DISSERTATION

IN VITRO AND *IN VIVO* STUDIES ON PRE-mRNA SPLICING IN PLANTS

Submitted by

Mohammed M. Albaqami

Graduate Degree Program in Cell and Molecular Biology

In partial fulfillment of the requirements

For the Degree of Doctor of Philosophy

Colorado State University

Fort Collins, Colorado

Summer 2017

Doctoral Committee:

Advisor: A. S. N. Reddy

Jeffrey Wilusz
Asa Ben-Hur
Tai Montgomery

ABSTRACT

IN VITRO AND *IN VIVO* STUDIES ON PRE-mRNA SPLICING IN PLANTS

In the processes of eukaryotic gene expression, the nascent precursor messenger RNAs (pre-mRNAs) must undergo extensive processing including splicing. Pre-mRNA splicing is the process of removing introns from the transcript and ligation of exons together. Splicing is catalyzed in two-step trans-esterification reaction by a dynamic ribonucleoprotein (RNP) machine termed the spliceosome that is comprised of five different snRNPs along with other non-snRNP proteins. Since most of the eukaryotic genes contain multiple introns, pre-mRNA splicing is a vital step in the posttranscriptional regulation of gene expression.

As in other eukaryotes, the majority of plant protein-coding genes contain introns that are subjected to splicing during gene expression. Plant pre-mRNA splicing is fundamental for controlling gene expression, and recent studies underscore the importance of this process in regulating plant growth and developmental processes in response to intrinsic and environmental signals. Several lines of evidence suggest that plants have unique pre-mRNA splicing regulatory mechanisms; however, these mechanisms are poorly understood and have not received attention equivalent to those of animals and yeast. Thus, understanding how plant introns are recognized and processed requires innovative *in vitro* and *in vivo* approaches that will enhance our understanding of splicing regulatory mechanisms in plants.

Arabidopsis SR45, one of the serine/arginine-rich (SR) proteins, is a well-characterized splicing factor that controls multiple biological processes. However, little is known about SR45-regulated global changes in gene expression and alternative splicing (AS) that are likely to mediate its

biological functions. To address this gap, we performed transcriptome analysis in wild-type and *sr45* loss-of-function mutant using high-throughput RNA-seq to identify SR45-dependent changes in gene expression and AS. By comparing the transcriptomes of *sr45* mutant and wild-type, we identified 1,345 differentially expressed (DE) genes and 927 differentially spliced (DS) genes. Further analysis of AS events distribution revealed that the choice of 3' alternative splicing sites (A3'ss) was the predominant AS type in DS genes, suggesting that SR45 has a major role in selecting the 3'ss. Gene Ontology (GO) enrichment analysis of DE and DS genes revealed that they are highly enriched in terms associated with hormonal signaling and response to abiotic stresses, including heat stress. Our phenotypic and molecular analysis with the *sr45* mutant confirmed that this splicing factor is required for basic thermotolerance. *SR45* pre-mRNA is alternatively spliced to generate two splice variants that encode two distinct functional proteins, SR45.1 (long) and SR45.2 (short), which differ in eight amino acids. Interestingly, complementation of *SR45* mutant with each isoform independently revealed that SR45.1 but not SR45.2 is a positive regulator of thermotolerance. Furthermore, protein phosphorylation analysis using Phos-tag SDS-PAGE indicated both SR45 isoforms are phosphorylated in response to high temperature. Since there are multiple experimentally confirmed and predicted phosphorylation sites both in the common and unique regions of SR45 isoforms, the Phos-tag SDS-PAGE results did not allow us to address the role of putative phosphorylation sites unique to the long form. To address the role of phosphorylation sites unique to the long isoform, we used complemented lines expressing mutated proteins in which the phosphorylation sites of only the long isoform are affected. Analysis of heat shock response in these transgenic lines revealed that threonine 218 is likely to be critical for thermotolerance in *Arabidopsis*. These findings establish the importance

of a specific splice isoform of a splicing factor in heat stress response and indicate that a particular amino acid in the long isoform is important for its function in thermotolerance.

In vitro splicing systems using nuclear or cytoplasmic extracts from mammalian cells, yeast, and *Drosophila* have provided a wealth of mechanistic insights into eukaryotic pre-mRNA splicing. A corresponding plant-derived *in vitro* splicing system has long been awaited; therefore, we present here an effort toward developing such a system from plants. We show that nuclear extract (NE) derived from dark-grown (etiolated) *Arabidopsis* seedlings is capable of converting a truncated *LIGHT-HARVESTING CHLOROPHYLL B-BINDING PROTEIN 3* (*LHCB3*) pre-mRNA substrate with a single intron into the expected size of mRNA. Based on several lines of evidence, we suggest that this is an authentic *in vitro* splicing reaction. Supporting evidence include: i) generation of an RNA product that corresponds to the size of the expected mRNA, ii) generation of RNA species that migrate within the size range expected for splicing intermediates, iii) indications from a junction-mapping assay using S1 nuclease that the two exons are linked together, iv) remarkable similarities between plant and non-plant *in vitro* splicing assay reaction conditions, such as requirements for ATP and Mg^{+2}, and v) finally, more importantly, mutations in conserved donor and acceptor sites abolished the production of the putative spliced product. Unlike mammalian *in vitro* splicing assays, the optimal incubation temperature for splicing with plant extract was lower ($24^{o}C$), within the optimal growth temperature range of *Arabidopsis* seedlings. Collectively, our results suggest that *Arabidopsis* NE is capable of splicing pre-mRNA substrate and that the *in vitro* reaction conditions are similar to those found with non-plant extracts. This is the first step toward developing a plant-derived *in vitro* pre-mRNA splicing assay. Further confirmation of these results with additional approaches and optimization of this assay would lead to development of a robust *in vitro* splicing assay and open new avenues to

investigate spliceosome assembly and composition, splicing regulatory mechanisms specific to plants, and thereby enhance the overall understanding of post-transcriptional gene regulatory mechanisms in plants.

ACKNOWLEDGMENTS

Here is the end of my journey that started years ago when I came from the Kingdom of Saudi Arabia to the United States of America, looking to achieve a dream that I wished to make real: Dr. Mohammed. What I have accomplished to date is because of the help, support, monitoring, and encouragement of a number of people who walked with me along the path to my PhD: advisors, colleagues, friends, and family. I am deeply thankful to all of them.

Dr. A. S. N. Reddy, Dr. Jeffrey Wilusz, Dr. Asa Ben-Hur, and Tai Montgomery, you four have provided me with extensive personal and professional guidance that pushed me to the furthest of my potential in scientific research. A very special thank-you goes out to my advisor, Dr. A. S. N. Reddy, for his excellent guidance, caring, patience, and for providing me with an excellent atmosphere for doing my PhD study and research. My advisory committee, sincerely, you gave more than I could ever thank you enough for here.

Reddy's lab members, you are my second family here at Colorado State University, and I will never forget the pleasant time that I spent with you during my PhD study. I am really grateful to Dr. Salah Abdel-Ghani, Dr. David Xing, Dr. Saiprasad Palusa, and Dr. Prasad Kasavajhala for their encouragement, insightful comments, and hard questions.

I really appreciate the opportunity of collaborating with Dr. Asa Ben-Hur and lab members Mike Hamilton and Mark Rogers, as well as with Dr. David Xing to perform the bioinformatics analysis of RNA-Seq data. This enabled me to learn a really important part of scientific research.

I must also thank Dr. Mount SM (University of Maryland) and Dr. Zhang XN (Saint Bonaventure University) for providing me with the seeds of transgenic lines of *Arabidopsis* SR45.

I want to extend a heart-felt thank you to the government of the Kingdom of Saudi Arabia and Umm Al-Qura University for their generous financial support throughout my PhD study. I am also grateful to the National Science Foundation for generous financial support toward conducting this research program over the years.

Words cannot express how grateful I am to my mother (Munira), and father (Munahi) for all of the sacrifices that you have made on my behalf. Your prayer for me was what sustained me thus far. I would also like to say a heartfelt thank you to my sisters (Hussa, Fawziah, and Samiyah) and brothers (Abdullah and Majed) for always believing in me and encouraging me to achieve my goals. Although it has been years since my brother Faisal passed, I still remember his words (FOLLOW YOUR DREAM) when I was leaving the country to go to the United States. You left fingerprints of grace on our lives. You shan't be forgotten.

Last but not the least, I am grateful to Latifah, my wife, who accompanied me on my trip to the United States looking forward to achieve my dream, I know it was a difficult time for you. However, being with me during my PhD study has made everything possible, and made this the best time of my life. Thank you for being there for me all the time.

DEDICATION

To the next generation of scientists, researchers, and thought leaders for the Kingdom of Saudi Arabia

TABLE OF CONTENTS

CHAPTER I

Transcriptome-wide analysis of SR45-regulated changes in gene expression and alternative splicing revealed a role for this splicing factor in stress responses

SUMMARY

Pre-mRNA splicing is one of the fundamental mechanisms that significantly increases eukaryotic transcriptome complexity and proteome diversity. *Arabidopsis* SR45, one of the serine/arginine-rich (SR) proteins, is a well-characterized splicing factor that regulates multiple biological processes. However, little is known about SR45-regulated global changes in gene expression and alternative splicing (AS) that are likely to mediate its biological functions. To address this gap, we performed transcriptome analysis in wild-type and *sr45* loss-of-function mutant using a high-throughput RNA-Seq to identify SR45-dependent changes in gene expression and AS. By comparing the transcriptomes of *SR45* mutant and wild-type, we identified 1,345 differentially expressed (DE) genes and 927 differentially spliced (DS) genes. Further analysis of AS events distribution revealed that the choice of 3' alternative splicing sites (A3'ss) was the predominant AS type in DS genes, suggesting that SR45 has a major role in selecting the 3'ss. Gene Ontology (GO) enrichment analysis of DE and DS genes revealed that they are highly enriched in terms associated with hormonal signaling and response to abiotic stresses, including heat stress. Our phenotypic and molecular analysis with the *sr45* mutant confirmed that this splicing factor is required for basic thermotolerance. *SR45* pre-mRNA is alternatively spliced to generate two splice variants that encode two distinct functional proteins, SR45.1 (long) and SR45.2 (short), which differ in eight amino acids. Interestingly, complementation of *SR45* mutant with each isoform independently revealed that SR45.1 but not SR45.2 is a positive regulator of

thermotolerance. Furthermore, protein phosphorylation analysis using Phos-tag SDS-PAGE indicated both SR45 isoforms are phosphorylated in response to high temperature. Since there are multiple experimentally confirmed and predicted phosphorylation sites both in the common and unique regions of SR45 isoforms, the Phos-tag SDS-PAGE results did not allow us to address the role of putative phosphorylation sites unique to the long form. To address the role of phosphorylation sites unique to the long isoform, we used complemented lines expressing point mutants in which the phosphorylation sites of only the long isoform are affected. Analysis of heat shock response in these transgenic lines revealed that threonine 218 is likely to be critical for thermotolerance in *Arabidopsis*. These findings establish the importance of a specific splice isoform of a splicing factor in heat stress response and indicate that a particular amino acid in the long isoform is important for its function in thermotolerance.

INTRODUCTION

The properties and biological functions of each cell type are determined by the cell type-specific expression of information stored in genes. While certain genes are expressed to produce functional proteins, other genes are responsible for generating regulatory RNA molecules such as microRNAs (miRNAs) and long non-coding RNAs (lncRNAs). Land plants, like other organisms, have the ability to change the expression pattern of their genes in response to intrinsic as well as external signals. Thus, gene expression is the most vital mechanism by which the genotype influences the phenotype.

The process of gene expression in eukaryotes comprises of several steps: transcription, co/post-transcriptional processing of primary transcripts, transport of mRNA to cytoplasm, RNA decay and translation. Transcription is the generation of messenger RNA (mRNA) via the enzyme RNA polymerase, using DNA as a template. Transcription is both the first and a central step in

the pathway from DNA to protein, and plays a fundamental role in controlling gene expression to dictate cell function. Transcription involves several steps, such as transcription initiation, elongation, and termination, and all of them are tightly regulated mechanisms (reviewed in Singh, 1998; Wu, 2014). For example, sequence-specific DNA-binding proteins, termed transcription factors, play fundamental roles in regulating transcription initiation and transcript levels. The regulation of gene expression at the transcriptional stage largely controls when and how often a given gene is transcribed. Transcription of genes in plants and other eukaryotes results in precursor mRNAs (pre-mRNA) that are further processed (See "Pre-mRNA processing" section below) into mature mRNA. Following processing, mRNA is exported out of the nucleus to the cytoplasm where it is translated into a protein. While a cell regulates expression of its genes at the transcriptional level, it can also control expression at the pre-mRNA processing or post-transcriptional level. Gene regulation at the pre-mRNA processing level largely influences transcriptome diversity and increases the proteome complexity. Translation is the synthesis of protein using the information that was transcribed into mRNA. Post-transcriptional regulatory mechanisms include transcript export, localization, mRNA stability, translation, posttranslational modifications of proteins, and protein stability and degradation (reviewed in Floris et al., 2009; Glisovic et al., 2008).

Pre-mRNA Processing

Eukaryotic pre-mRNA processing involves three major steps towards generating functional mRNA: addition of a 5' cap, addition of a 3' polyadenylate tail, and splicing. First, during the synthesis of pre-mRNA, a 7-methylguanosine (m^7G) cap is linked to the 5′ end of the growing transcript via an uncommon 5'-to-5' triphosphate linkage. In addition to protecting the nascent transcript from degradation, mRNA m^7G capping plays a vital role in the coordination of

mRNA-associated processes such as splicing, mRNA export, and translation (reviewed in Cowling and Cole, 2010; Lewis and Izaurralde, 1997). Once transcription elongation is completed, the nascent pre-mRNA is cleaved downstream from the AAUAAA conserved sequence. Following cleavage, poly(A) polymerase, a component of the polyadenylation machinery, adds adenine residues to form a poly(A) tail at the new 3' end of the RNA. The poly(A) tail plays an essential role in mRNA stability, nuclear export, and translation (reviewed in Garneau et al., 2007; Guhaniyogi and Brewer, 2001; Proudfoot, 2011). During Pre-mRNA splicing, an important processing event, the introns are removed and exons are ligated together to form a mature mRNA. Pre-mRNA splicing is an incredible mechanism in molecular biology that influences transcriptome complexity and proteome plasticity. Since the focus of my research is investigating the role of *Arabidopsis thaliana* splicing regulator SR45, pre-mRNA splicing mechanisms will be discussed in greater detail below.

Introns

Phillip Allen Sharp and Richard J. Roberts discovered interrupted or split genes independently in 1977, for which they shared the Nobel Prize in Physiology or Medicine in 1993 (Berget et al., 1977; Chow et al., 1977). In 1978, Walter Gilbert suggested the terms introns for non-expressed regions, and exons for expressed regions (Gilbert, 1978). While originally observed in viruses, pre-mRNA splicing was rapidly confirmed to occur in eukaryotic organisms as well. Furthermore, with the complete sequencing of many plant and animal genomes, it has become clear that introns represent a large fraction of genomic DNA in eukaryotes. However, the percentage of intron-containing genes varies enormously across the genomes of different organisms. About 4% of genes in the lower eukaryote *Saccharomyces cerevisiae* contain introns, compared to about 80% in model organisms *Drosophila melanogaster, Caenorhabditis elegans*,

and *Arabidopsis*, and more than 90% in human, rat, and mouse genomes (reviewed in Atambayeva et al., 2008). Intron density is not related to complexity levels of organisms; for instance, *Drosophila melanogaster* and *Caenorhabditis elegans* genomes contain comparable fractions of intronic DNA (29.1% and 30.4% respectively) and numbers of introns per gene (4.22 and 5.46 respectively) (Morello and Breviario, 2008). Even though introns are ubiquitous in eukaryotic genomes, their origin and evolution remain unclear (Rogozin et al., 2012).

Introns cannot be considered junk DNA, due to the accumulated evidence of their vital functions in controlling gene expression. One of the well-understood functions of introns in eukaryotes is the enhancement of protein expression of the intron-bearing gene. This function is particularly prevalent in plants, and is used in genetic engineering to increase the expression of engineered proteins (Clark et al., 1993; Morello and Breviario, 2008; Rose, 2008). In addition to directly affecting expression, some transcribed introns are processed after splicing to generate regulatory non-coding RNAs including miRNAs (Rearick et al., 2011). Introns also play an essential role in pre-mRNA splicing and alternative splicing (AS), which will be discussed in more detail later.

Spliceosomal Introns

Following the discovery of introns, extensive analysis of sequence and structural features of introns led to the identification of four main types: introns in nuclear protein-coding genes that are spliced via the spliceosome (spliceosomal introns), introns in nuclear and archaeal tRNA genes that are fully catalyzed by protein enzymes (tRNA introns), self-splicing group I introns, and self-splicing group II introns (reviewed in Irimia and Roy, 2014). Types other than spliceosomal introns are beyond the scope of my research, so we will keep the discussion focus only on the spliceosomal introns.

Spliceosomal introns are present in a high percentage of eukaryotic protein-coding genes. For example, sequencing analysis of *Arabidopsis* and rice (*Oryza sativa*) genomes shows about 80% of coding genes contains spliceosomal introns, with similar intron densities of about 4 introns per gene (Morello and Breviario, 2008). Spliceosomal introns share a conserved splicing mechanism and contain conserved regulatory sequences. First, the removal of spliceosomal introns is catalyzed via a spliceosome, large ribonucleoprotein machinery that will be discussed later in great detail. Second, these introns have conserved *cis*-regulatory elements that are recognized and bound by core components of the spliceosome and by other splicing regulatory proteins. Most spliceosomal introns have a short conserved 5′ splice site (5′ss) border, a short conserved 3′ splice site (3′ss) border, a branch point (BP) with a conserved adenine, and a polypyrimidine tract region (PPT) between the BP and 3′ss (Figure 1) (reviewed in Irimia and Roy, 2014). When these conserved splicing sites are used to produce a single mature mRNA, this is identified as constitutive splicing (CS). Using different sets of splice sites to generate various mature mRNAs from a gene is termed alternative splicing (AS) (see below).

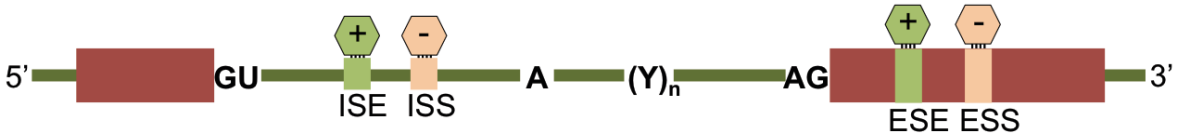


Figure 1: *Cis*-regulatory elements that contribute to regulation of CS and AS. Schematic diagram shows the sequence features in pre-mRNA that drive AS by trans-acting regulatory factors. The *cis*-regulatory elements including splice sites at the beginning and end of each intron (GU and AG), branch point (A), polypyrimidine tract (Yn), exonic splicing enhancer (ESE), intronic splicing enhancer (ISE), exonic splicing suppressor (ESS), and intronic splicing suppressor (ISS). (adopted from Syed et al., 2012).

Alternative Splicing (AS)

Pre-mRNAs from most intron containing genes produce multiple distinct transcripts. For example, about 95% of human multi-exonic genes are alternatively spliced, and more than 60% intron-containing genes in *Arabidopsis* also produce two or more splice isoforms (Pan et al., 2008; Syed et al., 2012; Wang et al., 2008). Eukaryotes have five major types of AS events (Figure 2). First, exon skipping or cassette exons, in which an exon may be spliced out of pre-mRNA together with its flanking introns. Second, mutually exclusive exons, in which nearby exons are processed in such a manner that only one of them is contained at a time in the mRNA. Third, alternative donor sites, in which either a proximal or distal 5'ss is used. Fourth, alternative acceptor sites, in which either a proximal or distal 3'ss is used. Finally, intron retention, in which an intron may be retained in the mRNA. Remarkably, intron retention is the most predominant class of AS event in plants (reviewed in Mastrangelo et al., 2012; Reddy, 2007; Reddy et al., 2013; Syed et al., 2012).

AS is a way to expand transcriptome and proteome complexity. Generation of multi-transcripts with distinct functions from a single gene could explain how higher organisms accomplish their complexity despite having comparable gene numbers (Reddy et al., 2013). Different studies have demonstrated several functions of AS. First, AS generates different protein isoforms with distinct biological functions. For example, *Arabidopsis* SR45 is alternatively spliced to generate two isoforms that differ by 8 amino acids, SR45.1 and SR45.2, and these isoforms have different roles in plant development. While SR45.1 controls flower development, SR45.2 regulates root growth (Zhang and Mount, 2009). A second function of AS is to down-regulate gene expression via generating non-functional transcripts, or by nonsense-mediated mRNA decay (NMD)

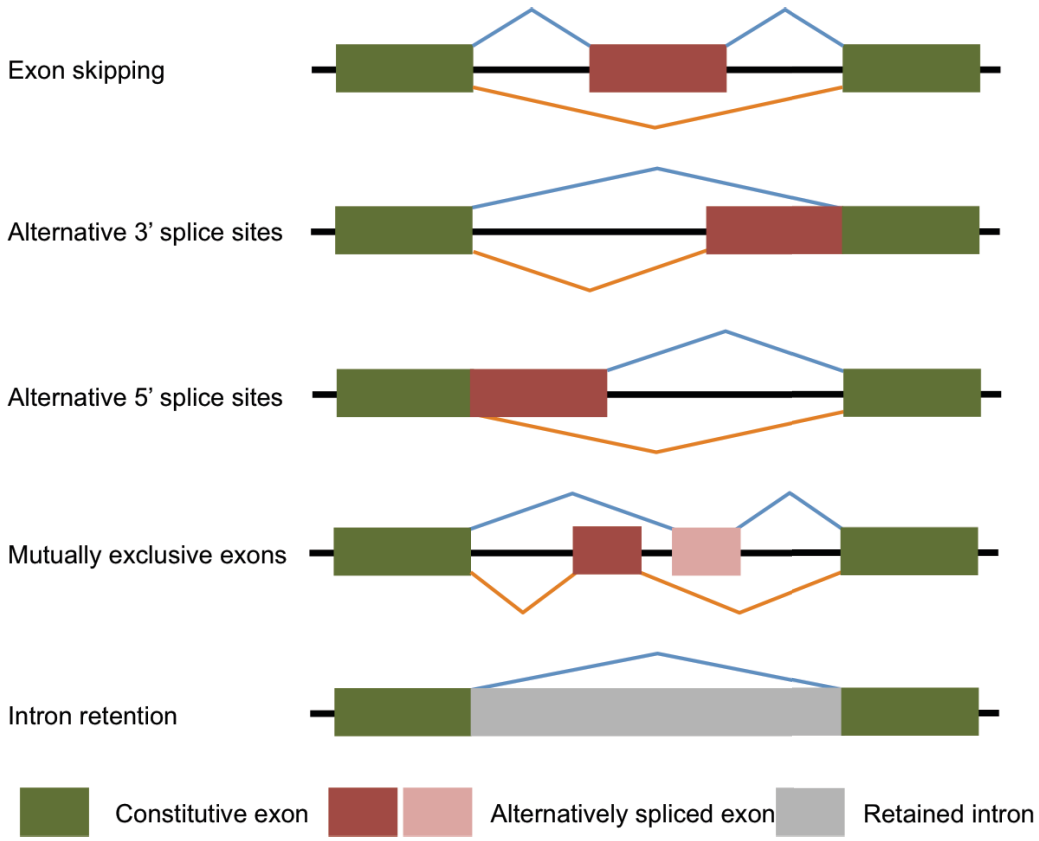


Figure 2: Types of AS events. Schematic diagram of different AS events in eukaryotes: exon inclusion or skipping, alternative splice-site selection, mutually exclusive exons, and intron retention. Exons (colored boxes), introns (solid lines) and spliced segments (triangles). (adopted from Edwalds-Gilbert, 2010).

(Kalyna et al., 2012). Third, AS regulates mRNA and protein intracellular localizations via incorporating different localization sequences (Kabran et al., 2012; Stoss et al., 1999). For example, *Arabidopsis* Zinc Induced Facilitator-like1 (ZIFL1) is alternatively spliced, creating two isoforms with a dinucleotide difference. The full-length protein is located to the plasma membrane and plays a regulatory role in auxin signaling transduction pathway, while the trimmed protein is located to the tonoplast membrane and plays a role in drought response (Remy et al., 2013). In addition to these three examples, AS can affect the recruitment of mRNA to ribosomes, translation productivity, protein stability, enzyme function, and posttranslational changes (reviewed in Reddy et al., 2013; Stamm et al., 2005).

Advances in high-throughput sequencing technologies have accelerated our understanding of AS. RNA sequencing (RNA-Seq) has emerged as a powerful method for characterization and profiling of genome-wide AS (Goodwin et al., 2016). A transcriptome-wide study using RNA-Seq has revealed that in *Arabidopsis*, more than 60% of intron-containing genes undergo AS to generate multiple mRNA isoforms (Syed et al., 2012). Furthermore, transcriptome-wide analyses in plants indicate extensive changes of AS profiles at various developmental stages or under different conditions (reviewed in Staiger and Brown, 2013). Condition-specific AS (regulated splicing) is providing significant insights into plant development mechanisms and how plants communicate with their environments (reviewed in Staiger and Brown, 2013).

