JIPB Journal of Integrative Plant Biology

Alternative splicing enhances transcriptome complexity in desiccating seeds

Arunkumar Srinivasan[1,2], José M. Jiménez-Gómez[1,3], Fabio Fornara[4], Wim J. J. Soppe[1] and Vittoria Brambilla[1,4,5]*

[1]Department of Plant Breeding and Genetics, Max Planck Institute for Plant Breeding Research, Cologne, Germany, [2]Open Analytics, Antwerp, Belgium, [3]Institut Jean-Pierre Bourgin, Institut National de la Recherche Agronomique, Versailles, France, [4]University of Milan, Department of Biosciences, Milano 20133, Italy, [5]University of Milan, Department of Agricultural and Environmental Sciences, via Celoria 2, 20133 Milano, Italy.
*Correspondence: vittoria.brambilla@unimi.it

Abstract Before being dispersed in the environment, mature seeds need to be dehydrated. The survival of seeds after dispersal depends on their low hydration in combination with high desiccation tolerance. These characteristics are established during seed maturation. Some key seed maturation genes have been reported to be regulated by alternative splicing (AS). However, so far AS was described only for single genes and a comprehensive analysis of AS during seed maturation has been lacking. We investigated gene expression and AS during *Arabidopsis thaliana* seed development at a global level, before and after desiccation. Bioinformatics tools were developed to identify differentially spliced regions within genes. Our data suggest the importance and shows the peculiar features of AS during seed desiccation. We identified AS in 34% of genes that are expressed at both timepoints before and after desiccation. Most of these AS transcript variants had not been found before in other tissues. Among the AS genes some seed master regulators could be found. Interestingly, 6% of all expressed transcripts were not transcriptionally regulated during desiccation, but only modified by AS. We propose that AS should be more routinely taken into account in the analysis of transcriptomic data to prevent overlooking potentially important regulators.

Keywords: Alternative splicing; *Arabidopsis thaliana*; seed desiccation
Citation: Srinivasan A, Jiménez-Gómez JM, Fornara F, Soppe WJJ, Brambilla V (2016) Alternative splicing enhances transcriptome complexity in desiccating seeds. **J Integr Plant Biol** 58: 947–958 doi: 10.1111/jipb.12482
Edited by: Martin A. J. Parry, Rothamsted Research, UK
Received Feb. 16, 2016; **Accepted** Apr. 20, 2016
Available online on Apr. 28, 2016 at www.wileyonlinelibrary.com/journal/jipb

High-Impact Article

Free Access

INTRODUCTION

The survival of plant species and their ecological success depends on the quality of their seeds. As the main source of calories required for human nutrition worldwide, seeds also have a primary economical value. The quality of seeds is related with their hydration levels and their tolerance to desiccation, which are established during seed development and maturation. Therefore, it is of great importance to understand how seed maturation and desiccation are regulated.

Seed development starts with ovule fertilization and ends when the seed is ready for dispersal. In *Arabidopsis thaliana*, seed development takes 20 d and can be divided into three major phases: (i) embryo morphogenesis (day 1 to 10), (ii) early seed maturation (day 11 to 14) and (iii) late seed maturation, seed ripening and desiccation (day 14 to 20) (Baud et al. 2002). During the first 10 d after pollination (DAP) the embryo develops and the seed grows. Once morphogenesis is completed, the seed has reached its final shape and size. Nevertheless, some crucial processes still need to take place before a viable seed can be dispersed. During early seed maturation from 11 to 13 DAP chlorophyll is degraded and storage compounds like fatty acids and proteins are accumulated. Later on, from 14 to 20 DAP the seed progressively dehydrates, becomes fully ripened, and dormancy is induced. Freshly harvested 20 DAP seeds have less than 10% water content, which is much lower compared to any other plant tissues (van Zanten et al. 2011). To survive this extremely dehydrated status the embryo cells are filled with storage protein vacuoles and oil bodies (Goldberg et al. 1994; Baud et al. 2002; Fait et al. 2006; Finch-Savage and Leubner-Metzger 2006; Jenik et al. 2007). During seed maturation, the chromatin of embryonic nuclei becomes increasingly compacted and nuclear size decreases (van Zanten et al. 2011). For these reasons, dry seed cells are physiologically dissimilar from other plant tissues, and they share common features only with vegetative tissues of desiccation tolerant plants. Several studies have elucidated a framework of seed developmental regulators. Four transcription factors, *ABSCISIC ACID INSENSITIVE 3 (ABI3)*, *FUSCA3 (FUS3)*, *LEAFY COTYLEDON 1 (LEC1)* and *LEC2*, play major roles in seed maturation. *LEC1* and *LEC2* are expressed since early embryo morphogenesis followed by *FUS3* and finally by *ABI3*, which is primarily involved in seed maturation. *ABI3*, *FUS3* and *LEC2* belong to the AFL (ABI3-FUS3-LEC2) subgroup of the B3 transcription factor family (Holdsworth et al. 2008; Suzuki and McCarty 2008; Swaminathan et al. 2008). The B3 domain of AFL transcription factors recognizes the RY motif of its target DNA, that has been shown to be involved in seed specific expression (Reidt et al. 2000; Mönke et al. 2004). Many direct, indirect and also common targets of LEC1, ABI3, FUS3 and LEC2 have been identified (Wang and Perry 2013). One common function of AFL genes during seed maturation is the control of storage compounds accumulation. Phytohormones, especially abscisic acid (ABA) and gibberellins (GA), have an important role in

the seed development regulatory network (Gazzarrini et al. 2004; Santos-Mendoza et al. 2008; Wang and Perry 2013). They also integrate molecular and environmental cues to best time germination (Joosen et al. 2012). ABA is required for seed development, for the establishment of desiccation tolerance during seed maturation and for its re-establishment in germinated seeds (Maia et al. 2014). Hormonal responses are in part mediated by the seed development master regulators. ABI3 is required, together with the bZIP transcription factor ABI5 and in parallel to ABA, to induce seed dormancy in an ABA-dependent manner (Lopez-Molina et al. 2002). FUS3 is required to induce ABA and repress GA biosynthesis by direct interaction with the promoter of GA biosynthetic genes (Tsai and Gazzarrini 2012). *ABI3*, *LEC1* and *FUS3* homologues have also been characterized in distantly related species as pea, rice, maize, barley and Selaginella (Nakagawa et al. 1996; Miyoshi et al. 2002; Moreno-Risueno et al. 2008; Gagete et al. 2009; Shen et al. 2010; Kirkbride et al. 2013). In *Arabidopsis thaliana*, tomato, pea, rice and wheat, *ABI3/VP1* genes have been shown to be regulated by AS (McKibbin et al. 2002; Fan et al. 2007; Gagete et al. 2009; Sugliani et al. 2010; Gao et al. 2013). In *Arabidopsis thaliana ABI3*, AS is developmentally regulated. Two transcript variants have been detected: the full-length coding *ABI3-α* and the *ABI3-β* transcript that contains a cryptic intron within its first exon causing a frame-shift shortly after. At the beginning of seed maturation only *ABI3-α* transcript could be detected, but from 16 DAP *ABI3-β* was also detectable and became the most abundant protein isoform in mature seeds (Sugliani et al. 2010). Similarly, in tomato, two *SlABI3* transcripts that encode for a full-length and a truncated protein are accumulated. In addition, in this case, the truncated SlABI3 protein results from the AS of a cryptic intron within the first exon of *SlABI3*. Only the full length SlABI3 protein is able to activate its target genes, while the truncated SlABI3 has possibly a regulatory function. Another interesting example comes from rice, where the AS of *ABI3* homologue *OsVIVIPAROUS 1 (OsVP1)* has been phenotypically linked to seed dormancy and pre-harvest-sprouting. As in *Arabidopsis thaliana*, rice OsVP1 can interact with OsABI5. Two OsABI5 distinct proteins derived from AS transcripts have different binding affinity to OsVP1 and transactivation of their targets (Zou et al. 2007).

AS is also specifically affecting seed maturation and dormancy through two regulators of this process: *PHYTOCROME INTERACTING FACTOR 6 (PIF6)* and *DELAY OF GERMINATION1 (DOG1)*. *PIF6* has two splice variants, one of them originating from an out of frame exon skipping AS event that creates a premature stop codon and a protein isoform that lacks the DNA binding domain (Penfield et al. 2010). Overexpression of the *PIF6* AS variant, but not of the full length coding sequence, reduces seed dormancy (Penfield et al. 2010). *DOG1* presents 5 AS transcript variants that are all functional, but unstable if not expressed in combination with different AS isoforms (Bentsink et al. 2006; Nakabayashi et al. 2012, 2015).

