Figure 3C and Supplementary Figure S5). Although this does not exclude the possibility that the long DOG1 form is expressed but not detected, this result clearly showed that use of the proximal

**Figure 3.** Comparison of the two DOG1 protein isoforms

**A** Alignment of DOG1 genomic sequences from selected plant species used to construct the protein sequence alignment beneath. Nucleotides and amino acids shared with *A. thaliana* are highlighted by grey and red shading, respectively. Black box highlights peptide unique to shDOG1 identified below.

**B** Purification of DOG1 from green siliques extract of plants expressing a GFP-DOG1 fusion under the control of the native promoter and Col-0 (Wt) plants, using GFP Nano-trap beads. Isolated proteins were analysed with western blot using GFP specific antibodies (left panel) and on a SDS-PAGE gel followed by silver staining (right panel). GFP-DOG1 band is marked with an arrow.

**C** Purified proteins were subjected to mass spectrometry analysis. Short DOG1 specific peptide fragmentation with MS spectra of fragments derived, shown as intensity (Y-axis) sorted by mass-to-charge ratio (m/z) shown on X-axis.

**D** Short and long DOG1 show nuclear localization. Arabidopsis protoplasts were transformed with constructs encoding N- and C-terminal fusions of GFP with long and short DOG1, then analyzed using a confocal microscope. The green colour represents GFP and red the chlorophyll signal.
polyadenylation site leads to the production of a short *DOG1* mRNA that is translated *in vivo*.

A previous study examining Arabidopsis lines constitutively overexpressing DOG1 suggested nuclear localization of this protein (Nakabayashi et al., 2012). However, this experiment did not distinguish between the short and long DOG1 forms. Therefore, we compared the localization of these DOG1 proteins. Both N- and C-terminal GFP fusions of the shDOG1 and IgDOG1 protein showed predominantly nuclear localization when transiently expressed in Arabidopsis protoplasts (Figure 3D).

We conclude that the short alternatively polyadenylated *DOG1* mRNA isoform is translated *in vivo*, thus producing a short nuclear localized DOG1 protein. The fact that the shDOG1 is conserved whereas the IgDOG1 is less together with our inability to find long DOG1 specific peptides suggested that the shDOG1 maybe the functional DOG1 protein in respect to seed dormancy.

**Short DOG1 is sufficient to inhibit germination**

To directly test the ability of the short and long DOG1 proteins to control seed dormancy in Arabidopsis we created *dog1-3* mutant plants with constructs carrying the shDOG1 and IgDOG1 sequences driven by the DOG1 promoter (Figure 4A). Only the shDOG1 construct was able to partially complement the weak seed dormancy phenotype of *dog1-3* (Figure 4A), even though both showed a similar DOG1 expression as measured by qRT-PCR (Supplementary Figure S6). This indicated that shDOG1 protein is sufficient for seed dormancy establishment. To further support this conclusion we selected a *dog1-5* T-DNA insertion mutant allele. In this mutant a T-DNA is inserted in exon 3 of the *DOG1* gene (Figure 4B). RT-qPCR analysis with primers located at the exon 2-exon 3 junction, upstream of the T-DNA insertion, confirmed that IgDOG1 is no longer expressed in this mutant (Figure 4C). In contrast, levels of the shDOG1 transcript appear slightly increased in *dog1-5*. Accordingly, the *dog1-5* mutant exhibited an increase in the DOG1 protein level, and increased rather than reduced seed dormancy (Figure 4D and Figure 4E). These findings confirmed that the short version of the *DOG1* transcript generated through APA is translated and functional in controlling seed dormancy.
Figure 4. Short DOG1 is required for seed dormancy.

(A) Germination rate (%) of freshly harvested seeds of the *dog1-3* mutant and *dog1-3* complemented with long (# lg1, # lg2) and short (# sh3, # sh4) versions of the *DOG1* gene controlled by its native promoter. Data represent the means of at least 4 biological replicates with error bars showing SD.