Spliceosome

The spliceosome machinery is a large complex of ribonucleoproteins (RNPs) containing five small nuclear RNAs (U1, U2, U4, U5, and U6 snRNA) and up to 300 proteins (Matera and Wang, 2014; Nilsen, 2003; Rino and Carmo-Fonseca, 2009; Ritchie et al., 2009; Wahl et al., 2009). During the splicing reaction, both conformation and composition of the spliceosome are

highly dynamic, resulting in accuracy and flexibility of this process. The spliceosome follows a stepwise assembly process on its target pre-mRNA substrate (Figure 3). Assembly begins with the ATP-independent interactions of the U1 snRNP to the pre-mRNA 5′ss, and of the U2 auxiliary factor (U2AF) to the BP and the PPT. These collectively form the early complex (E complex). Subsequently, the E complex is converted into the pre-spliceosomal A complex through binding of the U2 snRNP to the pre-mRNA BP in an ATP-dependent manner. After formation of the A complex, the pre-catalytic B complex is formed via association of U4/U6 and U5 as a preassembled tri-snRNP. The B complex undergoes several conformational and compositional changes, resulting in its activation as the catalytic B* complex. This B* complex catalyzes the first transesterification step of the splicing reaction, in which the 2′ OH group of the conserved adenosine in the BP of the intron carries out a nucleophilic attack on the 5′ss, resulting in cleavage at the 5′ss and lariat formation. This leads to the formation of complex C and activated C* complex, which catalyzes the second step of the splicing reaction. In this step, the 3′ss is attacked by the 3′ OH group of the 5′ exon; this results in the ligation of exons and release of introns to form mature mRNA (reviewed in Hoskins and Moore, 2012; Lee and Rio, 2015; Matera and Wang, 2014; Papasaikas and Valcarcel, 2016; Shefer et al., 2014; Will and Luhrmann, 2011).

Interestingly, in addition to the previously described spliceosomal introns, eukaryotes have a second type spliceosomal intron, termed minor or U12 intron. While both types are spliced in the same manner, the minor spliceosome has four distinct snRNAs (U11, U12, U4atac, and U6atac), and additional specific regulatory proteins. The functions of these snRNAs are identical with their counterparts (U1, U2, U4, and U6) in the major spliceosome. The other components of both spliceosomes are similar (reviewed in Turunen et al., 2013).

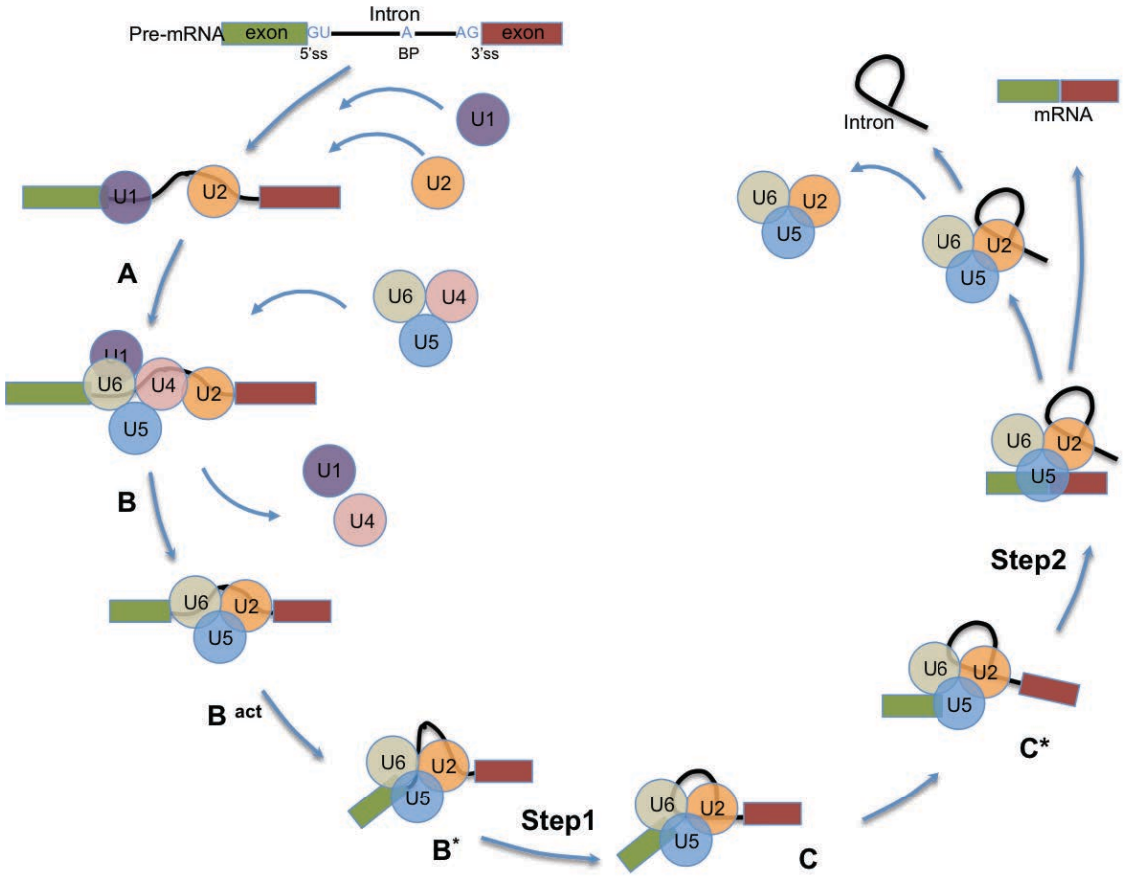


Figure 3: Schematic representation of step-wise assembly of spliceosome in eukaryotes. Spliceosome assembly is initiated with binding of U1 snRNA to 5′ss, followed by recruiting U2 snRNA to the BP to form A complex. Then, the pre-catalytic B complex is formed via association of U4/U6 and U5 as a preassembled tri-snRNP. B complex is activated via additional factors as the catalytic B complex. Next, the first step of splicing takes place at B* complex. In the next step, C and activated C* complexes are formed to catalyze the second step of splicing. Subsequently, the intron is released from the spliceosomal complex in the form of a lasso (lariat) and the snRNPs are recycled for subsequent rounds of splicing. Colored boxes indicate exons; thin black lines show intron and intron lariat (adapted from Will and Luhrmann, 2011).

Plant Spliceosome

Plants, like other eukaryotes, have both types of spliceosomal introns. However, the plant spliceosome assembly pathway and its protein composition are yet to be elucidated experimentally. Despite the lack of a plant-derived *in vitro* splicing assay, efforts have been made to illustrate distinct plant splicing-related mechanisms. Computational analysis using sequence similarities has identified the core components of plant spliceosomes, including snRNAs and several orthologous of known spliceosomal proteins (Koncz et al., 2012; Lorković et al., 2000; Reddy et al., 2013; Ru et al., 2008; Wang and Brendel, 2004). Moreover, the splicing consensus *cis*-elements (5′ss, 3′ss, and BP) are similar between plants and animals (Reddy, 2007). These similarities are clear evidence for comparable mechanisms of intron processing in plants and animals. However, several indications of plant-specific splicing regulatory mechanisms have been reported. First, animal pre-mRNAs cannot be processed to mature mRNAs by plant systems (Barta et al., 1986). Second, orthologous analysis of splicing-regulatory factors in *Arabidopsis* indicates that there is almost twice the number of plant splicing factors compared to the number in humans (Koncz et al., 2012; Reddy et al., 2013). Third, the average size of plant introns (~180 bp) is shorter than that of animal introns (~3300) (Reddy et al., 2013). Fourth, intron retention is the predominant mode of pre-mRNA AS in plants (Ner-Gaon et al., 2004; Syed et al., 2012), whereas exon skipping is the predominant mode in animals (Kim et al., 2007). Fifth, serine/arginine-rich (SR) proteins, a family of fundamental splicing regulatory proteins conserved in eukaryotes, are more numerous in plants compared to other eukaryotes. For example, *Arabidopsis* has 19 SR proteins and rice contains 24, while there are only 11 in animal and 7 in *C. elegans* (Figure 4) (Kalyna and Barta, 2004; Reddy, 2007). Several SR proteins are conserved and others are species specific, suggesting similarities and differences

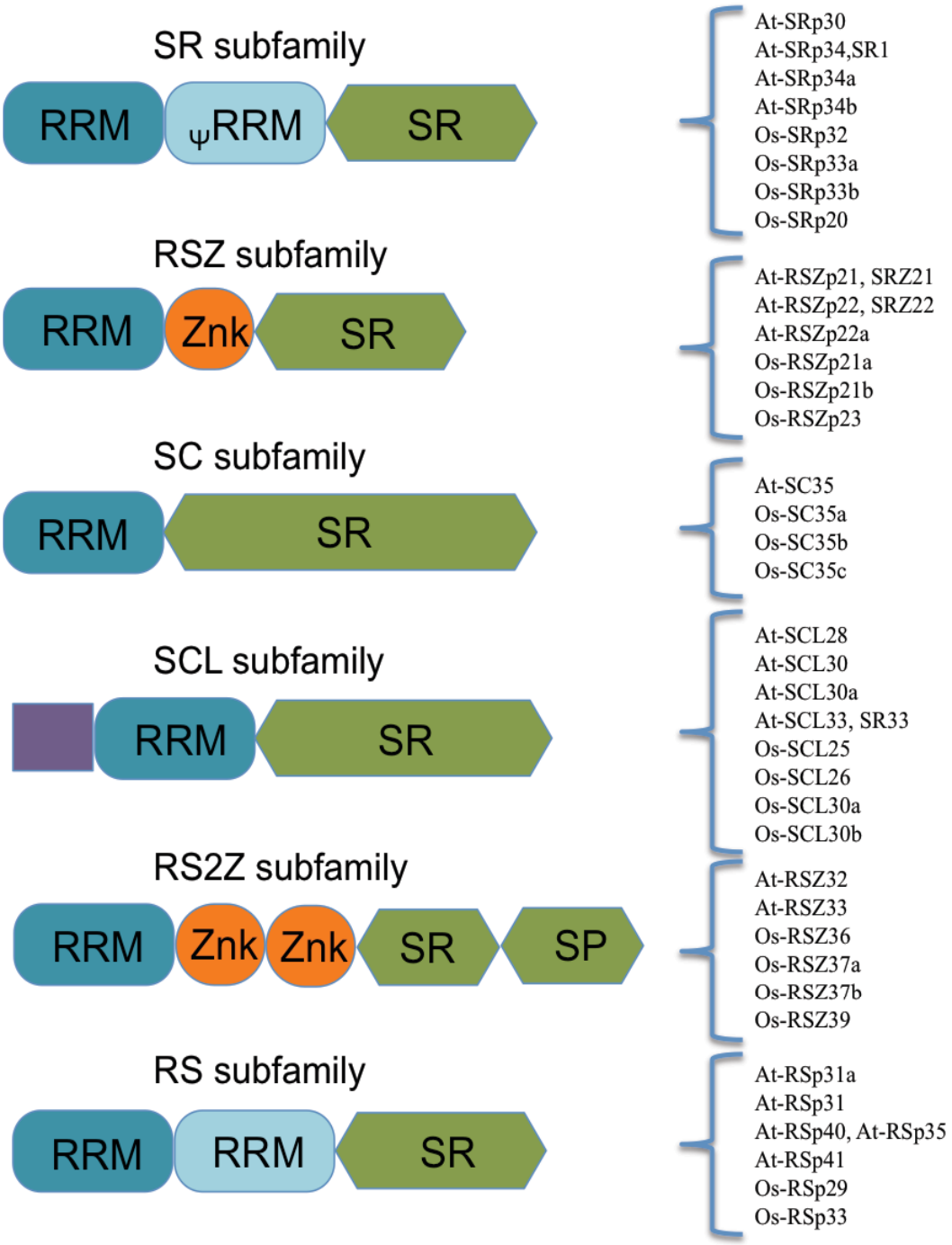


Figure 4: Schematic diagram representing structural features of *Arabidopsis* and rice serine/arginine-rich (SR) proteins. Abbreviations: RRM, RNA recognition motif; RS, arginine/serine-rich domain; Z, zinc knuckle. (adopted from Barta et al., 2010).

in pre-mRNA splicing mechanisms between different organisms. Taken together, these indicate that the process of intron detection and regulated splicing in plants may vary from yeast and animals, and likely include plant-specific splicing regulatory factors.

Regulation of Splicing

AS is due to differential usage of splice sites. However, canonical splicing *cis* element are short consensus sequences (5′ss, 3′ss, BP, and PPT), and by themselves are not sufficiently informative to direct splice site selection during intron removal. Therefore, additional splicing regulatory sequences are needed to control CS and AS. These include exonic splicing enhancers (ESEs) or silencers (ESSs) and intronic splicing enhancers (ISEs) or silencers (ISSs) (Figure 1 and 5) (Chen and Manley, 2009; Fu and Ares, 2014; Syed et al., 2012). UA-richness in plant introns is another feature that is essential for efficient splicing in plants, and has also been observed in animals (Brown et al., 2002; Syed et al., 2012). Finally, GC-rich exons also influence splice site usage (Lorković et al., 2000). These regulatory elements play central roles in splicing regulation and splice site selection in a context-dependent manner.

Regulation of splicing depends on interactions between sequence elements (*cis*-elements) and a large number of proteins (*trans*-acting factors) (Figure 5). *Trans*-acting factors involve members of well-characterized SR proteins and heterogeneous nuclear ribonucleoprotein (hnRNP) families (reviewed in Lin and Fu, 2007; Long and Caceres, 2009; Martinez-Contreras et al., 2007; Meyer et al., 2015). These proteins recognize and bind to *cis*-elements in a concentration-dependent manner to recruit and stabilize spliceosomal components (core components) (Matlin et al., 2005; Reddy et al., 2013). *Trans*-acting factors may be splicing activators or inhibitors, and some can act as either depending on the *cis*-elements and position of the target site in the pre-mRNA (Figure 5) (Ule et al., 2006).

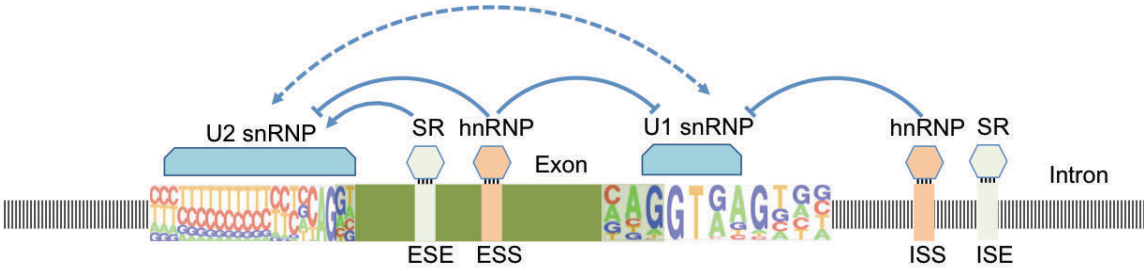


Figure 5: Splice site selection is regulated through *cis*-elements and *trans*-acting factors. Four types of *cis*-elements, termed ESEs, ISEs, ESSs or ISSs, regulate splicing. These splicing cis-elements function by recruiting splicing factors, which either promote or inhibit recognition of adjacent splice sites. Common splicing regulators are SR proteins (promote splicing), and heterogeneous nuclear ribonucleoproteins (hnRNPs) (inhibit splicing), which affect the function of U2 and U1 snRNPs at the early stage of spliceosome assembly. The conserved sequences of splice sites are displayed in the colored character. The height of each letter represents nucleotide frequency in each position. The formation of spliceosomal complex across exon is indicated by a dashed line. (adapted from Matera and Wang, 2014).

Serine/Arginine-Rich Proteins (SR Proteins)

The SR proteins are master regulators of CS and AS (reviewed in Graveley, 2000; Long and Caceres, 2009; Reddy, 2004). SR proteins are highly conserved in animals and plants, with one or two RNA recognition motifs (RRMs) at the N terminus and a C-terminal domain rich in arginines and serines (RS domain) (Figure 4). While the RRMs recognize and bind to *cis*-elements, RS domains regulate protein-protein interaction and assembly of spliceosomal components. RS domains of SR proteins contain subcellular localization signals, and these proteins are mostly found in the nucleus. However, several studies have shown that some SR proteins shuttle between the nucleus and cytoplasm (reviewed in Twyffels et al., 2011).

SR proteins also exhibit altered expression and AS under different conditions and developmental stages. In *Arabidopsis*, most of the genes encoding SR proteins are alternatively spliced, collectively generating about 95 transcripts; this dramatically increases the complexity of the SR gene family transcriptome (Palusa et al., 2007). Importantly, expression and differential splicing of SR proteins are regulated in a developmental-, tissue-, hormone-, and abiotic stresses-dependent manner (Palusa et al., 2007). While the functions of most *SR* splice variants are not known, the majority of them contain a premature stop codon (PTC) and are targeted for degradation by NMD (Palusa and Reddy, 2010).

In addition to splicing regulation, other roles have been documented for SR proteins in RNA metabolism and gene regulation. Several studies in animal systems have shown that SR proteins influence mRNA transfer, stability, and translation; chromatin binding; transcription elongation; subcellular localization of transcripts; genome stability and foundation of cellular RNA granules; and miRNA processing (reviewed in Long and Caceres, 2009). However, roles in RNA

metabolism and gene regulation beyond pre-mRNA splicing remain to be illustrated in plants (Reddy et al., 2013).

Phosphorylation and dephosphorylation influence the function of SR proteins in pre-mRNA splicing. SR-specific protein kinase (SRPK) and Clk/Sty (LAMMER-type kinases) are well-characterized protein kinases that phosphorylate SR proteins in both animals and plants (Misteli, 1999; Savaldi-Goldstein et al., 2003). Phosphorylation analysis of SR proteins shows that serine residues of the RS domain are highly phosphorylated; these modifications regulate subcellular localization and activity of SR proteins (reviewed in Lin and Fu, 2007; Reddy et al., 2012; Zhou and Fu, 2013). Additionally, in animals the interaction between splicing regulator SF2/ASF and other splicing factors, such as U1-70K, is mediated by phosphorylation of the RS domain (Xiao and Manley, 1997). In plants, the distribution, localization, and mobility of SR proteins are altered upon inhibition of phosphorylation (reviewed in Reddy et al., 2012). *In vitro* phosphorylation assays have confirmed that plant SR proteins are phosphorylated by LAMMER-type protein kinases (Golovkin and Reddy, 1999). *In vivo* phosphoproteomic analysis has shown that *Arabidopsis* SR proteins are highly phosphorylated (de la Fuente van Bentem et al., 2006). In contrast, dephosphorylation of some SR proteins during splicing is needed for splicing catalysis to continue (Cao et al., 1997). Furthermore, signal transduction pathways have been shown to regulate AS; for instance, the phosphatase PP1 dephosphorylates SRp38 in response to heat shock, which results in a general inhibition of splicing (Shi and Manley, 2007; Shin et al., 2004). Thus, the phosphorylation state of SR proteins appears to be a vital post-translation modification that modulates SR protein functions during splicing.

Genetic studies of SR proteins have revealed their roles in plant development and stress responses. In *Arabidopsis*, overexpression of SR30 and RS2Z33 affect the splicing patterns of

their own pre-mRNAs and several endogenous *Arabidopsis* genes; this results in pleiotropic developmental and morphological changes (Kalyna et al., 2003; Lopato et al., 1999). Loss of function mutations in SC35 and SC35-like (SCL) proteins influence plant development and flowering via controlling the splicing and transcription of FLOWERING LOCUS C FLC (FLC) (Yan et al., 2017b). Interestingly, overexpression of LAMMER-type protein kinases that phosphorylates SR proteins in *Arabidopsis* results in changes in expression and splicing profiles of several SR genes as well as other *Arabidopsis* genes, and causes many developmental abnormalities (Savaldi-Goldstein et al., 2003). Expression analyses in *Arabidopsis* also show promising involvement of SRs in abscisic acid (ABA)-mediated stress responses (Cruz et al., 2014). Furthermore, mutations of plant-specific SR proteins *RS40* and *RS4* result in ABA and salt stress hypersensitive phenotypes (Chen et al., 2013). These studies illustrate the key roles of SR proteins to plant development and performance.

Arabidopsis Splicing Regulator SR45

Arabidopsis SR-like splicing proteins exhibit essential roles in both CS and AS of pre-mRNAs. *Arabidopsis* SR45, unlike other SR proteins, has two RS domains, hence it is consider as an SR-like proteins, and shows homology to animal RNA binding protein S1 (RNPS1) (Ali et al., 2007; Golovkin and Reddy, 1999; Mayeda et al., 1999). This protein is localized to nuclear speckles, and its mobility is dependent on ATP, phosphorylation, and transcription (Ali et al., 2003; Ali et al., 2008; Ali and Reddy, 2006). It has been established as an essential pre-mRNA splicing factor (Ali et al., 2007) and has been shown to interact with several spliceosomal proteins and other SRs such as U1-70K, AFC2 kinase, U2AF35, SC35-like (SR33), RSZ21, SR34, and SR34a (Day et al., 2012; Golovkin and Reddy, 1999; Zhang et al., 2014). Furthermore, *in vitro* mechanistic analysis suggests that SR45 recruits U1snRNP and U2AF to 5′ and 3′ splice sites, respectively,

by interacting with the pre-mRNA (10^{th} intron of SR30), U1-70K, and U2AF35, and thus modulates AS (Day et al., 2012).

Arabidopsis SR45 exhibits fundamental regulatory functions in plant development and abiotic stress responses. The knockout mutant of SR45 (*sr45*) displays abnormal developmental phenotypes including narrow leaves and petals, altered number of petals and stamens, and delayed root growth and flowering (Ali et al., 2007). In addition, it shows hypersensitivity to glucose (Glc) and abscisic acid (ABA) during early seedling development, indicating that SR45 has a negative role in plant sugar response in early seedling growth (Carvalho et al., 2010). Recently, it has been suggested that the glucose-responsive function of SR45 is involved in modulating the degradation of the SnRK1 energy sensor in response to sugars (Carvalho et al., 2016). Furthermore, we have recently shown that *sr45* is highly sensitive to high temperature and high salinity stresses at different developmental stages. These phenotypes can be caused by an SR45 mutation disrupting the expression and splicing of several stress-responsive genes (Albaqami, 2013). Another study has shown that SR45 functions in the establishment and maintenance of DNA methylation in *Arabidopsis* via small interfering RNA-mediated DNA methylation and gene silencing (Ausin et al., 2012). These studies indicate the biological significance of SR45 as a splicing regulator in plant development and environmental responses.

As with many *Arabidopsis* SR proteins, the SR45 gene is alternatively spliced to produce two splice variants, *SR45.1* and *SR45.2*. The *SR45.1* isoform has an in-frame additional 21-nucleotide sequence as a result of an alternative 3′ss selection event at the beginning of the seventh exon (Palusa et al., 2007; Zhang and Mount, 2009). Because the AS conserves the reading frame, both isoforms are translated into functional proteins with a single arginine in the SR45.2 form replaced by eight amino acids (TSPQRKTG) in SR45.1 (Zhang and Mount, 2009). Interestingly,

isoform-specific complementation of *sr45* indicates that the two isoforms have different biological functions (Zhang and Mount, 2009). While SR45.1 functions only in flower development, SR45.2 plays a role in root growth (Zhang and Mount, 2009). Similarly, SR45.1 is the only isoform that is capable of complementing the mutant's heat and salt stress phenotypes (Albaqami, 2013). However, in the case of Glc and ABA response phenotypes, both SR45.1 and SR45.2 are able to complement the mutant phenotypes (Carvalho et al., 2010). A recent study showed that a single phosphorylation site at threonine 218 in SR45.1, within the alternatively spliced region, promotes the function of SR45.1 in flower development (Zhang et al., 2014). In addition, another study identified about 4000 transcripts associated with SR45.1 using RNA immunoprecipitation (RIP) followed by high-throughput sequencing (RIP-seq), and Gene Ontology (GO) analysis revealed them to be enriched in hormone and stress signaling pathways (Xing et al., 2015). These assays indicate that SR45 regulates plant development and abiotic stress response in an isoform-dependent manner.

Although SR45 is a well-characterized SR-like protein in plants and it has well-known functions in both pre-mRNA splicing and several biological processes, how its biological functions are mediated by SR45-regulated AS events remains to be explored. Towards this end, we aimed to identify the genes regulated by SR45, directly or indirectly. By comparing the whole transcriptomes of *sr45* and wild-type (WT) using high-throughput RNA-Seq, we identified thousands of SR45-regulated genes that were either differentially expressed (DE) or differentially spliced (DS). Furthermore, analyses of AS events distribution along the gene bodies of DS genes indicated that SR45 primarily affects the choice of 3′ss within the coding DNA sequence (CDS) region. Gene Ontology (GO) enrichment analyses of those genes revealed that they are involved in both previously suggested and novel biological functions of SR45.

Importantly, we successfully confirmed the role of SR45 in heat stress response at the time of seed germination in *Arabidopsis,* and provided evidence that SR45 confers thermotolerance in an isoform-dependent manner at this stage, where SR45.1 but not SR45.2 rescues the mutant heat stress phenotype. Finally, our work suggests that phosphorylation events may mediate the function of SR45.1 in thermotolerance, as a point mutation the unique phosphorylation site at threonine 218 of SR45.1 fails to rescue the mutant phenotype in response to heat stress.

MATERIALS AND METHODS

Plant Materials and Growth Conditions

Arabidopsis thaliana ecotype Columbia-0 (Col-0) was used as the wild-type (WT) in all experiments. A homozygous line for the *sr45-1* T-DNA insertion mutant (*sr45*) was isolated and described previously (Ali et al., 2007). Complementation lines (*35S:SR45.1-GFP* and *35S:SR45.2-GFP*) were kindly provided by Dr. Xiao-Ning Zhang (Zhang and Mount, 2009). Lines overexpressing substitution mutations of predicted phosphorylation sites of SR45.1 (*SR45.1-T218A-GFP*, *SR45.1-S219A-GFP*, and S*R45.1-T218A+S219A-GFP*) were described previously (Zhang et al., 2014; Zhang and Mount, 2009). For all genotypes, seeds used in this study were harvested at the same time from plants grown on soil under the same conditions (22 °C; long-day-16-h/8-h light/dark photoperiod; 120 mmol/m^2/s white fluorescent light). Mutant and transgenic lines were genotyped by RT-PCR using gene-specific primers, and by immunoblotting using SR45 or GFP antibodies. For immunoblotting, seedlings were prepared as follows: seeds (50 mg) were surface-sterilized with 70% ethanol followed by 15% bleach and stratified for 3 days at 4 °C (to break dormancy). Then the seeds were placed into 100 mL of Murashige and Skoog (MS) medium (1x MS basal salt, 1 mL/L MS vitamin solution, and 1% sucrose, pH 5.7) in a 250 mL flask and grown in a growth chamber. The growth condition was

22 °C with a 16-h/8-h light/dark photoperiod, and flasks kept on a shaker at 150 rpm. One-week-old seedlings were harvested, washed three times with Nanopure water, and Kimwipes were used to remove excessive water. Afterwards, the seedlings were weighed, directly frozen in liquid N_2, and stored at -80°C. For RNA expression analysis under heat stress treatment, seedlings were prepared as above.

