LSM PROTEINS PROVIDE ACCURATE SPLICING AND DECAY OF SELECTED TRANSCRIPTS TO ENSURE NORMAL ARABIDOPSIS DEVELOPMENT

Carlos Perea-Resa¹, Tamara Hernández-Verdeja¹, a, Rosa López-Cobollo¹, M. Mar Castellano¹ and Julio Salinas²

Departamento de Biología Ambiental, Centro de Investigaciones Biológicas (CIB-CSIC), Ramiro de Maeztu, 9, 28040 Madrid, Spain

¹ These authors contributed equally to this work

a Current address: Centro de Biotecnología y Genómica de Plantas UPM/INIA, Campus de Montegancedo, 28223 Madrid, Spain (THV and MMC); Imperial College London, South Kensington Campus, London SW7 2AZ, UK (RLC)

² Address correspondence to salinas@cib.csic.es

Running Title: LSMs control Arabidopsis development

Estimated length of the article: 16 pages

The author responsible for distribution of materials integral to the findings presented in this article in accordance with the policy described in the Instructions for Authors (www.plantcell.org) is: Julio Salinas (salinas@cib.csic.es)

Synopsis: This study describes the molecular and functional characterization of Arabidopsis LSM proteins. Results demonstrate that they are organized in two heptameric complexes, one nuclear and another cytoplasmic, that play a critical role in Arabidopsis development by ensuring the appropriate developmental-related gene expression through the control of mRNA splicing and decay, respectively.
ABSTRACT

In yeast and animals, Sm-like (LSM) proteins typically exist as heptameric complexes and are involved in different aspects of RNA metabolism. Eight LSM proteins, LSM1-8, are highly conserved and form two distinct heteroheptameric complexes, LSM1-7 and LSM2-8, that function in mRNA decay and splicing, respectively. A search of the Arabidopsis thaliana genome identifies eleven genes encoding proteins related to the eight conserved LSMs, the genes encoding the putative LSM1, LSM3 and LSM6 proteins being duplicated. Here, we report the molecular and functional characterization of the Arabidopsis LSM gene family. Our results show that the eleven LSM genes are active and encode proteins that are also organized in two different heptameric complexes. The complex LSM1-7 is cytoplasmic and is involved in P-body formation and mRNA decay by promoting decapping. The complex LSM2-8 is nuclear and is required for pre-mRNA splicing through U6 snRNA stabilization. More important, our results also reveal that these complexes are essential for the correct turnover and splicing of selected developmental-related mRNAs, and for the normal development of Arabidopsis. We propose that LSMs play a critical role in Arabidopsis development by ensuring the appropriate developmental-related gene expression through the control of mRNA splicing and decay.
INTRODUCTION

During the last years, an increasing body of evidence indicates that post-transcriptional regulation plays an important role in modulating gene expression during development in eukaryotes (Halbeisen et al., 2008). Most eukaryotic genes are transcribed as precursors (pre-mRNAs) containing intron sequences. In order to yield correct translation products, introns need to be excised to generate mature mRNAs. This process, known as pre-mRNA splicing, is fundamental in both constitutive and regulated gene expression. Pre-mRNA splicing is precisely and efficiently carried out by the spliceosome, a large ribonucleoprotein (RNP) complex machinery composed of five small nuclear RNP particles (snRNPs, U1, U2, U4/U6, U5) and more than two hundred polypeptides not tightly associated with snRNPs (Wahl et al., 2009). In many cases, however, the splicing process is flexible enough to allow the generation of alternative transcripts from a single gene by differential use of splicing sites. Site use may depend on the cell type, developmental stage or physiological condition, thereby affecting protein diversity and transcript levels (Matlin et al., 2005). The general mechanism of splicing has been well studied in humans and yeast, being largely conserved between these organisms. In plants, the splicing process remains comparatively poorly understood although the basic mechanisms of spliceosome assembly and intron excision appear to be as in the rest of eukaryotes (Lorković et al., 2000; Reddy, 2001). Consistent with this, the analysis of the Arabidopsis thaliana genome for the presence of known spliceosomal proteins indicated that the core of spliceosomal machinery is conserved between plants and animals (Wang and Brendel, 2004). Nonetheless, despite this conservation, incorrect splicing of mammalian pre-mRNAs in plant cells and vice versa denotes the existence of plant-specific splicing regulatory mechanisms requiring plant-specific splicing factors (Lorković et al., 2000; Reddy, 2001; Lorković, 2009). The characterization of different plant splicing proteins, including some glycine-rich RNA-binding proteins, SR proteins, RNA helicases and other RNA-binding proteins, have revealed that they are essential for the accurate progress of diverse plant developmental processes (Raab and Hoth, 2007; Barta et al., 2008; Lorković, 2009; Deng et al., 2010; Zhang et al., 2011).

The control of mRNA turnover is another critical aspect in the regulation of eukaryotic gene expression. Two major pathways exist in yeast and mammals for mRNA decay, both of them being initiated by deadenylation through the CCR4/POP2/NOT1-5 complex (Meyer et al., 2004; Parker and Song, 2004). Subsequently, transcripts can be
processed by the 3’ to 5’ or the 5’ to 3’ decay pathways. In the first pathway, the deadenylated mRNA is degraded by a complex of proteins known as exosome (Anderson and Parker, 1998). In the second pathway, the mRNA is decapped by the DCP1/DCP2 enzyme, making the mRNA susceptible to the exonuclease XRN1 (Beelman et al., 1996; Dunckley and Parker, 1999). Decapping, therefore, is an important node in the control of mRNA lifespan and is modulated by a set of different proteins (Bonnerot et al., 2000; Coller et al., 2001). The decapping machinery accumulates in discrete cytoplasmic foci named processing bodies (P-bodies), which have been suggested to be functionally involved not only in mRNA decapping (Sheth and Parker, 2003; Cougot et al., 2004), but also in nonsense-mediated mRNA decay (Unterholzner and Izaurralde, 2004; Sheth and Parker, 2006), mRNA storage (Brengues et al., 2005), general translation repression (Coller and Parker, 2005) and microRNA-mediated repression (Bhattacharyya et al., 2006). Although the existence of both the 5’ to 3’ and the 3’ to 5’ decay pathways has been documented and their core components identified (Xu et al., 2006; Goeres et al., 2007; Belostotsky and Sieburth, 2009; Lange and Gagliardi, 2010), the governing principles of mRNA decay in plants, as in the case of the splicing process, are still poorly known. Moreover, genetic analyses have also uncovered plant-specific functional features in mRNA degradation pathways that are associated with plant-specific factors (Belostotsky and Sieburth, 2009; Xu and Chua, 2011). In Arabidopsis, for instance, no XRN1-like gene has been identified. Instead, the cytoplasmic 5’ to 3’ exoribonuclease activity is carried out by XRN4 (Kastenmayer and Green, 2000; Souret et al., 2004). Plant P-bodies seem to function as yeast and human P-bodies. However, they also contain their own distinct protein components (Xu and Chua, 2011). Plants affected in mRNA turnover display severe developmental perturbations, indicating that proteins related to mRNA decapping and decay play important roles in controlling gene expression during plant development (Xu et al., 2006; Goeres et al., 2007; Belostotsky and Sieburth, 2009; Xu and Chua, 2009; Xu and Chua, 2011).

The LSMs constitute a large family of SM-like proteins that function in multiple aspects of RNA metabolism. In yeast and animals, there are eight highly conserved LSM proteins (LSM1-LSM8) that form two different heptameric ring complexes, LSM1-7 and LSM2-8, localized in the cytoplasm and nucleus, respectively. LSM1 and LSM8 define and confer the specificity to each complex, while the other proteins (LSM2-
LSM7) participate in both cytoplasmic and nuclear complexes. The LSM1-7 cytoplasmic complex binds to oligoadenylated mRNAs promoting their decapping and subsequent degradation by the 5’ to 3’ pathway, and accumulates in P-bodies. The LSM2-8 nuclear complex binds to and stabilizes the U6 snRNA, forms the core of the U6 snRNP, and functions in pre-mRNA splicing (see Beggs, 2005 and Tharun, 2009 for reviews). In silico approaches have allowed to identify potential plant homologs of LSM proteins. Arabidopsis have homologs for the eight conserved LSMs, and three of them (LSM1, LSM3 and LSM6) are duplicated (Wang and Brendel, 2004). To date, however, plant LSMs have not been functionally characterized and their role in RNA metabolism remains to be determined. Only Arabidopsis LSM5 (At5g48870) and LSM4 (At5g27720) genes have been experimentally studied, both of them being related to ABA and osmotic stress signaling (Xiong et al., 2001; Deng et al., 2010; Zhang et al., 2011). Here, we report the molecular and functional characterization of the Arabidopsis LSM gene family. Our results indicate that Arabidopsis LSM proteins are also organized in two different heptameric complexes localized in the cytoplasm and nucleus. Whereas the cytoplasmic complex (LSM1-7) is involved in P-body formation, mRNA decapping and, therefore, accurate mRNA decay, the nuclear one (LSM2-8) is required for U6 snRNA stabilization and, consequently, proper pre-mRNA splicing. Genetic and molecular analyses reveal that LSM1-7 and LSM2-8 complexes are essential for the correct turnover and splicing of selected developmental-related mRNAs, respectively. Consistently, the absence of LSM1 and LSM8 proteins causes severe perturbations in Arabidopsis development, which correlates with alterations in developmental-regulated gene expression. We conclude that LSMs play a critical role in Arabidopsis development by ensuring the appropriate developmental-related gene expression through the control of mRNA splicing and decay.
RESULTS

The Arabidopsis genome contains eleven genes encoding the eight highly conserved LSM proteins

Sequence comparisons and motif searches allowed to identify 11 genes in the Arabidopsis genome encoding proteins related to the eight highly conserved proteins that in yeast and animals constitute the heptameric LSM complexes, LSM1-7 and LSM2-8, three of them, the putative LSM1, LSM3 and LSM6, being duplicated (Wang and Brendel, 2004). The predicted Arabidopsis proteins contain the Sm1 and Sm2 motifs that are separated by a nonconserved linker region of variable length, and conform the Sm bipartite domain typical of LSM proteins (Tharun, 2009) (Supplemental Figure 1 online). LSM proteins have also been found in the genomes of different plant species (Proost et al., 2009; Goodstein et al., 2012). A phylogenetic analysis was performed with the LSM proteins from Arabidopsis and other representative plant species, including Glycine max, Populus trichocarpa, Oryza sativa, and Zea mays. Results unveiled that all plant genomes analyzed contain genes encoding LSMs related to the eight conserved proteins from yeast and animals, and that many of them are present in more than one copy as in the case of Arabidopsis LSM1, LSM3 and LSM6. The human LSM proteins were also included in the analysis as an internal control (Supplemental Figure 2 online).