(B) Schematic description of *DOG1* mutants: *dog1-3* (SALK_000867), *dog1-4* (SM_3_20886) and *dog1-5* (SALK_022748).

(C) Relative expression level of long and short *DOG1* mRNAs in freshly harvested seeds from Col-0 and *dog1-5* plants determined by RT-qPCR. Values were normalized to the level of *UBC* mRNA. Expression in the mutant is normalized to the Col-0 value. Data represent the means of three biological replicates; error bars show SD.

(D) Germination rate (%) of freshly harvested seeds of *dog1* mutants. The graph shows the means from three biological replicates, with error bars showing SD. The experiment was repeated at least three times.

(E) Western blot analysis of the DOG1 protein level in freshly harvested seeds from Col-0, *dog1-5*, *dog1-4* and *dog1-3* plants. The Coomassie blue-stained gel is shown as a loading control. Three biological replicates were examined for Col-0 and *dog1-5* plants.
DISCUSSION

Alternative polyadenylation is widespread in eukaryotes. Whole genome polyadenylation site mapping has revealed that the vast majority of Arabidopsis genes harbour alternative polyadenylation sites, but the biological significance of this phenomenon is mostly unknown. In this study, we have shown that the Arabidopsis DOG1 gene is subjected to alternative polyadenylation. The proximally polyadenylated short DOG1 transcript is translated and functional, in contrast to the long DOG1. Our data clearly demonstrate that expression of the short DOG1 isoform is sufficient to promote seed dormancy, thus providing an example where alternative polyadenylation is important in seed development.

**DOG1 mRNA is alternatively polyadenylated**

DOG1 is a major QTL for seed dormancy in Arabidopsis and an important regulator of this process in agriculturally important plants. Here, we show that the Arabidopsis DOG1 gene produces two alternatively polyadenylated RNAs. The production of the short isoform is sensitive to mutations in RNA 3’ processing complex components. RNA 3’ processing complex mutants with defects in selection of the proximal polyadenylation site on DOG1 show weak seed dormancy, suggesting that the proximal polyadenylation is required for DOG1 function. This is corroborated by the finding that the RNA 3’ processing mutant fy-2 displays low DOG1 protein levels, but no change, or increase, in the level of the full-length DOG1 transcript. Taken together, these data indicate that the short alternatively polyadenylated DOG1 isoform is functional in seed dormancy regulation. The most well characterized, developmental processes that involve a critical change in the polyadenylation site, is the global switch to the usage of proximal polyadenylation sites observed during mammalian brain development or in tumors, which has been proposed to cause a release from miRNA mediated control due to the loss of miRNA binding sites (Elkon et al., 2013; Di Giammartino et al., 2011). This is unlikely to be the underlying reason for the switch in polyadenylation site of DOG1 since there is no dramatic difference in RNA stability between the long and short DOG1 isoforms.

**Short DOG1 protein is functional in seed dormancy establishment**
A reanalysis of published data together with our results show that the DOG1 locus produces two alternatively polyadenylated forms of DOG1 mRNA. Interestingly, the DOG1 genes in *A. lyrata* and *L. sativum*, two species quite closely related to Arabidopsis, harbour mutations, including stop codons in exon 3, that encode the specific C-terminal region of the long DOG1 isoform. This shows that these species express only the short DOG1 protein isoform encoded by exons 1 and 2 plus part of intron 2.

Our complementation analysis demonstrated that while the long DOG1 is unable to complement the dog1-3 mutation, the short version of DOG1 is functional. In addition, we identified a novel mutant dog1-5, carrying a T-DNA insertion in exon 3, a region that is only present in *lgDOG1*. This mutant lacks the long version of the DOG1 transcript but it expresses the proximally polyadenylated *shDOG1* at higher level than wild type and exhibits stronger seed dormancy. This confirms that the *shDOG1* isoform produced by alternative polyadenylation in intron 2 is sufficient for seed dormancy.