Generation of RNA-Seq Data

RNA-Seq analysis was performed with 2-week-old seedlings of WT and *sr45* grown on MS medium under long-day conditions. Two biological replicates were used for RNA-Seq analysis. Total RNA was extracted and purified using the RNAeasy Plant Mini kit (Qiagen, Hilden, Germany). On-column DNase digestion was performed according to the manufacturer's protocol. Prior to generating the RNA-Seq library, RNA quality for all samples was evaluated with an Agilent 2100 Bioanalyzer using the RNA 6000 Nano kit (Agilent Technologies, Santa Clara, CA). All RNA samples had a high RNA integrity (RIN) value of around 8. Poly(A) RNA was isolated from total RNA using oligo-dT beads and sheared to about 200 nt under elevated temperature (80^oC); cDNA was synthesized using random primers. RNA-Seq libraries were generated from cDNA using TruSeq kit (Illumina, San Diego, CA) for adapter ligation and PCR amplification. Single-end 75 nt reads were generated on the Illumina GAIIx platform (IGSP core resource, Duke University).

RNA-Seq Analyses

The RNA-Seq data was first evaluated using FastQC, then mapped to the TAIR10 version of *Arabidopsis* genome using either TopHat or MapSplice with default parameters (Trapnell et al., 2009; Wang et al., 2010). Sequences mapped using MapSplice were analyzed using SpliceGrapher (Rogers et al., 2012) to generate splice graphs. Sequences mapped using TopHat

were analyzed using Cufflinks package version 2.0 (Trapnell et al., 2012). Briefly, an *in silico* annotation file was first generated using Cufflinks to best explain the mapped sequences. Annotation files from each sample and TAIR10 were then merged together using Cuffmerge. Using the merged file as a reference, the mapped sequences from each sample were compared and Cuffdiff used to identify DE genes. Differential splicing analysis was performed using MISO (Katz et al., 2010) and iDiffIR (https://bitbucket.org/comp_bio/idiffir).

GO Enrichment Assays

DE and DS genes were pooled together and enriched GO terms identified using GENECODIS (Carmona-Saez et al., 2007). Analyses were conducted with these parameters: reference list, all annotated genes in the *Arabidopsis* genome; statistical method, hypergeometric; method to correct *p* values for multiple hypothesis testing, false discovery rate.

Analysis of Heat Shock Response

Seed heat tolerance assays were performed as previously described (Silva-Correia et al., 2014). In brief, after seed sterilization and stratification of all lines, seeds were placed into 1.5 mL microfuge tubes containing 500 mL sterilized water and either exposed to heat stress using a heating block at 50^{o}C for 60 min or kept at 22^{o}C as an experimental control. Subsequently, both heat-stressed and control seeds were germinated on MS medium in long-day (16h/8h light/dark photoperiod, 120 mmol/m^{2}/s white fluorescent light) at 22°C. After 10 days, pictures were taken and germination rate (radicle emergence and formation of green cotyledons) calculated. The germination frequency was determined in relation to total sown seeds of each genotype.

RNA Extraction and RT-PCR Analyses

For validation of DE and AS results, total RNA from two week-old seedlings of WT and *sr45* grown on MS plates under long-day conditions was extracted using the RNeasy Plant Mini Kit

(Qiagen, Hilden, Germany). RNA samples were quantified using a NanoDrop 1000 spectrophotometer (Thermo Fisher Scientific, Waltham, MA) and treated with DNase I (Fermentas, Hanover, MD) to remove residual genomic DNA. For RT-PCR, 1.5 µg of DNase-treated total RNA was used to synthesize first-strand cDNA with an oligo(dT) primer and SuperScript III (Invitrogen, Carlsbad, CA) in a 20 mL reaction. The 1st strand cDNA was further amplified by regular PCR using gene-specific primers. The PCR reaction mixture was heated at 94°C for 3 min, cooled to 55°C, and 2 units of Ex-Taq polymerase (Takara Bio, Inc., Kusatsu, Japan) was added to initiate the amplification reaction. Thirty cycles of amplification were performed on a Mastercycler Gradient (Eppendorf, Hamburg, Germany), each consisting of denaturation at 94°C for 30 sec, annealing at 55°C for 30 sec, and extension at 72°C for a time period based on the expected length of the PCR product. The amplified products were resolved by electrophoresis and gel pictures captured using ChemiDocTM XPS+ (BioRAD, Hercules, CA). *ACT2* was used as an internal control to show equal amount of template in different samples.

To determine expression levels of *SR45* isoforms under heat stress, WT seedlings were grown in liquid MS medium as described above. One-week-old seedlings in a 250 mL flask containing 100 mL liquid MS medium were exposed to heat stress (38^{o}C) for 3 hours or kept at 22°C as control plants. For the recovery assay, one flask was moved back to 22°C for 3 hours after heat stress. Subsequently, the seedlings were washed three times with Nanopure water and wrapped in a few sheets of Kimwipes to remove excessive water. The seedlings were weighed, immediately frozen in liquid N_2, and stored at -80°C for total RNA isolation. RNA extraction and cDNA synthesis were performed as described earlier. The 1st strand cDNA was further amplified using either regular or quantitative PCR with SR45 isoform-specific primers (Table 1).

Table 1: Primers used for RT-PCR.

Gene ID	Forward Primer	Reverse Primer
AT1G16610	AAGTCCTGCTGGACCTGCTA	CCTTCTTCGAACAGGACTGC
AT1G16610.1	CATCTCCTCAACGGAAAACA	AACGGCCTCTAGATGGTGAT
AT1G16610.2	CGCCCAAGAGAGAGGCTTTC	AACGGCCTCTAGATGGTGAT
AT2G26150	GCTTTGTGGTGTGGGATTCT	CATCCCAGATCCTTGCTGAT
AT5G37850	TTCCACAAGGACCACAATCA	CCCCAACAATTGAAGAGGAA
AT2G33380	CGGAACGATTTGGAGGAAAC	AGTATCCATTCAACTTTGTTT
AT1G45249	CGAGAATCAGCTGCAAGGTC	AAGGTCCCGACTCTGTCCTC
AT3G18780	GGCAAGTCATCACGATTGG	CAGCTTCCATTCCCACAAAC
AT4G31877	GAGAAACGCATAGAAACTGA	GAATCGGAGCCGGAATCT

For quantitative PCR, amplification reactions were conducted as described in (Albaqami, 2013) using LightCycler 480 SYBR Green 1 master mix on a LightCycler 480 (Roche Applied Science, Penzberg, Germany). A reaction of 20 µl comprises 10 µl of Roche master mix, 0.5 µmol of forward and reverse primers, and 1 ul cDNA template of the original RT reaction. The q-PCR settings were 95°C for 4 min followed by 40 cycles of 95°C for 10 s, annealing temp (estimated as Tm -5°C for each primer pair) for 10 s, and 72°C for 30 s. The threshold (CT) value was determined and used to compare gene expression from different samples; CT values for all examined genes were standardized to the CT value of the housekeeping gene (*ACT2*), by withdrawing the CT value of *ACT2* from the CT value of the examined gene.

Nuclear Protein Isolation

For immunoblot and Phos-tag western blotting assays, nuclear proteins were extracted as previously described (Xing et al., 2015). Briefly, for each sample 3 g of one-week-old seedlings, grown in liquid MS as described above, were ground into a fine powder in liquid nitrogen. Subsequently, each sample was homogenized in 25 mL of Honda buffer (1.25% Ficoll 400, 2.5% Dextran T40, 0.44M sucrose, 10mM $MgCl_2$, 0.5% Triton X-100, 20mM HEPES KOH, pH 7.4, 5 mM DTT, 1 mM PMSF, and 1% protease inhibitor cocktail [P9599; Sigma-Aldrich]). The homogenate was filtered through two layers of Miracloth into a 30 mL Corex tube and the filtrate centrifuged at 2000 *g* for 15 min at 4°C. The supernatant was discarded and the pellet resuspended in 1 mL Honda buffer, then transferred to a new 1.5 mL microcentrifuge tube and centrifuged at 1500 *g* for 10 min at 4°C. This washing step was repeated two more times. The pellet was then resuspended in 500 mL nuclei lysis buffer (50 mM Tris-HCl, pH8.0, 10 mM EDTA, 1% SDS, 1 mM PMSF, and 1% protease inhibitor cocktail) by pipetting up and down and sonicated using a Covaris M220 Focused-ultrasonicator for 1 min at 7°C (peak power, 75;

duty factor, 20; cycles/burst, 300). The extract was then centrifuged for 30 min at maximum speed (16,000 *g*) at 4°C and the supernatant transferred to a new 1.5 mL microcentrifuge tube. Concentrations of the nuclear protein were evaluated by a standard Lowry assay to confirm similar concentrations through all samples (Lowry et al., 1951). Aliquots of 100 mL for each sample were made and frozen in liquid N_2, and stored at -80 °C.

SDS Polyacrylamide Gel Electrophoresis (SDS-PAGE)

Nuclear proteins were separated using denaturing sodium dodecyl sulphate polyacrylamide (SDS-PAGE) gels (Laemmli, 1970). 10% SDS-PAGE was suitable for resolving SR45 proteins. The recipe for 8 mL of 10% separating gels is: 3.2 mL ddH_2O, 2 mL 30% acrylamide, 2.67 mL 2M Tris pH 8.8, 80 μL 10% SDS, 80 μL 10% APS, and 8 μL TEMED. Stacking gel (6%, 5 mL) was prepared with: 2.6 mL ddH_2O, 1 mL 30% acrylamide, 1.25 mL 2M Tris pH 8.8, 50 μL 10% SDS, 50 μL 10% APS, and 5 μL TEMED. Protein samples were boiled for 5 minutes at 100°C in 1X SDS sample buffer before loading. The gels were run in a mini-gel apparatus (Bio-Rad, Hercules, CA) in 1X SDS electrophoresis buffer (25 mM Tris, 192 mM glycine, 0.1% SDS) at 150 V for 90 minutes.

Western Blot Analysis

For regular western blot analysis, electrophoresed proteins were transferred onto a methanol-activated PVDF membrane by wet blot procedure using a Bio-Rad Mini Trans-Blot cell (Bio-Rad, Hercules, CA). The transfer was performed at 200 V for 2 h using transferring buffer containing 25mM Tris, 190mM glycine, and 20% methanol. Subsequently, the membrane was rinsed 2-3 times with a TBS-Tween solution (20 mM Tris pH 7.5, 150 mM NaCl, and 0.1% Tween 20) for 5 min each. The membrane was then blocked for 1 h at RT in a TBS-Tween solution containing 5% fat-free milk powder to prevent nonspecific interactions between the

antibody and the membrane. Following that, the antibody was added to blocking solution in the designated dilution and incubated at 4^{o}C overnight. On the next day, the membrane was washed 3 times with TBS-Tween for 5 min each; then the second antibody (HRP-conjugated secondary antibody) was added and incubated for 1 h at RT. After that, the previous washing steps were repeated to prepare the membrane for colorimetric detection by alkaline phosphatase reaction.

Phos-tag Western Blotting

For SR45 protein phosphorylation analysis, seedlings of overexpression lines (*sr45::SR45.1* and *sr45::SR45.2*) were grown in liquid MS as described above (plant materials and growth conditions). One-week-old seedlings in a 250 mL flask containing 100 mL liquid MS medium were exposed to heat stress (38^{o}C) for indicated times. Subsequently, nuclear proteins were isolated as described above. Phos-tag™ SDS-PAGE was prepared according to manufacturer's protocol (Wako Pure Chemical Industries, Ltd., Osaka, Japan) (Kinoshita et al., 2009). In brief, standard 10% SDS-PAGE was prepared as described above with the addition of 50 mM Mn^{2+}-Phos-tag. Following electrophoresis, gels were soaked for 10 min in a general transfer buffer containing 1 mM EDTA, with gentle agitation to eliminate excess manganese ion and improve transfer efficiency. After that, the standard western blotting procedure was performed as described above.

RESULTS

Transcriptome Analysis of *Arabidopsis SR45* Mutant

The T-DNA insertion mutant of *Arabidopsis SR45* (*sr45*) was isolated and described previously (Figure 6A) (Ali et al., 2007). Defects in root growth and late flowering phenotypes of *sr45* are shown in (Figure 7). Before using *sr45* in the transcriptome assay, we further confirmed the knockout at both RNA and protein levels using RT-PCR and Western blotting respectively. RT-

PCR amplification with *SR45*-specific primers showed that *SR45* was expressed normally in WT plants but not in T-DNA insertion mutants (Figure 6B). Nuclear proteins isolated from *Arabidopsis* WT and mutant plants and probed with SR45 antibody confirmed expression of SR45 protein in WT but not in mutant plants (Figure 6C).

To identify global changes in gene expression and AS in *sr45*, we performed RNA-Seq analysis on *Arabidopsis* WT and the *SR45* mutant. The steps of the RNA-Seq experiment are shown in (Figure 8). Total RNA from two biological replicates for WT and *sr45* were sequenced using an Illumina GAIIx platform. For each sample, about 22 million reads were obtained (Table 2). Quality assessment of the reads using FastQC tools revealed that the average base quality score was between 36 and 40 (with the exception of the last base whose average quality score was around 30) and the average read quality score was around 39, indicating very high quality sequences (Figure 9). Importantly, sequence qualities are very consistent across all biological replicates of both genotypes (Figure 9AB), allowing an un-biased down-stream analysis.

Mapping of RNA-Seq Reads to the *Arabidopsis* Reference Genome

Sequence reads were mapped to the TAIR10 *Arabidopsis* reference genome using two different mapping tools, TopHat and MapSplice (Trapnell et al., 2009; Wang et al., 2010). Mapped reads were rather evenly distributed across the whole genome except in centromeric regions, which are known for lower gene density; this suggests there was no genome-level bias of the sequencing data (Figure 10). Both programs gave high rates of mapping. With MapSplice, 95% of total reads were mapped to the genome, while the mapping rate was 90% with TopHat (Table 2). The discrepancy between the two aligners was largely attributed to the greater sensitivity of MapSplice for spliced reads (reads aligning to exon junctions); MapSplice detected about 24% of spliced reads while TopHat detected about 20% of spliced reads (Table 2). MapSplice also

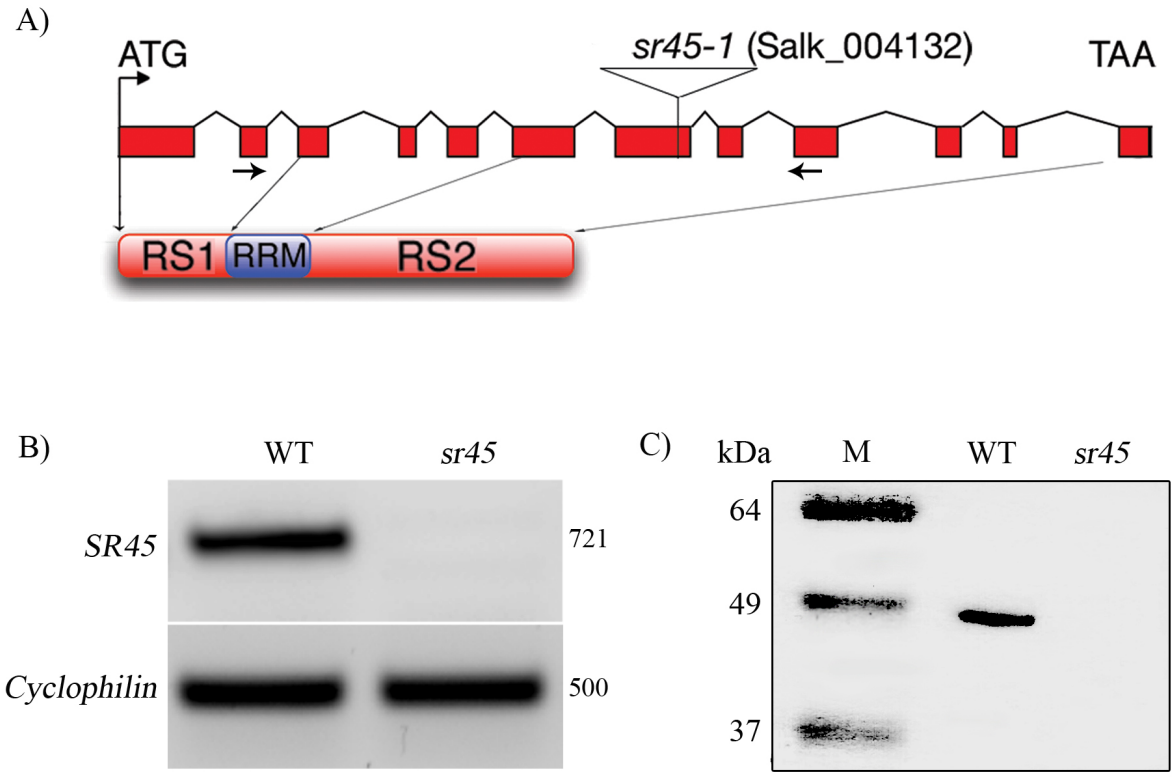


Figure 6: Genotypic characterization of *Arabidopsis* SR45 T-DNA insertion mutant. (A) The upper diagram shows the gene organization of *SR45* (AT1G16610). Exons are indicated by red rectangles; introns are indicated by black lines; ATG and TAA are start and stop codons, respectively; the T-DNA insertion in the 7th exon are shown by reversed open triangle. The lower illustration shows SR45 protein structure and domain arrangement. The downward arrows indicate corresponding gene parts coding for the N-terminal arginine/serine-rich (RS1) domain, the central RNA recognition motif (RRM), and the C-terminal RS2 domain (adapted from Ali et al., 2007). Arrows underneath the second and ninth exons show the positions of the forward (F) and reverse primers (R) used in RT-PCR. (B) RT-PCR detection of the *SR45* transcript. Total RNA from two-week-old seedlings of WT and *sr45* was used for cDNA synthesis. The full-length transcript of *SR45* was detected only in WT plant. *ACT2* was used as an input control. (C) Immunodetection of SR45 protein. Nuclear proteins from 7day-old-seedlings of WT and *sr45* were electrophoresed on a denaturing polyacrylamide gel and probed with the SR45 antibody. The SR45 protein was detected as a ~ 45kD band only in WT plants.

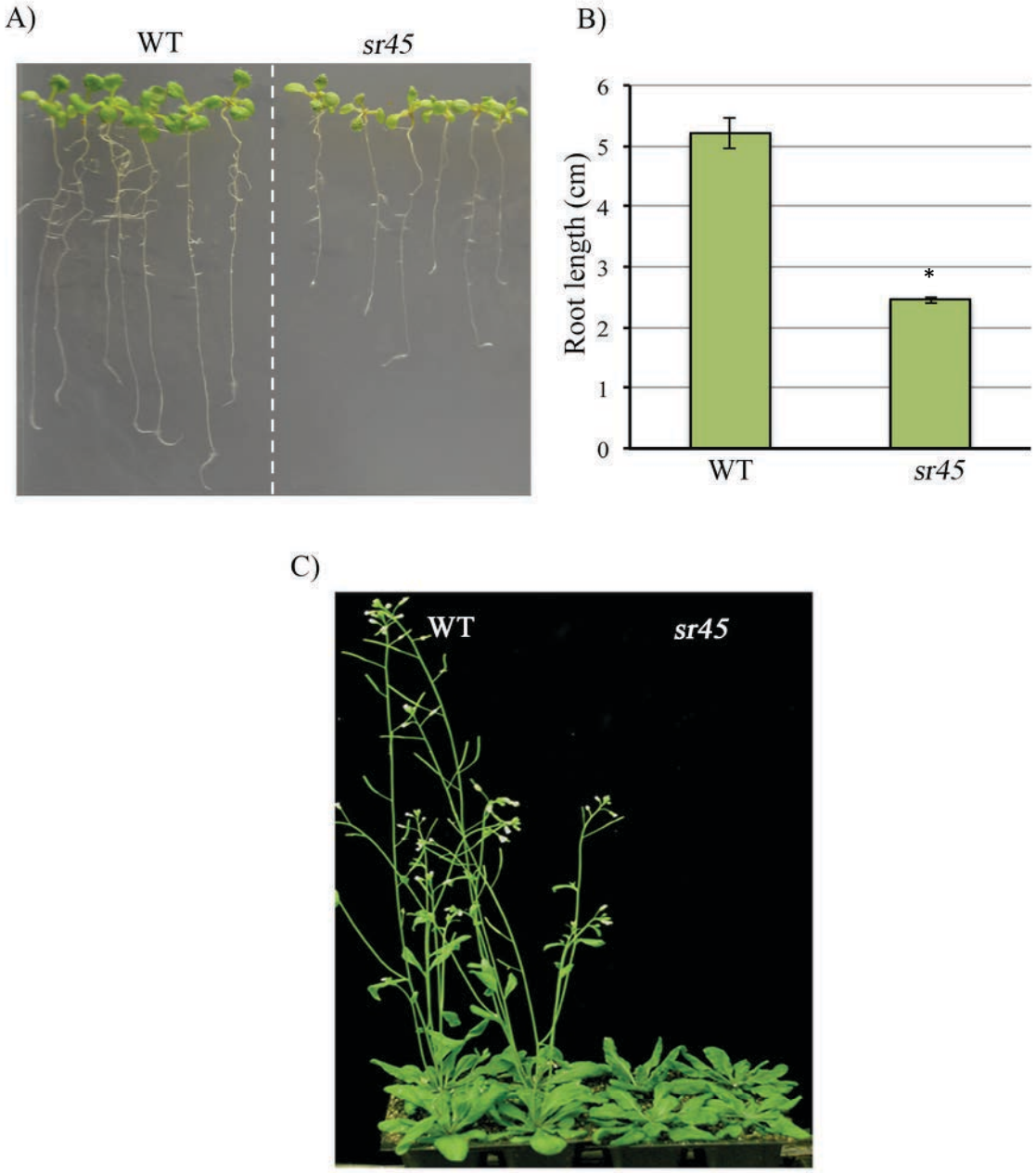


Figure 7: Phenotype of *sr45* plants. (A) A representative photograph shows the difference in root growth between WT and *sr45*. Seeds of WT and *sr45* were germinated and grown under the same conditions on Murashige and Skoog medium for two weeks. (B) Root length quantification of WT and *sr45*. Statistical significance (*t*-test $P<0.05$) is indicated by (*). (C) *sr45* plants show delayed flowering as compared to WT plants. Plants were grown on soil under long-day condition (16h:8h – light:dark) for 37 days (adapted from Ali et al., 2007).

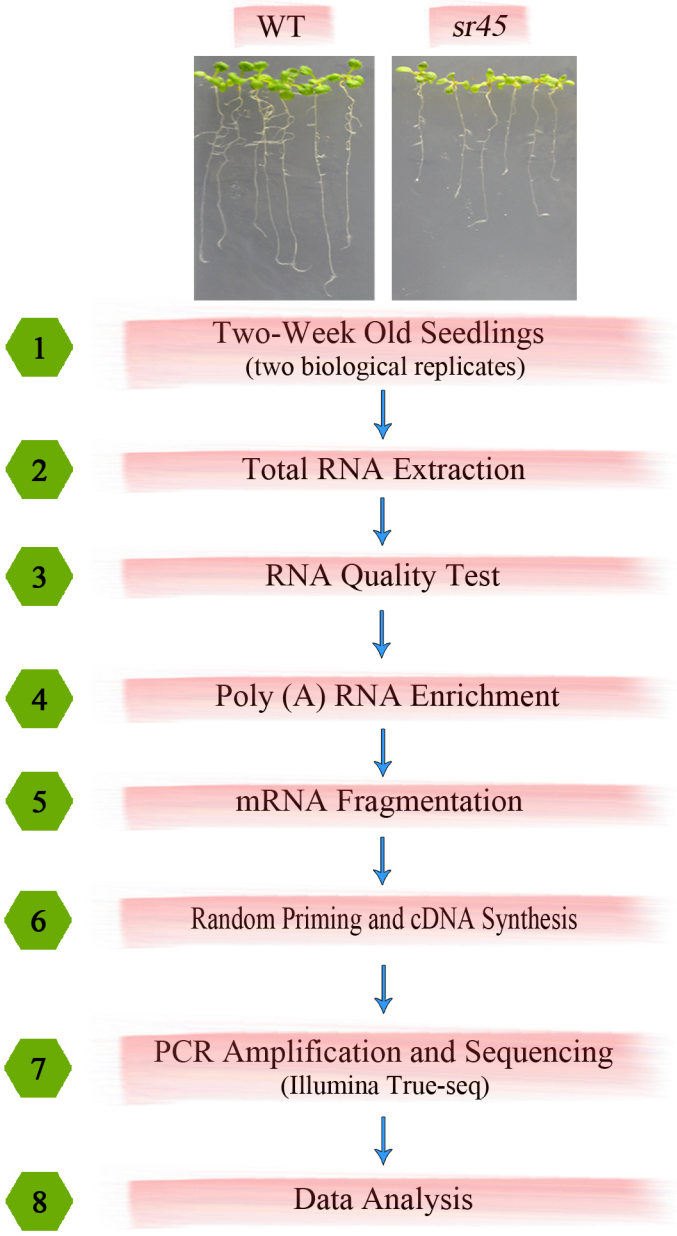


Figure 8: Schematic overview of RNA sequencing (RNA-Seq) experiment of two-week-old WT and *sr45* seedlings.