RNA gel blot analysis revealed that the 11 Arabidopsis LSM genes are expressed in all organs tested, including leaves, roots, flowers and stems. Each pair of duplicated genes exhibited the same expression pattern (Figure 1A). To determine the expression of LSM genes at the tissue level, transgenic Arabidopsis plants containing fusions between all LSM promoters (LSMPRO) and the GUS reporter gene were generated and assayed for GUS activity. Consistent with the results obtained from the northern-blot experiments, nearly constitutive GUS activity was observed in all cases. In leaves and cotyledons, GUS staining was preferentially detected in the vascular tissues. As representative examples, the expression of LSM8PRO-GUS, LSM1APRO-GUS and LSM1BPRO-GUS is shown (Figures 1B-1E; Supplemental Figure 3 online). These results demonstrate that the 11 Arabidopsis LSM genes are active and ubiquitously expressed.

Subcellular localization of Arabidopsis LSM proteins
To investigate the subcellular localization of *Arabidopsis* LSM proteins, transgenic *Arabidopsis* expressing genomic LSM-green fluorescent protein (LSM-GFP) fusions driven by the corresponding LSM*PRO* were obtained and analyzed. We first examined the subcellular localization of LSM1A, LSM1B and LSM8, the *Arabidopsis* putative homologs of yeast and animal LSM proteins that differentiate the cytoplasmic and nuclear complexes, respectively. In root cells from seedlings expressing LSM1APRO-LSM1A-GFP or LSM1BPRO-LSM1B-GFP, green fluorescence suggested a cytoplasmic localization of LSM1A and LSM1B (Figure 2A). Conversely, in seedlings expressing the LSM8PRO-LSM8-GFP fusion, green fluorescence was specifically localized in nuclei (Figure 2A). We also investigated the subcellular localization of *Arabidopsis* LSM3A, LSM3B and LSM4, whose related yeast and animal proteins participate in both cytoplasmic and nuclear LSM complexes. In root cells from seedlings expressing LSM3APRO-LSM3A-GFP or LSM3BPRO-LSM3B-GFP, green fluorescence was detected in both nuclei and cytoplasm, indicating that LSM3A and LSM3B proteins simultaneously localize to these subcellular compartments (Figure 2A). Similar results were obtained when studying the subcellular localization of the LSM4-GFP fusion protein in seedlings expressing LSM4PRO-LSM4-GFP (Figure 2A). These data strongly suggest that *Arabidopsis* LSM proteins subcellularly localize as the LSM proteins from other eukaryotes.

Yeast and humans LSM1-7 proteins have been described to accumulate in P-bodies (Ingelfinger et al., 2002; Sheth and Parker, 2003). We therefore examined whether *Arabidopsis* LSM proteins belonging to the cytoplasmic complex also localize in these cytoplasmic foci. P-bodies are rarely observed in plants growing under control conditions, whereas its number and size markedly increase under conditions that are associated to high levels of mRNA turnover such as hypoxic or heat stress (Weber et al., 2008). When seedlings expressing LSM1APRO-LSM1A-GFP or LSM1BPRO-LSM1B-GFP were exposed to heat stress, LSM1A-GFP and LSM1B-GFP were largely localized to discrete cytoplasmic spots (Figure 2B). Under heat stress conditions, LSM3A-GFP, LSM3B-GFP and LSM4-GFP fusion proteins also localized to cytoplasmic foci in root cells from seedlings expressing LSM3APRO-LSM3A-GFP, LSM3BPRO-LSM3B-GFP or LSM4PRO-LSM4-GFP, respectively (Figure 2B). Following cycloheximide treatment, which in yeast and humans results in the loss of P-bodies (Sheth and Parker, 2003), no cytoplasmic foci were observed in any case (Figure 2C), suggesting that the detected
cytoplasmic spots of LSM-GFP fusion proteins corresponded to P-bodies. Consistent with its specific nuclear localization, LSM8-GFP did not accumulate in cytoplasmic spots in \textit{LSM8PRO-LSM8-GFP} expressing seedlings exposed to heat stress (Figure 2B) or to heat stress plus cycloheximide (Figure 2C). To confirm that the foci defined by \textit{Arabidopsis} cytoplasmic LSM proteins corresponded to P-bodies, we further analyzed their colocalization with DCP1, a protein that belongs to the \textit{Arabidopsis} decapping complex and accumulates in P-bodies (Xu et al., 2006; Goeres et al., 2007). The examination of seedlings expressing \textit{LSM1APRO-LSM1A-GFP} or \textit{LSM1BPRO-LSM1B-GFP} cotransformed with a \textit{35S-red fluorescent protein (RFP)-DCP1} fusion revealed that, in fact, LSM1A-GFP and LSM1B-GFP colocalized with RFP-DCP1 in root cells grown at room temperature (20ºC) or exposed to 37ºC (Figure 2D). Taken together, these observations evidence that \textit{Arabidopsis} cytoplasmic LSM proteins accumulate in P-bodies.

DCP2 and VCS, as DCP1, also belong to the \textit{Arabidopsis} decapping complex and accumulates in P-bodies (Xu et al., 2006; Goeres et al., 2007). Accordingly, GFP-DCP2 and GFP-VCS fusion proteins localized to P-bodies within the cytoplasm of root cells from wild-type (WT) seedlings containing \textit{35S-GFP-DCP2} or \textit{35S-GFP-VCS} constructs, respectively, exposed to heat stress (Figure 2E). When these constructs, were introduced into an \textit{Arabidopsis} mutant defective in LSM1 proteins (\textit{lsm1a lsm1b}; see below) under the same stress conditions, the P-bodies were lost and the GFP-DCP2 and GFP-VCS signals were mostly dispersed in the cytosol (Figure 2E). From these results we conclude that, in addition to accumulate in P-bodies, LSM1 proteins are required for P-body formation in \textit{Arabidopsis}.

**Organization of \textit{Arabidopsis} LSM proteins**

As mentioned above, yeast and animal LSM proteins typically exist as highly organized ring-shaped heptameric complexes (Figure 3A) (Tharun, 2009). Once established that \textit{Arabidopsis} LSM proteins subcellularly localize as in other eukaryotes, we decided to study how they are organized. For this, we conducted in vivo LSM-LSM interactions by means of bimolecular fluorescence complementation (BiFC) (Hu et al., 2002; Walter et al., 2004) in \textit{Nicotiana benthamiana} leaves. In yeast and animal LSM complexes, LSM1 and LSM8 are flanked by LSM2 and LSM4 (Figure 3A). Our experiments unveiled that a significant proportion of cells cotransformed with \textit{LSM1A-nGFP} and
LSM2-cGFP or LSM4-cGFP, and LSM8-nGFP and LSM2-cGFP or LSM4-cGFP displayed green fluorescence (Figure 3B). Identical results were obtained cotransforming LSM1B-nGFP with LSM2-cGFP or LSM4-cGFP (Supplemental Figure 4 online), indicating that Arabidopsis LSM1A, LSM1B and LSM8 are capable to interact in vivo with LSM2 and LSM4. Consistent with the typical cytoplasmic localization of LSM1 proteins in Arabidopsis (Figure 2A), LSM1(A or B)-LSM2 and LSM1(A or B)-LSM4 interactions mainly appeared in the cytoplasm of the Nicotiana cells (Figure 3B; Supplemental Figure 4 online). Conversely, interactions between LSM8 and LSM2 and LSM4 were essentially detected in the nucleus (Figure 3B), which is consistent with the characteristic nuclear localization of Arabidopsis LSM8 protein (Figure 2A). The specificity of all these interactions was demonstrated by the fact that, as expected from their different subcellular localization, we could not observe interaction between LSM1 proteins and LSM8 (Figure 3B; Supplemental Figure 4 online). Interactions between LSM2 and LSM4 proteins could not be noticed either (Figure 3B), in agreement to what has been proposed in yeast and animal LSM complexes (Figure 3A) (Beggs, 2005). We detected interactions, however, between LSM2 and LSM3 (A or B), LSM3 (A or B) and LSM6 (A or B), LSM6 (A or B) and LSM5, LSM5 and LSM7, and LSM7 and LSM4 (Figure 3C; Supplemental Figure 4 online). These interactions parallel those proposed in LSM complexes from other eukaryotes (Beggs, 2005) and were observed simultaneously in both cytoplasm and nucleus (Figure 3C; Supplemental Figure 4 online), consistently with the subcellular localization of the corresponding LSMs (see above). In addition, according also to the interactions assumed in other LSM complexes (Beggs, 2005), we did not detect interactions between LSM2 and LSM7, LSM4 and LSM6 (A or B), LSM6 (A or B) and LSM7, and LSM5 and LSM3 (A or B) (Figure 3C; Supplemental Figure 4 online). All these data indicate that Arabidopsis LSMs are organized in two heptameric ring complexes localized in cytoplasm (LSM1-7) and nucleus (LSM2-8).

In yeast and animals, cytoplasmic and nuclear LSM complexes are determined by the presence of LSM1 and LSM8 proteins, respectively (Tharun, 2009). The occurrence of a similar structural requirement in Arabidopsis complexes was examined by analyzing the subcellular distribution of LSM4, a protein marker of both Arabidopsis cytoplasmic and nuclear LSM complexes, in plants deficient in LSM1 and LSM8 proteins. As described above, in root cells from Arabidopsis seedlings containing the LSM4PRO-
**LSM4-GFP** construct, the LSM4-GFP fusion protein was simultaneously detected in both cytoplasm and nucleus (Figure 3D). Interestingly, however, in mutant seedlings for **LSM1** and **LSM8** (lsm1a lsm1b and lsm8, respectively) bearing the same construct the fusion protein preferentially localized in nuclei or cytoplasm, respectively (Figure 3D). These observations strongly support the notion that LSM1 and LSM8 proteins are essential for the formation of the cytoplasmic and nuclear LSM complexes, respectively, in *Arabidopsis*.