The fact that the *fy*-2 mutant shows a lower level of the DOG1 protein without any reduction in the *lgDOG1* mRNA strongly suggests that this transcript does not produce a protein in this mutant. Examination of currently available proteomic datasets revealed DOG1 peptides. However all of them correspond to DOG1 exons 1 and 2, and none are specific to either the short or long DOG1 protein isoforms (Baerenfaller et al., 2008; Castellana et al., 2008). Nevertheless, using targeted mass spectrometry analysis of immunoprecipitated DOG1-GFP, *shDOG1*-specific peptides were clearly detectable, but we did not identify peptides specific for *lgDOG1*.

The short DOG1 protein can be derived either through the use of a proximal polyadenylation site or by splicing of the long DOG1 transcript at an alternative splice site (*DOG1*-beta mRNA isoform) (Bentsink et al., 2006). Analysis of splice sites in the DOG1 gene reveals weak conservation of the exon 3 acceptor splice site used by the *DOG1*-beta mRNA isoform. This, together with the strong DOG1 protein signal in mutant dog1-5 that lacks the *lgDOG1* mRNA isoforms, suggests that the *shDOG1* protein is predominantly produced from the proximally polyadenylated short DOG1 transcript isoform.

The function of the long DOG1 transcript and *lgDOG1* protein is currently not clear. Lack of conservation of the *lgDOG1* protein specific sequence encoded by exon 3, together with the inability of *lgDOG1* to complement *dog1* mutant suggest...
that long DOG1 protein may be not functional. However, extensive examples of alternative polyadenylation sites competition, including proximal and distal FCA polyadenylation sites (Simpson et al., 2003), imply that the DOG1 distal polyadenylation site might acts as a decoy for transcripts that fail to terminate earlier. If so, then it could be expected that the lgDOG1 mRNA is rapidly degraded. Due to technical limitations we were unable to perform RNA stability assays on intact seeds. However, mRNA half-life measurements in seedlings showed no clear difference between the stability of the long and short DOG1 transcripts.

It is interesting to note, that the dog1-5 mutant shows upregulation of both the proximally polyadenylated shDOG1 mRNA isoform and the DOG1 protein, suggesting that the lgDOG1 mRNA may act as a negative regulator of shDOG1 mRNA expression.

In summary, our findings provide several lines of evidence demonstrating that DOG1 nascent transcript is subjected to alternative polyadenylation and that the short form of this transcript is functional in seed dormancy establishment. Given that DOG1 is subject to elaborate transcriptional elongation control it would be interesting to test whether polyadenylation site selection is controlled by the rate of PolII elongation on DOG1, as is the case for splice site selection (Dolata et al., 2015).
METHODS

Plant materials and growth conditions

*Arabidopsis thaliana* plants were grown on soil in a greenhouse with a 16-h light/8-h-dark photoperiod at 22°C/18°C. For expression analysis during seed development, flowers were emasculated and pollinated the following day. Plant material was collected at the indicated times representing days since manual pollination. The termination complex mutants used in this study have been described previously: *fy-2* (Simpson et al., 2003), *esp1-2* (SALK_078793), *pcsf4-1* (SALK_102934c) (Xing et al., 2008). Col-0 was used as Wt control. The *DOG1* T-DNA insertion mutants used were: *dog1-3* (SALK_000867) (Bentsink et al., 2006), *dog1-4* (SM_3_20886) and *dog1-5* (SALK_022748).

Germination assay

About 100-150 freshly harvested Arabidopsis seeds were sown on blue germination paper (Anchor) supported by 2 layers of thick fabric saturated with water. Plates were sealed and transferred to a growth chamber with 16h/8h light/dark at 22°C/18°C. Plates were photographed daily for 7 days and germinating seeds counted. At least three biological replicates were performed for each experiment. Data show a time point when the *dog1-3* mutant seed had fully germinated. Seed dormancy strength is strongly dependent on external conditions during seed maturation (Kendall et al., 2011) we therefore always included a set of controls: *dog1-3* and WT in all experiments. Small external condition fluctuations between separate experiments may therefore explain the difference in seed germination strength between Col-0 plants in Figure 1A and 1B.