Table 2: Mapping Statistics of RNA-Seq data with TopHat and MapSplice.

Source	Reads	Aligned Reads		
		Ungapped	Spliced	Total
Wild-type				
Replicate1 Replicate2 TopHat Total	22,538,960 21,637,979 44,176,939	15,506,079 (68.8%) 15,095,551 (69.8%) 30,601,630 (69.3%)	4,741,752 (21.0%) 4,488,085 (20.7%) 9,229,837 (20.9%)	20,247,831 (89.8%) 19, 583,636 (90.5%) 39,831,467 (90.2%)
Replicate1 Replicate2 MapSplice Total	22,538,960 21,637,979 44,176,939	15,912,625 (70.6%) 15,481,294 (71.5%) 31,393,919 (71.1%)	5,524,452 (24.5%) 5,224,188 (24.1%) 10,748,640 (24.3%)	21,437,077 (95.1%) 20,705,482 (95.7%) 42,142,559 (95.4%)
SR45 Mutant				
Replicate1 Replicate2 TopHat Total	18,322,494 20,504,063 38,826,557	12,771,565 (69.7%) 14,296,063 (69.7%) 27,067,628 (69.7%)	3,692,480 (20.2%) 4,157,370 (20.3%) 7,849,850 (20.3%)	16,464,045 (89.9%) 18,453,433 (90.0%) 34,917,478 (89.9%)
Replicate1 Replicate2 MapSplice Total	18,322,494 20,504,063 38,826,557	13,146,617 (71.8%) 14,695,749 (71.7%) 27,842,366 (71.7%)	4,313,690 (23.5%) 4,857,766 (23.7%) 9,171,456 (23.6%)	17,460,307 (95.3%) 19,553,515 (95.4%) 37,013,822 (95.3%)

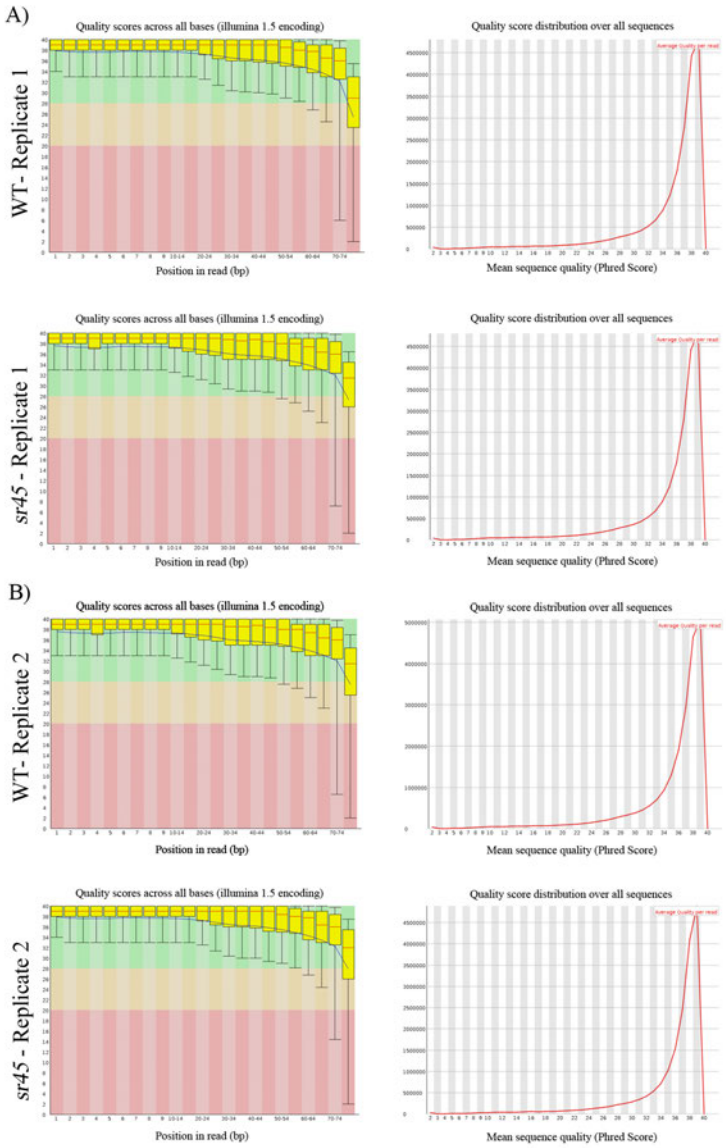


Figure 9: Assessment RNA-Seq reads quality using the FastQC tool. On the left, box-and-whisker plots showing per base sequence quality of WT and *sr45* (A-replicate 1, B-replicate 2), respectively. The yellow box represents the inter-quartile range (25-75%), the upper and lower whiskers represent the 10% and 90% points, and the blue line represents the mean quality. The central red line in the yellow box represents the median value. The y-axis shows the quality scores, with the higher scores having the better base call. The background colors on the graphs indicate good quality calls (green), calls of reasonable quality (orange), and calls of poor quality (red). On the right, per-sequence quality scores distribution of all sequences of WT and *sr45*-1 (A-replicate 1, B-replicate 2). The x-axis on the graph shows the mean quality of sequences (Phred score) and Y-axis shows the number of reads.

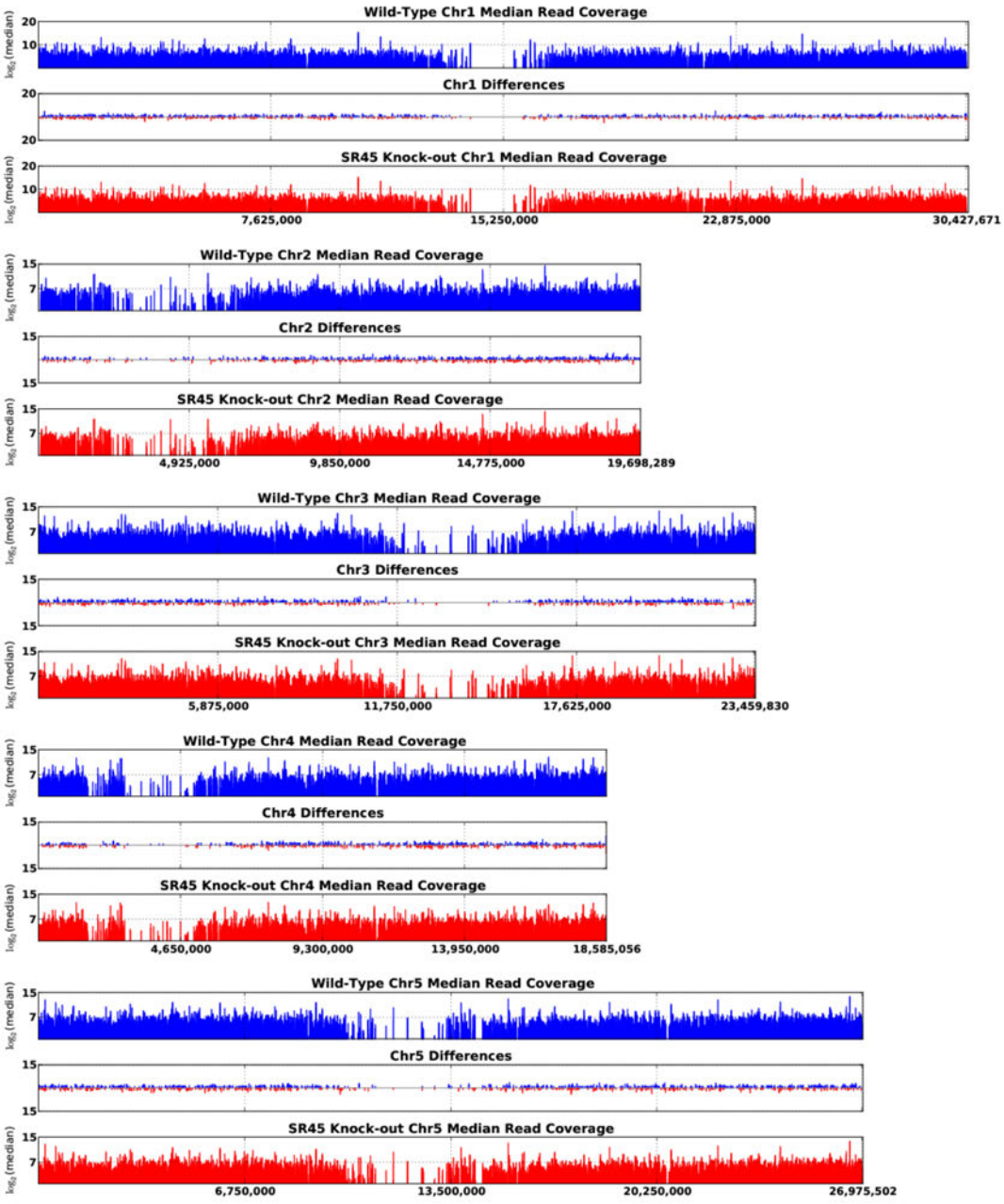


Figure 10: The distribution of RNA-Seq reads of WT and *sr45* across the genome. Reads for the replicates were combined and mapped on the chromosomes labeled 1-5. WT, blue; *sr45*, red; the read difference between WT and *sr45* for a 1kb window across each chromosome (Chr) was presented in the middle panel for each chromosome.

performed better in terms of re-capitulating the splice junctions annotated on TAIR10, identifying more than twice as many annotated splice junctions as TopHat (Table 3). We thus determined differential AS between WT and *sr45* using the data aligned with MapSplice coupled with SpliceGrapher (Rogers et al., 2012). In analyzing differential gene expression, we used TopHat alignments coupled with the Cufflinks package (Trapnell et al., 2012).

For most genes, the RNA-Seq data accurately re-capitulated known exon-intron structures. Of all genes supported by the RNA-Seq data, around 17.2% were found to be single-exon genes, lower than the proportion of 23.7% indicated in TAIR10 gene models; conversely, 82.8% mapped as multi-exon genes, higher than the 76.3% in TAIR10 annotations (Table 4). To some extent these discrepancies likely reflect the categorical differences of genes expressed at seedling stage versus the overall genome composition. Alternatively, with increased sequence depth, previously unknown cryptic introns within some annotated single-exon genes might be revealed, therefore reducing the observed proportion of single-exon genes. Among multi-exon genes, 93.8% had only U2 type splicing sites (93.9% in *sr45*), 0.1% of genes had only U12 type, and 6.1% contained both U2 and U12 splicing sites. These results are rather similar to the ratios of 94%, 0.2%, and 5.0%for each respective category in TAIR10 annotations (Table 4).

Differentially Expressed Genes in *sr45*

To rule out a potential genome-wide, non-specific effect of SR45 on gene expression, the density of genes with varied expression was plotted against expression levels for both WT and *sr45*. The gene density distribution curve for *sr45* was almost identical to that of WT, suggesting that the effect of SR45 on gene expression, if any, was not expression-level dependent (Figure 11). This conclusion was further confirmed by plotting expression levels between WT and *sr45* for each individual gene. The regression line for all genes largely overlapped the theoretical line, which

Table 3: Comparison of spliced alignment statistics for TopHat and MapSplice.

Source		Covered Genes		Recapitulated Junctions					
				Known		Novel		All	
		Total	Unique	Total	Unique	Total	Unique	Total	Unique
sr45	TopHat	8,162	13	45,477	64	1,654	18	47,131	82
	MapSplice	16,822	8,673	98,448	53,035	4,705	3,069	103,153	56,104
WT	TopHat	8,206	14	46,370	63	1,956	33	48,326	96
	MapSplice	16,929	8,737	99,829	53,522	5,333	3,410	105,162	56,932

Table 4: Summary of spliceosome activity found in RNA-Seq data for the WT and the *sr45* data sets.

Gene Spliceosome Statistics					
Source	Gene Statistics		Multi-Exon Genes		
	Single-Exon	Multi-Exon	U2 Only	U12 Only	U2&U12
Wild-type SR45 mutant	3,593 (17.2%) 3,578 (17.2%)	17,288 (82.8%) 17,181 (82.3%)	16,214 (93.8%) 16,126 (93.9%)	22 (0.1%) 21 (0.1%)	1,052 (6.1%) 1,034 (6.0%)
TAIR10 Gene Models	6,746 (23.7%)	21,750 (76.3%)	20,436 (94.0%)	33 (0.2%)	1,281 (5.9%)

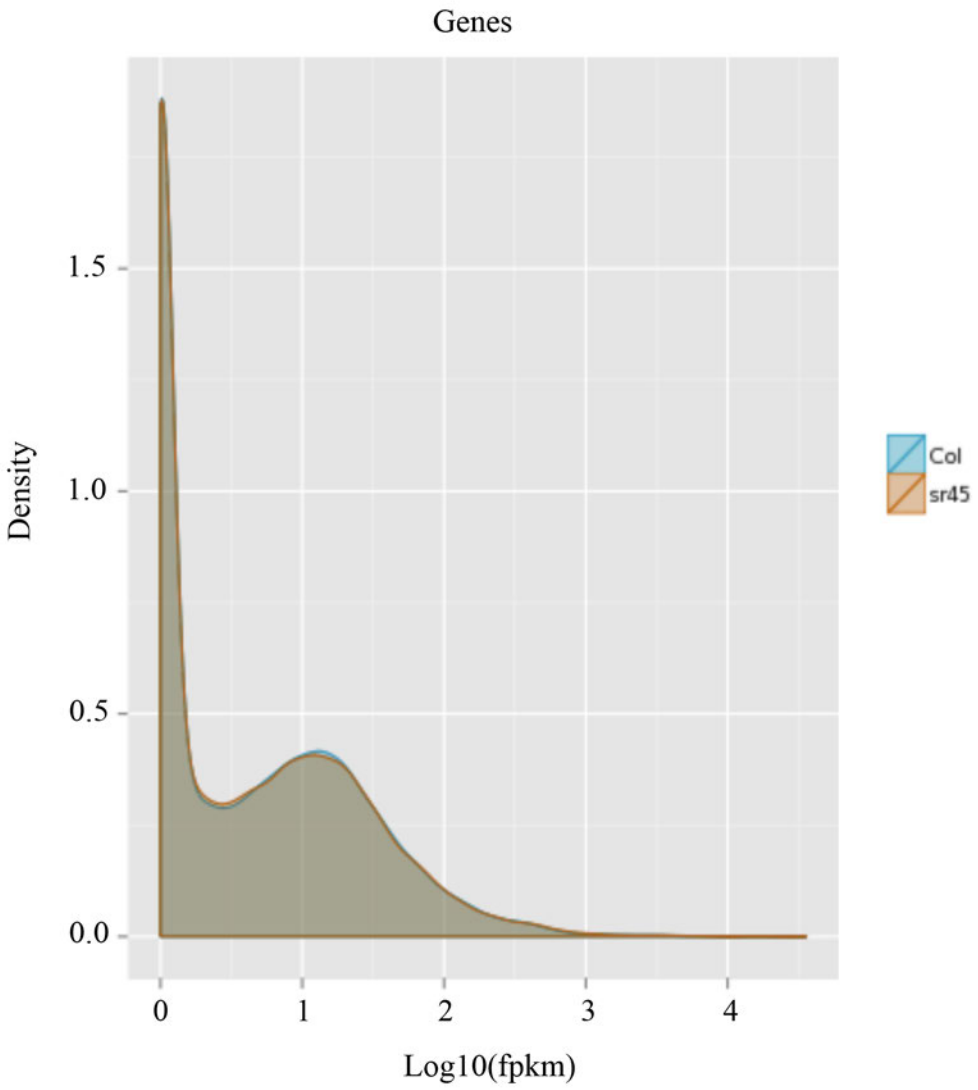


Figure 11: Expression density plot. All annotated *Arabidopsis* genes (*y*-axis) and their expression level are presented in the *x*-axis as log10 (FPKM) values. Blue line represents WT, and orange line represents *sr45*. The distributions of the genes with varied abundance (log10 (FPKM)) are almost identical between WT and *sr45*, suggesting no sequencing bias for the two samples.

assumed no overall expression difference (Figure 12). However, for genes with extremely high abundance, their expression was slightly higher in WT than in *sr45* (Figure 12), likely due to an authentic effect of SR45 on the expression of a few genes, though it could also be simple sampling error. In short, there is no obvious evidence for an overall expression bias in either *sr45* or WT. Any differential expression between WT and *sr45* for a given gene is most likely attributable to the specific effects of SR45.

We identified 1345 differentially expressed (DE) genes between WT and *sr45*, defined as having >2-fold expression change at a false discovery rate of 0.05 (Figures 13 and 14). Among them, 571 were up-regulated and 774 down-regulated. These DE genes included 6 genes encoding microRNAs (Table 5), and 19 genes not previously annotated, designated here as novel genes.

As expected, among the DE genes was *SR45*, whose 3′ end transcripts were totally eliminated and whose 5′ end transcripts preceding the T-DNA insertion was barely present (Figure 15). This finding was further verified using RT-PCR (Figure 15), and is consistent with previous results (Ali et al., 2007). To validate the identified DE genes, we performed RT-PCR analyses for randomly selected, 12 each, up- and down-regulated genes. Expression of all 24 genes was consistent with the RNA-Seq read depth (Figure 16 and 17).

Of the six DE microRNA genes, three (MIR156C, MIR398B, MIR163) are known to play a role in response to phosphate starvation (Hsieh et al., 2009; Lundmark et al., 2010), and MIR156C has been recently reported to function in *Arabidopsis* heat stress response (Stief et al., 2014). MIR167A is involved in nitrogen response, while MIR824A and MIR850A had no known function (Gifford et al., 2008). Five out of the six have known or putative targets (Hsieh et al., 2009; Jones-Rhoades and Bartel, 2004; Lundmark et al., 2010; Wang et al., 2014). Significantly, at least one target gene for each of four micro RNAs is also differentially expressed between WT

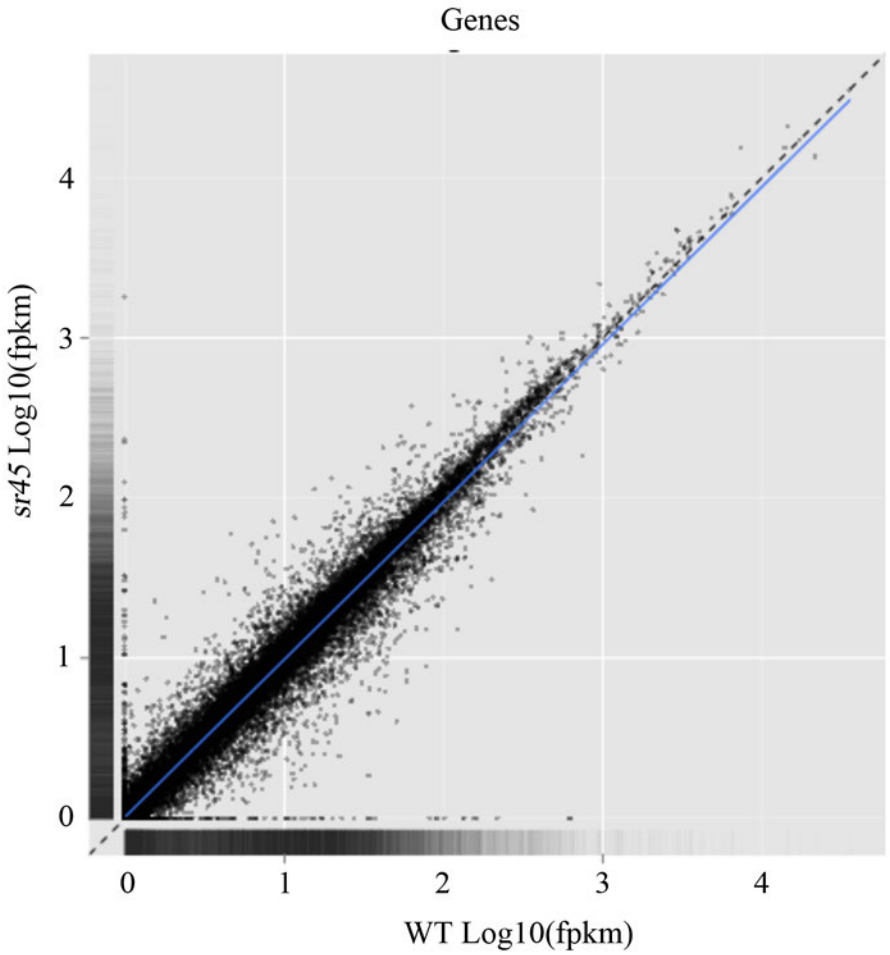


Figure 12: A scatter plot showing the expression of genes in WT and *sr45*.

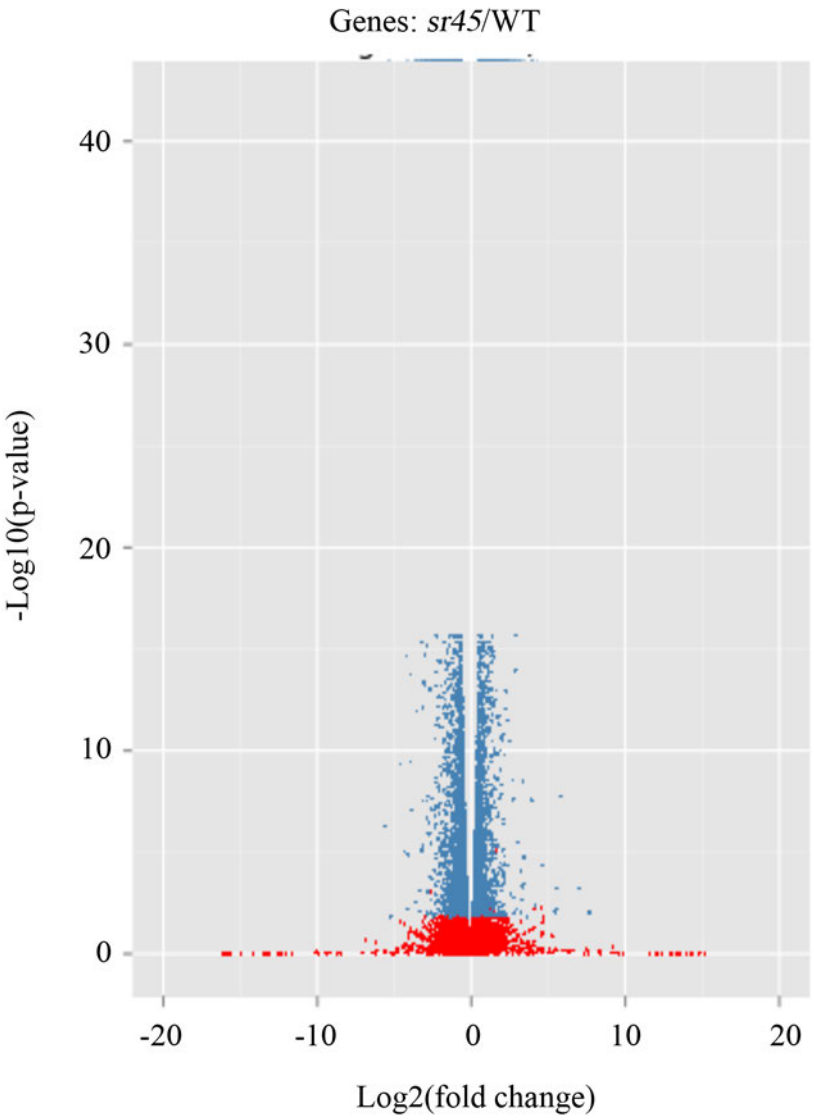


Figure 13: Volcano plot showing the differentially expressed genes in WT and *sr45*. The *x*-axis shows the log2 fold change in gene expression between WT and *sr45*, and the y-axis shows the statistical significance of the differences (log10 (p-value)). Dots represent different genes. Genes that have similar expression in both WT and *sr45* are indicated by red dots, and the differentially expressed genes are indicated by blue dots (p-value < 0.05).

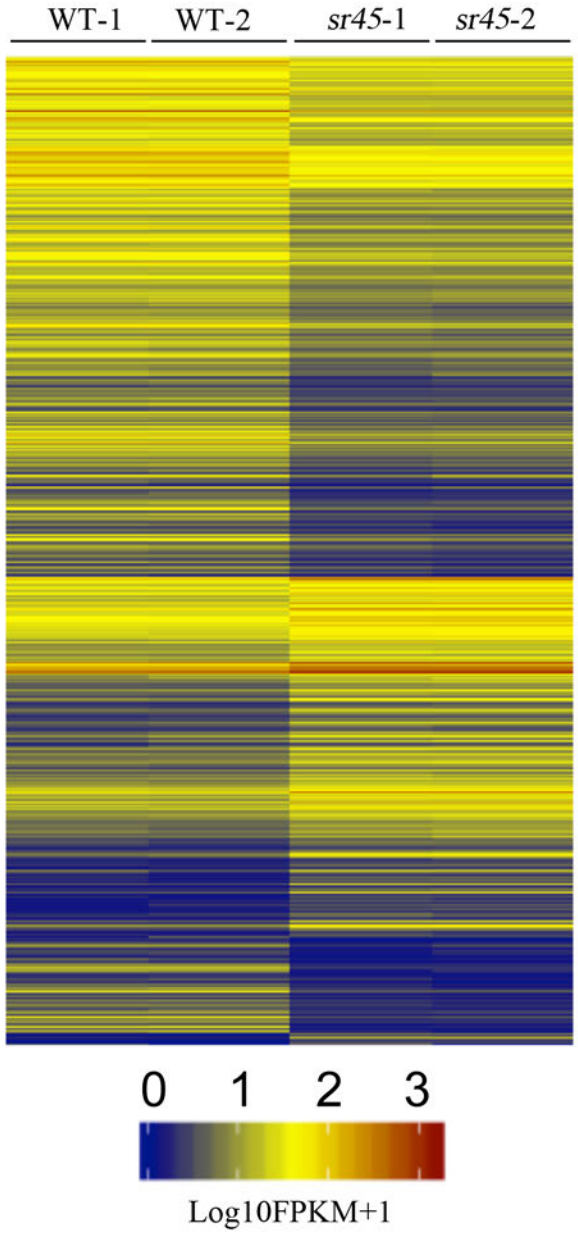


Figure 14: The Heat map of transcript levels of differentially expressed genes in WT and *sr45* (all biological replicates). Columns in the heat maps represent samples, and rows represent genes. Color scale (below the heat map) indicates gene expression level. Red indicates high expression while blue indicates low expression.