**LSM1- and LSM8-deficient Arabidopsis mutants display severe developmental alterations**

The results described above indicated that *Arabidopsis* LSMs are also organized in cytoplasmic and nuclear complexes determined by the presence of LSM1 and LSM8, and suggested similar functions as the complexes from yeast and animals. To test this assumption, we first searched for T-DNA insertion mutants in **LSM1** and **LSM8** genes. Plants containing single T-DNA insertions located in the fourth exon of **LSM1A** or in the second intron of **LSM1B** were identified (Figure 4A). **LSM1A** or **LSM1B** mRNAs were undetectable in homozygous plants for the insertions (Figure 4B), revealing that these new **LSM1A** and **LSM1B** alleles (lsm1a and lsm1b) were null or highly hypomorphic. Intriguingly, lsm1a and lsm1b plants did not present any obvious morphological or developmental abnormality, being indistinguishable from their corresponding wild-type plants, No-0 and Col-0 ecotypes, respectively (Supplemental Figure 5 online).

Since LSM1A and LSM1B are 80% identical (Supplemental Figure 1 online), they might be functionally redundant, which would explain the WT phenotypes exhibited by lsm1a and lsm1b single mutant plants. We, therefore, decided to obtain the lsm1a lsm1b double mutant that was subsequently backcrossed four times with Col-0 to have both mutations within this genetic background. As expected, lsm1a lsm1b plants did not accumulate **LSM1A** and **LSM1B** mRNAs (Figure 4B). Remarkably, in contrast to single mutants, the lsm1a lsm1b double mutant showed severe developmental alterations. Seed germination in lsm1a lsm1b was delayed compared to the WT and disturbed, producing epinastic, chlorotic and small cotyledons (Figures 4C and 4D; Supplemental Figure 6A online). Cotyledonary veins were disorganized with disruptions, preventing to form closed loops as in WT veins (Figure 4E; Supplemental Figure 6C online). lsm1a lsm1b
rosette and cauline leaves were smaller than WT leaves, more serrated, and presented an abnormal venation phenotype and smaller petioles (Figures 4F and 4G; Supplemental Figures 6D-6F online). The root system was also altered in lsm1a lsm1b plants, the root length and the number of secondary roots being reduced (Figure 4H; Supplemental Figures 6G and 6H online). On the other hand, the elongation of primary and secondary inflorescences ceased prematurely in the double mutant, altering plant architecture and given rise dwarf plants (Figure 4I). lsm1a lsm1b plants flowered earlier than WT plants under both long and short day photoperiods, though this phenotype was much more pronounced under non inductive photoperiodic conditions (Supplemental Figures 6I and 6J online). Finally, mutant plants produced few siliques that were shorter and contained less seeds than those of WT plants (Figure 4J; Supplemental Figures 6K and 6L online). These seeds, moreover, were small and frequently presented morphological alterations (Figure 4K). lsm1a lsm1b plants transformed with either LSM1APRO-LSM1A-GFP (c-lsm1a) or LSM1BPRO-LSM1B-GFP (c-lsm1b) recovered all above phenotypes (Figures 4C-4K; Supplemental Figures 6 and 7 online), confirming that LSM1A and LSM1B are, in fact, functionally redundant and that the mutant phenotypes displayed by the double mutant were due to the absence of LSM1A and LSM1B expression.

In addition, two transgenic lines were identified that contained single T-DNA insertions located in the fifth exon of LSM8 (Figure 5A). In homozygous plants for the insertions, LSM8 mRNA was undetectable, indicating that these new LSM8 alleles (lsm8-1 and lsm8-2) were null or highly hypomorphic (Figure 5B). lsm8-1 and lsm8-2 mutants also exhibited developmental defects (Figure 5; Supplemental Figure 6 online). Both of them showed the same phenotypes, but they were more pronounced in lsm8-1. Seeds from lsm8 mutants germinated as WT seeds, although a significant percentage of mutant seedlings exhibited alterations in the shape and number of their cotyledons, and veins formed more closed loops in lsm8 than in WT cotyledons (Figures 5C and 5D; Supplemental Figures 6A-6C online). lsm8-1 and lsm8-2 rosette leaves had short petioles and were smaller and flatter than WT leaves, but their vasculature and margins were normal (Figure 5E; Supplemental Figures 6D-6F online). Regarding the radicular system, the root length and the number of secondary roots were reduced in lsm8 mutants compared to the WT (Figure 5F; Supplemental Figure 6G and 6H online). The length of primary and secondary inflorescences was not affected in the mutants (Figure 5G). Nonetheless, they flowered significantly earlier than WT plants under short day
photoperiods (Supplemental Figures 6I and 6J online). Although the number of siliques produced in \textit{lsm8} mutants was as in the WT, they were shorter and contained less seeds that frequently aborted (Figures 5H and 5I; Supplemental Figures 6K and 6L online). \textit{lsm8-1} mutant plants transformed with the construct \textit{LSM8\textsubscript{PRO}-LSM8-GFP} (\textit{c-lsm8}) exhibited WT phenotypes (Figures 5C-5I; Supplemental Figure 6 online), confirming that their mutant phenotypes were due to the lack of \textit{LSM8} expression. Altogether, these data provide direct evidence that LSM1 and LSM8 proteins are required to ensure correct developmental transitions in \textit{Arabidopsis}, from germination to flowering, and also in seed formation.

\textbf{Accumulation of capped transcripts and mRNA stability is affected in \textit{lsm1a lsm1b} mutants}

The possibility that the \textit{Arabidopsis} LSM1-7 cytoplasmic complex functions in mRNA degradation, as described in yeast and animals (Bouveret et al., 2000; Tharun et al., 2000), was tested by analyzing the decay rates of several mRNAs that have been reported to be unstable transcripts, such as \textit{EXPL1}, \textit{ATHSPRO2}, \textit{JAZ6}, \textit{NIA2}, \textit{JAZ1}, and \textit{RAV1} (Gutiérrez et al., 2002), in \textit{lsm1a lsm1b} and WT plants. As a control, we also analyzed the turnover of \textit{EIF4A1} mRNA, which is considered a stable transcript (Gutiérrez et al., 2002). Decay rates were assayed by comparing relative levels of mRNAs following cordycepin-induced transcriptional arrest (Gutiérrez et al., 2002). Our results confirmed the unstability of the former mRNAs and the stability of the latter in the WT (Figure 6A). In \textit{lsm1a lsm1b}, however, the steady state levels of all unstable transcripts analyzed were higher than in WT plants and their rates of decay clearly reduced, their estimated half-lives (the time required for an mRNA to be reduced to half its initial value) being at least two times longer (Figures 6A and 6B). As expected, the steady state levels and the decay rate of \textit{EIF4A1} RNA were similar in mutant and WT plants (Figures 6A and 6B). The analysis of the stability of \textit{EXPL1}, \textit{ATHSPRO2}, \textit{JAZ6} and \textit{EIF4A1} transcripts in \textit{c-lsm1a} and \textit{c-lsm1b} plants confirmed that LSM1A and LSM1B are functionally redundant, and demonstrated that the increased mRNA stability noticed in \textit{lsm1a lsm1b} was caused by the simultaneous absence of \textit{LSM1A} and \textit{LSM1B} expression (Figures 6C and 6D; Supplemental Figures 8A and 8B online). We also examined the stability of \textit{EXPL1}, \textit{JAZ6} and \textit{EIF4A1} mRNAs in \textit{lsm1a}, \textit{lsm1b} and \textit{lsm8-1} single mutants. As presumed, all mRNAs showed similar turnover in cordycepin-treated WT and mutant plants (Supplemental Figures 8C-8H online),
confirming again the functional redundancy of LSM1A and LSM1B, and establishing that the *Arabidopsis* LSM2-8 nuclear complex has not a role in cytoplasmic mRNA degradation.

We next assessed whether the reduction of mRNA decay observed in *lsm1a lsm1b* could be due to a deficiency in its mRNA decapping capacity. For this, a rapid amplification of cDNA ends (RACE)-PCR that allow detecting capped forms of specific mRNAs was used. PCR experiments with low and high number of cycles were carried out. In both cases, we found that *EXPL1*, *ATHSPRO2* and *JAZ6* mRNAs accumulated in their capped form in the *lsm1a lsm1b* mutant compared with the WT (Figure 6E). These effects were corrected by *LSM1A* and *LSM1B* transgenes in c-*lsm1a* and c-*lsm1b* plants, respectively (Figure 6E). The cap forms of mRNAs corresponding to the above genes were found not changed in *lsm1a*, *lsm1b* and *lsm8-1* mutants (Supplemental Figure 8I online). These results indicated that the *Arabidopsis* LSM1-7 complex operates in cytoplasmic mRNA degradation by promoting decapping.

**Loss of LSM8 influences U6 snRNA stability and results in pre-mRNA splicing defects**

In yeast and animals, the LSM2-8 nuclear complex acts in pre-mRNA splicing by stabilizing the spliceosomal U6 snRNA (Beggs, 2005). To determine whether the *Arabidopsis* LSM nuclear complex has a similar function, we first analyzed the effects of LSM8 on pre-mRNA splicing at genome-wide level using tiling arrays (Affymetrix *Arabidopsis* Tiling 1.0R) and total RNAs from WT and *lsm8-1* mutant plants. Two-week-old plants were selected for these experiments as they represent an intermediate stage of development. We searched for introns with significantly higher hybridization signals in mutant than in WT plants. Thus, we identified 469 introns, belonging to 453 genes, with increased hybridization signals in *lsm8-1* (See Supplemental Table 1 online). The increased hybridization signals detected in *lsm8-1* should reflect intron retention since hybridization signals in other introns and exons of the genes did not differ between WT and *lsm8-1* plants. These results were validated by RT-PCR for a subset of genes appertaining to different ontology categories, including protein metabolism (*AT1G17960*), intracellular transport (*AT3G59390*), developmental processes (*PRMT4A*) or signal transduction (*CKL5*, *AME3*), in both *ls8-1* and *ls8-2* mutants (Figure 7A; Supplemental Figure 9 online). The intron retention events in these
genes, however, were not detected in c-lsm8 and lsm1a lsm1b plants (Figure 7A), confirming that the splicing defects unveiled in lsm8 mutants were specifically due to the loss of LSM8 function, and that the Arabidopsis LSM1-7 cytoplasmic complex is not involved in pre-mRNA splicing. As a control, tiling array data were also validated by analyzing the retention of an intron of ARP4, a gene that did not display any intron retention event in the array, in lsm8-1 and lsm8-2 mutants. As expected, the intron was not retained in these plants (Figure 7A).