These observations indicate the importance of AS in the regulation of seed maturation in both dicots and monocots. Transcriptomic analyses by RNA-seq and Tiling Arrays have been used to study the transcriptome complexity, also taking into account AS (Yoshimura et al. 2011; Leviatan et al. 2013). Thanks to these data, an increasing number of genes have been predicted to be alternatively spliced. In plants, AS is currently estimated to affect about 61% of intron-containing genes in *Arabidopsis thaliana* (Marquez et al. 2012) and 31% in rice (Filichkin et al. 2010). These numbers are likely to increase with the investigation of AS in different environmental conditions, developmental stages and tissues (Syed et al. 2012; Loraine et al. 2013; Reddy et al. 2013).

Despite the importance of AS in seed maturation, this process has not been investigated at a transcriptomic level yet. This paper provides the first comprehensive description of AS during late seed maturation when desiccation is achieved. In addition, we developed a new pipeline to detect and annotate AS events from RNA-seq data. Our analysis also expands the number of known splicing variants in the transcriptome in general and for seeds in particular.

RESULTS

Whole transcriptome profiling of *Arabidopsis thaliana* seeds before and after desiccation

In order to understand the seed transcriptome changes prior to and after seed desiccation, we examined gene expression and splicing patterns using high-throughput RNA sequencing. Poly-adenylated RNA from developing seeds of the *Arabidopsis thaliana* accession Columbia-0 was obtained at 14 d after pollination (DAP) and from mature dry seeds at 20 DAP. We then constructed cDNA libraries. We retained information about the direction of transcription thanks to the application of the dUTP method to the second strand cDNA synthesis (Levin et al. 2010). We deep sequenced each library (72.3 to 83.6 million 95 bp paired end (PE) reads each) to study also low abundant transcripts and splicing variants (Table 1 raw number). Three independent biological replicates were sequenced for each time point (named 14 DAP-1, 14 DAP-2, 14 DAP-3; 20 DAP-1, 20 DAP-2, 20 DAP-3). Between 90.9% and 98.9% of the reads mapped uniquely to the genome in all

Table 1. Number of raw, mapped and processed reads per sample sequenced

Sample name	Raw number	Processed number	Processed/ raw %	Mapped number	Mapped/ processed %
14 DAP-1	83564710	80857827	96.76	79992843	98.93
14 DAP-2	82414264	79986955	97.05	77681534	97.12
14 DAP-3	72270682	67677610	93.64	61486031	90.85
20 DAP-1	80650756	78945942	97.89	76691712	97.14
20 DAP-2	76094330	74330212	97.68	71603236	96.33
20 DAP-3	83162760	81114639	97.54	78693439	97.02

libraries (Table 1-mapped/processed). Between 93.2% and 97.0% of the uniquely mapped reads from all the libraries were first strand specific. We considered for our analysis a total of 15,670 genes that presented RPKM greater than 1 in at least three out of the six samples (Table S1). From these, 14,519 genes were expressed at 14 DAP and 12,925 genes at 20 DAP (Tables S2−3), with 11,774 genes, 75% of the total, being expressed in both developmental stages. (Figure 1A; Table S4). We found that 74% (8,695) of the 11,774 genes expressed at 14 DAP and 20 DAP were differentially expressed. 59% of the differentially expressed genes (5,156) were down-regulated between 14 and 20 DAP (Table S5), while 41% (3,539 genes) were up-regulated at 20 DAP compared to 14 DAP (Table S6). Therefore overall transcription was slightly reduced between 14 and 20 DAP (Figure 1B).

RNA-seq identified extensive alternative splicing during seed maturation

We developed a pipeline to detect alternative splicing (AS) events, including not previously annotated AS events. Six classes of AS were discriminated: alternative 3′ (A3P), exon skipping (ES), alternative 5′ (A5P), intron retention (IR), cryptic intron (CI) and cryptic exon (CE) (Figure 2A). We obtained a total of 8,927 AS events in 4,875 genes (Table S7), among which IR events were the most abundant (60.8% of all AS events), followed by A3P (21%) and A5P (10%, Figure 2B). The ratios between the different types of AS events were similar to those reported for other plant tissues (Chardon et al. 2004; Marquez et al. 2012; Syed et al. 2012; Kornblihtt et al. 2013). We also identified a 4% of CI, 1.8% of ES and 1.5% of CE (Figure 2B).

Out of the 8,927 total AS events (for the definition of alternative and canonical variant used in this work see methods) that we have found in the two seed developmental stages, 88% (7,856 AS events) were not present in the TAIR 10 gene model annotation. In particular, 92% of the IR events detected, 89% of the A5P, 77% of the A3P and 66% of the ES events were not described in the annotation (Figure 2C). None of the CI and CE that we have identified were previously reported. Among the junctions described, the canonical splice site GT-AG was found in 97.75% of the total alternatively spliced transcripts (Table S8). The second most abundant type of donor-acceptor sequence is GC-AG (1.8%) followed by AT-AC (0.45%) (Figure 2D; Table S8). These numbers represent a 1% decrease in the novel GT-AG splice site reported in the TAIR10 annotation (98.76%), which is not observed when taking into account only the annotated events (Figure 2E). We also analyzed the type of di-nucleotide splice site sequence in relation to the different types of AS (Table 2). The ratio of canonical vs. non-canonical splice sites varies between the AS types and between novel and annotated splicing variants. All of CE and 92% of CI events had canonical AG-GT splice site junctions. IR events showed 97.6% and 98% of the novel and annotated splicing variants respectively having an AG-GT canonical splice site.

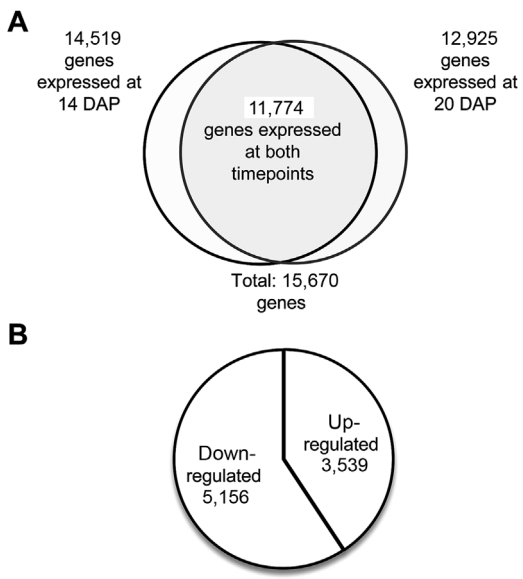

Figure 1. Reads mapping and expression analysis
(A) Venn diagrams of genes expressed at 14 d after pollination (DAP) and 20 DAP. **(B)** Differentially expressed genes between 14 DAP and 20 DAP.

Relationship between alternative splicing and differential gene expression during seed development

A total of 4,723 and 4,494 genes have AS variants in 14 and 20 DAP seeds, respectively. Since some genes are affected by more than one AS event, the total number of events per developmental stage is 8,567 at 14 DAP and 8,250 at 20 DAP. We asked whether AS plays a role in the regulation of genes involved in seed maturation. We found a total of 1,809 significantly differentially alternatively spliced (DAS) events between developmental stages in 1,408 genes (Table S9), representing about 20% of the genes that are alternatively spliced in the seed at 14 or 20 DAP. These genes either present an alternative splicing variant only at 14 or at 20 DAP or they show a different ratio between the canonical variant and an alternative variant at a specific splice site. The most common event again is IR, with 69.26% of total DAS events. A3P was 14.21%, followed by A5P at 8.46%, CI 6.69% and ES 1.33% (Figure 3A).

In order to assess the regulatory role of AS during seed maturation, we evaluated its relationship to differential gene expression.

Out of the 1,408 significantly differentially alternatively spliced genes, 688 are also differentially expressed. In particular, 468 are up-regulated and 220 are down-regulated from 14 DAP to 20 DAP (Tables S10–11). Interestingly, for 720 (51%) of the DAS genes found there is no significant difference in expression between 14 and 20 DAP (Figure 3B; Table S12). Considering the whole transcriptome of 14 and 20 DAP seeds, 6% of the genes are DAS but their overall expression is unchanged. These results suggest that AS is a form of transcriptome control in part independent from expression during seed desiccation in *Arabidopsis thaliana*.