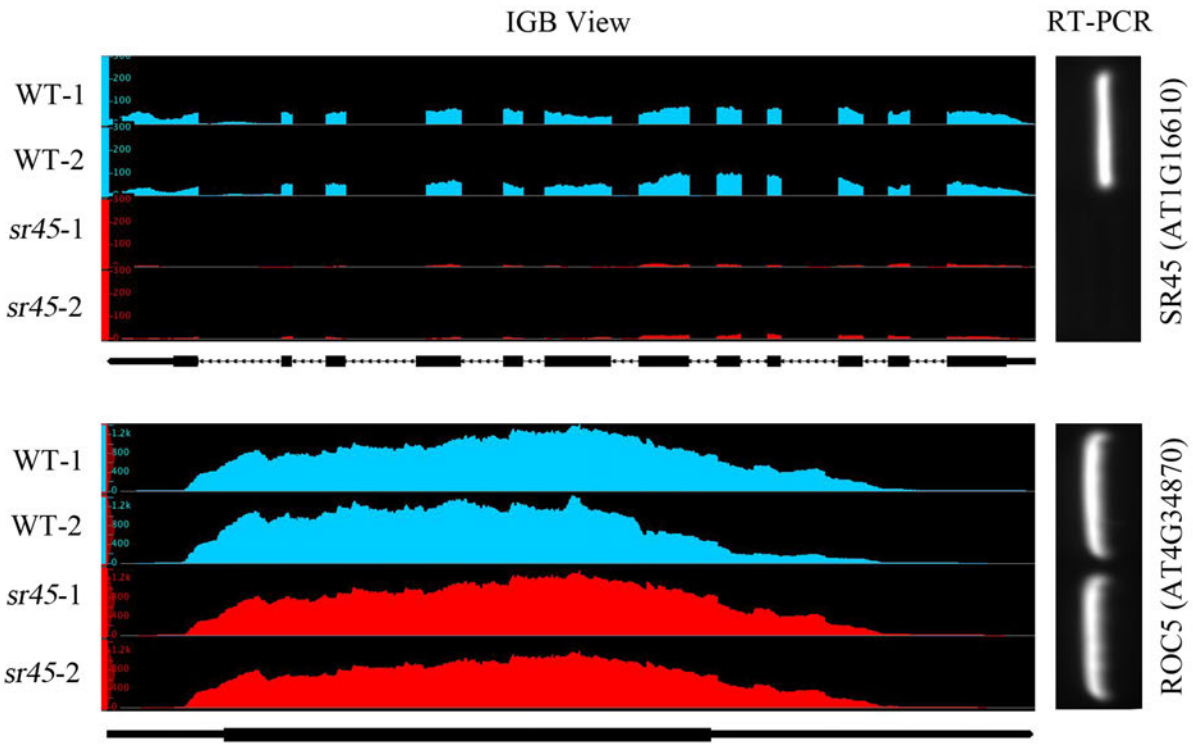
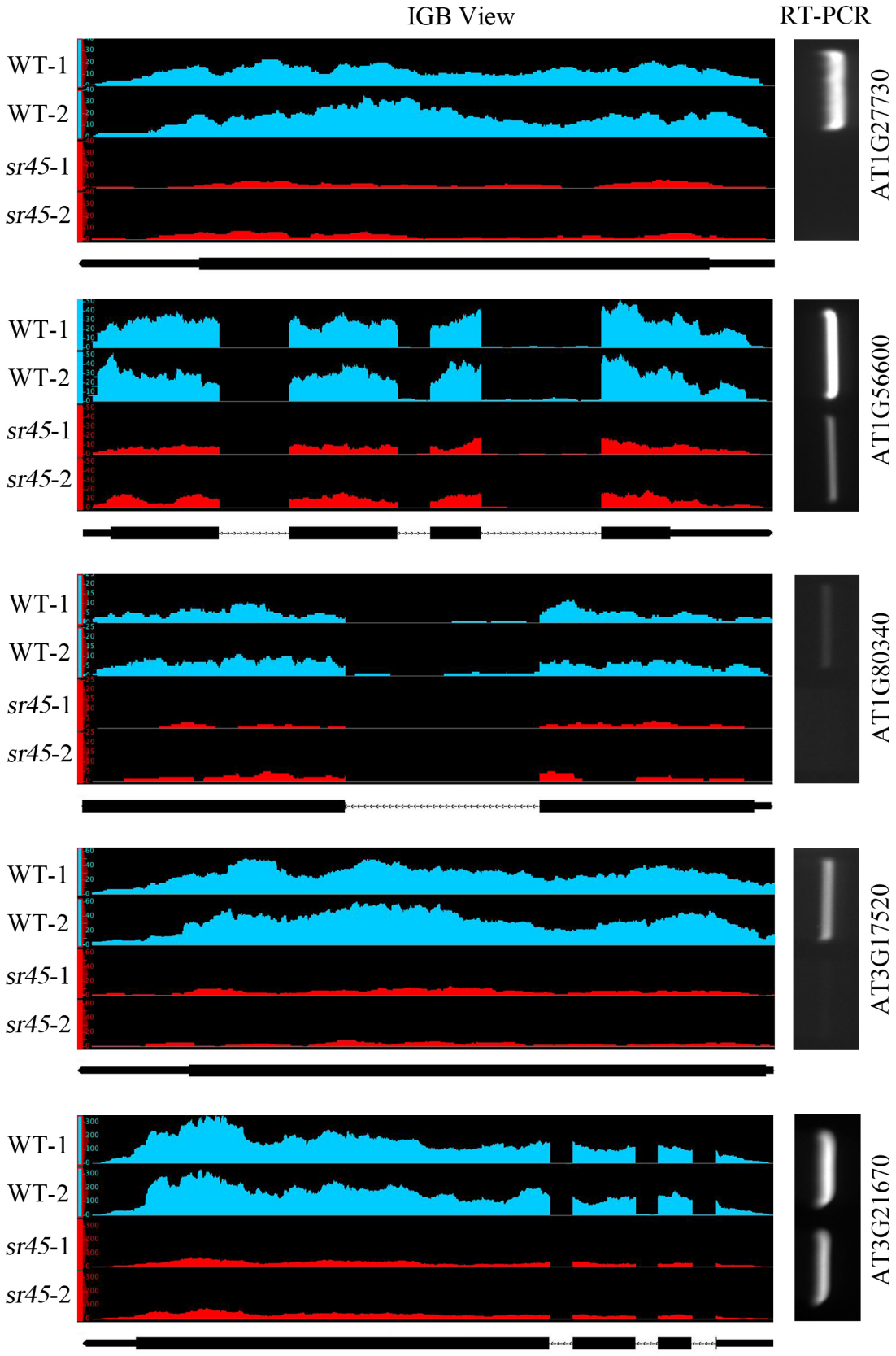
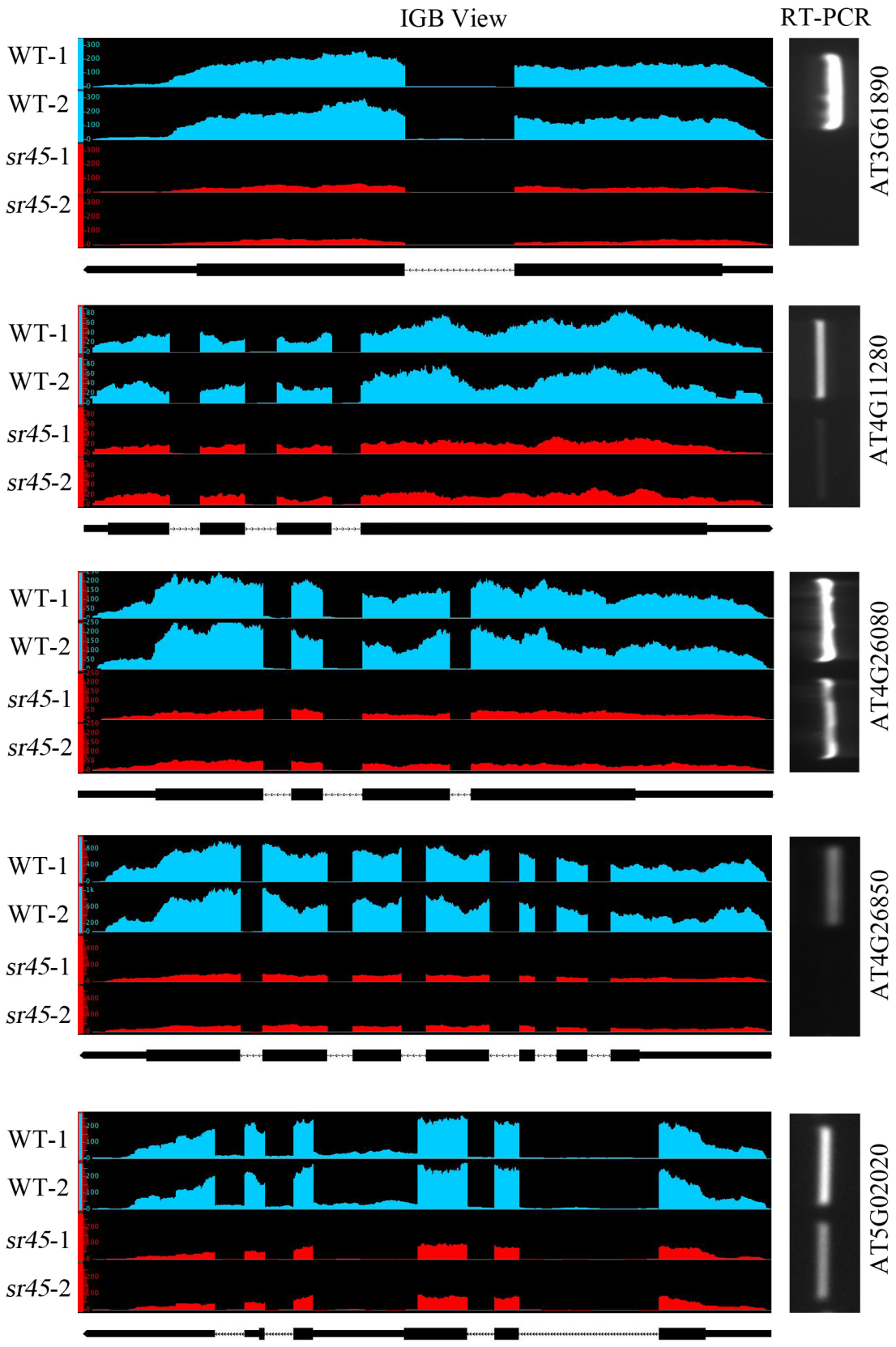


Figure 15: The Integrated Genome Bowser (IGB) view of sequence read depth in WT and *sr45* for *SR45* and *ROC5*. Panels on the left show the read depth in replicates both samples. The *y*-axis indicates read depth. The gene structure is displayed on the top of each IGB view. Thick boxes represent exons; lines represent introns, and thin boxes represent UTR. The expression of SR45 was abolished as expected in the knockout mutant of SR45 (*sr45*). *ROC5*, a non-differentially expressed gene, was used as a control. Panels on the right show RT-PCR results.


IGB View
RT-PCR
WT-1
WT-2
sr45-1
sr45-2
AT1G27730
AT1G56600
AT1G80340
AT3G17520
AT3G21670


IGB View
RT-PCR
WT-1
WT-2
sr45-1
sr45-2
AT3G61890
AT4G11280
AT4G26080
AT4G26850
AT5G02020

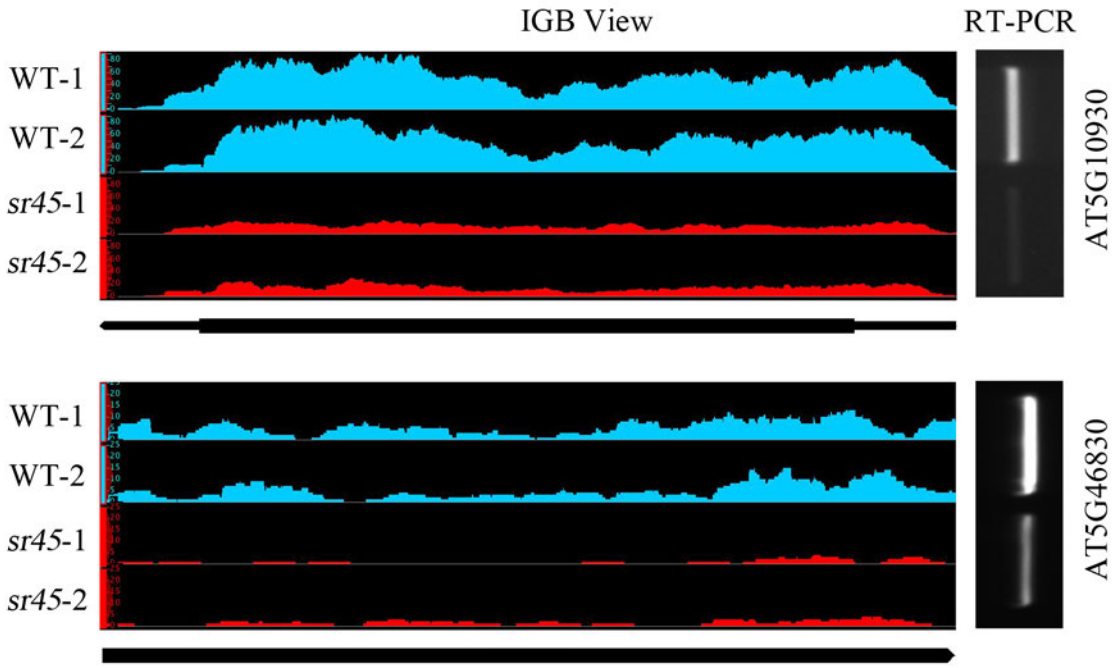
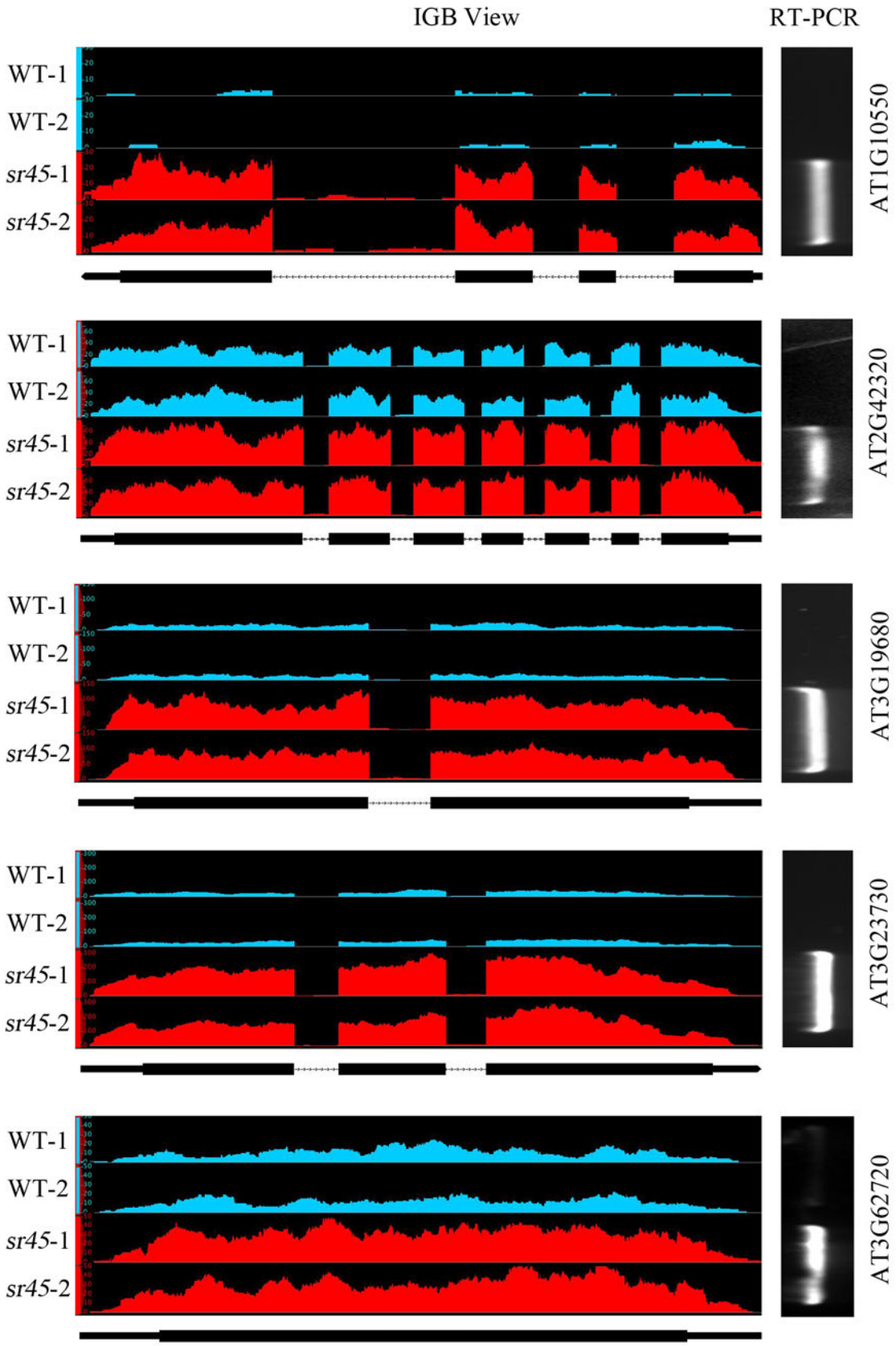
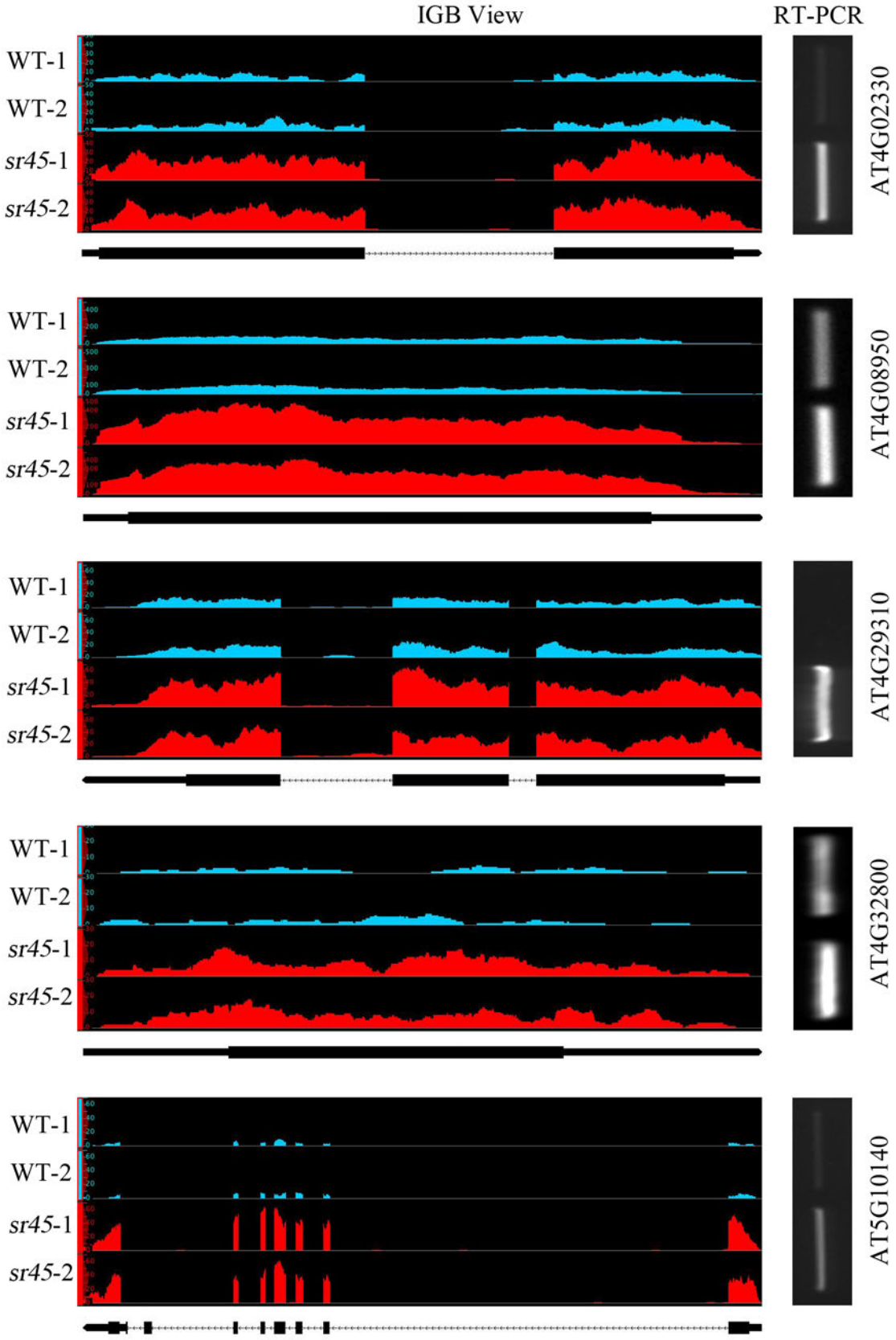


Figure 16: RT-PCR verification of down-regulated genes in *sr45*. Twelve down-regulated genes were randomly selected from the differentially expressed gene list. Panels on the left show the read depth graphs of both replicates of WT and *SR45* mutant. The *y*-axis indicates read depth for each gene. The gene structure is displayed underneath of each IGB display. Thick boxes represent exons; lines represent introns, and thin boxes represent UTR. *ROC5*, which is expressed equally in both samples (Fig.15), was used as an input control. Panels on the right show validation of RNA-seq results using RT-PCR, Gene ID numbers are shown next to the RT-PCR panel.


IGB View
RT-PCR
WT-1
WT-2
sr45-1
sr45-2
AT1G10550
AT2G42320
AT3G19680
AT3G23730
AT3G62720


IGB View
RT-PCR
WT-1
WT-2
sr45-1
sr45-2
AT4G02330
AT4G08950
AT4G29310
AT4G32800
AT5G10140

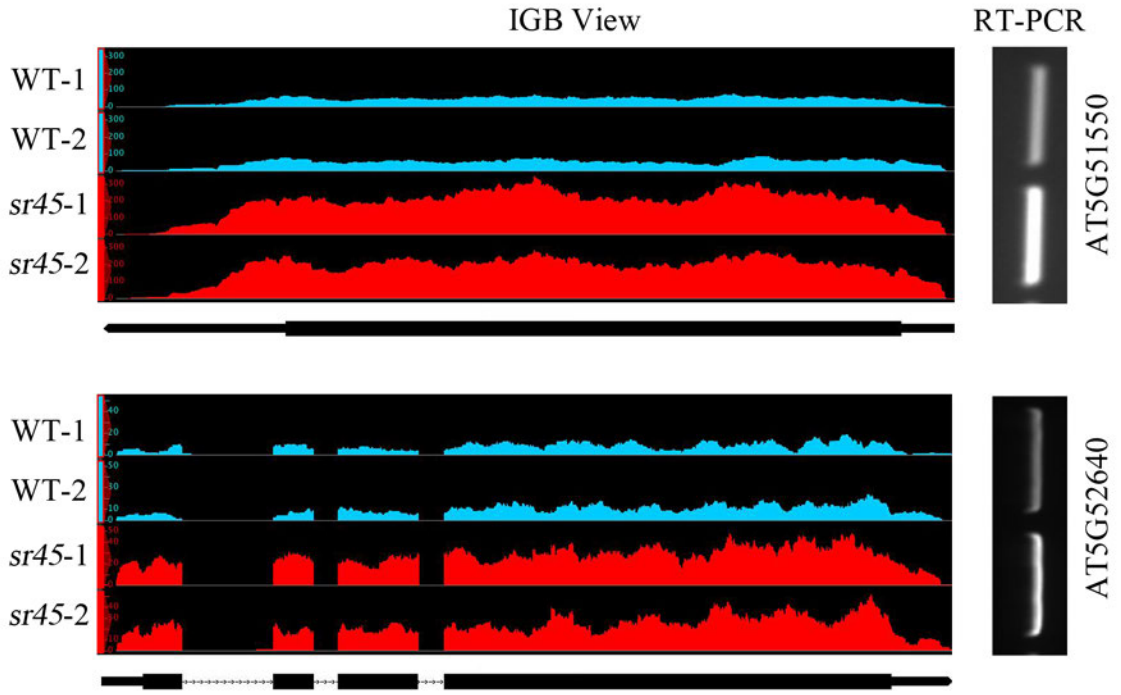


Figure 17: RT-PCR verification of up-regulated genes in *sr45*. Twelve up-regulated genes were randomly selected from the differentially expressed gene list. Panels on left show the read depth graphs of both replicates of WT and *SR45* mutant. The *y*-axis indicates read depth for each gene. The gene structure is displayed underneath of each IGB view. Thick boxes represent exons; lines represent introns, and thin boxes represent UTR. *ROC5*, which is expressed equally in both samples (Fig.15), was used as an input control. Panels on the right show validation of RNA-seq results using RT-PCR, Gene ID numbers are shown next to the RT-PCR panel.

and *sr45* (Table 6). To validate the differential expression of these micro RNA genes, we performed RT-PCR for one randomly selected micro RNA gene, MIR156C. Interestingly, the expression and AS of MIR156C was misregulated in *sr45* (Figure 18A). The down-regulation of MIR156C in *sr45* suggests that its targets will be up-regulated; indeed, the RNA-Seq data shows that the expression at least three of its known target genes is increased in *sr45* (Figure 18B).