Next, we investigated the possible role of Arabidopsis LSM nuclear complex in U6 snRNA stability by assessing the levels of this snRNA in cordycepin-treated lsm8 mutant and WT plants. Results revealed that the steady state levels of U6 snRNA were lower in mutants than in WT plants, and that after cordycepin treatment these levels were maintained in WT plants but decreased rapidly in lsm8-1 and lsm8-2 mutants (Figure 7B). Therefore, the stability of U6 snRNA is dependent on the presence of LSM8 and, consequently, on the LSM2-8 nuclear complex. The effect of LSM8 on U6 snRNA stability seems to be highly specific since the levels of U3 snoRNA, which is transcribed by RNA polymerase III as the U6 snRNA, and U4 snRNA, which is synthesized by RNA polymerase II, did not decrease in cordycepin-treated lsm8 mutants (Figure 7B). As expected, c-lsm8 and lsm1a lsm1b plants showed similar levels of U6 snRNA, U3 snoRNA and U4 snRNA as the WT before and after cordycepin treatment (Figure 7C; Supplemental Figure 10 online). Therefore, it was concluded that the Arabidopsis LSM2-8 nuclear complex is essential for adequate splicing of selected mRNAs through the stabilization of the spliceosomal U6 snRNA.

Arabidopsis mutants deficient in LSM1 or LSM8 proteins exhibit altered developmental-related gene expression

In an attempt to understand the function of LSM complexes in Arabidopsis development, we studied the global impact of lsm1 and lsm8 mutations on gene expression. The comparison of mRNA profiles from lsm1a lsm1b and WT was performed using Agilent Arabidopsis Oligo Microarrays v4 and total RNAs extracted from 2-week-old plants. Transcript levels of 358 genes were found to be higher, by at least twofold, in lsm1a lsm1b than in the WT (Supplemental Table 2 online). On the other hand, transcripts corresponding to 316 genes were reduced by more than twofold in lsm1a lsm1b compared with the WT (Supplemental Table 2 online). Gene ontology
analysis of deregulated genes in the double mutant unveiled that 72 of them were implicated in developmental processes, including seed germination, root and leaf development, inflorescence development, flowering and embryogenesis (Supplemental Table 2 online), which is consistent with its severe mutant phenotype (Figures 4C-4K). The microarray data were validated confirming the altered expression of several overexpressed and underexpressed genes related to different developmental processes in lsm1a lsm1b plants by qRT-PCR (Figures 8A and 8B). On the other hand, c-lsm1a and c-lsm1b plants exhibited WT expression patterns for all validated genes (Figures 8A and 8B), demonstrating that the LSM1-7 cytoplasmic complex is required for the accurate expression of developmental-related genes in Arabidopsis.

Since the Arabidopsis LSM cytoplasmic complex functions in mRNA degradation by promoting decapping (see above), the high levels of some developmental-related mRNAs detected in the absence of LSM1 proteins might be due to a selective stabilization of the corresponding transcripts as a result of the retention of their 5’ cap. This possibility was first examined by measuring the degradation rates of five developmental-related mRNAs (YLS9, UGT87A2, ATEXP14, MEE14 and ATHB12), whose levels were elevated in the lsm1a lsm1b double mutant, in cordycepin-treated WT and lsm1a lsm1b plants. Interestingly, the decay of all transcripts, except ATHB12, was significantly much slower in the mutant than in WT plants (Figure 8C). In addition, all transcripts, but not ATHB12, retained their 5’ cap in lsm1a lsm1b (Figure 8D), providing evidence that, in fact, the Arabidopsis LSM1-7 cytoplasmic complex is essential for a correct developmental-related gene expression by controlling the decapping and, therefore, the stabilization of specific developmental-related transcripts. Accordingly, the degradation rates and cap levels of YLS9, UGT87A2, ATEXP14, MEE14 and ATHB12 transcripts in c-lsm1a and c-lsm1b plants were as in the WT (Figures 8C and 8D).

The effect of lsm8 mutations on gene expression at a genome-wide level was determined analyzing the above-mentioned tiling arrays, which, in addition to allow splicing analysis, constitute a robust platform for detection of transcriptional activity (Laubinger et al., 2008). Compared to the WT, 65 and 193 annotated genes were found to be at least twofold up- and down-regulated, respectively, in the lsm8-1 mutant (See Supplemental Table 3 online). Gene ontology categorization of these deregulated genes
revealed that a representative number (17 upregulated and 50 downregulated) was related to developmental processes throughout the Arabidopsis life cycle (See Supplemental Table 3 online), which could explain the mutant phenotypes exhibited by lsm8 mutants (Figures 5C-5I). Microarray results were validated by assaying the expression of a group of deregulated genes implicated in different developmental processes in WT, lsm8-1, lsm8-2 and e-lsm8 plants by qRT-PCR (Figures 9A and 9B). These data demonstrated that the Arabidopsis LSM2-8 nuclear complex is also crucial for an appropriate developmental-related gene expression.

Considering that the LSM2-8 nuclear complex regulates developmental-related gene expression and functions in pre-mRNA splicing (see above), it was presumed that a number of genes involved in development might display splicing defects. Remarkably, 65 out of the 453 genes that showed intron retention events resulted to be related to different developmental processes (See Supplemental Table 1 online). The inefficient splicing of some of these genes, including ASU1/DCL1, OLI2, EMB2785, EMB2016 and GUT2 in the absence of LSM8 was confirmed by RT-PCR analysis with appropriate primers (Figure 9C). As expected, the splicing of other developmental-related genes, such as ANT, was not affected (Figure 9C). These findings indicate that the Arabidopsis LSM2-8 nuclear complex ensures the accurate splicing of specific developmental-related mRNAs, allowing a correct developmental-related gene expression.
DISCUSSION

Although LSM-related proteins have been found in the genomes of different plant species (Proost et al., 2009; Goodstein et al., 2012), they have not yet been biochemically characterized and their function in RNA metabolism remained to be established. In this study, we used genetic, molecular, cell biology and biochemical studies to demonstrate that *Arabidopsis* LSMs are organized in two heptameric complexes. More important, our results reveal that these complexes are essential for normal *Arabidopsis* development, and this role seems to be carried out by controlling the proper turnover and splicing of selected developmental-related mRNAs which, in turn, ensures the appropriate gene expression during plant development.

Subcellular localization and BiFC experiments strongly support the idea that *Arabidopsis* LSM proteins assemble into two heteroheptameric complexes that differ by a single subunit, LSM1A/B or LSM8, and localize in cytoplasm (LSM1-7) and nucleus (LSM2-8). First, in *Arabidopsis*, as in other eukaryotes (Beggs, 2005), LSM1 proteins (LSM1A and LSM1B) specifically accumulate in the cytoplasm, while LSM8 has a nuclear localization and the rest of LSM proteins are simultaneously localized in cytoplasm and nucleus. Second, *Arabidopsis* LSM proteins do not interact promiscuously with each other. Instead, each LSM specifically interacts with two other LSM proteins following the same pattern of interaction as in the yeast and human heptameric complexes (Beggs, 2005). Consistent with their different subcellular localization and with the assumption that they define the two *Arabidopsis* LSM complexes, LSM1 and LSM8 proteins do not interact with each other. Moreover, while all interactions involving LSM1 proteins take place in the cytoplasm, those involving LSM8 occur into the nucleus and those implying LSM2-7 proteins occur simultaneously in both subcellular compartments. Third, LSM1 and LSM8 proteins are required for the formation of the *Arabidopsis* cytoplasmic and nuclear LSM complexes, respectively.

Our genetic and molecular analyses allowed us to establish the function of *Arabidopsis* LSM complexes. In *Arabidopsis* plants deficient in LSM1 proteins, several transcripts accumulate in their capped forms and show a reduced degradation rate with the corresponding increase in their half-lives, indicating that the *Arabidopsis* LSM1-7 complex function in the 5’ to 3’ pathway of mRNA decay as an activator of decapping. As expected from their high amino acid identity (80%), LSM1A and LSM1B are
functionally redundant. *lsm1a* and *lsm1b* single null mutants are not perturbed in mRNA decapping and decay, and LSM1A and LSM1B, individually, are able to complement the alterations in mRNA decapping and decay displayed by the *lsm1a lsm1b* double mutant. On the other hand, *Arabidopsis* plants lacking LSM8 are affected in the stability of the spliceosomal U6 snRNA which, accordingly, results in pre-mRNA splicing defects. Compared to WT, at least 469 intron retention events distributed among 453 genes were detected in the *lsm8-1* mutant under our experimental conditions, evidencing that the LSM2-8 complex regulates genome-wide pre-mRNA splicing. Although intron retention constitutes the most frequent splicing defect in plant genes (Syed et al., 2012), it is probable that other mRNA splicing defects, including exon skipping, alternative 5’ splicing and alternative 3’ splicing, also occur in the absence of LSM8. Unfortunately, however, the detection of these defects is unreliable when using tiling arrays to analyse pre-mRNA splicing at global level (Ner-Gaon and Fluhr, 2006). The existence of splicing defects has also been disclosed in some genes of a postembryonic lethal mutant for *LSM4* (Zhang et al., 2011). Nonetheless, only one of these genes (*AT1G28060*) has been found in our tiling analysis of the *lsm8-1* mutant, in all likelihood because of the plants used being at different developmental stages and the different methods of analysis being used. Furthermore, only a few genes were analyzed for splicing defects in the *lsm4* mutant (Zhang et al., 2011). The fact that not all mRNAs exhibit reduced degradation rates in *lsm1a lsm1b* plants nor splicing defects in *lsm8* mutants, indicates that the cytoplasmic and nuclear LSM complexes from *Arabidopsis*, as described for other components of the *Arabidopsis* machineries involved in mRNA degradation and processing (Xu et al., 2006; Goeres et al., 2007; Xu and Chua, 2009; Kim et al., 2010; Rymarquis et al., 2011), act on selected targets. How selected mRNAs are targeted to these complexes remains largely unknown. According to their relevant function in mRNA decapping and degradation, the lack of LSM1 proteins has a deep impact on *Arabidopsis* gene expression, the levels of more than 600 transcripts being significantly altered, 358 increased and 316 reduced, in *lsm1a lsm1b* plants. Similarly, the expression of at least 250 genes is significantly affected in null mutants for *LSM8*. Consistent with the implication of *Arabidopsis* LSM8 in pre-mRNA splicing, in this case the number of down-regulated (193) genes is much higher than that of up-regulated (65) ones. These data indicate that, as other factors involved in the *Arabidopsis* decapping 5’ to 3’ decay pathway, including DCP2, DCP5 and XRN4, or in pre-mRNA splicing, such as STA1 (Lee et al., 2006; Xu et al., 2006; Goeres et al., 2007; Xu and Chua, 2009; Rymarquis et
al., 2011), the *Arabidopsis* LSM1-7 and LSM2-8 complexes also play a major task in maintaining appropriate levels of gene expression. Interestingly, however, the result of the absence of these factors on *Arabidopsis* gene expression seems to be highly specific.