AS impact on protein function

In order to evaluate the possible impact of AS on protein function, we predicted the protein sequence resulting from the AS events identified. We focused this analysis only on IR, which represents the most common type of AS (61% of the AS events, 5,431 out of 8,927). The majority of IR events resulted

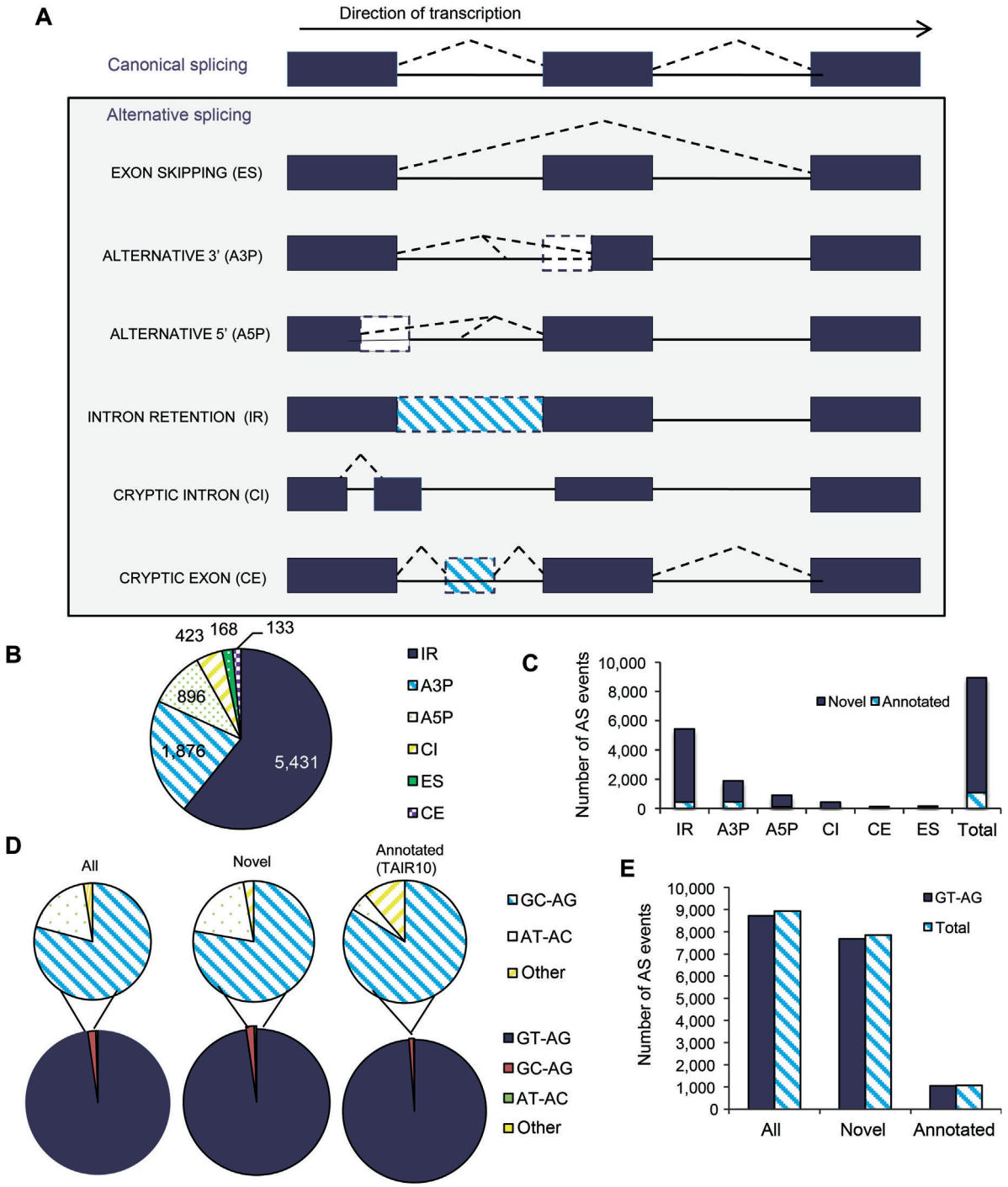

Figure 2. Alternative splicing in seeds

(**A**) Schematic representation of the different types of AS events with respect to a hypothetical canonical form (top). For details on the definition of canonical introns see Methods (alternative splicing- generation of representative introns). Full boxes: canonical exons; dashed boxes: exons (or part of) affected by AS; dashed bars: coordinates of splicing event. For A3P/A5P two different types of events are illustrated, one where the A3P/A5P event happens within an exon (in this case the AS transcript will be longer than the canonical- starting point) or in an intron (in this case the longer transcript will be the canonical). White dashed boxes: intronic ending/starting point of an A3P/A5P event. Striped boxes: retained intron between exons/ retained cryptic exon within an intron. (**B**) Number of AS events identified at 14 DAP and 20 DAP, arranged per type of event according to the previous categories. (**C**) Number of novel and annotated (TAIR10) events per each type. (**D**) Type of di-nucleotide splice site signals found among -all AS events, -novel events, -annotated (TAIR10) events only. (**E**) Number of canonical splice site signals compared to all signals in the same categories of AS event.

Table 2. Type of di-nucleotide sequence in relation to the different types of AS

AS type	Splice site signal	All AS events	Novel AS events	Annotated AS events	Total novel	Total annotated	% GT-AG among annotated	% GT-AG among novel
A3P	GT-AG	1,859	1,422	437	1,436	440	99.3	99
A3P	GC-AG	16	13	3				
A3P	AT-AC	1	1	0				
A5P	GT-AG	859	762	97	793	103	94.2	96.1
A5P	GC-AG	37	31	6				
CE	GT-AG	133	133	0	133	0		100
CI	GT-AG	390	390	0	423	0		92.2
CI	GC-AG	17	17	0				
CI	AT-AC	11	11	0				
CI	other	5	5	0				
ES	GT-AG	165	64	101	65	103	98.1	98.5
ES	GC-AG	3	1	2				
IR	GT-AG	5,320	4,905	415	5,006	425	97.6	98
IR	GC-AG	86	78	8				
IR	AT-AC	25	23	2				
	Total	8,927	7,856	1,071				

in the loss of frame and the inclusion of a premature translation termination signal. Out of the 5,431 IR events, only 420 (7.7%) presented the canonic termination signals (Table S13).

To see if some gene functional categories were preferentially regulated by AS during seed maturation, we analyzed the list of significantly DAS genes following the Gene Ontology (GO) biological process categorization (Ashburner et al. 2000; Gene et al. 2014) (Figure 4).

The most enriched (4.19 and 3.95%) functional categories include genes involved in mRNA catabolic processes. Genes specifically involved in mRNA splicing are enriched by 2.57%. These data are in agreement with previous observations that genes related to RNA metabolism and in particular splicing factors are in general more affected by AS in response to development or environmental cues (Syed et al. 2012; Reddy et al. 2013).

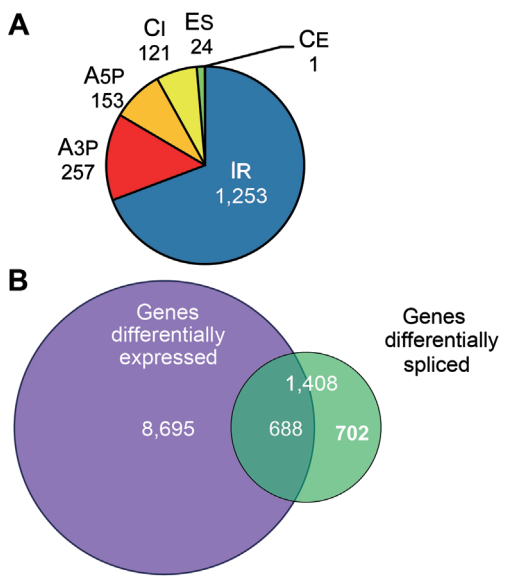

Figure 3. Differential alternative splicing and expression
(A) Number of differentially alternatively spliced (DAS) events across developmental stages, divided by type of event. **(B)** Number of genes differentially spliced against number of genes differentially expressed.

Experimental validation of the computational predictions

In order to validate the results of our AS and DAS pipelines, we chose 21 events from all different types of predicted AS. Twenty events were not previously reported in other tissues or experimental conditions, while one of them was already annotated (IR in AT1G55350 3′UTR) (Table S14). We evaluated the relative abundance of each splicing event by qRT-PCR using primers specific for each splicing variant or when not possible specific to one variant and all transcripts. All events were validated on three independent biological samples per time point. For all genes, the computationally predicted splicing variant could be detected. We calculated the ratio between 14 and 20 DAP for each variant and we found consistency between qRT and RNA-seq data. The variant more represented at one time point in our RNA-seq dataset was also confirmed by qRT (Figure 5A).