Alternative Splicing (AS)

For a given gene, if there was more than one transcript detected in the pool of both WT and *sr45* samples, the gene was defined as an AS gene. We detected 6,820 AS events derived from 4,881 genes (Table 7), about 20% of annotated multi-exon genes. When broken down by type of AS event, we found that intron retention (IR) accounted for 51% of AS transcripts, followed by alternative 3′ splicing sites (A3′ss, 27%), alternative 5′ splicing sites (A5′ss, 13%) and exon skipping (ES, 9%). This distribution of AS event types remains consistent when WT and *sr45* are analyzed individually (Figure 19; Table 8).

Differentially Splicing in *sr45*

Qualitative difference (present in one sample but not in the other) in transcript isoforms from a gene between WT and *sr45* is defined as DS. Pre-mRNAs from a gene may undergo one or more differential AS events. Using SpliceGrapher, we identified a total of 927 DS genes. To validate DS events, we performed RT-PCR analyses for four randomly selected genes. Pre-mRNAs from all four genes were validated as differentially spliced in *sr45* (Figure 20). The proportions of types of AS events among DS genes were very close to the genome-wide proportions of AS genes either in WT, *sr45*, or in the combined data; however, the proportions of IR and A3′ss in DS genes were slightly lower and higher, respectively, as compared to the AS events in the genome (Figure 19 and 21A). However, when we examined the distribution of DS events, we found that DS events were biased to A3′ss,

Table 5: The DE microRNAs.

Micro RNA						
Gene	TAIR-ID	Locus	WT (RPKM)	*sr45* (RPKM)	log2(fold_change)	q_value
MIR156C	AT4G31877	4:15413161-15415898	0.472418	3.61654	2.93647	0.034059
MIR167A	AT3G22886	3:8108020-8109338	0.999193	4.09948	2.03661	4.03E-07
MIR398B	AT5G14545	5:4690938-4694086	3.51711	8.70034	1.30668	0.00284476
MIR824A	AT4G24415	4:12623529-12626432	4.23015	1.38813	-1.60756	0.00433161
MIR850A	AT4G13493	4:7842966-7846899	152.103	58.1566	-1.38703	0.0311145
MIR163	AT1G66725	1:24883930-24892046	19.6925	4.07464	-2.2729	4.73E-10

Table 6: The DE microRNAs and their DE targets.

Micro RNAs Target				
Micro RNAs	Gene-ID	Gene-name	Log2 (Fold-change)	Q-value
MIR156C	AT5G43270	SPL2	1.12472	2.26E-11
	AT5G50570	SPL13A	1.51974	1.83E-11
	AT5G50670	SPL13B	1.47982	0
	AT2G42200	SPL9	-1.24685	1.76E-07
MIR167A	AT4G23980	ARF9	-0.602337	6.25E-07
	AT2G46530	ARF11	-0.949295	2.42E-07
MIR824A	AT4G11880	AGL14	-1.10266	0.000113046
	AT2G45660	AGL20 (SOC1)	-1.0964	0
MIR163	AT3G44860	FAMT	-1.52081	1.40E-06

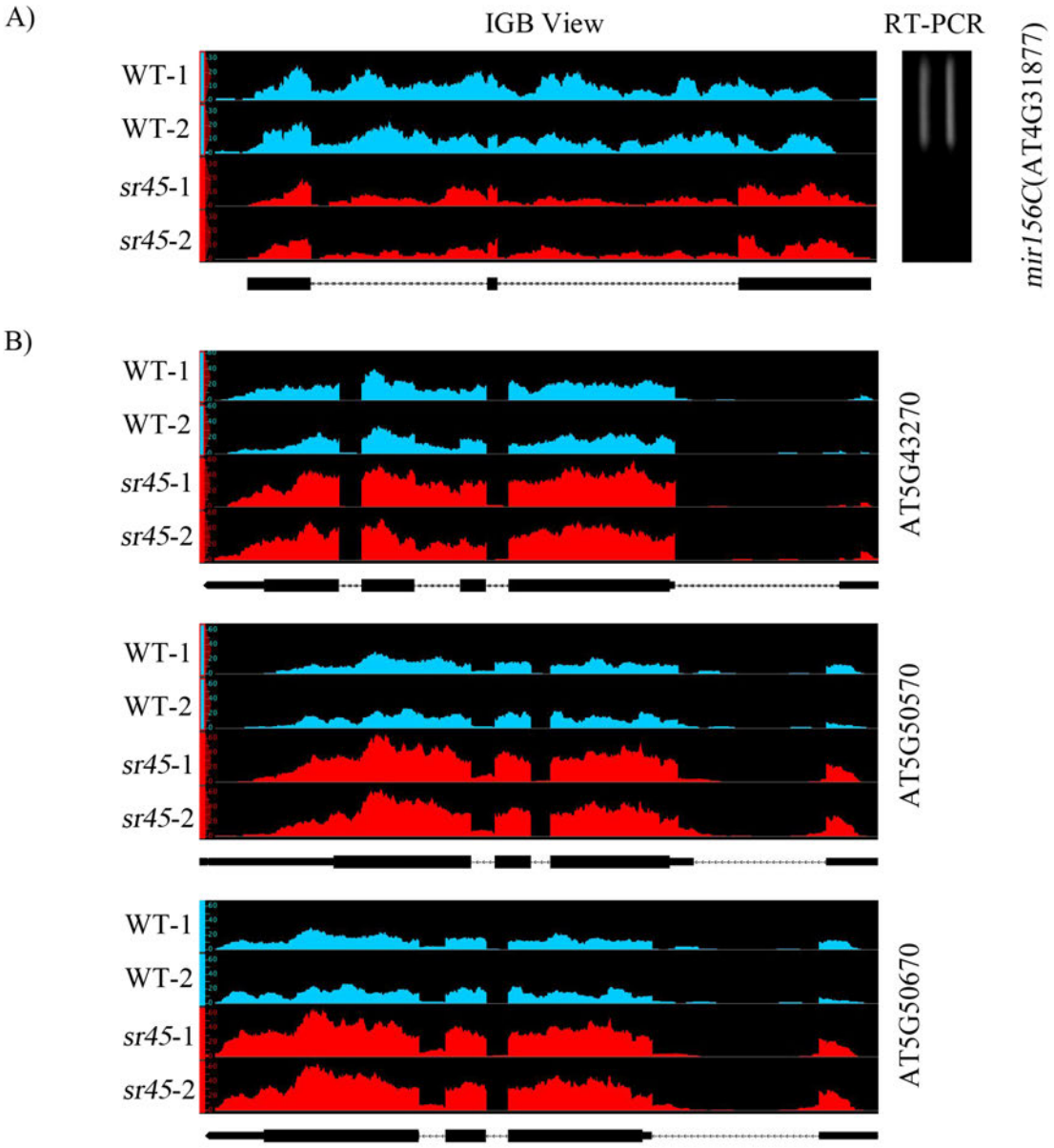


Figure 18: IGB view and RT-PCR verification of differentially expressed microRNA in *sr45*. (A) Mir156C (AT4G31877), one of the differentially expressed microRNAs in *sr45*, was selected for further verification. (B) Expression profile of selected mir156C target genes. Panels on the lift for A and B show the RNA-seq read depth from IGB of all replicates. The y-axis indicates read depth for each gene. The gene structure was displayed underneath of each IGB view. The box represents exon; line represents intron, and the thin box represents UTR. *ROC5*, a non-differentially expressed gene, was used as an input control (Fig.15). The RT-PCR verification of read depth differences of Mir156C is shown on the right of panel A. Panels on the right for A and B show genes names, and gene's ID numbers.

Table 7: AS statistics based on RNA-Seq data for WT and *sr45*.

Source	AS Genes	IR	A3'SS	A5'SS	ES	Total
Unique Wild-type	1379	1235	546	283	219	6820
Unique SR45 mutant	1100	1064	446	195	167	
In both WT& SR45	2402	1182	874	414	195	
Total	4881	3481	1866	892	581	
%		51	27	13	9	100

Table 8: The % AS events for each type of splicing.

Samples	IR	A3'SS	A5'SS	ES
Total	51.0	27.4	13.1	8.5
WT	48.8	28.7	14.1	8.4
sr45	49.5	29.1	13.4	8.0
DAS	46.8	30.8	13.6	8.9

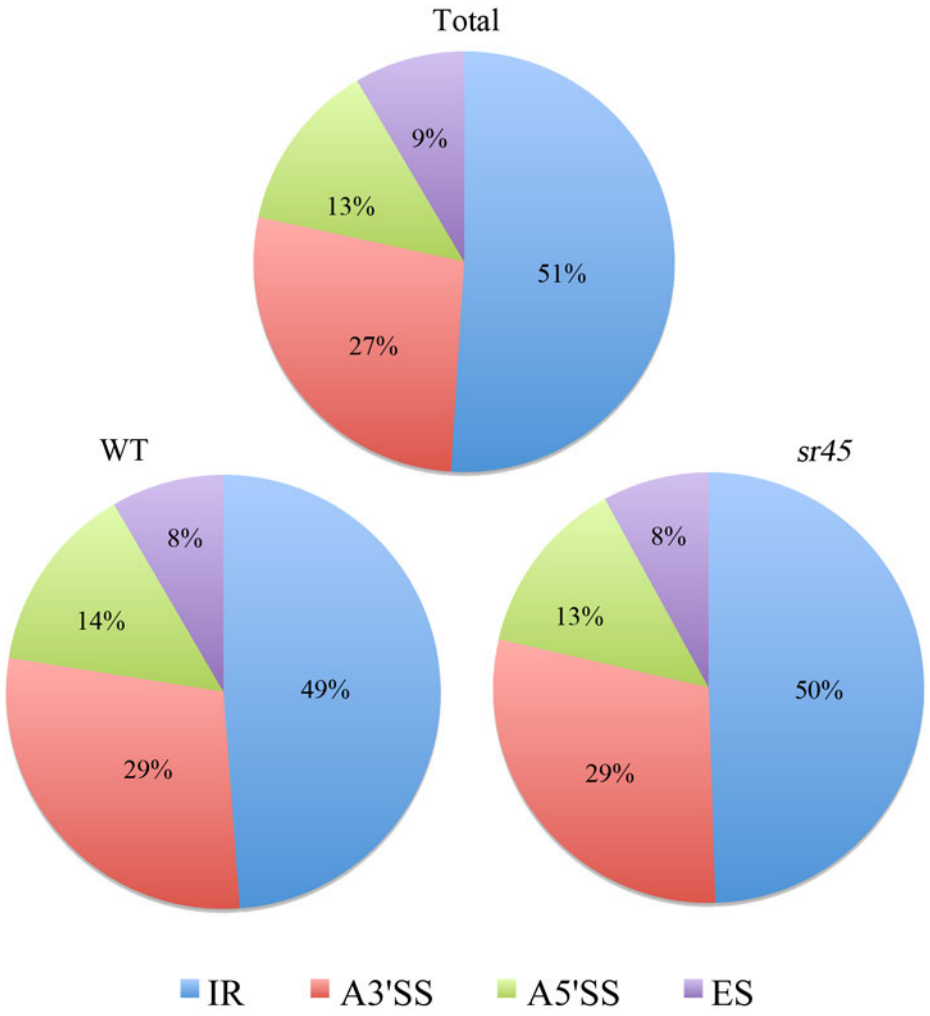


Figure 19: The percentage of different type of AS events. The upper graph (Total) represents the total AS events derived from all the mapped sequence reads of WT and *sr45* replicates (~ 80 million reads). Lower graphs (WT and *sr45*) show the percentage of AS events derived from WT and *sr45* individually. Intron retention (IR), alternative 3′ splice site (A3′ss), alternative 5′ splice site (A5′ss), and exon skipping (ES).

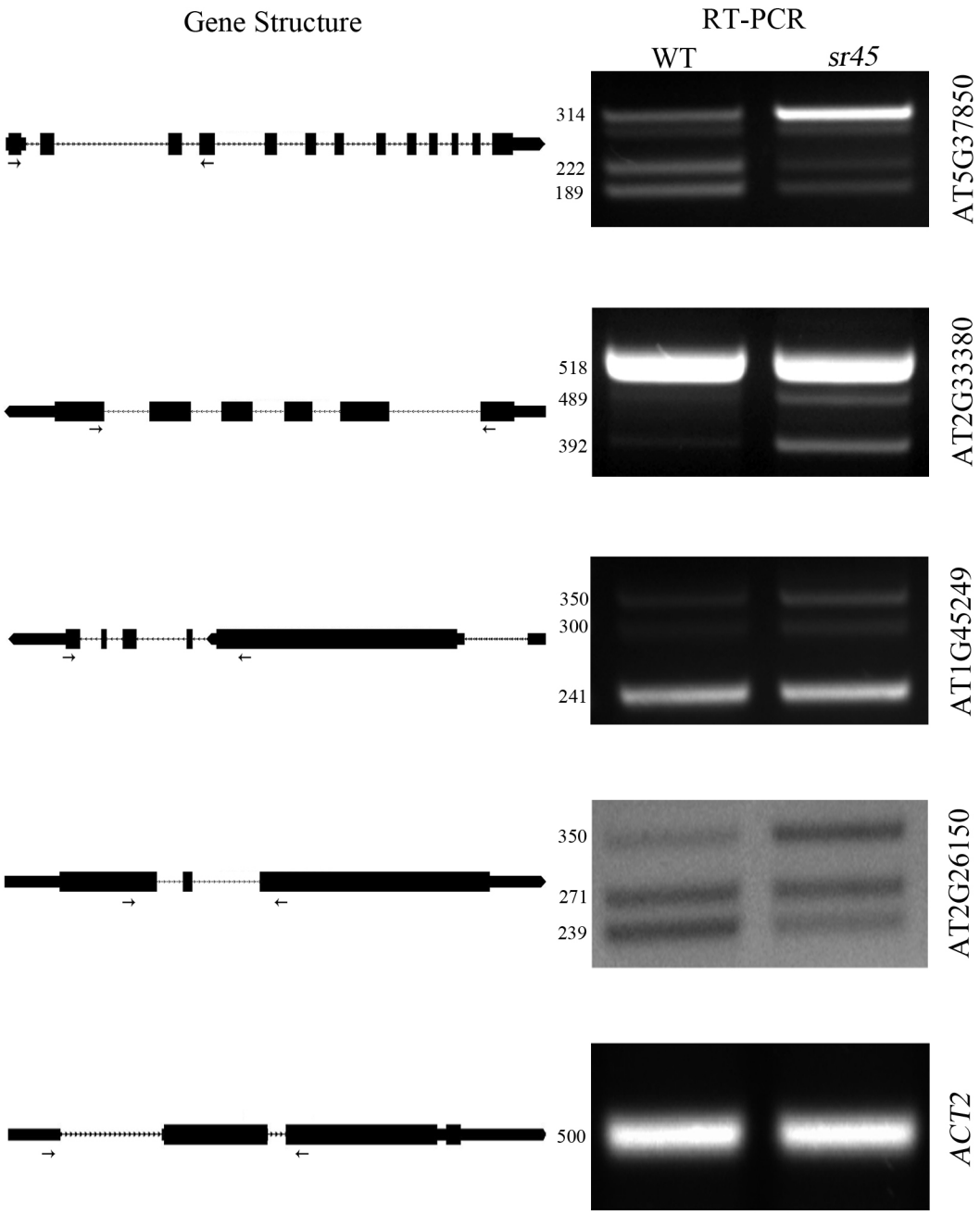


Figure 20: RT-PCR verification of selected differentially spliced genes in *sr45*. Left panels show the gene structure. Thick boxes represent exons; lines represent introns, and the thin boxes represent UTRs. *ACT2* was used as a control for cDNA template input. Right panels show RT-PCR verification of differential splicing, ID numbers. Arrowheads under the gene structure indicate the location of primers used in RT-PCR. Sizes of amplified products are shown next to the RT-PCR panel.

suggesting that SR45 preferentially affects the choice of 3′ splice sites (Figure 21B). When the distribution of the DS events was broken down along the gene body, the A3′ss within coding sequence (CDS) stood out as the most prevalent AS type, followed by A5′ss within CDS and 3′ UTR (Figure 22; Table 9). ES events were the least represented AS type in any part of the DS gene body (Figure 22; Table 9). Taken together, this indicates that SR45 largely regulates the AS of its targets by affecting the 3′ splicing sites within CDS regions.

SR45 Functions in Hormonal Signaling and Abiotic Stress Responses

Previous genetic and biochemical studies revealed that SR45 plays a role in plant development, probably through controlling pre-mRNA splicing, and also functions in abiotic stress responses (Albaqami, 2013; Ali et al., 2003; Ali et al., 2007; Carvalho et al., 2010; Zhang and Mount, 2009). To determine whether SR45-regulated genes mediate previously reported functions and/or have additional biological roles, we performed a Gene Ontology (GO) enrichment assay on DE and DS genes together or separately in *sr45* using GeneCoDis (Carmona-Saez et al., 2007). Pooled genes of DE and DS were significantly enriched in 46 GO biological process terms (Figure 23; Table 10). Remarkably, among enriched GO terms are response to abscisic acid stimulus, salt stress, and heat stress; these categories are in agreement with the previously suggested functions of SR45 (Albaqami, 2013; Carvalho et al., 2010). Most strikingly, 26 out of the 46 enriched GO terms are involved in either response to hormones or response to abiotic and biotic stresses (Table 10), suggesting that the major role of SR45 is in regulating *Arabidopsis* response to developmental and environmental stimuli. Similar results were obtained when GO analysis were performed with either DE or DS separately (Figure 24 and 25). Taken together, the enriched GO terms are consistent with SR45's biological and molecular functions and suggest that SR45 is a key regulator for the reprogramming of gene expression,

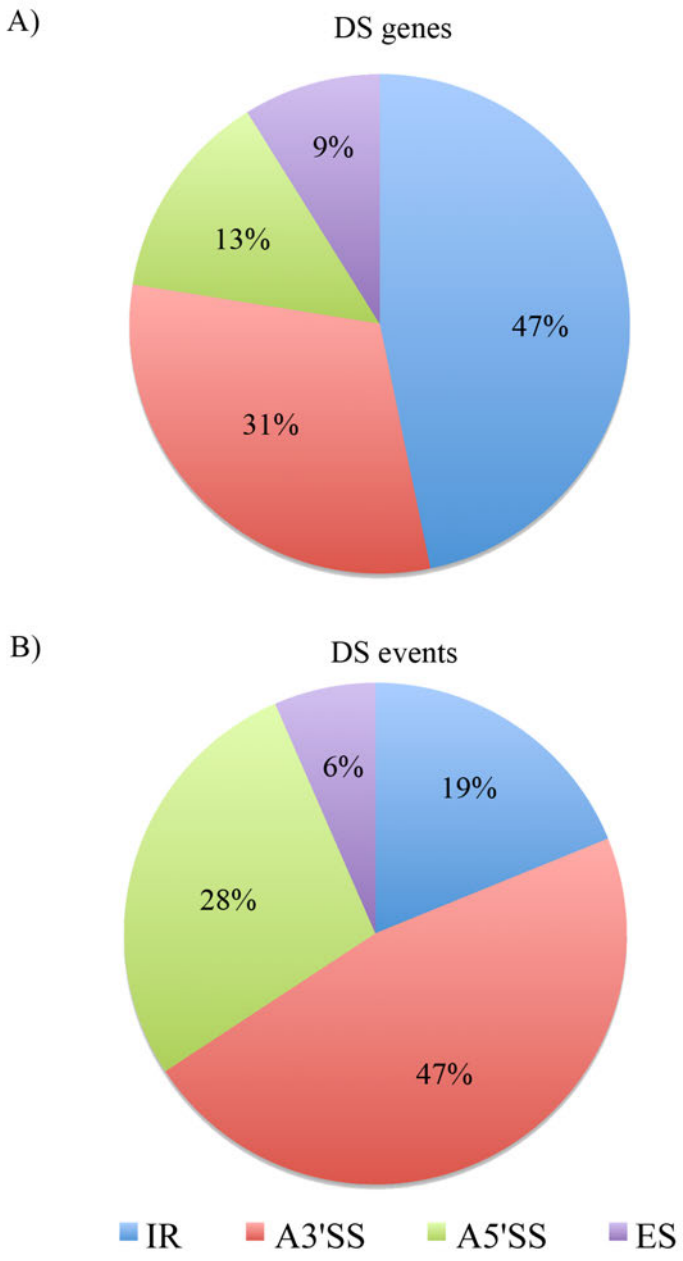


Figure 21: The percentage of differential AS events in *sr45*. (A) A pie chart showing the proportion of each type of AS event in differentially spliced genes. (B) The proportion of differential AS events between WT and *sr45*. Intron retention (IR), alternative 3′ splice site (A3′SS), alternative 5′ splice site (A5′SS), and exon skipping (ES).

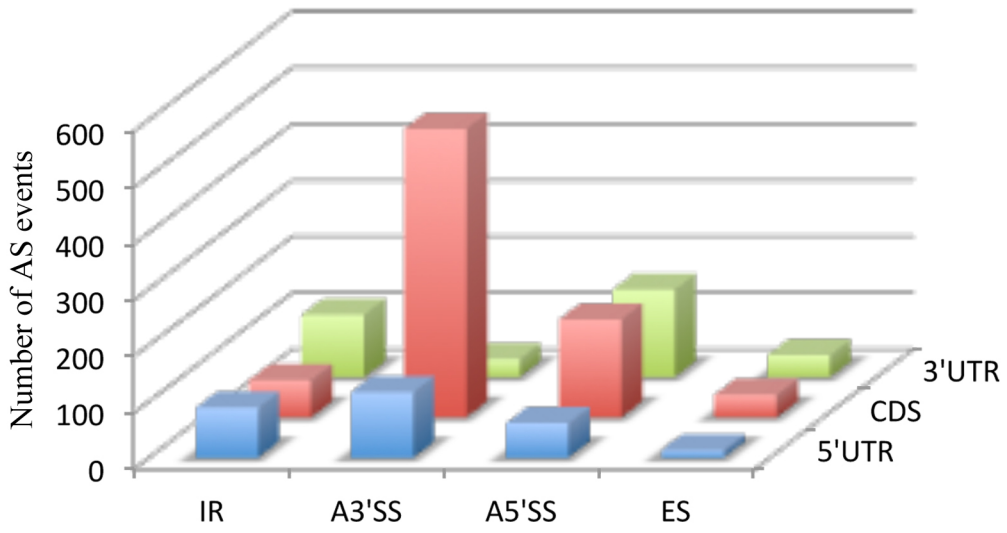


Figure 22: Distribution of differential AS events across the gene body. Intron retention (IR), alternative 3′ splice site (A3′SS), alternative 5′ splice site (A5′SS), and exon skipping (ES), in different parts of the transcript (*z*-axis). 5′ untranslated region (5′UTR), Coding sequence (CDS), 3′ untranslated region (3′UTR).

Table 9: The distribution of DS events along the gene body for each type of splicing.

	IR	A3'SS	A5'SS	ES
Overall	265 (19%)	660 (47%)	390 (28%)	92 (13%)
5'UTR	88	116	61	13
CDS	66	512	174	40
3'UTR	111	32	155	39

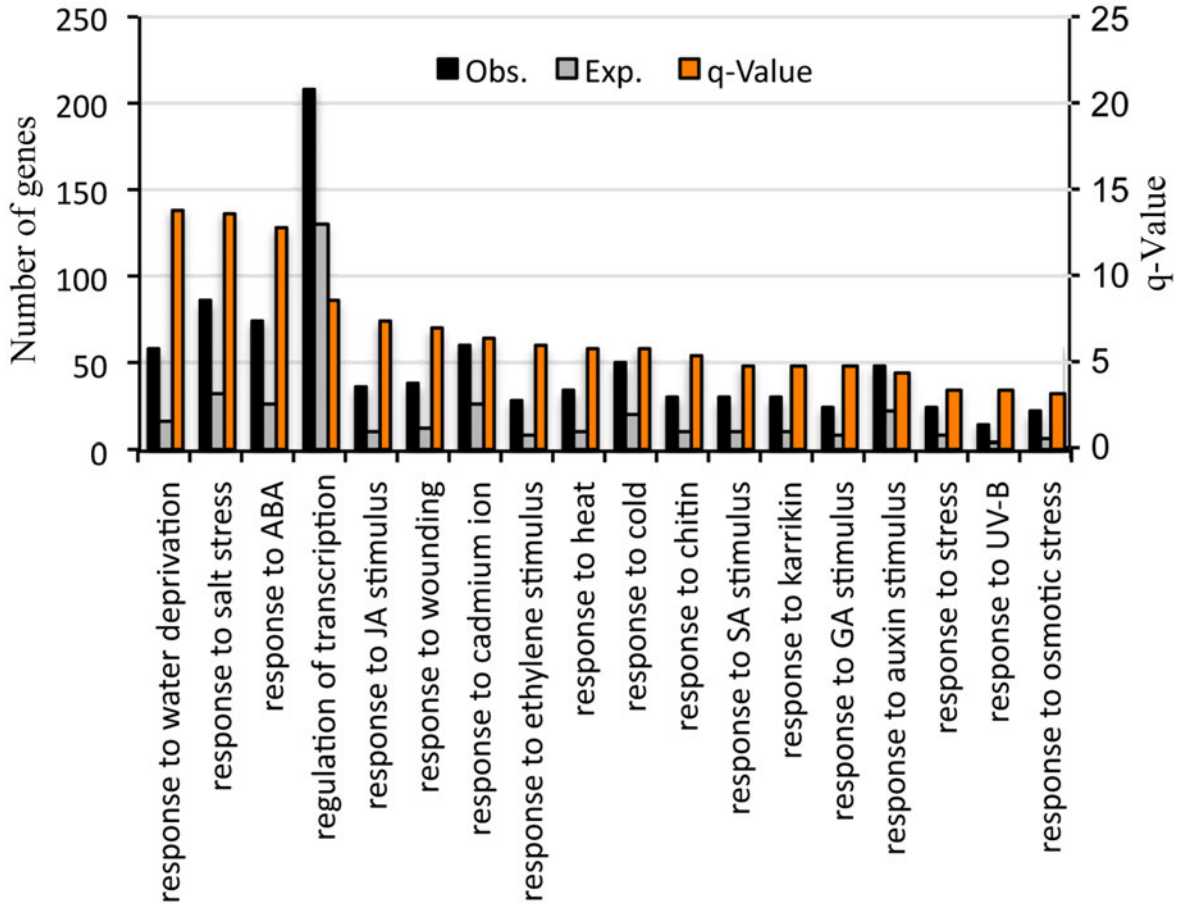


Figure 23: Gene Ontology term enrichment analysis of the differentially expressed and differentially spliced genes using GeneCoDis (http://genecodis.cnb.csic.es). Obs., the number of observed differentially expressed and differentially spliced genes; Exp., the number of expected differentially expressed genes; q-Value, the adjusted p-value calculated based on hypergeometric statistics and presented as "$-Log_{10}$(Hyp-c)".