In eukaryotic cells, P-bodies appear as cytoplasmic foci containing RNP complexes associated to translational repression, mRNA storage and cytoplasmic mRNA decay pathways (Xu and Chua, 2011). Under conditions promoting high levels of mRNA turnover, such as osmotic, hypoxic or heat stress conditions, P-bodies increase in number and size being more apparent (Teixeira et al., 2005; Weber et al., 2008). Nevertheless, it is not yet clear how P-bodies are formed and what is their function in eukaryotic cells. Human LSM4 localizes in P-bodies and loses this localization when mutations are introduced in residues involved in interacting with other LSM proteins (Ingelfinger et al., 2002). In yeast, LSM2 and LSM7 fail to localize to P-bodies in LSM1 deficient cells (Tharun et al., 2005), and it has been shown that LSM4 plays a role in the localization of the LSM1-7 complex in P-bodies and in P-body assembly (Decker et al., 2007; Reijns et al., 2008). Our findings demonstrate that the *Arabidopsis* LSM1-7 complex not only accumulates in P-bodies, which is consistent with its function in cytoplasmic mRNA decapping and decay, but is also essential for their formation. As expected from its specific nuclear localization, LSM8 does not localize in P-bodies. The implication of *Arabidopsis* LSM2-LSM7 proteins in P-body formation is difficult to assess due to the absence of viable *lsm2-lsm7* null mutants (see below). To date, only few proteins have been related with P-bodies in plants, including DCP5, an *Arabidopsis* protein indirectly implicated in regulating mRNA decapping that has a function in P-body formation (Xu and Chua, 2009), and DCP1, DCP2 and VCS, three proteins that constitute a decapping complex and colocalize with P-bodies in *Arabidopsis* (Xu et al., 2006; Goeres et al., 2007). In addition, *Arabidopsis* proteins XRN4, AtTZF1 and PMBs have also been found in plant P-bodies (Weber et al., 2008; Pomeranz et al., 2010; Stauffer et al., 2010). The identification of additional P-body components will certainly help to understand how they are formed and what is their function in plant cells.

It has been described that the proteins involved in mRNA decapping DCP1, DCP2, VCS and DCP5, as well as the splicing factors STA1, U11/U112-31K and EMU, play an essential role in *Arabidopsis* development, their absence being lethal (Lee et al.,
Arabidopsis plants deficient in LSM1 and LSM8 proteins also display quite severe development alterations but they are not lethal. \textit{lsm1a lsm1b} and \textit{lsm8} mutants are affected in both vegetative and reproductive developmental traits, indicating that cytoplasmic and nuclear LSM complexes are required for the normal development of \textit{Arabidopsis} throughout the different phases of its life cycle. Nevertheless, consistent with the different function of the two LSM complexes, the phenotypes exhibited by \textit{lsm1a lsm1b} and \textit{lsm8} mutants are different. Recently, T-DNA insertional mutants for \textit{LSM4} and \textit{LSM7} have been described to show postembryonic and embryonic lethality, respectively (Zhang et al., 2011; \url{http://www.seedgenes.org/}). We have observed the same lethal phenotype not only in the \textit{lsm4} and \textit{lsm7} null mutants but also in the \textit{lsm3a lsm3b} and \textit{lsm6a lsm6b} doubles, the corresponding single mutants exhibiting WT phenotypes, as well as in the \textit{lsm1a lsm1b lsm8} triple mutant (Perea-Resa et al., in preparation). Moreover, we have not found any insertion abolishing the expression of \textit{LSM2} and \textit{LSM5} genes in the available T-DNA collections, which suggests that, probably, \textit{lsm2} and \textit{lsm5} null mutants are also lethal. Weak mutant alleles for these genes, however, do not appear to be lethal. In fact, a point mutation in \textit{LSM5} (\textit{sad1}) that provokes the conversion of a glutamic acid residue to a lysine makes mutant plants much smaller than the WT but does not result in lethality (Xiong et al., 2001). Altogether, these data indicate that the presence of at least one LSM complex is essential in \textit{Arabidopsis}. In yeast, it has been proposed that LSM2-7 proteins might associate, in the apparent absence of LSM1 or LSM8, with other proteins, including related SM proteins, to form complexes that would remain at least partially active (Beggs, 2005). We can not exclude that this could be the case in \textit{lsm1a lsm1b} and \textit{lsm8} mutants. Further studies are required to understand how \textit{Arabidopsis} can develop and reproduce with just one LSM complex.

The results presented in this work demonstrate that post-transcriptional regulation has an important role in controlling gene expression related to plant development. In fact, we show that several selected genes involved in both vegetative and reproductive development are targets of the \textit{Arabidopsis} LSM complexes. Thus, the LSM1-7 cytoplasmic complex ensures the precise half-life of the transcripts corresponding to its targets, for instance \textit{UGT87A2} (floral transition), \textit{MEE14} (embryo development) or \textit{YLS9} (leaf development), and, consequently, their adequate temporal expression.
patterns. The LSM2-8 nuclear complex, in turn, guarantees the correct splicing of its targets, such as ASU1/DCL1 (flower development), OLI2 (leaf development) or EMB2785 (embryo development), and, therefore, the accurate translation of the corresponding transcripts. Furthermore, we also show that, consistent with their role in turnover and splicing of developmental-related mRNAs, the Arabidopsis LSM complexes control the expression levels of many genes that are implicated in different developmental processes, including seed germination, root development, leaf development, floral transition, flower development and embryogenesis. In particular, the expression levels of 72 and 67 specific genes involved in development were found to be altered in lsm1a lsm1b and lsm8-1 mutants, respectively. It is obvious, however, that these numbers should be considerably higher taking into account that only plants from one developmental stage (2-week-old) were analyzed by microarray experiments. We propose that the cumulative defects in gene expression are responsible for the abnormal developmental phenotypes observed in these plants.

In conclusion, the findings presented here unveil the organization and function of Arabidopsis LSM proteins, and demonstrate that these proteins are crucial for plant growth and development. Understanding the molecular mechanisms that regulate the function of LSMs and confer their target specificity constitutes an interesting challenge for the future.
METHODS

Plant material, constructs and growth conditions

*Arabidopsis* Columbia (Col-0) ecotype and mutants *lsm8-1* (Salk-025064) and *lsm8-2* (Salk-048010) were obtained from NASC. Mutant *lsmlb* is a Gabi-kat line from Max Plank Institute (GK 391E05). *Arabidopsis* Nossen (No-0) ecotype and mutant *lsmla* (12-2253-1) were obtained from Riken Institute. *lsmla* is a Ds-transposon insertion line in No-0 background. *lsmla* and *lsmlb* single mutants were crossed to generate a *lsmla lsm1b* double mutant that was subsequently backcrossed four times with Col-0 to have both mutations within this genetic background. Transgenic Col-0 plants containing the *35S-GFP-DCP2* and *35S-GFP-VCS* constructs (Goeres et al., 2007) were kindly provided by Leslie Sieburth (University of Utah, Salt Lake City, USA). These plants were crossed with *lsmla lsm1b* to obtain double mutants with the *35S-GFP-DCP2* and *35S-GFP-VCS* constructs in homozygosis. All mutant and transgenic lines were genotyped using the primers listed in Supplemental Table 4 online.

To obtain the *LSMsPRO-GUS* fusions, at least 1kb promoter fragment from each of the 11 *Arabidopsis* *LSM* genes was cloned into the pBI101 binary vector (Clontech). For the *LSMsPRO-LSM-GFP* fusions, genomic regions containing the *LSM1A, LSM1B, LSM3A, LSM3B, LSM4* and *LSM8* genes, including at least 1kb of the corresponding promoters, were cloned into the pGWB4 gateway binary vector (Nakagawa et al., 2007). All fusions were verified by sequencing and introduced in Col-0 via *Agrobacterium tumefaciens* C58C1 using the floral dip method (Clough and Bent, 1998). Fusions *LSM1APRO-LSM1A-GFP* and *LSM1BPRO-LSM1B-GFP*, and *LSM8PRO-LSM8-GFP* were also introduced in *lsmla lsm1b* and *lsm8-1* mutants, respectively. Similarly, the *LSM4PRO-LSM4-GFP* fusion was used to transform *lsmla lsm1b* and *lsm8-1* mutants. Finally, the fusion *35S-RFP-DCP1* (Weber et al., 2008), kindly provided by Markus Fauth (Johann Wolfgang Goethe-University Frankfurt, Germany), was introduced in transgenic lines containing *LSM1APRO-LSM1A-GFP* or *LSM1BPRO-LSM1B-GFP*. All transgenic lines were genetically determined to have the constructs integrated at a single locus in homozygosis. For BiFC assays, full-length cDNAs corresponding to the 11 *LSM* genes were amplified with appropriate primers (see Supplemental Table 4 online) to incorporate convenient restriction sites at their 5’ and 3’ ends. Fragments were cloned into the pSPYNE-35S and pSPYCE-35S binary vectors (Walter et al., 2004), kindly provided by Jörg Kudla (Westfälische Wilhelms-
Universität Münster, Germany), sequenced and introduced in *A. tumefaciens* C58C1 for subsequent agroinfiltration. Agroinfiltration was performed in leaves from 3-week-old plants of *N. benthamiana* grown at 25°C, essentially as described (English et al., 1997) without using a silencing suppressor. The expression of fusion proteins was subsequently assayed 3 days after agroinfiltration.