Strict filtering criteria to control false positives was applied to the pipeline. For this reason, some AS events that could be taking place and be biologically relevant might have been omitted. Therefore, to test the sensitivity of the method, we also tested an A3P event that was affecting *FUSCA3*, a master regulator of seed development. The event was not among the most statistically significant genes, but the read counts suggested a potential DAS event. *FUS3* expression decreased about 10 times from 14 to 20 DAP. In an opposite trend to the overall transcript, *FUS3* A3P variant is accumulated at 20 DAP, where it reaches about 35% of the canonic variant. Interestingly *FUS3* A3P encodes a truncated protein that lacks part of the B3 DNA-binding domain (Figure 5B–E), similarly to the *ABI3-β* transcript (Sugliani et al. 2010) and to PIF6 exon skipping (Penfield et al. 2010).

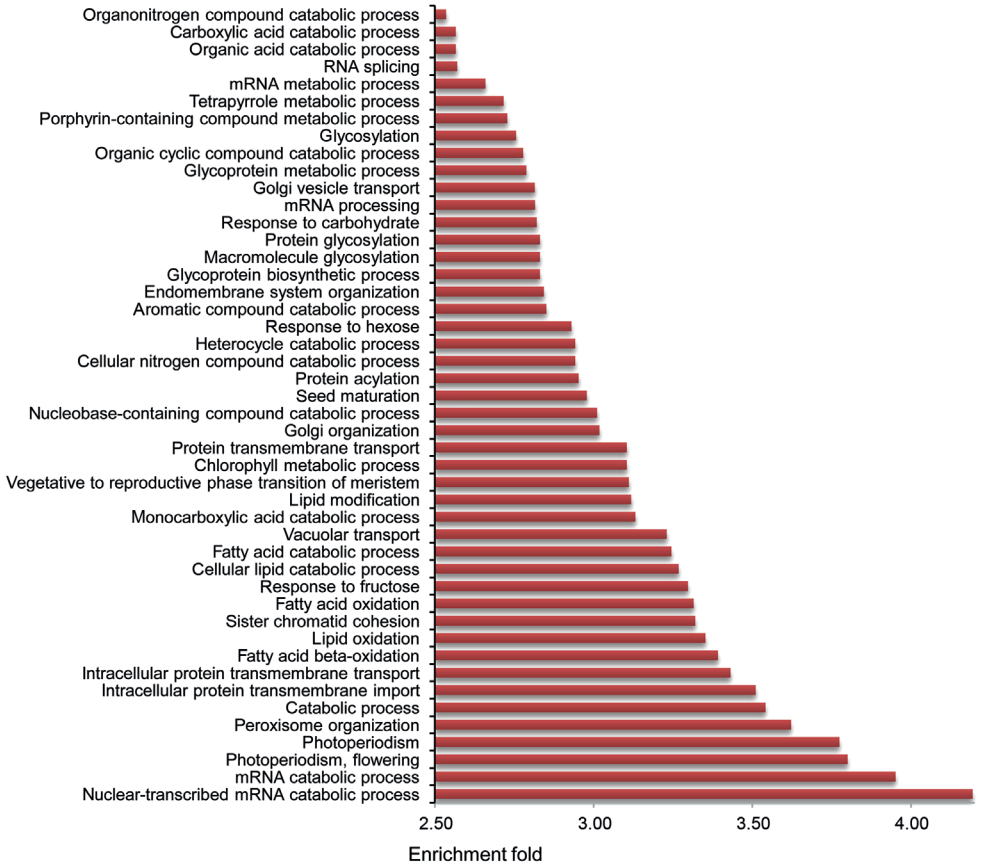

Figure 4. Gene ontology (GO) functional categorization of all differentially alternatively spliced (DAS) genes
Only those categories with enrichment fold higher than 2.5 were plotted.

DISCUSSION

Splicing dynamics in seed during acquisition of desiccation tolerance

AS is a powerful mechanism that controls gene expression and allows rapid changes in transcriptome and proteome complexity during development and upon environmental changes (Reddy et al. 2013). Here, we presented an extensive description of AS during the last phase of seed maturation (14–20 DAP), when the fully developed seed desiccates and becomes dormant. This information was lacking for this developmental stage and contributes to a growing list of AS events occurring during plant development (Barbazuk et al. 2008; Yoshimura et al. 2011; Leviatan et al. 2013; Chang et al. 2014). Mature dry seeds have a unique cellular organization due to their very low water potentials (Terrasson et al. 2013). It is therefore interesting to study transcriptomic changes including AS in this specific environment. Besides seeds, AS has been reported to be involved together with ABA in desiccation of leaves and roots from desiccation tolerant plants (Xiao et al. 2015).

Surprisingly, 88% of the AS events that we found in 14 and 20 DAP seeds were novel events. This might be due to the fact that no high-throughput studies had been performed on this tissue before.

To better understand the transcriptome dynamics during seed dehydration, we also analyzed gene expression. Interestingly gene expression levels and AS have a slightly opposite trend. Overall transcription was reduced between 14 and 20 DAP, while AS was increased. Transcription is probably affected by increasing desiccation and by the vital but quiescent state in which the seed is set until germination is induced (Holdsworth et al. 2008; Terrasson et al. 2013). Interestingly, AS affects a slightly higher percentage of genes at 20 DAP compared to 14 DAP, possibly representing a mechanism that allows the rapid end of seed maturation once desiccation is achieved.

Our dataset is unique, since it is the first published for 14 and 20 DAP seeds and analyzed for AS and DAS. AS datasets for other *Arabidopsis* tissues, and for other plant species are available, and to some extent present similar results as ours. The first comprehensive analysis of AS in *Arabidopsis* in the next-generation sequencing era comes from the valuable work of Marquez et al. (2012). Marquez and colleagues study the dynamics of the transcriptome in flowers and seedlings at different developmental stages. They find that 61% of multi-exonic genes are AS under normal growth conditions, which is a slightly higher percentage compared to what we find in our analysis (57%). Possibly this is the result of that we analyzed the transcriptome of more homogeneous tissues (seeds at 14 and 20 DAP), with less variability in AS patterns. Our results are