RNA extraction, cDNA synthesis and PCR analysis

Total RNA was isolated using a phenol-chloroform extraction procedure (Shirzadegan et al., 1991). The RNA was then treated using a TURBO DNA-free™ Kit (Life Technologies), according to the manufacturer’s protocol. The efficiency of DNA removal was monitored by *PP2A* PCR. Before reverse transcription, the RNA was examined by electrophoresis on a 1.2% agarose gel to determine its quality and quantified using a Nanodrop 2000 spectrophotometer. For cDNA synthesis, 2.5 µg of RNA were used with oligo(dT) primers and a RevertAid™ First Strand cDNA
Synthesis Kit (Fermentas). Diluted cDNA (10-fold) was used in qPCR (LightCycler® 480 Roche), with SYBR Green mix (Roche). A UBC gene was used for normalization (Czechowski et al., 2005).

**Protein extraction**
Freshly harvested Arabidopsis seeds (30 mg) were frozen in liquid nitrogen and pulverized in 100 µl of acetone. The suspension was centrifuged for 1 min at 1000 g. The pelleted material was rinsed with 96% EtOH and centrifuged again as above. The pellet was then resuspended in 50% Percoll and centrifuged for 10 min at 1000 g. Pelleted material was again washed in 96% EtOH and centrifuged for 1 min at 1000g. The final pellet was resuspended in 8 M urea for analysis by SDS-PAGE.

**Western blotting**
Proteins were extracted from freshly harvested seeds as described above. The protein concentration was determined using the Bradford method (Bradford, 1976) and samples were analyzed by standard SDS-PAGE. Proteins were then electrophoretically transferred to Hybond P membrane (Amersham) using a Trans-Blot Cell (Bio-Rad). A synthetic peptide (MGSSSKNIEQADSY) corresponding to amino acids 1 to 15 of DOG1 was synthesized and used to produce an affinity purified DOG1-specific rabbit polyclonal antibody (Eurogentec). DOG1-specific (Eurogentec) and Goat Anti-Rabbit IgG H&L (HRP) (Abcam b7090) antibodies were also used for western blotting. Western blot was developed using a 1:1 mixture of SuperSignal West Pico and Femto Chemiluminescent Substrates (Pierce).

**GFP-DOG1 Immunoprecipitation (IP) and mass spectrometry analysis**
pDOG1::GFP::DOG1g vector was constructed using primers shown in Supplementary Table 1 and a pCambia5408 modified to carry a Basta resistance under NOS promoter resulting in native DOG1 gene with GFP inserted at ATG. Plant extracts from mixture of green and mature siliques of transgenic plants were used for IP with GFP Nano-trap beads (ChromoTek). MS followed by analysis using Mascot software was performed as described previously (Dolata et al., 2015).

**Protoplast transformation**
The following plasmid constructs were prepared in the vectors pSAT6-eGFP-C1 and pSAT6A-eGFP-N1 using the short and long DOG1 cDNAs: 35S::sDOG1-YFP, 35S::YFP-sDOG1, 35S::lqDOG1-YFP and 35S::YFP-DOG1. To maximize the efficiency of protoplast isolation protoplasts were isolated from the leaves of 20-d-old late flowering Col-0 with backcrossed active FRI (Frigida) from SF2 ecotype (Michaels and Amasino, 1999), and transformed with 30 µg of purified construct DNAs according to a previously described method (Wu et al., 2009). Transformed protoplasts were incubated overnight under continuous light at 22°C prior to analysis by confocal microscopy.

Confocal microscopy imaging
Fluorescence images of transformed Arabidopsis protoplasts were visualized using a Nikon EZ-C1 laser scanning microscope mounted on an inverted Nikon TE 2000E epifluorescence microscope. Excitation was at 514 nm, and the emission signal was collected between 525 and 590 nm for YFP fluorescence, and between 622 and 700 nm for chlorophyll autofluorescence. Untransformed protoplasts were examined as a negative control.