Plants were grown at 20°C under long-day photoperiods (16h of cool-white fluorescent light, photon flux of 90µmol m⁻² s⁻¹) in pots containing a mixture of organic substrate and vermiculite (3:1 v/v) or in Petri dishes containing GM medium, MS supplemented with 1% sucrose and solidified with 0.8% (w/v) agar. Plants used to estimate flowering time in short-day conditions were grown under an 8h light regime.

**Gene expression analysis**
Total RNA was extracted using the Purezol reagent (Bio-Rad) according to the manufacturer protocol. RNA samples were treated with DNase I (Roche) and quantified with a Nanodrop spectrophotometer (Thermo Scientific). RNA-blot hybridizations were carried out according to standard procedures. Specific probes were obtained by PCR with the primers described in Supplemental Table 4 online and labelled with [α-³²P]dCTP using the Megaprime DNA labelling systems kit (GE Healthcare). Equal RNA loading in the experiments was monitored by rRNA staining. RNA samples for each experiment were analyzed in at least three independent blots, and each experiment was repeated at least twice. For real-time RT-PCRs, cDNAs were prepared with the iScript cDNA Synthesis Kit (Bio-Rad) and then amplified using the Bio-Rad iQ2 thermal cycler, the SsoFast EvaGreen Supermix (Bio-Rad), and gene-specific primers (Supplemental Table 4 online). The relative expression values were determined using the AT4G24610 gene as a reference (Czechowski et al., 2005). All reactions were realized in triplicate employing three independent RNA samples.

**Determination of GUS activity**
GUS activity in *Arabidopsis* transgenic plants containing the fusions *LSMsPRO-GUS* was detected and measured as previously described (Medina et al., 2001).

**Microscopic analysis**
Subcellular localization of fusion proteins in transgenic *Arabidopsis* was performed in
roots from 6-day-old seedlings grown in vertical position on GM medium. Heat treatment was carried out by transferring seedlings to 37°C for two hours. Treatment with cycloheximide was performed by incubating seedlings in liquid MS medium supplemented with 200µg/ml of cycloheximide for two hours at 37°C. Transient expression of fusion proteins in leaves of 3-week-old plants of *N. benthamiana* was assayed 3 days after agroinfiltration as described above. Microscopy images were collected using a Confocal Laser Spectral microscope TCS SP2 (Leica Mycrosystem, Wetzlar, Germany). The excitation lines for imaging GFP and RFP fusions were 488 and 561 nm, respectively.

**Cordycepin treatments, mRNA half-life estimations and capped mRNA analysis**

Six-day-old seedlings and 2-week-old plants were used for cordycepin treatment, essentially as described (Gutiérrez et al., 2002). Samples were collected at the indicated time points and total RNA was extracted by using the Purezol reagent (Bio-Rad). Gene expression was analysed by RNA-blot hybridizations or real time RT-PCR as described above. To examine U6 snRNA, U3 snoRNA and U4 snRNA decay, additional cordycepin was added to the samples at 9 and 24 hours to ensure transcriptional repression. For graphical representation of mRNA stability and mRNA half-life estimation, the hybridization bands were quantified with the ImageJ software (NIH) and values were normalized to WT time 0.

To determine if accumulating mRNAs were capped, RNA ligase-mediated (RLM) RACE was performed using the First Choice RLM-RACE kit (Ambion) following manufacturer specifications. RNAs were extracted from 6-day-old seedlings or 2-week-old plants with the RNeasy kit (Quiagen) and PCRs were performed by using low (20-25) or high (30-32) number of cycles. Specific primers for the 5’RACE Adapter and for the genes tested are described in Supplemental Table 4 online.

**Microarray analysis**

Total RNA from 2-week-old Col-0, *lsm1a lsm1b* and *lsm8-1* plants was extracted by using the RNeasy Kit (Quiagen), and 3 biological replicates were independently hybridized per transcriptomic comparison. For microarray analysis of the *lsm1a lsm1b* mutant, RNA amplification and labelling were carried out basically as described (Goda et al., 2008). Hybridization was performed on Agilent *Arabidopsis* Oligo Microarrays.
v4 (catalogue number G2519F-V4021169) in accordance with the manufacturer specifications. The statistical significance of the results was evaluated with the FIESTA software (http://bioinfogp.cnb.csic.es). Genes with an FDR-corrected $P$ value lower than 0.05 and a fold change of more or less than 2 were selected for consideration. Data from these microarray experiments have been deposited in the Gene Expression Omnibus database under accession number GSE39630.

For microarray analysis of the *lsm8-1* mutant, double stranded cDNAs were synthesized, processed and labeled with the GeneChip® Whole Transcript (WT) Double-Stranded Target Assay Kit (Affymetrix), following the manufacturer instructions. Labeled cDNAs were used to hybridize Affymetrix *Arabidopsis* Tiling 1.0R arrays (catalogue number 900594). Data were analyzed with Tiling Analysis Software (TAS) from Affymetrix using TAIR7 as reference annotation (BPMAP file). To detect altered gene expression, genes with at least one exon identified as significantly over- or under-expressed ($P$ value lower than 0.05 and a fold change of more or less than 2) were considered. A gene was accepted as differentially expressed when the 10% trimmed mean of the signals of all probes in its exons and UTRs was at least twofold higher or lower in the mutant than in the WT. For those genes with splicing variants, only the constitutive exons were considered. Similarly, introns with significantly higher signals in the mutant than in the WT were initially considered to be intron retention events. For high confidence, only the introns covered with a minimum of three probes and average signals over twofold were selected (Supplemental Table1 online). Data from these microarray experiments have been deposited in the Gene Expression Omnibus database under accession number GSE39617.

**Intron retention analysis**

Total RNA from 2-week-old plants was extracted with Purezol (Bio-Rad) and used for cDNAs generation with the iScript cDNA Synthesis Kit (Bio-Rad). Intron retention was revealed by RT-PCR using a pair of specific primers for each gene tested (Supplemental Table 4 online). One primer was situated inside the retained intron and the second one in an adjacent exon. All PCR reactions were performed using RNA with (+RT) or without (-RT) reverse transcriptase to detect genomic DNA contaminations. Genomic DNA was included in all reactions as a positive control, and *TUBULIN* expression level was used as a loading control.
Accession numbers
Sequence data from this article can be found in the Arabidopsis Genome Initiative or GenBank/EMBL data libraries under the accession numbers described in Supplemental Table 5 online. The microarray data were submitted to the Gene Expression Omnibus site (www.ncbi.nlm.nih.gov/geo) under accession numbers GSE39630 and GSE39617.

Supplemental Data
The following materials are available in the online version of this article.

Supplemental Figure 1. Sequence alignment of Arabidopsis LSM proteins.
Supplemental Figure 2. Phylogenetic analysis of plant LSM proteins.
Supplemental Figure 3. Expression patterns of LSM1A and LSM1B genes.
Supplemental Figure 4. Visualization of in vivo interactions between Arabidopsis LSM proteins by BiFC assays.
Supplemental Figure 5. Phenotypical analysis of lsm1a and lsm1b single mutants.
Supplemental Figure 6. Quantification of developmental phenotypes shown by lsm mutants.
Supplemental Figure 7. Complementation of the lsm1a lsm1b double mutant by LSM1B.
Supplemental Figure 8. mRNA stability and accumulation of capped transcripts in c-lsm1b, lsm1a, lsm1b and lsm8-1 plants.
Supplemental Figure 9. Tilling array hybridization signals in representative genes showing intron retention events in the lsm8-1 mutant.
Supplemental Figure 10. Stability of U6 snRNA in the lsm1a lsm1b double mutant.

Supplemental References.

Supplemental Table 1. Intron retention events in the lsm8-1 mutant.
Supplemental Table 2. Genes whose expression is up- and down-regulated in the lsm1a lsm1b double mutant.
Supplemental Table 3. Genes whose expression is up- and down-regulated in the lsm8-1 mutant.
Supplemental Table 4. Oligonucleotide sequences of primers used in this study.
Supplemental Table 5. Accession numbers of the sequence data presented in this study.
ACKNOWLEDGEMENTS

We thank M. Fauth for the 35S-RFP-DCP1 fusion, L. Sieburth for the transgenic plants containing the fusions 35S-GFP-DCP2 and 35S-GFP-VCS, J. Kudla for the pSPYNE-35S and pSPYCE-35S vectors, J. M. Franco, J. Froment and P. Gonzalez-Garcia for tiling array data analysis, and J. A. Jarillo and R. Solano for discussions and comments. This work was supported by grants CSD2007-00057, EUI2009-04074 and BIO2010-17545 from the Spanish Secretary of Research, Development and Innovation.

AUTHOR CONTRIBUTIONS

C. Perea-Resa and T. Hernández-Verdeja designed the research, performed research and analyzed data. R. López-Cobollo and M. M. Castellano performed research. J. Salinas designed the research, analyzed data and wrote the paper.
FIGURE LEGENDS

Figure 1. Expression patterns of Arabidopsis *LSM* genes.

(A) Expression analysis of *LSM* genes in different organs of *Arabidopsis* by RNA hybridization using specific probes. Total RNA (20 µg) from 4-week-old rosette leaves (L), roots (R), flowers (F), and stems (S) were used. *rRNA* levels are shown as a loading control.

(B) to (E) GUS activity in *Arabidopsis* plants containing the fusion *LSM8*<sub>PRO</sub>-GUS. Whole plant (B), root (C), cross-section of a stem (D), and flower (E).

Figure 2. Subcellular localization of *Arabidopsis* *LSM* proteins.

(A) to (C) Subcellular localization of different *LSM*-GFP proteins in root tip cells from 6-day-old *Arabidopsis* seedlings. Seedlings grown under control conditions (A), seedlings grown under control conditions and subsequently exposed 2h at 37ºC (B), and seedlings grown under control conditions and subsequently exposed 2h at 37ºC with cycloheximide (CHX) (C). Bars = 10 µm.