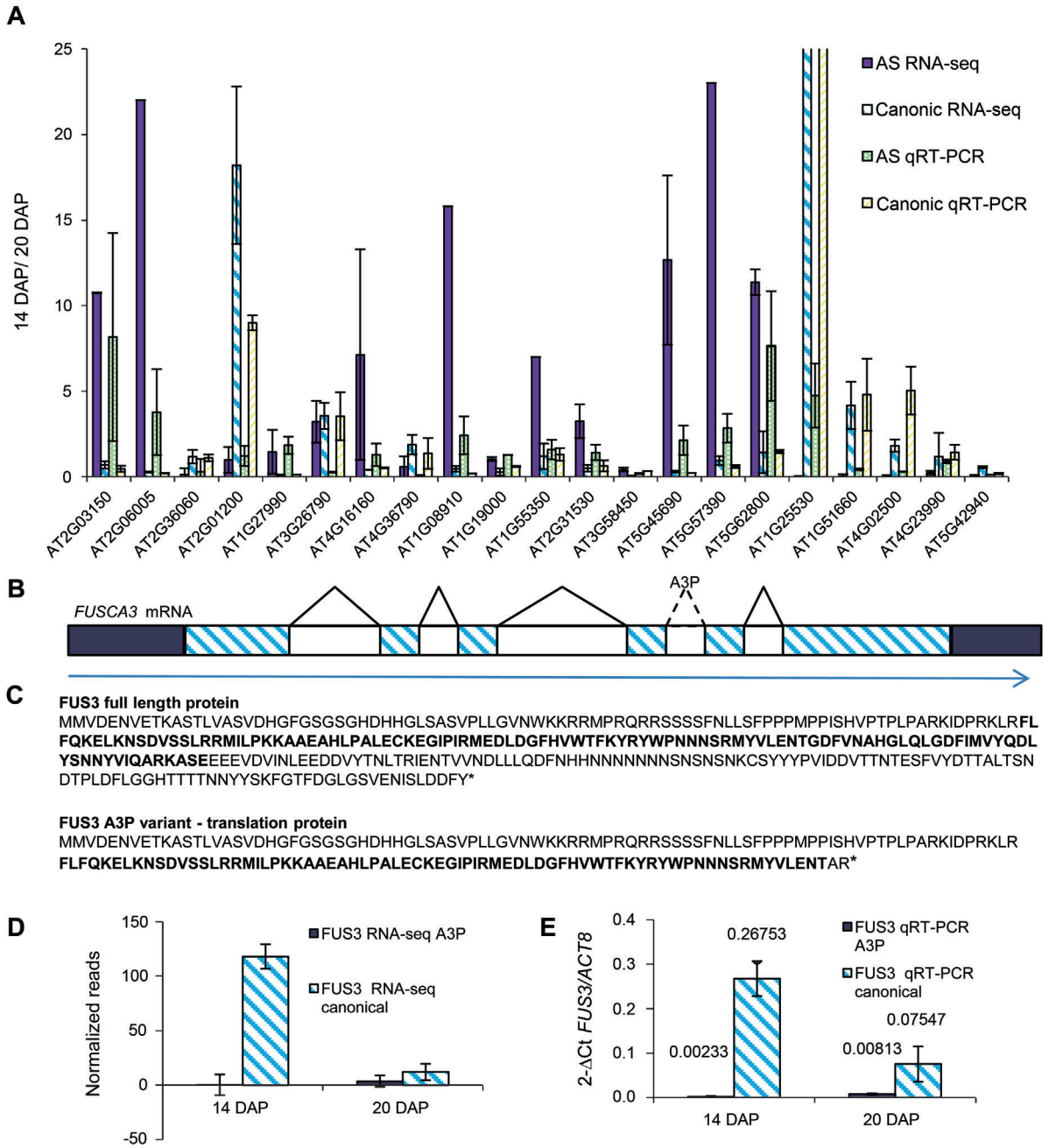

Figure 5. Experimental validation of RNA-seq data and AS bioinformatic pipelines by qRT(A)

(**A**) For the 21 as AS events in 21 genes chosen from the AS dataset we show the 14 DAP/20 DAP ratio of the AS and the canonical variant of each gene, calculated for RNA-seq and qRT data. Standard deviations for three independent replicas are shown. (**B–E**) The seed development master regulator *FUSCA3* undergoes an A3P event. (**B**) *FUS3* gene model. Dark full boxes: 5′ and 3′ UTRs; striped boxes: exons; bars: introns. Between exon 4 and exon 5 there is an A3P event indicated by dashed bars. The A3P splice variant is 29bp shorter than the canonical variant. Arrow indicates the direction of transcription. (**C**) A3P causes a frame shift in the *FUS3* coding sequence with the formation of a premature stop codon, 6 bp after the A3P event and within the B3 DNA binding domain of the gene (in bold in the protein sequence). The predicted protein resulting from the A3P event has a truncated B3 domain. (**D–E**) Quantification of *FUS3* splicing variants from RNA-seq data (D) and from qRT-PCR data on independent samples. Error bars indicate standard deviations between independent biological replicates.

in agreement with previous analyses on plant AS (reviewed in Syed et al. 2012; Carvalho et al. 2013; Reddy et al. 2013) where IR are always the most common AS events, and AG-GT are the by far most represented types of di-nulceotide splice junctions.

As with many other biological processes, AS is likely to have an important contribution to the correct progression towards seed desiccation, since it regulates some important regulators of seed development, maturation and dormancy as PIF6, ABI3 and DOG1, considering only *Arabidopsis* (Penfield et al. 2010; Sugliani et al. 2010; Nakabayashi et al. 2015). The importance of AS in the regulation of these genes emerged from experimental data during their functional characterization. Despite the relevance of AS in the regulation of their function, none of these events appear in our significantly DAS gene list, for different reasons. *ABI3* and *PIF6* are DAS between 14 and 20 DAP and they could be found in our rough data analysis before filtering for significance. But due to the low reads number supporting the different AS events at a transcriptomic level they could not be assigned to the list of most significant DAS events. Concerning DOG1, our pipeline was designed to specifically detect genes DAS between 14 and 20 DAP, and is therefore dependent on the expression levels of the genes and their splicing variants at these timepoints. But since the ratio between the different splicing forms of DOG1 is constant during seed maturation (also between 14 and 20 DAP) (Nakabayashi et al. 2015) it was not listed in our DAS genes table.

Prediction of AS on proteasome complexity

AS has the potential to rapidly expand proteasome complexity and had a crucial role during plant evolution (Tack et al. 2014; Vitulo et al. 2014; Xu et al. 2014; Li et al. 2015; Xiao et al. 2015). To gain information about the possible effects of AS during seed desiccation, we predicted the sequences of the protein isoforms resulting from the IR AS transcripts. Different from animals, IR is by far the most common type of AS event found in plants. This diversity reflects the different gene structure where plant introns are generally smaller and genes contain fewer exons compared to animals (Reddy et al. 2013). We compared the predicted protein sequence resulting from IR with that of "canonically" spliced genes. The majority of IR events resulted in premature termination of the proteins. Out of the 5,431 IR events, only 420 (7.7%) did not end prematurely. Premature termination codons (PTC) IR events are not necessarily directed to degradation through NMD, as only a small percentage of them follows this fate (Kalyna et al. 2012). In addition it has been proposed that some of them can turn into Micro Proteins, that lack some of the functional domain of the entire protein and have regulatory functions (Staudt and Wenkel 2011; Graeff and Wenkel 2012; Brandt et al. 2014). To this extent, IR events might have a prominent regulatory role. The actual presence of the predicted isoforms was not verified. Nevertheless the fact that the ABI3 truncated cryptic intron isoform accumulated at 20 DAP, provides an example of an AS protein isoform having a regulatory function. Candidate genes with different predicted isoforms could be further characterized to verify their function.

It is well known that splicing can rapidly modify an mRNA sequence and function in response to stimuli. To promptly modify a cascade of several AS genes, splicing regulators are often the primary targets of AS themselves (Ding et al. 2014). We can conclude that this general observation is also valid in seeds, where the most represented category of AS genes are RNA-binding proteins and other various components of the spliceosome.

MATERIALS AND METHODS

Plant growth conditions, RNA extraction and quantification of transcript abundance

Plants were grown in a growth chamber (16 h-light/8 h-dark; 22 °C/16 °C). Flowers were marked the day they pollinated and siliques were collected at 14 d and 20 d after pollination. Plants for RNA-seq and qRT-PCR validation were grown in independent experiments. Seeds were removed from the siliques in liquid nitrogen prior to RNA extraction. 14 DAP and 20 DAP seeds were grinded in a mortar in liquid nitrogen and RNA was extracted with the Ambion RNAqueous extraction kit supplemented with the RNA Isolation Aid as previously described by Nakabayashi et al. 2012. Retro Transcription was performed using oligo dT and the SuperScript II reverse transcriptase (Invitrogen). qRT-PCR was performed in an Eppendorf realplex^2. Primers sequences are listed in Table S15. Primers were either specific to both single variants or when not possible to only one and to both simultaneously. In the second case we ensured that primers had comparable good efficiency and that amplicon size was the same and were deducted by subtracting the amount of the second variant. Expression was normalized to *ACTIN 8* (At1g49240) with the primers described in Sugliani et al. (2010) and used also in Nakabayashi et al. (2012 and 2015). The ratio between 14 DAP and 20 DAP of the two variants was calculated for RNA-seq data using the average reads number per variant from the three RNA-seq biological replicates. For qRT-PCR we used the average of three independent experiments.

Mapping and expression analysis

Sequencing was performed on the Illumina GAII platform. Each replica was run on a single flow cell. RNA-seq reads were examined using fastQC (http://www.bioinformatics.babraham.ac.uk/projects/fastqc) to inspect read qualities and extent of adapter sequence contamination. Adapter sequences were clipped and reads with qualities of less than 30 were trimmed using in-house Perl scripts. Filtered reads shorter than 50 bp were discarded. Reads from a pair where only one read passed filtering criteria were retained as single-end (SE) reads. After pre-processing, between 93.6% and 97.9% of the reads were retained. Reads were then mapped to the *Arabidopsis thaliana* reference genome TAIR10 using Tophat2 v2.0.9 (Trapnell et al. 2009) allowing a maximum of five mismatches, insertions and deletions each, with a total edit distance of 10 (these and following packages would be available on request).