3' RACE
RACE was performed as described previously (Swiezewski et al., 2009). Total RNA from seedlings was ligated to 3' RACE oligo adapter (Gene Race kit, Invitrogen) using T4 RNA Ligase 1 (New England Biolabs). After phenol/chloroform extraction RNA was precipitated and subjected to reverse transcription using adapter-specific primer and Superscript III (Invitrogen). 3' ends of DOG1 mRNA were amplified using adapter specific primer and DOG1 specific RACE primer. PCR products were purified from agarose gel and cloned to pJet1.2 vector and sequenced using plasmid specific primers in both directions.

RNA stability assay
Cordycepin dependent RNA stability assay was performed as previously described (Golisz et al., 2013). In brief, Col-0 plants were grown on ½ MS medium containing 1% sucrose in a 16-h light/8-h-dark cycle at 22°C. About 100 two-week-old seedlings were transferred to 40 ml of buffer (1 mM PIPES pH 6.25, 1 mM Trisodium citrate, 1
mM KCl, 15 mM sucrose) and incubated for 30 min with shaking. Cordycepin was then added to a final concentration of 150 µg/ml and vacuum infiltration performed for 30 s. At each time point thereafter, seedlings representing ~0.5 g were collected and frozen in liquid nitrogen. Samples were analyzed in triplicate. RNA extraction was performed as described above and RT-qPCR analysis with EIF4A and At3g45970 was used mRNAs with respectively high and low stability (Golisz et al., 2013)

**dog1-3 complementation**

For complementation with IgDOG1, dog1-3 mutant was transformed with a DOG1 genomic fragment with the sequence of the second intron deleted. For complementation with shDOG1, the dog1-3 mutant was crossed with Col-0 plants expressing a construct containing the genomic fragment spanning DOG1 exon 1-intron 1 and exon 2. The DOG1 genes in both constructs are driven by the endogenous DOG1 promoter. Two independent transformants were analyzed for each construct.

**Accession numbers**

The sequences of genes examined in this study can be found in The Arabidopsis Information Resource (TAIR) data library under the following accession numbers: DOG1 (At5g45830), FY (At5g13480), ESP1 (At1g73840), PCSF4 (At4g04885).
Supplemental Data

Supplemental Figure S1. Western blot and Coomassie-stained gel used for Figure 1D.

Supplemental Figure S2. DOG1 polyadenylation site identified using Direct RNA sequencing data (Sherstnev et al., 2012).

Supplemental Figure S3. RT-PCR to examine the expression of full-length shDOG1 and lgDOG1 mRNA transcripts in Col-0 and fy-2 seeds.

Supplemental Figure S4. RNA stability assay of shDOG1 and lgDOG1 transcripts.

Supplemental Figure S5. Characterization of short DOG1-specific peptides.

Supplemental Figure S6. DOG1 expression in freshly harvested seeds of Col-0 (WT) and dog1-3 or dog1-3 mutants complemented with lgDOG1 (# lg1, # lg2) or shDOG1 (# sh3, # sh4) expressing constructs.

Supplemental Table 1. List of primers.

ACKNOWLEDGMENTS

We thank Maciej Kotliński for help with analysis of Masspectrometry data and Krzysztof Kokoszka for excellent technical assistance. We are also very grateful to Dorota Konopka for critical reading of the manuscript, and all members of the Swiezewski laboratory for stimulating discussions.

AUTHOR CONTRIBUTIONS

S.S. and F.L. designed the project and planned the program of research. H.F., M.C. and A.S. performed the germination tests and expression level measurements. H.F. and M.C. performed the western blots. L.B. did the RACE experiment. S.S. and S.K. performed the bioinformatics analysis. A.C. and H.F. examined localization. G.Y. performed the DOG1-GFP IP. L.B. and A.S. performed the RNA stability assay. K.K and Z.P. prepared the short and long DOG1 lines for complementation. S.S. and M.C. wrote the manuscript.