(D) Co-localization of *LSM1A*-GFP and *LSM1B*-GFP with RFP-DCP1 in root tip cells from 6-day-old *Arabidopsis* seedlings grown under control conditions (upper panel) and subsequently exposed 2h at 37ºC (lower panel). Bars = 10 µm.

(E) Subcellular localization of GFP-DCP2 and GFP-VCS in root tip cells from 6-day-old wild-type (WT) and *lsm1a lsm1b* *Arabidopsis* seedlings grown under control conditions and subsequently exposed 2h at 37ºC. Bars = 10 µm.

Figure 3. Organization of *Arabidopsis* *LSM* proteins.

(A) Cellular model showing cytoplasmic and nuclear heptameric *LSM* complexes as described in yeast and humans.

(B) and (C) Visualization of *in vivo* interactions between *Arabidopsis* *LSM* proteins by BiFC assays. The corresponding *LSM*-nGFP/*LSM*-cGFP proteins were pairwise tested by *Agrobacterium*-mediated transformation in *Nicotiana benthamiana* leaves. Interactions between *LSM1A/LSM2*, *LSM1A/LSM4*, *LSM1A/LSM8*, *LSM8/LSM2*, *LSM8/LSM4*, and *LSM2/LSM4* (B), and *LSM2/LSM3A*, *LSM6A/LSM3A*, *LSM6A/LSM5*, *LSM5/LSM7*, *LSM7/LSM4*, *LSM2/LSM7*, *LSM4/LSM6A*, *LSM6A/LSM7*, and *LSM3A/LSM5* (C) are presented. Bars = 20 µm.
(D) Subcellular localization of LSM4-GFP in root tip cells from 6-day-old wild-type (WT), lsm8-1 and lsm1a lsm1b Arabidopsis seedlings grown under control conditions. Bars = 10 µm.

**Figure 4. Phenotypical analysis of lsm1a lsm1b double mutant.**

(A) Schematic representation of lsm1a and lsm1b T-DNA insertions in LSM1A and LSM1B genes, respectively. Boxes symbolize exons.

(B) Expression analysis of LSM1A and LSM1B genes in 2-week-old wild-type (WT), lsm1a, lsm1b and lsm1a lsm1b Arabidopsis plants by RNA hybridization using specific probes. rRNA levels are shown as a loading control.

(C) to (K) Morphological phenotypes of WT, lsm1a lsm1b and c-lsm1a plants. Three-day-old seedlings (C), 5-day-old seedlings (D), cotyledon vein patterns (E), rosette leaves (F), cauline leaves (G), 12-day-old seedlings (H), 6-week-old plants (I), siliques (J), and seeds (K).

**Figure 5. Phenotypical analysis of lsm8 mutants.**

(A) Schematic representation of lsm8-1 and lsm8-2 T-DNA insertions in the LSM8 gene. Boxes symbolize exons.

(B) Expression analysis of LSM8 in 2-week-old wild-type (WT), lsm8-1 and lsm8-2 Arabidopsis plants by RNA hybridization using a specific probe. rRNA levels are shown as a loading control.

(C) to (I) Morphological phenotypes of WT, lsm8-1, lsm8-2 and c-lsm8 plants. Five-day-old seedlings (C), cotyledon vein patterns (D), rosette leaves (E), 12-day-old seedlings (F), 6-week-old plants (G), siliques (H), and seeds (I).

**Figure 6. mRNA stability and accumulation of capped transcripts in the lsm1a lsm1b double mutant.**

(A-D) Transcript accumulation in lsm1a lsm1b and c-lsm1a plants. Levels of several transcripts in 6-day-old Arabidopsis seedlings of wild-type (WT) and lsm1a lsm1b (A-B), and of WT and c-lsm1a (C-D), at different minutes (min) after cordycepin treatment. (A and C) RNA hybridizations using specific probes. rRNA levels were used as a loading control. The estimated half-life (min) of mRNAs is shown to the right of each panel (WT/analyzed genotype). (B and D) Normalized quantification of the hybridization bands corresponding to genes of A (B) and C (D).
(E) Accumulation of capped transcripts corresponding to different genes in 6-day-old WT, *lsm1a lsm1b, c-lsm1a* and *c-lsm1b* Arabidopsis seedlings by RACE-PCR. RACE-PCR products obtained by using low (left panel) and high (right panel) number of cycles are shown. The products of *EIF4A1*, also derived from RACE-PCR, were used as a loading control.

**Figure 7. Intron retention and U6 snRNA stability in lsm8 mutants.**

(A) Validation of intron retention events in different genes identified by tiling arrays in the *lsm8-1* mutant. RT-PCR were performed with total RNA from 2-week-old wild-type (WT), *lsm8-1, lsm8-2, c-lsm8* and *lsm1a lsm1b* Arabidopsis plants and specific pairs of primers for each gene. In all cases, one primer was situated inside the retained intron and the other in an adjacent exon. Genomic DNA (Genomic) was used as a control. +RT indicates reactions with reverse transcriptase (RT). Control reactions without RT (-RT) were also performed. *TUBULIN* expression is shown as a loading control.

(B) and (C) Stability of U6 snRNA in *lsm8-1, lsm8-2* and *c-lsm8* plants. Levels of U6 snRNA, U3 snoRNA and U4 snRNA in 6-day-old Arabidopsis seedlings of WT, *lsm8-1* and *lsm8-2* (B), and of WT and *c-lsm8* (C), at different hours (h) after cordycepin treatment, as shown by RNA hybridization using specific probes. rRNA levels were used as a loading control.

**Figure 8. Accumulation of developmental-related transcripts in the lsm1a lsm1b double mutant.**

(A) and (B) Expression levels of different developmental-related genes detected in the microarray with altered expression in *lsm1a lsm1b*. The relative levels of 12 RNAs that in the microarray were increased (A) or decreased (B) are shown. Real-time RT-PCR analyses were performed with total RNA from 2-week-old wild-type (WT), *lsm1a lsm1b, c-lsm1a* and *c-lsm1b* Arabidopsis plants and specific pairs of primers for each gene.

(C) Accumulation of transcripts corresponding to several developmental-related genes detected in the microarray with increased expression in *lsm1a lsm1b*. In all cases, the relative transcript levels were determined by Real-time RT-PCR analysis, as described above, in WT, *lsm1a lsm1b, c-lsm1a* and *c-lsm1b* Arabidopsis plants at different minutes (min) after cordycepin treatment. Values are relativized to the control values obtained for each genotype.
(D) Accumulation of capped transcripts corresponding to genes analyzed in (C) in 2-week-old WT, *lsm1a lsm1b*, *c-lsm1a* and *c-lsm1b* Arabidopsis plants by RACE-PCR. RACE-PCR products obtained by using low (left panel) and high (right panel) number of cycles are shown. The products of *EIF4A1*, also derived from RACE-PCR, were used as a loading control.

**Figure 9. Intron retention in developmental-related genes in *lsms* mutants.**

(A) and (B) Expression levels of different developmental-related genes detected in the tiling array with altered expression in *lsms* mutants. The relative levels of 11 RNAs that were increased (A) or decreased (B) in the array are shown. Real-time RT-PCR analyses were performed with total RNA from 2-week-old wild-type (WT), *lsms*-1, *lsms*-2 and *c-lsms* Arabidopsis plants and specific pairs of primers for each gene.

(C) Validation of intron retention events in some developmental-related genes identified by tiling arrays in *lsms*-1. RT-PCR were performed with total RNA from 2-week-old wild-type (WT), *lsms*-1, *lsms*-2 and *c-lsms* Arabidopsis plants and specific pairs of primers for each gene. In all cases, one primer was situated inside the retained intron and the other in an adjacent exon. Genomic DNA (Genomic) was used as a control. +RT indicates reactions with reverse transcriptase (RT). Control reactions without RT (-RT) were also performed. *TUBULIN* expression is shown as a loading control.
REFERENCES

Anderson, J.S.J., and Parker, R. (1998). The 3' to 5' degradation of yeast mRNAs is a general mechanism for mRNA turnover that requires the SKI2 DEVH box protein and 3' to 5' exonucleases of the exosome complex. EMBO J. 17, 1497-1506.


Bonnerot, C., Boeck, R., and Lapeyre, B. (2000). The two proteins Pat1p (Mrt1p) and Spb8p interact in vivo, are required for mRNA decay, and are functionally linked to Pab1p. Mol. Cell Biol. 20, 5939-5946.


Lorković, Z.J. (2009). Role of plant RNA-binding proteins in development, stress


### Figure 1

#### A

<table>
<thead>
<tr>
<th>Protein</th>
<th>L</th>
<th>R</th>
<th>F</th>
<th>S</th>
</tr>
</thead>
<tbody>
<tr>
<td>LSM1A</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>LSM1B</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>LSM2</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>LSM3A</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>LSM3B</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>LSM4</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>LSM5</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>LSM6A</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>LSM6B</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>LSM7</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>LSM8</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>rRNA</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

#### B

[Image of plant structure]

#### C, D, E

[Images of different plant structures]
Figure 2

A 20°C  B  37°C  C  37°C+CHX

- LSM1A-GFP
- LSM1B-GFP
- LSM8-GFP
- LSM3A-GFP
- LSM3B-GFP
- LSM4-GFP

D 20°C

- LSM1A-GFP
- LSM1B-GFP
- LSM8-GFP
- LSM3A-GFP
- LSM3B-GFP

E 37°C

- WT
- GFP-DCP2
- GFP-VCS

- lsm1a lsm1b

- WT
- GFP-DCP2
- GFP-VCS

- lsm1a lsm1b
Figure 3

A

B

C

D

WT

lsm8-1

lsm1a lsm1b
Figure 4

A

B

C

D

E

F

G

H

I

J

K

LSM1A

LSM1B

Ism1a

Ism1b

rRNA

WT lsm1a lsm1b c-lsm1a

WT lsm1a lsm1b c-lsm1a

WT lsm1a lsm1b c-lsm1a

WT lsm1a lsm1b c-lsm1a

WT lsm1a lsm1b c-lsm1a

WT lsm1a lsm1b c-lsm1a

WT lsm1a lsm1b c-lsm1a
Figure 5
Figure 7

A

<table>
<thead>
<tr>
<th>Gene</th>
<th>WT</th>
<th>Ism8-1</th>
<th>Ism8-2</th>
<th>c-lsm8</th>
<th>Genomic</th>
</tr>
</thead>
<tbody>
<tr>
<td>AT2G43850</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CKL5</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>AT2G33205</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>AT1G19340</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>AME3</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PRMT4A</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>AT1G17960</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>AT3G59390</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ARP4</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>TUBULIN</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