Read counts for each of the six libraries were obtained using in-house R-package readCounts, which internally uses Bioconductor packages GenomicRanges, IRanges (Lawrence et al. 2013) and Rsamtools (http://bioconductor.org/packages/release/bioc/html/Rsamtools.html), and the CRAN data.table package (Extension of Data.frame. R package version 1.9.5 https://github.com/Rdatatable/data.table/wiki). (M Dowle,

T Short, S Lianoglou, A Srinivasan with contributions from R Saporta and E Antonyan). RPKM values (Mortazavi et al. 2008) were used for filtering genes with low expression. They were computed for each gene and library from the raw read counts and using the gene lengths derived from the representative gene models. Those genes with RPKM greater than one in at least three out of the six samples were considered to be expressed and therefore retained. Raw counts for those genes that were retained were used to analyze differential gene expression. Clcbio (www.clcbio.com) was also used for read mapping and visualization of the AS events.

Generation of representative introns

In order to efficiently detect also novel AS events from RNA-seq data, we first computed a set of representative intron coordinates. This enabled us to detect splicing variants of genes with multiple transcripts that may contain introns (and exons) that overlap with one another across multiple transcripts. The set of representative intron coordinates are generated using in house R package gffutils on the *Arabidopsis thaliana* TAIR10 gene models. When multiple introns had identical start, end or both coordinates, only the shortest intron was retained. If introns were overlapping, but neither start nor end coordinate was identical, then all those introns were retained. We refer to these representative intron coordinates as canonical introns. An example of this method is illustrated in Figure 6. Out of a total of 127,896 introns in the annotation, 124,399 canonical introns were obtained using gffutils.

Splice junctions

We estimated splice junction coordinates from the alignment of each individual library using custom methods. At the first stage of filtering, splice junctions with three or more reads in at least two out of the six libraries were retained. The MMES score (Wang et al. 2010) of each of the retained spliced read were computed, and only those splice junctions with >50% of the reads having MMES score >5 were retained. If a read extended to more than one gene, and those genes were not overlapping genes, those reads were discarded as erroneous. This resulted in a set of high quality splice junctions. Read counts across these filtered splice junctions were then calculated. Similarly median coverage across all of the retained splice junction was also computed separately and only those junctions with a median coverage ≥3 in at least two out of six samples were retained. These operations were accomplished using the in-house R package splicerutils. There were a total of 145,550 splice junctions from all the six libraries combined. Out of those, 143,093 junctions occurred within an annotated genic region, and the remaining 2,457 junctions were in the intergenic region. Out of the 143,093 splice junctions in the genic regions, 98,666 junctions were annotated as an intron in the gene model. The remaining 44,427 splice junctions were not annotated in the TAIR10 gene model and were classified as potential *novel* AS events.

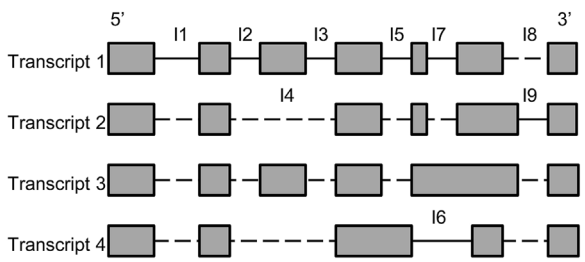

Figure 6. Definition of canonical intron

Representation of a hypothetical gene with four transcripts and a total of nine unique introns, marked from I1 to I9. Boxes: exons, lines: introns. Intron I1 has identical coordinates across all four transcripts. Introns I2 and I3 overlap with I4 have identical start and end coordinates respectively. Only the shorter introns are considered and intron I4 is not considered as a canonical intron. Introns I5, I7 overlap with I6, but do not have identical start or end coordinates and are all considered canonical introns. Introns I8 and I9 share identical end coordinates; the shorter intron I9 alone is considered canonical. Canonical introns: solid lines, non canonical introns: dashed lines.

Alternative splicing events

All splice junctions with identical coordinates as canonical introns were classified as canonical (splice) junctions, CJ. The other splice junctions that overlap with CJ could be automatically classified as alternatively spliced AS. These AS junctions consist of union of the set of annotated AS junctions, i.e., present in the TAIR10 gene model, and novel AS junctions. They are classified as A3P, ES, A5P, IR, CE or CI following this set of rules: AS junctions where start coordinates matched a CJ, but not the end coordinates were classified as A3P events. AS junctions where start coordinate matched a CJ, and the end coordinate matched another CJ within the same gene were classified as ES events. AS junctions where end coordinates matched a CJ, but not the start coordinates were classified as A5P events. Canonical junctions where unspliced reads had a median coverage ≥3 were considered to be IR events. AS junctions, which occurred within an annotated exon are marked as CI events. Two AS junctions where the start of the first AS junction and the end of the second correspond to a CJ implies that the end coordinate of the first AS and the start coordinate of the second occurred within that CJ. There must be therefore an exon within this CJ that has not yet been annotated. Those events are marked as CE events and are removed from A3P or A5P events. The DAS pipeline workflow is schematically represented in Figure S1. Our pipeline was specifically designed to detect these rare types of events since ABI3 function and seed maturation was strongly affected by a CI AS event (Sugliani et al. 2010).

Identification of differential alternative splicing (DAS) events

The pipeline to detect DAS events is implemented in the in-house R package splicer and the normalized read counts from canonical (CJ) and alternative spliced junctions (AS). For each AS event, the corresponding overlapping CJ is identified. This resulted in a total of 12 normalized read count values per AS event – three each corresponding to AS and CJ in developmental stages 14 DAP and 20 DAP. Once the normalized read counts were extracted, a negative binomial generalized linear model, NB-GLM, was fitted using the R package MASS (Venables and Ripley 2002) for each AS event by modeling a two-way interaction between time point and AS

type (type) with *normalized* read counts as the response variable. The binary variable timepoint (TP) takes two values corresponding to each of the time points, 14 DAP and 20 DAP. The binary variable type takes two values as well corresponding to whether the read counts come from alternatively spliced or canonical junctions, AS and CJ. *P*-values corresponding to the interaction term were extracted and adjusted to correct for multiple testing by using the Benjamini-Hochberg procedure using the R-package multtest (Pollard et al. 2005) at a false discovery rate (FDR) of 5%. AS events with $q \leq 0.05$ were considered as undergoing significant differential alternative splicing.

Protein variant prediction

For each gene, we extracted and computed the length of the protein sequence for the representative gene model. Following that, we incorporated the IR event on to the representative gene model and computed the protein sequence under all six frames and chose the longest. Those events where the intron retained transcripts resulted in longer protein sequence than the representative model were considered not to end prematurely.

Differential expression analysis

Strand specific raw read counts from genes that were retained after filtering using RPKM > 1 as explained above were inputted to the bioconductor package DESeq v1.14.0 to detect the genes that are differentially expressed between the time points TP14 and TP20. Genes with FDR corrected *P*-values (or q-values), $q \leq 0.05$, were considered to be significantly differentially expressed.

Gene onthology functional categorization

Functional over-representation for genes containing DAS events was calculated in comparison with all genes expressed using AmiGO 1.8 release 01.08.2015 (http://amigo.geneontology.org/amigo) and only those categories represented by more than 40 genes and enriched at least 2.5 times compared to occurrence in the entire genome were taken into account.

ACKNOWLEDGEMENTS

We are grateful to Professor Maarten Koornneef for guidance and support. This work was supported by the Alexander von Humboldt Foundation, Germany, through a research fellowship to Vittoria Brambilla.

AUTHOR CONTRIBUTIONS

V.B., J.M.J-G. and W.J.J.S. designed the experiment. V.B. carried out the lab experiments. A.S. developed the bioinformatic tools and performed the data analysis. V.B., A.S., J.M.J-G., W.J.J.S. and F.F. contributed to writing the paper.