Table 10-1: The enriched GO terms in DE and DS genes.

GO-term-Description	Support	List size	Reference Support	Reference size	Hyp*	Hyp-c*
response to water deprivation (BP)	58	2822	201	33602	2.32E-17	1.16E-14
response to salt stress (BP)	86	2822	382	33602	1.93E-17	1.93E-14
response to abscisic acid stimulus (BP)	75	2822	324	33602	4.39E-16	1.46E-13
regulation of transcription, DNA-dependent (BP)	209	2822	1563	33602	8.41E-12	2.10E-09
response to jasmonic acid stimulus (BP)	37	2822	139	33602	1.73E-10	3.46E-08
response to wounding (BP)	38	2822	151	33602	5.79E-10	9.64E-08
response to cadmium ion (BP)	61	2822	327	33602	2.80E-09	3.99E-07
response to ethylene stimulus (BP)	29	2822	106	33602	7.96E-09	9.94E-07
response to heat (BP)	34	2822	139	33602	9.95E-09	1.10E-06
response to cold (BP)	50	2822	255	33602	1.37E-08	1.36E-06
response to chitin (BP)	31	2822	127	33602	4.61E-08	4.19E-06
response to salicylic acid stimulus (BP)	30	2822	127	33602	1.64E-07	1.26E-05
response to karrikin (BP)	30	2822	127	33602	1.64E-07	1.26E-05
response to gibberellin stimulus (BP)	25	2822	95	33602	1.91E-07	1.36E-05
response to auxin stimulus (BP)	48	2822	267	33602	4.09E-07	2.72E-05
response to stress (BP)	25	2822	114	33602	7.10E-06	0.000394254
response to UV-B (BP)	15	2822	49	33602	7.04E-06	0.000413891
response to osmotic stress (BP)	22	2822	93	33602	6.84E-06	0.000426965
metabolic process (BP)	97	2822	760	33602	2.38E-05	0.00124877
galactolipid biosynthetic process (BP)	5	2822	7	33602	7.57E-05	0.0034361
jasmonic acid and ethylene-dependent systemic resistance (BP)	5	2822	7	33602	7.57E-05	0.0034361
positive regulation of transcription, DNA-dependent (BP)	27	2822	145	33602	6.97E-05	0.003482
defense response to bacterium (BP)	33	2822	196	33602	9.43E-05	0.00409403
negative regulation of transcription, DNA-dependent (BP)	21	2822	106	33602	0.000174856	0.00671852
cold acclimation (BP)	9	2822	26	33602	0.00017101	0.00683358
protein folding (BP)	35	2822	219	33602	0.000170969	0.00711659
defense response to fungus (BP)	24	2822	130	33602	0.00019655	0.00727237
response to high light intensity (BP)	11	2822	40	33602	0.000335833	0.0115689
cellular response to phosphate starvation (BP)	18	2822	88	33602	0.000328009	0.0117029
abscisic acid mediated signaling pathway (BP)	15	2822	68	33602	0.000432105	0.0143891

Table 10-2: The enriched GO terms in DE and DS genes.

GO-term-Description	Support	List size	Reference Support	Reference size	Hyp*	Hyp_c*
carbohydrate metabolic process (BP)	51	2822	377	33602	0.000504034	0.0162429
protein phosphorylation (BP)	96	2822	820	33602	0.000590094	0.0173383
response to red light (BP)	12	2822	49	33602	0.000585651	0.0177292
syncytium formation (BP)	6	2822	14	33602	0.00057955	0.0180928
response to oxidative stress (BP)	35	2822	235	33602	0.000660267	0.0188459
response to virus (BP)	9	2822	31	33602	0.000750907	0.0208377
response to temperature stimulus (BP)	6	2822	15	33602	0.000897447	0.0235934
defense response to insect (BP)	6	2822	15	33602	0.000897447	0.0235934
microtubule-based movement (BP)	11	2822	45	33602	0.000997823	0.0255596
glucosinolate catabolic process (BP)	5	2822	11	33602	0.00124883	0.0297044
cellular response to heat (BP)	5	2822	11	33602	0.00124883	0.0297044
unidimensional cell growth (BP)	18	2822	98	33602	0.00124772	0.0311619
cellular glucan metabolic process (BP)	6	2822	17	33602	0.00191647	0.0445246
heat acclimation (BP)	7	2822	23	33602	0.00214844	0.0466586
trehalose biosynthetic process (BP)	7	2822	23	33602	0.00214844	0.0466586
hyperosmotic salinity response (BP)	11	2822	49	33602	0.0020981	0.0476364

* Hyp = Hypergeometric p-value, Hyp-c = corrected Hypergeometric p-value.

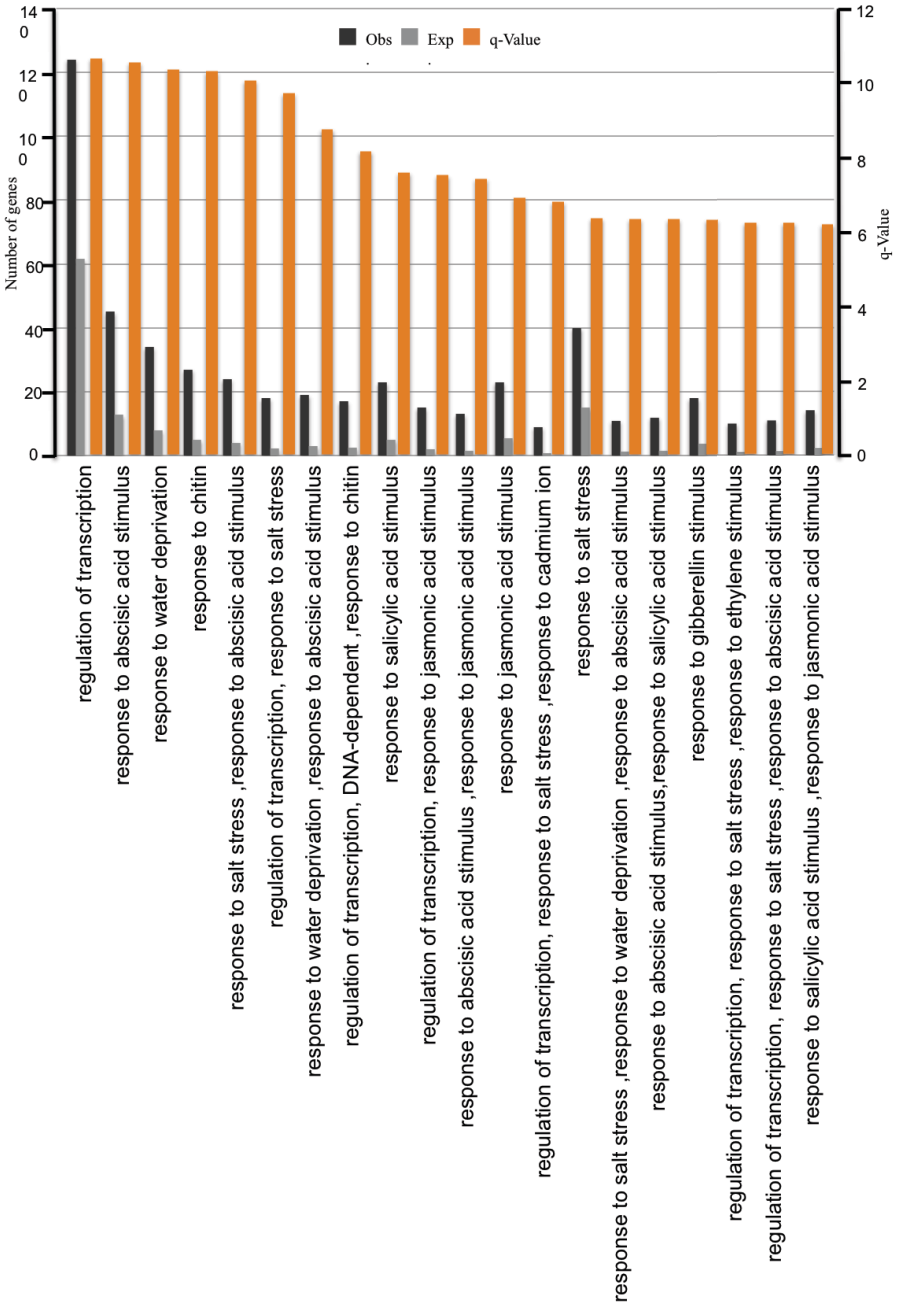


Figure 24: Gene Ontology term enrichment analysis of the differentially expressed genes using GeneCoDis (http://genecodis.cnb.csic.es). Obs., the number of observed differentially expressed and differentially spliced genes; Exp., the number of expected differentially expressed genes; q-Value, the adjusted p-value calculated based on hypergeometric statistics and presented as "-Log_{10}(Hyp-c)".

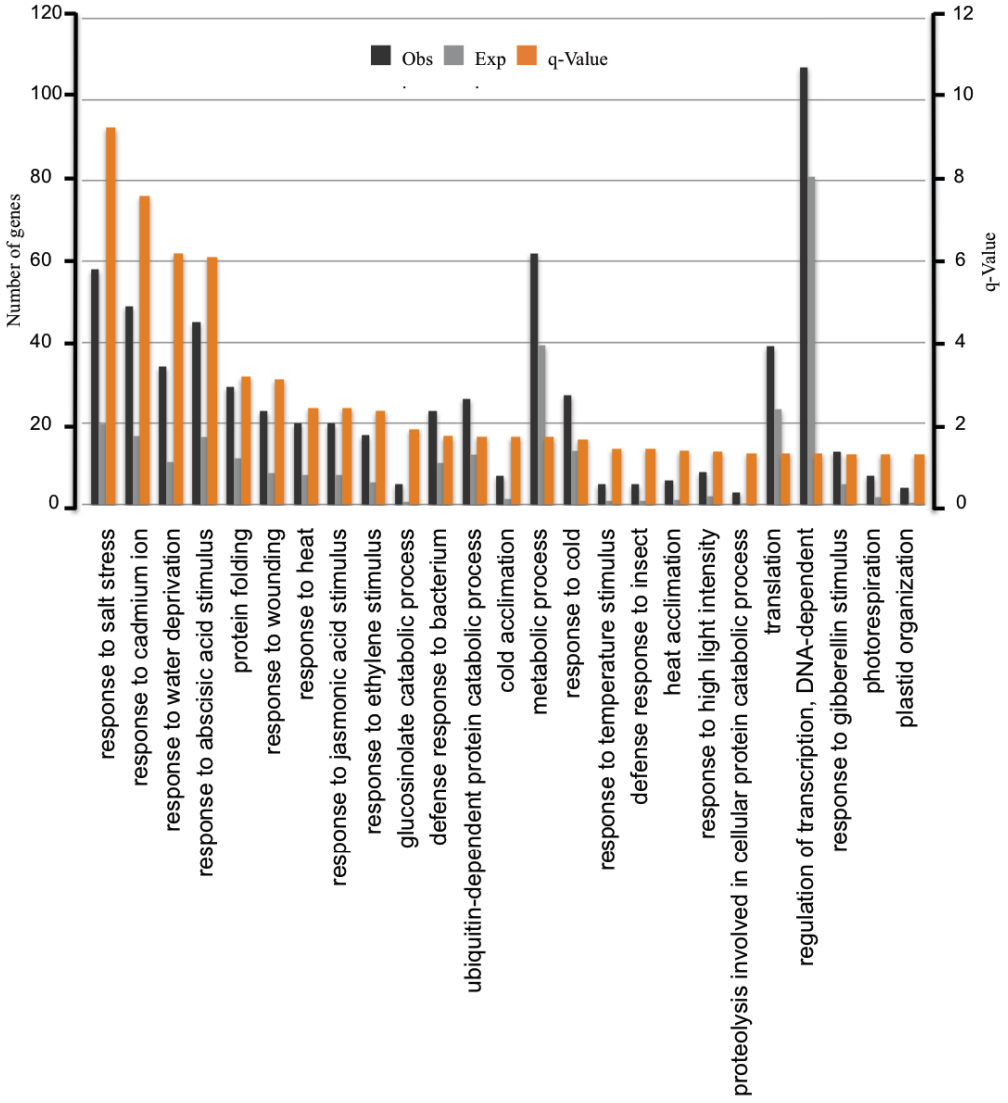


Figure 25: Gene Ontology term enrichment analysis of the differentially spliced genes using GeneCoDis (http://genecodis.cnb.csic.es). Obs., the number of observed differentially expressed and differentially spliced genes; Exp., the number of expected differentially expressed genes; q-Value, the adjusted p-value calculated based on hypergeometric statistics and presented as "-Log_{10}(Hyp-c)".

potentially through AS of a variety of transcription factors, in response to a variety of environmental stimuli.

SR45 Plays a Role in Heat Stress Response

The GO term response to heat stress merited further investigation for several reasons: (1) high temperature widely threatens plant growth and production (Bita and Gerats, 2013); (2) AS of pre-mRNAs has been shown to play a central role in thermotolerance (Guerra et al., 2015); (3) it was one of the significantly enriched GO terms for SR45-regulated genes (Figure 23); and (4) We recently showed that SR45 functions in response to heat stress (Albaqami, 2013).

Furthermore, SR45 itself is alternatively spliced, generating two isoforms, SR45.1 and SR45.2 (Figure 26A) (Palusa et al., 2007; Zhang and Mount, 2009). Both isoforms have different functions in plant development and abiotic stresses (Albaqami, 2013; Zhang and Mount, 2009), and we determined in a previous study that these isoforms also function differently in *Arabidopsis* response to heat stress; SR45.1 but not SR45.2 positively regulates thermotolerance (Albaqami, 2013).

As SR45.1 is the only isoform that rescues the mutant heat stress phenotype, we hypothesized that SR45.1 is more highly expressed than SR45.2, and determined the expression profile of each isoform in response to high temperature (Figure 27AB). When WT plants were exposed to 38^{o}C, q-PCR expression analysis of both isoforms showed that *SR45.1* was, in fact, more highly expressed than *SR45.2* (Figure 27C). However, three hours of recovery at normal temperature (24^{o}C) was sufficient to return expression of SR45.1 to normal, unstressed levels (Figure 27C). These results support my previous finding that SR45.1 plays a positive role in thermotolerance.

My previous work also demonstrated the role of SR45.1 in thermotolerance at different developmental stages, including early growing seedlings and 35-day-old rosette leaves (Albaqami, 2013). However, its importance in maintaining the normal germination of

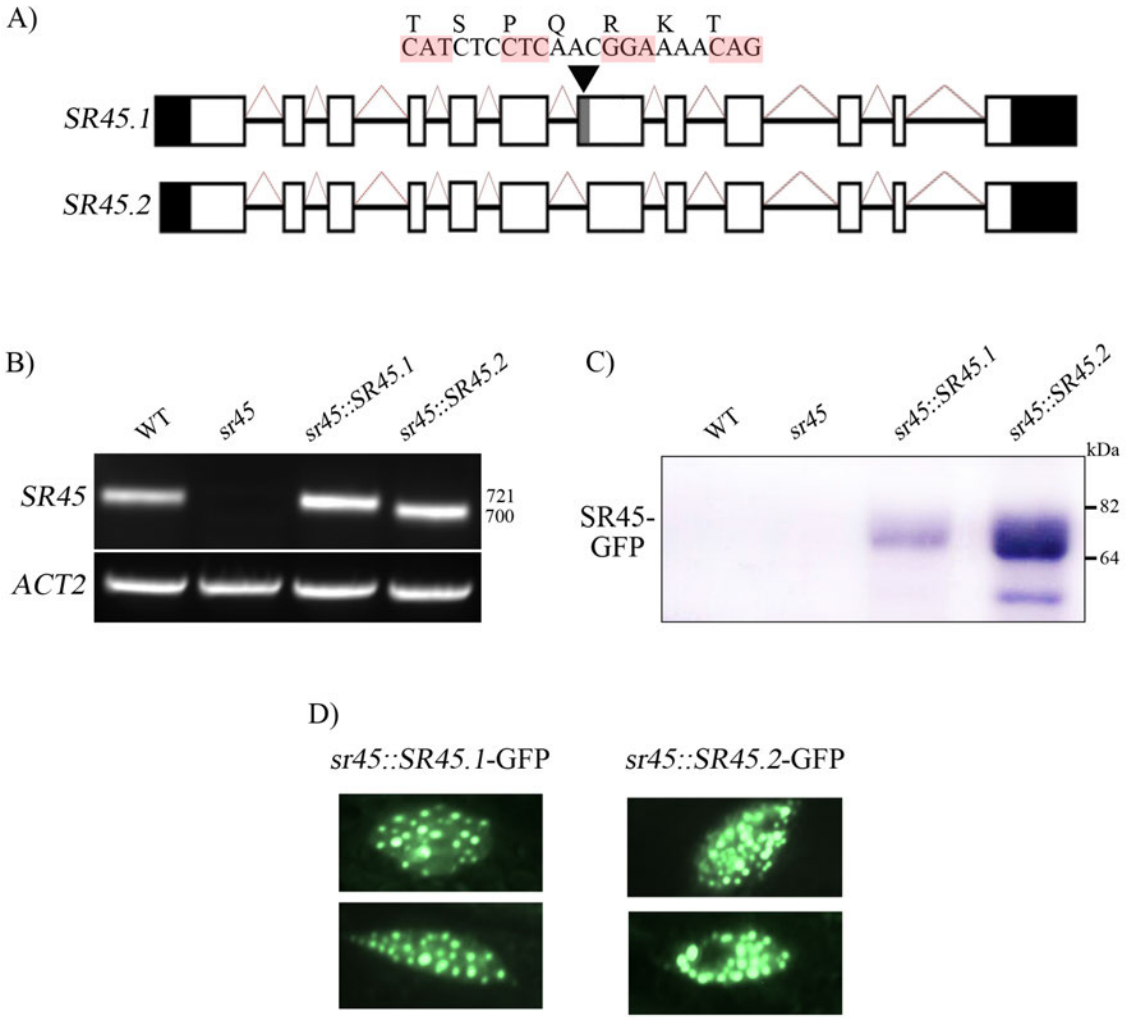


Figure 26: Overexpression of long (*SR45.1*) or short (*SR45.2)* isoforms of *SR45* in *sr45*. (A) Schematic diagrams showing AS of SR45 pre-mRNA that produces two splice isoforms: long (*SR45.1)* and short (*SR45.2)*. White boxes indicate exons, black lines indicate introns, black boxes indicate untranslated regions, and splicing is indicated by open triangles. Alternatively spliced region is indicated by a downward arrowhead. Seven amino acids and their corresponding nucleotides included in SR45.1 but missing in SR45.2 are shown on the top of the alternatively spliced region. (B) Confirmation of genotypes of WT, *sr45* and *sr45* expressing either long (*sr45*::*SR45.1-GFP)* or short isoform (*sr45*::*SR45.2-GFP*) using RT-PCR. *ACT2* was used as an input control (C) Immunodetection of overexpressed long and short isoform of SR45 using a GFP antibody. (D) Visualization of GFP-tagged long (*sr45*::*SR45.1-GFP)* and short isoform (*sr45*::*SR45.2-GFP*) in the nucleus as speckles using a fluorescent microscope (two pictures for each).

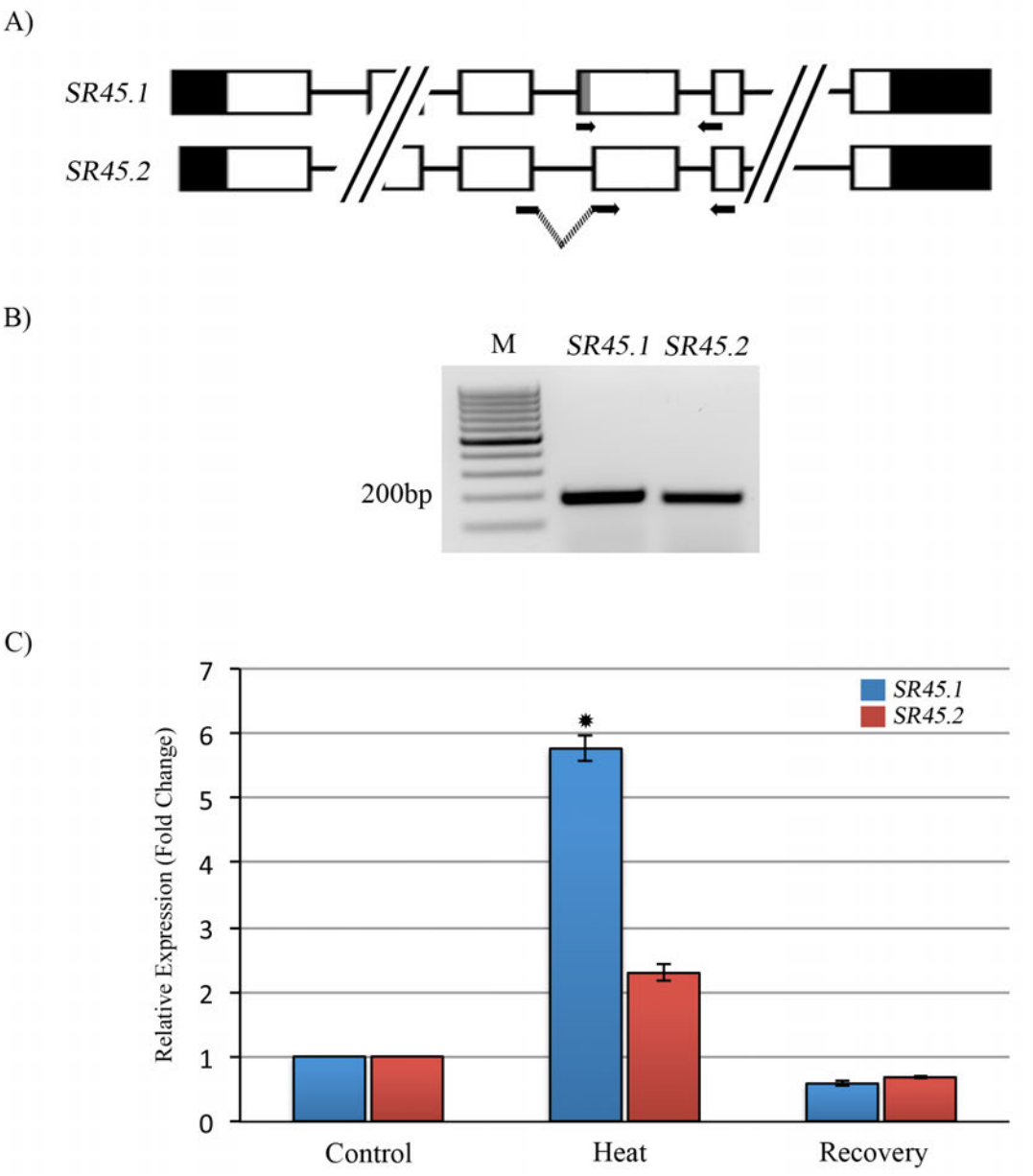


Figure 27: Expression of *SR45.1* and *SR45.2* splice isoforms in response to heat shock. (A) Schematic diagram showing the location of isoform-specific primers used in RT-PCR. (B) RT-PCR analysis to test the specificity of primers and the size of amplified product from *SR45.1* and *SR45.2* transcript in WT plants. (C) Quantification of expression of *SR45.1* and *SR45.2* isoform in control (24^{o}C) and heat-treated (38^{o}C) seedlings and in recovered seedlings after heat stress for 3hrs at 24^{o}C. Statistical significance (*t*-test P<0.05) is indicated by *). *SR45.1* is more abundant than *SR45.2* under heat stress (38^{o}C). This data represents one of three biological replicates experiments.

Arabidopsis seeds under high temperature is unclear. Therefore, we evaluated the effects of heat stress on germination in WT, *sr45* mutant, and SR45 isoform overexpression lines. Consistent with my previous results, *sr45* and the line overexpressing SR45.2 exhibited significantly lower germination under heat stress as compared WT and the line overexpressing SR45.1. No differences in germination were observed among these under normal temperature (Figure 28AB). Taken together, these results show that SR45.1 plays a positive role in thermotolerance at different developmental stages in *Arabidopsis*.