B

<table>
<thead>
<tr>
<th>RNA</th>
<th>WT</th>
<th>Ism8-1</th>
<th>Ism8-2</th>
<th>c-lsm8</th>
</tr>
</thead>
<tbody>
<tr>
<td>U6 snRNA</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>U3 snoRNA</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>U4 snRNA</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>rRNA</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

C

<table>
<thead>
<tr>
<th>RNA</th>
<th>WT</th>
<th>c-lsm8</th>
</tr>
</thead>
<tbody>
<tr>
<td>U6 snRNA</td>
<td></td>
<td></td>
</tr>
<tr>
<td>U3 snoRNA</td>
<td></td>
<td></td>
</tr>
<tr>
<td>U4 snRNA</td>
<td></td>
<td></td>
</tr>
<tr>
<td>rRNA</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Figure 8

(A) 

(B) 

(C) 

(D)
Figure 9

**A**

Relative Expression

- WT
- lsm8-1
- lsm8-2
- c-lsm8

**B**

Relative Expression

- WT
- lsm8-1
- lsm8-2
- c-lsm8

**C**

Genomic

- + RT
- - RT

- + RT
- - RT

- + RT
- - RT

- + RT
- - RT

- + RT
- - RT

- + RT
- - RT
Supplemental Figure 1. Sequence alignment of *Arabidopsis* LSM proteins.

Alignment of the eleven *Arabidopsis* LSM proteins. Sm1 and Sm2 domains are shown. Black and gray shading indicates identical or similar residues, respectively, in at least half of the sequences. Sequence alignment was generated using CLUSTALW software (Thompson et al., 1994) and edited with BioEdit software (Hall, 1999).

<table>
<thead>
<tr>
<th>Sequence alignment of <em>Arabidopsis</em> LSM proteins.</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Supplemental Figure 1. Sequence alignment of <em>Arabidopsis</em> LSM proteins.</strong></td>
</tr>
<tr>
<td><strong>Sm1</strong></td>
</tr>
<tr>
<td>LSM1A</td>
</tr>
<tr>
<td>LSM1B</td>
</tr>
<tr>
<td>LSM2</td>
</tr>
<tr>
<td>LSM3A</td>
</tr>
<tr>
<td>LSM3B</td>
</tr>
<tr>
<td>LSM4</td>
</tr>
<tr>
<td>LSM5</td>
</tr>
<tr>
<td>LSM6A</td>
</tr>
<tr>
<td>LSM6B</td>
</tr>
<tr>
<td>LSM7</td>
</tr>
<tr>
<td>LSM8</td>
</tr>
<tr>
<td><strong>Sm2</strong></td>
</tr>
<tr>
<td>LSM1A</td>
</tr>
<tr>
<td>LSM1B</td>
</tr>
<tr>
<td>LSM2</td>
</tr>
<tr>
<td>LSM3A</td>
</tr>
<tr>
<td>LSM3B</td>
</tr>
<tr>
<td>LSM4</td>
</tr>
<tr>
<td>LSM5</td>
</tr>
<tr>
<td>LSM6A</td>
</tr>
<tr>
<td>LSM6B</td>
</tr>
<tr>
<td>LSM7</td>
</tr>
<tr>
<td>LSM8</td>
</tr>
</tbody>
</table>

**Supplemental Figure 1. Sequence alignment of *Arabidopsis* LSM proteins.**

Alignment of the eleven *Arabidopsis* LSM proteins. Sm1 and Sm2 domains are shown. Black and gray shading indicates identical or similar residues, respectively, in at least half of the sequences. Sequence alignment was generated using CLUSTALW software (Thompson et al., 1994) and edited with BioEdit software (Hall, 1999).
Supplemental Figure 2. Phylogenetic analysis of plant LSM proteins.

The amino acid sequences of LSM proteins from Arabidopsis thaliana (At), Glycine max (Gm), Populus trichocarpa (Pt), Oryza sativa (Os), Zea mays (Zm) and Homo sapiens (Hs) were aligned with the MAFFT software version 6 (Katoh and Toh, 2008), and the unrooted cladogram was constructed by using the minimum linkage method with the MAFFT software. A, B, C and D suffixes indicate different isoforms of a given LSM protein. Plant protein sequences were retrieved from Phytozome (http://www.phytozome.net, Goodstein et al., 2012) using Arabidopsis LSM sequences to perform BLAST. The accession numbers for the represented proteins are described in Supplemental Table 5.
Supplemental Figure 3. Expression patterns of\nLSM1A and LSM1B genes.
GUS activity in whole Arabidopsis plants containing\nthe fusion $LSM1A_{PRO}$-GUS (A) or $LSM1B_{PRO}$-GUS (B). Bars = 10 mm.
Supplemental Figure 4. Visualization of in vivo interactions between Arabidopsis LSM proteins by BiFC assays.

The corresponding LSM-nGFP/LSM-cGFP proteins were pairwise tested by Agrobacterium-mediated transformation in Nicotiana benthamiana leaves. Interactions between LSM1B/LSM2, LSM1B/LSM4, LSM1B/LSM8, LSM2/LSM3B, LSM3B/LSM6A, LSM6B/LSM5, LSM4/LSM6B, LSM6B/LSM7 and LSM3B/LSM5 are presented. Bars = 20 µm.
Supplemental Figure 5. Phenotypical analysis of *Ism1a* and *Ism1b* single mutants. (A) and (B) Morphological phenotypes of No-0, *Ism1a*, Col-0 and *Ism1b* plants. Five-day-old seedlings (A) and 6-week-old plants (B).
Supplemental Figure 6. Quantification of developmental phenotypes shown by *lsm* mutants.

(A) Germination

(B) Abnormal seedlings

(C) Cotyledonary veins

(D) Rosette size

(E) Petiole length

(F) Leaf area

(G) Root length

(H) Secondary roots

(I) Flowering time (LD)

(J) Flowering time (SD)

(K) Siliques length

(L) Seed number

Supplemental Figure 6. Quantification of developmental phenotypes shown by *lsm* mutants.

(A) to (L) Quantitative data on the developmental phenotypes exhibited by wild-type (WT), *lsm1a lsm1b*, c-*lsm1a*, c-*lsm1b*, *lsm8-1*, *lsm8-2* and c-*lsm8* plants. Percentage of germination 5 days after stratification. At least 300 seeds of each genotype were analyzed (A). Percentage of seedlings with abnormal shape or number of cotyledons. A minimum of 500 seedlings of each genotype were scored (B). Percentage of cotyledons showing closed areoles. Data were collected from at least 50 seedlings of each genotype (C). Area of rosettes from 4-week-old plants. A minimum of 25 plants of each genotype were measured (D). Petiole lengths of the 1st and 2nd leaves from 10-day-old plants. At least 25 petioles of each genotype were measured (E). Area of the 3rd and 4th leaves from 15-day-old plants. A minimum of 20 leaves of each genotype were measured (F). Length of the main root (G) and number of secondary roots (H) in 11-day-old plants. At least 60 plants of each genotype were analyzed. Flowering time in long-day (LD) (I) and short-day (SD) (J) photoperiods scored as total leaf numbers. A total of 24 plants of each genotype were scored. Length of the 6th and 7th siliques of the main stem. A minimum of 25 siliques of each genotype were measured (K). Number of seeds from the 6th and 7th siliques of the main stem. Seeds from at least 20 siliques of each genotype were counted (L). In all cases, data represent mean ± SD. Asterisks indicate significantly different data from WT according to a t-test (*P<0.05, **P<0.01, ***P<0.001).
Supplemental Figure 7. Complementation of the *Isma Ismb* double mutant by LSM1B.

(A) to (E) Morphological phenotypes of wild-type (WT), *Isma Ismb* and *c-Ismb* plants. Five-day-old seedlings (A), rosette leaves (B), siliques (C), seeds (D), and 6-week-old plants (E).
Supplemental Figure 8. mRNA stability and accumulation of capped transcripts in c-lsm1b, lsm1a, lsm1b and lsm8-1 plants

(A) to (H) Transcript accumulation in c-lsm1b, lsm1a, lsm1b and lsm8-1 plants. Levels of several transcripts in 6-day-old Arabidopsis seedlings of Col-0 and c-lsm1b (A-B), No-0 and lsm1a (C-D), Col-0 and lsm1b (E-F), and Col and lsm8-1 (G-H) at different minutes (min) after cordycepin treatment. (A, C, E and G) RNA hybridization using specific probes. rRNA levels were used as a loading control. The estimated half-life (min) of mRNAs is shown to the right of each panel (WT/analyzed genotype).

(B, D, F and H) Normalized quantification of the hybridization bands corresponding to genes of A (B), C (D), E (F) and G (H).

(I) Accumulation of capped transcripts corresponding to different genes in 6-day-old No-0, lsm1a, Col-0, lsm1b and lsm8-1 Arabidopsis seedlings by RACE-PCR using a high number of cycles. The products of EIF4A1, also derived from RACE-PCR, were used as a loading control.
Supplemental Figure 9. Tiling array hybridization signals in representative genes showing intron retention events in the *lsm8-1* mutant.

Vertical bars represent the signal intensity values obtained by comparing the signals for all the array probes along the selected genes between the *lsm8-1* mutant and WT. The structure of each gene (TAIR 7) is also represented: black boxes, narrow lines and border boxes symbolize exons, introns and untranslated regions, respectively. Orange boxes indicate intron retention events. Genome graphs displaying probe intensity data all over the gene structure were generated with the Integrated Genome Browser (Nicol et al., 2009).
Supplemental Figure 10. Stability of U6 snRNA in the *lsma lsmb* double mutant.
Levels of U6 snRNA, U3 snoRNA and U4 snRNA in 6-day-old wild-type (WT) and *lsma lsmb* Arabidopsis seedlings at different hours (h) after cordycepin treatment, as shown by RNA hybridization using specific probes. rRNA levels were used as a loading control.
Supplemental References