REFERENCES

Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, Cherry JM, Davis AP, Dolinski K, Dwight SS, Eppig JT, Harris MA, Hill DP, Issel-Tarver L, Kasarskis A, Lewis S, Matese JC, Richardson JE, Ringwald M, Rubin GM, Sherlock G (2000) Gene ontology: Tool for the unification of biology. The Gene Ontology Consortium. **Nat Genet** 25: 25–29

Barbazuk WB, Fu Y, McGinnis KM (2008) Genome-wide analyses of alternative splicing in plants: Opportunities and challenges. **Genome Res** 18: 1381–1392

Baud S, Boutin JP, Miquel M, Lepiniec L, Rochat C (2002) An integrated overview of seed development in *Arabidopsis* thaliana ecotype WS. **Plant Physiol Biochem** 40: 151–160

Bentsink L, Jowett J, Hanhart CJ, Koornneef M (2006) Cloning of DOG1, a quantitative trait locus controlling seed dormancy in *Arabidopsis*. **Proc Natl Acad Sci USA** 103: 17042–17047

Brandt R, Cabedo M, Xie Y, Wenkel S (2014) Homeodomain leucine-zipper proteins and their role in synchronizing growth and development with the environment. **J Integr Plant Biol** 56: 518–26

Carvalho RF, Feijão CV, Duque P (2013) On the physiological significance of alternative splicing events in higher plants. **Protoplasma** 250: 639–650

Chang CY, Lin WD, Tu SL (2014) Genome-wide analysis of heat-sensitive alternative splicing in *Physcomitrella patens*. **Plant Physiol** 165: 826–840

Chardon F, Virlon B, Moreau L, Falque M, Joets J, Decousset L, Murigneux A, Charcosset A (2004) Genetic architecture of flowering time in maize as inferred from quantitative trait loci meta-analysis and synteny conservation with the rice genome. **Genetics** 168: 2169–2185

Ding F, Cui P, Wang Z, Zhang S, Ali S, Xiong L (2014) Genome-wide analysis of alternative splicing of pre-mRNA under salt stress in *Arabidopsis*. **BMC Genomics** 15: 431

Fait A, Angelovici R, Less H, Ohad I, Urbanczyk-Wochniak E, Fernie AR, Galili G (2006) *Arabidopsis* seed development and germination is associated with temporally distinct metabolic switches. **Plant Physiol** 142: 839–854

Fan J, Niu X, Wang Y, Ren G, Zhuo T, Yang Y, Lu BR, Liu Y (2007) Short, direct repeats (SDRs)-mediated post-transcriptional processing of a transcription factor gene *OsVP1* in rice (*Oryza sativa*). **J Exp Bot** 58: 3811–3817

Filichkin SA, Priest HD, Givan SA, Shen R, Bryant DW, Fox SE, Wong W, Mockler TC (2010) Genome-wide mapping of alternative splicing in *Arabidopsis thaliana*. **Genome Res** 20: 45–58

Finch-Savage WE, Leubner-Metzger G (2006) Seed dormancy and the control of germination. **New Phytol** 171: 501–523

Gagete AP, Riera M, Franco L, Rodrigo MI (2009) Functional analysis of the isoforms of an ABI3-like factor of *Pisum sativum* generated by alternative splicing. **J Exp Bot** 60: 1703–171

Gao Y, Liu J, Zhang Z, Sun X, Zhang N, Fan J, Niu X, Xiao F, Liu Y (2013) Functional characterization of two alternatively spliced transcripts of tomato ABSCISIC ACID INSENSITIVE3 (ABI3) gene. **Plant Mol Biol** 82: 131–145

Gazzarrini S, Tsuchiya Y, Lumba S, Okamoto M, McCourt P (2004) The transcription factor FUSCA3 controls developmental timing in *Arabidopsis* through the hormones gibberellin and abscisic acid. **Dev Cell** 7: 373–385

Gene T, Consortium O, Gene T, Go O (2014) Gene ontology consortium: Going forward. **Nucleic Acids Res** 43: D1049–D1056

Goldberg RB, de Paiva G, Yadegari R (1994) Plant embryogenesis: Zygote to seed. **Science** 266: 605–614

Graeff M, Wenkel S (2012) Regulation of protein function by interfering protein species. **Biomol Concepts** 3: 71–78

Holdsworth MJ, Bentsink L, Soppe WJJ (2008) Molecular networks regulating *Arabidopsis* seed maturation, after-ripening, dormancy and germination. **New Phytol** 179: 33–54

Holdsworth MJ, Finch-Savage WE, Grappin P, Job D (2008) Post-genomics dissection of seed dormancy and germination. **Trends Plant Sci** 13: 7–13

Jenik PD, Gillmor CS, Lukowitz W (2007) Embryonic patterning in *Arabidopsis thaliana*. **Annu Rev Cell Dev Biol** 23: 207–236

Joosen RVL, Arends D, Willems LAJ, Ligterink W, Jansen RC, Hilhorst HWM (2012) Visualizing the genetic landscape of *Arabidopsis* seed performance. **Plant Physiol** 158: 570–589

Kalyna M, Simpson CG, Syed NH, Lewandowska D, Marquez Y, Kusenda B, Marshall J, Fuller J, Cardle L, McNicol J, Dinh HQ, Barta A, Brown JW (2012) Alternative splicing and nonsense-mediated decay modulate expression of important regulatory genes in *Arabidopsis*. **Nucleic Acids Res** 40: 2454–2469

Kirkbride RC, Fischer RL, Harada JJ (2013) LEAFY COTYLEDON1, a key regulator of seed development, is expressed in vegetative and sexual propagules of Selaginella moellendorffii. **PLoS ONE** 8: e67971

Kornblihtt AR, Schor IE, Alló M, Dujardin G, Petrillo E, Muñoz MJ (2013) Alternative splicing: A pivotal step between eukaryotic transcription and translation. **Nat Rev Mol Cell Biol** 14: 153–65

Lawrence M, Huber W, Pagès H, Aboyoun P, Carlson M, Gentleman R, Morgan MT, Carey VJ (2013) Software for computing and annotating genomic ranges. **PLoS Comput Biol** 9: e1003118

Leviatan N, Alkan N, Leshkowitz D, Fluhr R (2013) Genome-wide survey of cold stress regulated alternative splicing in *Arabidopsis thaliana* with tiling microarray. **PLoS ONE** 8: e66511

Levin JZ, Yassour M, Adiconis X, Nusbaum C, Thompson DA, Friedman N, Gnirke A, Regev A (2010) Comprehensive comparative analysis of strand-specific RNA sequencing methods. **Nat Methods** 7: 709–715

Li P, Tao Z, Dean C (2015) Phenotypic evolution through variation in splicing of the noncoding RNA COOLAIR. **Genes Dev** 29: 696–701

Lopez-Molina L, Mongrand S, McLachlin DT, Chait BT, Chua NH (2002) ABI5 acts downstream of ABI3 to execute an ABA-dependent growth arrest during germination. **Plant J** 32: 317–328

Loraine AE, McCormick S, Estrada A, Patel K, Qin P (2013) RNA-seq of *Arabidopsis* pollen uncovers novel transcription and alternative splicing. **Plant Physiol** 162: 1092–1109

Maia J, Dekkers BJW, Dolle MJ, Ligterink W, Hilhorst HWM (2014) Abscisic acid (ABA) sensitivity regulates desiccation tolerance in germinated *Arabidopsis* seeds. **New Phytol** 203: 81–93

Marquez Y, Brown JWS, Simpson C, Barta A, Kalyna M (2012) Transcriptome survey reveals increased complexity of the alternative splicing landscape in *Arabidopsis*. **Genome Res** 22: 1184–1195

McKibbin RS, Wilkinson MD, Bailey PC, Flintham JE, Andrew LM, Lazzeri PA, Gale MD, Lenton JR, Holdsworth MJ (2002) Transcripts of Vp-1 homeologues are misspliced in modern wheat and ancestral species. **Proc Natl Acad Sci USA** 99: 10203–10208

Miyoshi K, Kagaya Y, Ogawa Y, Nagato Y, Hattori T (2002) Temporal and spatial expression pattern of the *OSVP1* and *OSEM* genes during seed development in rice. **Plant Cell Physiol** 43: 307–313

Mönke G, Altschmied L, Tewes A, Reidt W, Mock H-P, Bäumlein H, Conrad U (2004) Seed-specific transcription factors ABI3 and FUS3: Molecular interaction with DNA. **Planta** 219: 158–166

Moreno-Risueno MA, González N, Díaz I, Parcy F, Carbonero P, Vicente-Carbajosa J (2008) FUSCA3 from barley unveils a common transcriptional regulation of seed-specific genes between cereals and *Arabidopsis* **Plant J** 53: 882–894

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

Nakabayashi K, Bartsch M, Xiang Y, Miatton E, Pellengahr S, Yano R, Seo M, Soppe WJJ (2012) The time required for dormancy release in *Arabidopsis* is determined by DELAY of GERMINATION1 protein levels in freshly harvested seeds. **Plant Cell** 24: 2826–2838

Nakabayashi K, Bartsch M, Ding J, Soppe WJJ (2015) Seed dormancy in *Arabidopsis* requires self-binding ability of DOG1 protein and the presence of multiple isoforms generated by alternative splicing. **PLoS Genet** 11: e1005737

Nakagawa H, Ohkura E, Ohmiya K, Hattori T (1996) The seed-specific transcription factor VP1 (OSVP1) is expressed in rice suspension-cultured cells. **Plant Cell Physiol** 37: 355–362

Penfield S, Josse E-M, Halliday KJ (2010) A role for an alternative splice variant of *PIF6* in the control of *Arabidopsis* primary seed dormancy. **Plant Mol Biol** 73: 89–95

Pollard K, Dudoit S, Laan M (2005) Multiple testing procedures: The multtest package and applications to genomics. In: Gentleman R, Carey VJ, Huber W, Irizarry RA, Dudoit S, eds. *Statistics for Biology and Health*. Springer, New York. pp. 249–271

Reddy ASN, Marquez Y, Kalyna M, Barta A (2013) Complexity of the alternative splicing landscape in plants. **Plant Cell** 25: 3657–3683

Reidt W, Wohlfarth T, Ellerström M, Czihal A, Tewes A, Ezcurra I, Rask L, Bäumlein H (2000) Gene regulation during late embryogenesis: The RY motif of maturation-specific gene promoters is a direct target of the FUS3 gene product. **Plant J** 21: 401–408

Santos-Mendoza M, Dubreucq B, Baud S, Parcy F, Caboche M, Lepiniec L (2008) Deciphering gene regulatory networks that control seed development and maturation in *Arabidopsis*. **Plant J** 54: 608–620

Shen B, Allen WB, Zheng P, Li C, Glassman K, Ranch J, Nubel D, Tarczynski MC (2010) Expression of ZmLEC1 and ZmWRI1 increases seed oil production in maize. **Plant Physiol** 153: 980–987

Staudt AC, Wenkel S (2011) Regulation of protein function by "microProteins". **EMBO Rep** 12: 35–42

Sugliani M, Brambilla V, Clerkx EJM, Koornneef M, Soppe WJJ (2010) The conserved splicing factor SUA controls alternative splicing of the developmental regulator ABI3 in *Arabidopsis*. **Plant Cell** 22: 1936–1946

Suzuki M, McCarty DR (2008) Functional symmetry of the B3 network controlling seed development. **Curr Opin Plant Biol** 11: 548–553

Swaminathan K, Peterson K, Jack T (2008) The plant B3 superfamily. **Trends Plant Sci** 13: 647–655

Syed NH, Kalyna M, Marquez Y, Barta A, Brown JWS (2012) Alternative splicing in plants – coming of age. **Trends Plant Sci** 17: 616–623

Tack DC, Pitchers WR, Adams KL (2014) Transcriptome analysis indicates considerable divergence in alternative splicing between duplicated genes in *Arabidopsis thaliana*. **Genetics** 198: 1473–1481

Terrasson E, Buitink J, Righetti K, Ly Vu B, Pelletier S, Zinsmeister J, Lalanne D, Leprince O (2013) An emerging picture of the seed desiccome: Confirmed regulators and newcomers identified using transcriptome comparison. **Front Plant Sci** 4: 1–16

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

Tsai AY-L, Gazzarrini S (2012) Overlapping and distinct roles of AKIN10 and FUSCA3 in ABA and sugar signaling during seed germination. **Plant Signal Behav** 7: 1238–1242

van Zanten M, Koini MA, Geyer R, Liu Y, Brambilla V, Bartels D, Koornneef M, Fransz P, Soppe WJJ (2011) Seed maturation in

Arabidopsis thaliana is characterized by nuclear size reduction and increased chromatin condensation. **Proc Natl Acad Sci USA** 108: 20219–20224

Venables WN, Ripley BD (2002) *Modern Applied Statistics with S Fourth edition*. Springer, New York

Vitulo N, Forcato C, Carpinelli EC, Telatin A, Campagna D, D'Angelo M, Zimbello R, Corso M, Vannozzi A, Bonghi C, Lucchin M, Valle G (2014) A deep survey of alternative splicing in grape reveals changes in the splicing machinery related to tissue, stress condition and genotype. **BMC Plant Biol** 14: 99

Wang F, Perry SE (2013) Identification of direct targets of FUSCA3, a key regulator of *Arabidopsis* seed development. **Plant Physiol** 161: 1251–1264

Wang L, Xi Y, Yu J, Dong L, Yen L, Li W (2010) A statistical method for the detection of alternative splicing using RNA-seq. **PLoS ONE** 5: e8529

Xiao L, Yang G, Zhang L, Yang X, Zhao S, Ji Z, Zhou Q, Hu M, Wang Y, Chen M, Xu Y, Jin H, Xiao X, Hu G, Bao F, Hu Y, Wan P, Li L, Deng X, Kuang T, Xiang C, Zhu JK, Oliver MJ, He Y (2015) The resurrection genome of *Boea hygrometrica*: A blueprint for survival of dehydration. **Proc Natl Acad Sci USA** 112: 5833–5837

Xu P, Kong Y, Song D, Huang C, Li X, Li L (2014) Conservation and functional influence of alternative splicing in wood formation of *Populus* and *Eucalyptus*. **BMC Genomics** 15: 780

Yoshimura K, Mori T, Yokoyama K, Koike Y, Tanabe N, Sato N, Takahashi H, Maruta T, Shigeoka S (2011) Identification of alternative splicing events regulated by an *Arabidopsis* serine/arginine-like protein, atsr45a, in response to high-light stress using a tiling array. **Plant Cell Physiol** 52: 1786–1805

Zou M, Guan Y, Ren H, Zhang F, Chen F (2007) Characterization of alternative splicing products of bZIP transcription factors OsABI5. **Biochem Biophys Res Commun** 360: 307–313

SUPPORTING INFORMATION

Additional Supporting Information may be found online in the supporting information tab for this article: http://onlinelibrary.wiley.com/doi/10.1111/jipb.12482/suppinfo

Figure S1. Schematic representation of the DAS pipeline workflow

Two transcripts are present (**A** and **B**). Seven exons (numbered 1 to 7) are assumed on the gene model. Only one sample and one AS event are shown as an example. The statistical test assesses if the ratio between condition 1 and condition 2 is statistically different. Numbers on bars: read counts.

AS, alternative splicing; A3P, alternative 3′; A5P, alternative 5′; CE, cryptic exon; CJ, canonic junction; ES, exon skipping; IR, intron retention.

Table S1. List of all expressed genes

Table S2. List of all genes expressed at 14 d after pollination (DAP).

Table S3. List of all genes expressed at 20 DAP

Table S4. List of all genes co expressed at 14 and 20 DAP

Table S5. List of all down-regulated genes between 14 and 20 DAP

Table S6. List of all up-regulated genes between 14 and 20 DAP

Table S7. List of all alternatively spliced genes at 14 and 20 DAP

Table S8. Di-nucleotide splice site junctions

Table S9. List of all differential AS events between 14 and 20 DAP

Table S10. List of differential AS events between 14 and 20 DAP in genes whose expression is also up-regulated between 14 and 20 DAP

Table S11. List of differential AS events between 14 and 20 DAP in genes whose expression is also down-regulated between 14 and 20 DAP

Table S12. List of differential AS events between 14 and 20 DAP in genes that are not differentially expressed

Table S13. List of IR events that do not form a premature termination codon

Table S14. List of AS events independently verified by qRT-PCR

Table S15. List of primers used in qRT

Here below the abbreviations used in the Tables S1–14. seqname, chr = name of chromosome; start, end = beginning and ending coordinates of that feature; 14 DAP = 14 days after pollination; 20 DAP = 20 days after pollination; baseMeanA = mean of counts of all samples from 14DAP; baseMeanB = mean of counts of all samples from 20 DAP; foldChange = ratio of mean counts from 20 DAP to 14 DAP; log2FoldChange = log of Fold Change to the base 2; pval = p-value; padj = adjusted p-value (corrected for multiple testing); seid = start,end id meaning the span of a junction; jtype = junction type (one of ES, A3P, A5P, IR, CE, CI); spos, epos = start and end junction coordinates and different only for ES/CE events, otherwise both should be identical; 14DAP reads_AS, 20 DAP reads_AS = alternative spliced read count corresponding to time points; 14DAP reads_CJ, 20DAP reads_CJ = canonical spliced read counts corresponding to time points; jun1, jun2 = start and end dinucleotide junction nucleotides; 14.DAS.AS, 20.DAS.AS = mean alternative spliced reads; 14.DAS.CJ, 20.DAS.CJ = mean canonical spliced reads; rid = rna id; orig_trans_len = representative transcript length; ir_longest_tran_len = length of longest IR transcript.