Phosphorylation-Mediated Regulation of SR45.1

Arabidopsis phosphoproteome studies showed that both SR45 protein isoforms are highly phosphorylated under normal conditions (Figure 29) (Reiland et al., 2009 and Arabidopsis Protein Phosphorylation Site Database (PhosPhAt)). The additional seven amino acids (TSPQRKT) in SR45.1 but not in SR45.2 contain potential phosphorylation sites at threonine 218 (T218) and serine 219 (S219) (Figure 29) (Reiland et al., 2009; Zhang and Mount, 2009 and PhosPhAt). In fact, a previous study showed that phosphothreonine 218 is essential for the function of SR45.1 in controlling normal flower petal growth in *Arabidopsis* (Zhang and Mount, 2009). Moreover, sequence similarity analysis showed that S219 is highly conserved among different organisms; however, T218 appears to be a semi-conserved amino acid, and in some cases it is substituted with another S that is likely to be phosphorylatable (Figure 30). However, whether these SR45.1 specific-phosphorylation sites play a biological role in thermotolerance remain unknown. To investigate this, we isolated SR45 proteins from heat-stressed seedlings of overexpression lines and detected phosphorylated forms using SDS-PAGE-Phos-tag. This method differentiates phosphorylated protein species compared to their non-phosphorylated counterparts by the interaction of protein phosphogroups with a polyacrylamide-bound dinuclear

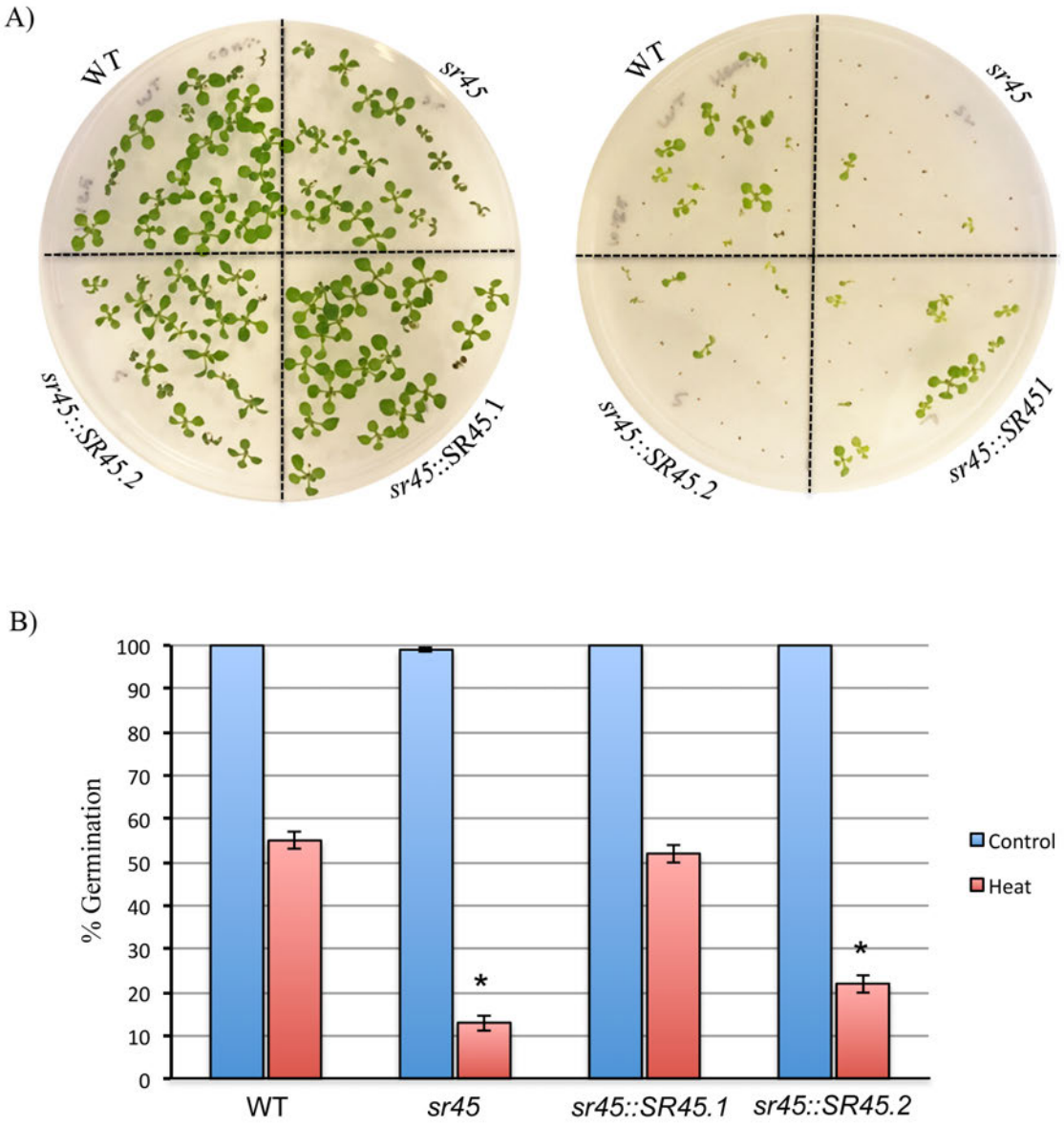


Figure 28: Basal heat tolerance of WT, *sr45* and *sr45* complemented with either the long (*sr45*::SR45.1) or short (*sr45*::SR45.2) isoform. (A) Surface sterilized seeds were stratified for three days at 4°C, kept at 24°C as a control or subjected to a 50°C heat shock for 60 min. Control (left) and treated (right) seeds of all lines were then germinated and grown under the same conditions on M.S medium. (B) The effect of heat stress on germination of seeds from WT, *sr45* and complemented lines. Percent germination was calculated after 10 days. Statistically significant differences (*t*-test, $P<0.05$) are indicated by asterisks. *sr45* is highly sensitive to heat stress than WT. Only the SR45.1 rescued the germination to WT level under heat stress. This data represents one of three biological replicates experiments.

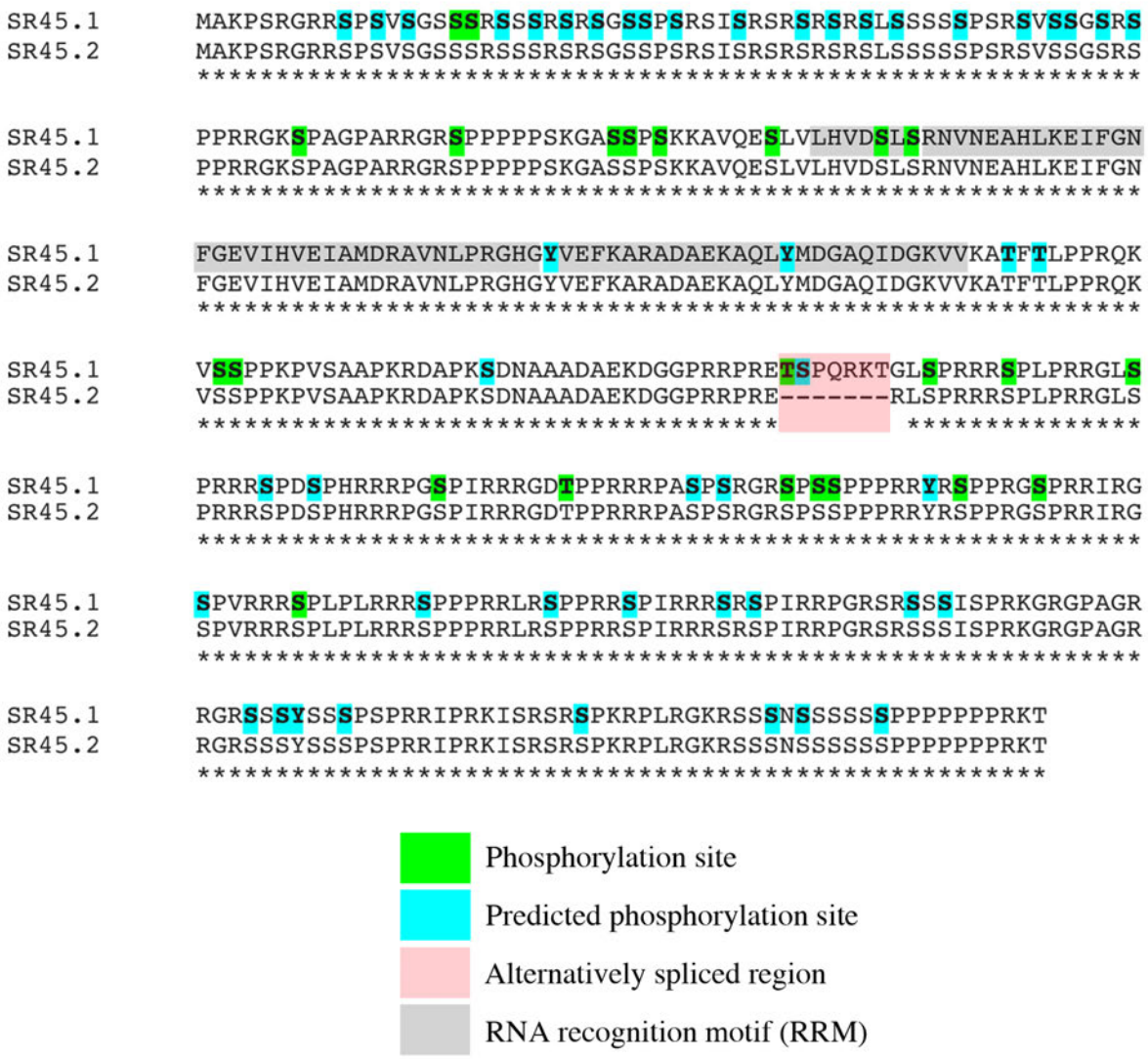


Figure 29: Predicted and experimentally verified phosphorylation sites in SR45 isoforms. Protein sequence of both isoforms (SR45.1 and SR45.2) was extracted from UniProt (Universal Protein Resource) database. Experimentally verified phosphorylation sites (green, Reiland et al., 2009)and predicted phosphorylation sites (blue) were from the *Arabidopsis* Protein Phosphorylation Site Database (PhosPhAt 4.0). Pink color shows the difference in amino acid sequences between two splice isoforms. Protein part highlighted in gray indicates RNA recognition motif.

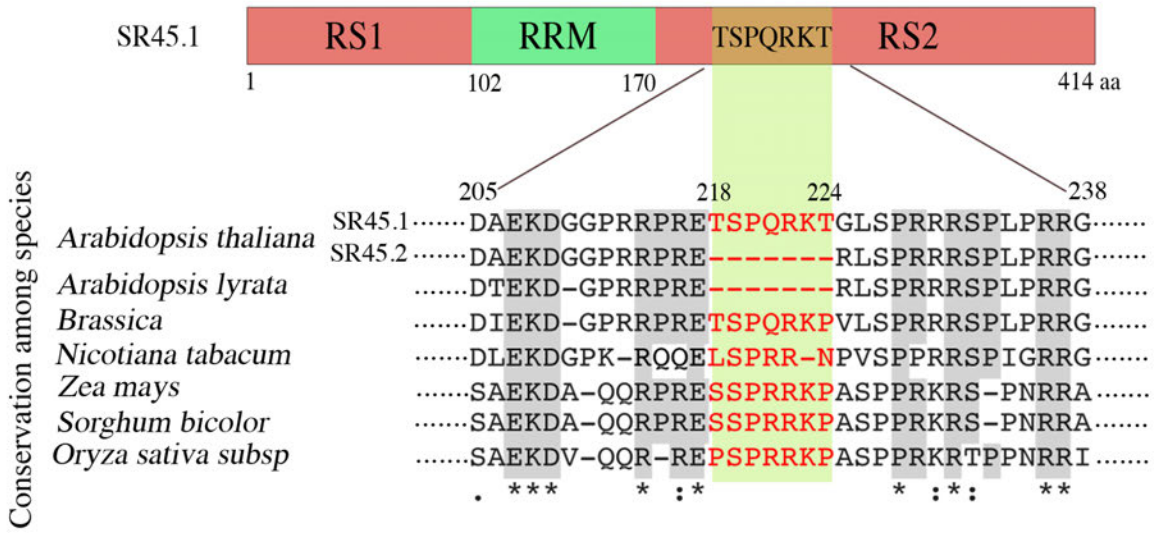


Figure 30: Conservation of alternatively spliced region across species. Top panel: A schematic diagram showing the domain organization of SR45.1. N-terminal arginine/serine-rich (RS1) domain, the middle RNA recognition motif (RRM), and the C-terminal RS2 domain are indicated. Bottom panel: sequences of SR45 homologues from six species were aligned with Clustal Omega. Amino acids 218 to 224 (TSPQRKT) in RS2 domain that are present in SR45.1 but missing in SR45.2 are highlighted in light green. Protein sequences and domain annotation were extracted from UniProt (Universal Protein Resource) database.

Mn^{2+} complex (Mn^{2+}-Phos-tag) (Kinoshita et al., 2009). Both SR45 isoforms exhibited a phosphorylation mobility shift in response to high temperature, and this shift increased with incubation time (Figure 31). Also, we noted that the stability of both SR45 isoforms was reduced under heat stress (Figure 31). However, this phosphorylation detection method was not sensitive enough to distinguish phosphorylation differences between the two protein isoforms, since the only differences between them are two possibly phosphorylated amino acids. Therefore, to investigate potential functional differences between SR45.1 and SR45.2 in response to heat stress, we used previously reported genetically modified SR45.1 lines with point mutations at the SR45.1-specific sites T218 and S219 (Zhang et al., 2014; Zhang and Mount, 2009). Several lines with different point mutations were used: (1) substitution of T218 with non-phosphorylatable alanine (SR45.1T218A); (2) substitution of S219 with non-phosphorylatable alanine (SR45.1S219A); and (3) substitution of both T218 and S219 with non-phosphorylatable alanines (SR45.1 T218A+S219A) (Figure 32A). Figure 32B is a representative picture of the seedling growth of these lines and the SR45.1 overexpression line on MS medium under normal conditions. When seeds were heat-stressed before germination, lines overexpressing SR45.1T218A and SR45.1T218A+S219A exhibited significantly lower germination capacity than lines overexpressing SR45.1 and SR45.1S219A; no difference was observed under normal temperature (Figure 33AB). These results indicate that SR45.1 plays a positive and vital role in thermotolerance, likely mediated by the phosphorylation state of T218 but not S219.

DISCUSSION

Even though AS events are highly prevalent in plant transcriptomes (Filichkin et al., 2010; Marquez et al., 2012; Wang and Brendel, 2006), the regulatory mechanisms behind them are yet to be elucidated. To address the role of an SR protein (SR45) in AS regulation in *Arabidopsis*,

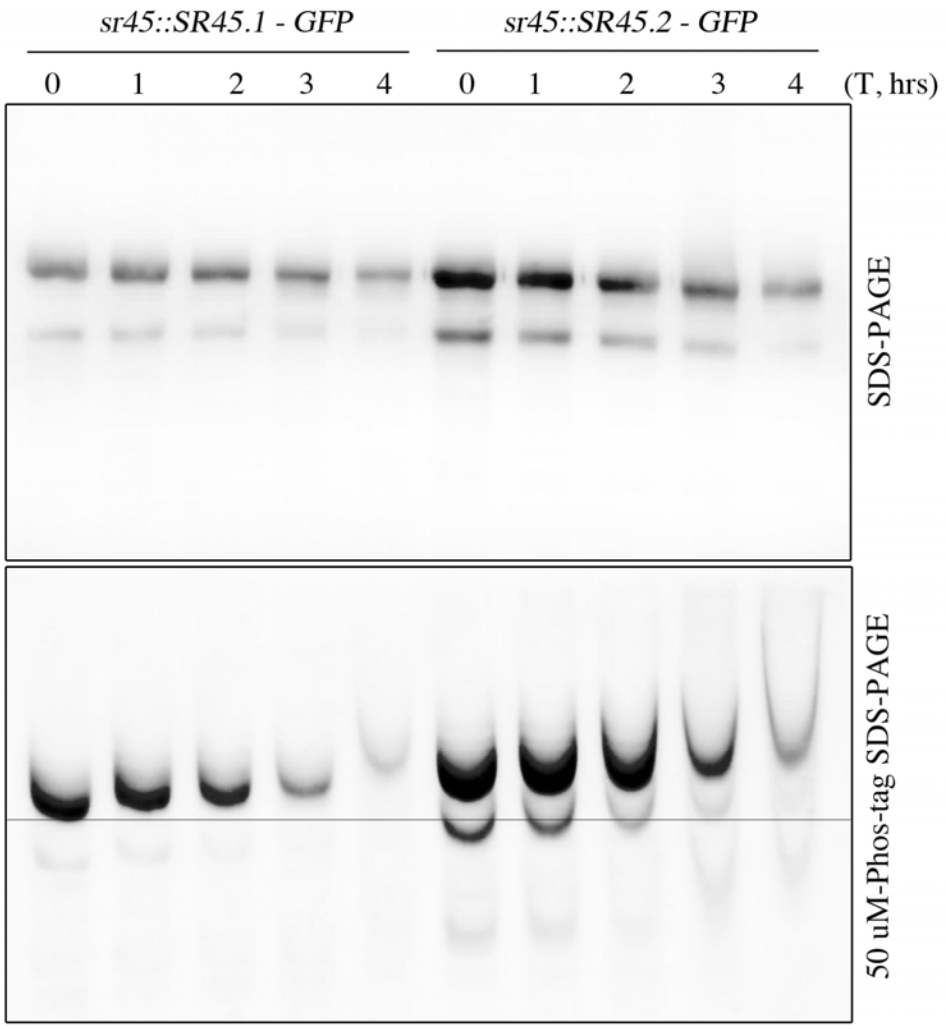


Figure 31: Phosphorylation status of SR45 splice isoforms in response to heat shock. Complemented lines expressing either long or short isoform were grown for seven days, seedlings were heat-treated (38^{o}C) and tissue was collected at different time points (0, 1, 2, 3, and 4 hrs.). Nuclear protein from these seedlings was the used for immunodetection using SR45 antibody. Top, immunoblot of nuclear proteins separated on a denaturing gel (SDS-PAGE). Bottom, immunoblot of the same samples separated on a Phos-tag gel (50mM Phos-tag™ SDS-PAGE). T, heat treatment time points in hours. The thin horizontal line on the second gel is to reflect different mobility of proteins.

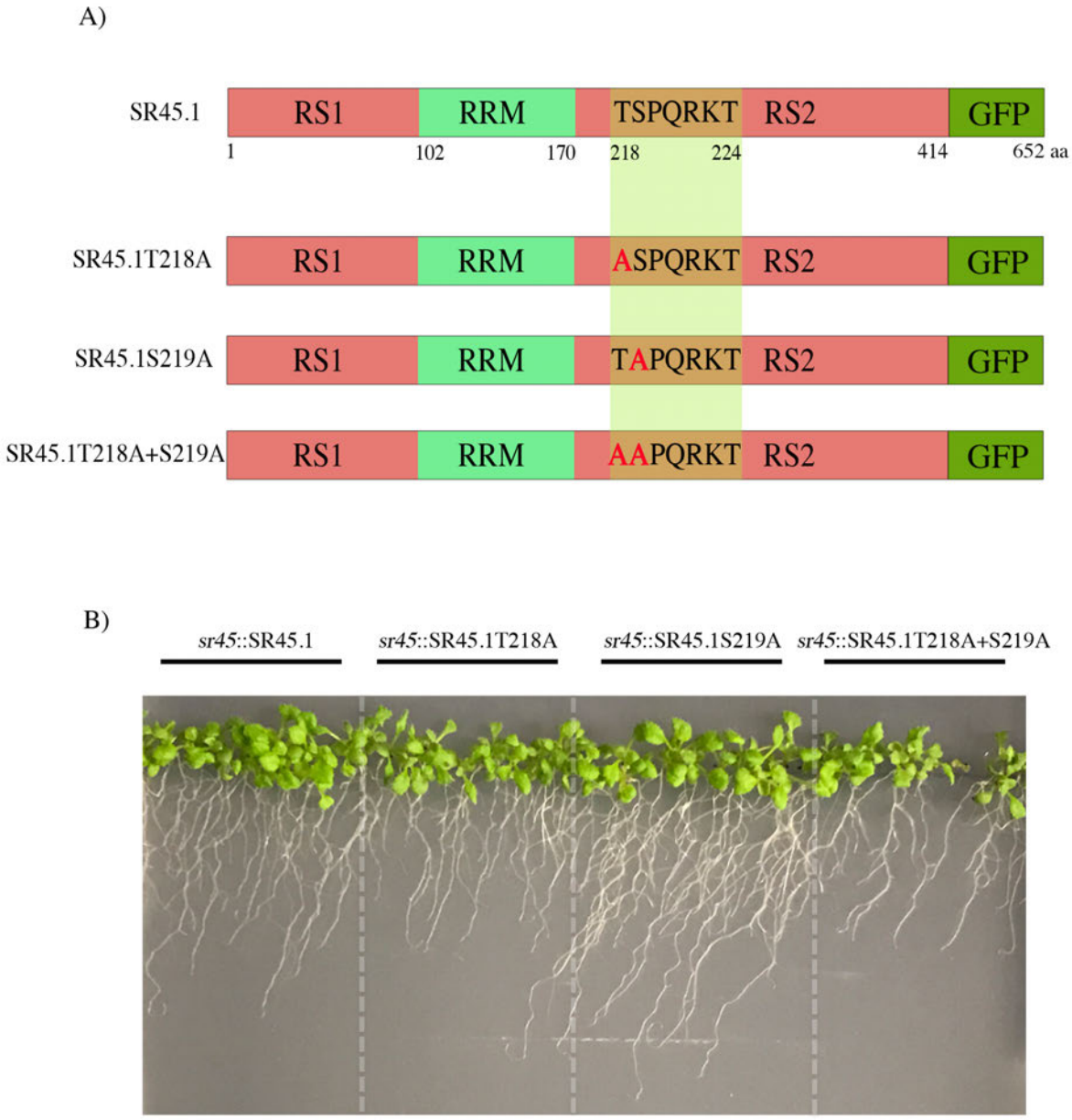


Figure 32: Substitutions of threonine 218 and serine 219 residues of SR45.1 with a non-phosphorylatable amino acid, alanine. (A) A schematic diagram showing the domain organization of SR45.1 as described before (Fig.28) tagged with green fluorescent protein (GFP). Amino acids substitutions in SR45 long isoform: SR45.1T218A, SR45.1S219A, and SR45.1T218A+SR45.1S219A are highlighted in red. (B) Seedling growth of transgenic plants overexpressing SR45.1, SR45.1T218A, SR45.1S219A, and SR45.1T218A+SR45.1S219A in *sr45* mutant background. Seedlings were grown under the same conditions on M.S. medium for two weeks.

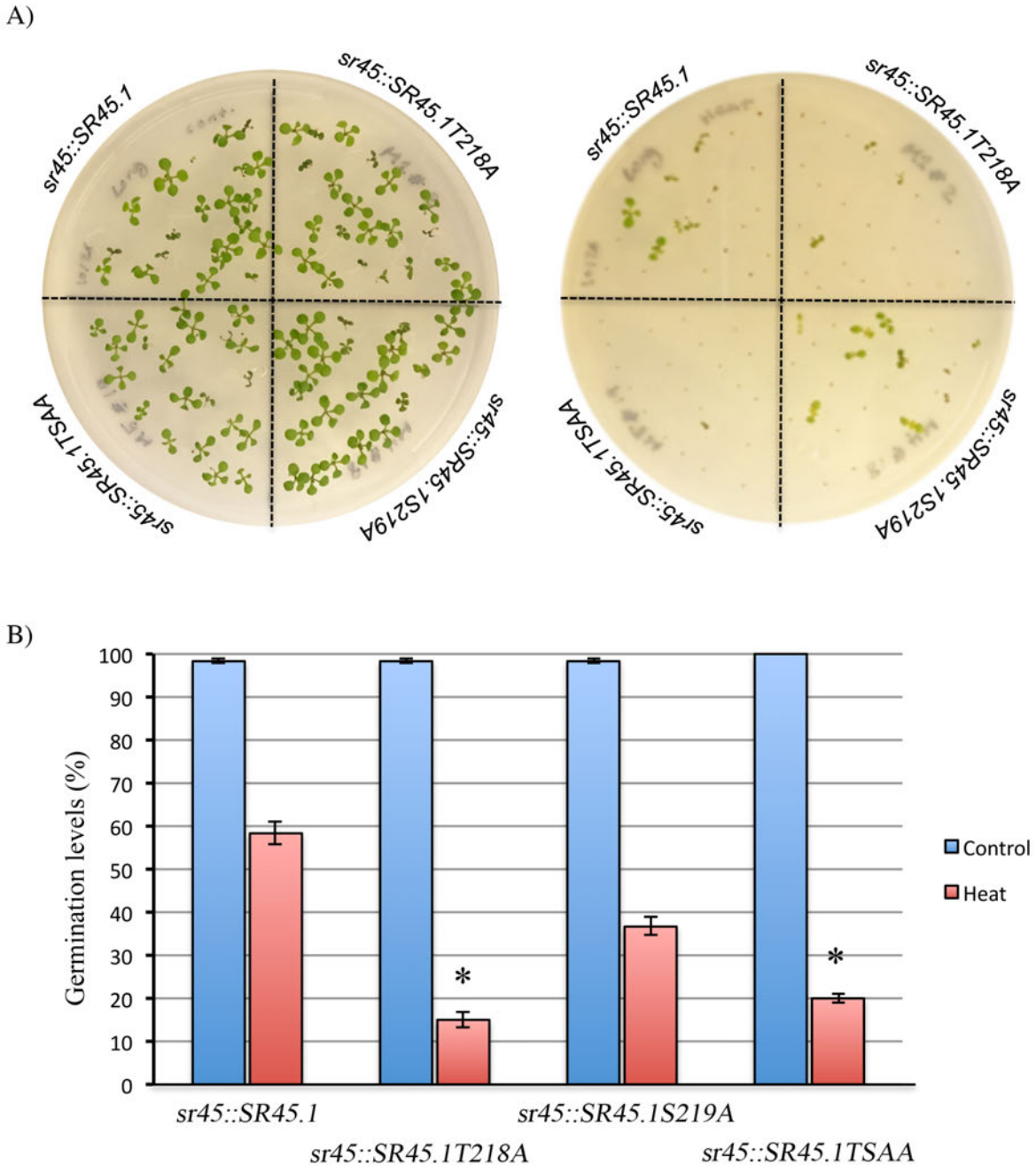


Figure 33: Basal heat tolerance of transgenic plants overexpressing different substitution mutations in SR45.1 (long isoform) (A) Seeds were surface sterilized, stratified at 4^{o}C for three days at 24^{o}C as a control (left) or subjected to a 50°C heat shock for 60 min (right). Following treatment, seeds of all lines were germinated and grown under the same conditions on M.S. medium. (B) Germination percentage was quantified after 10 days. Asterisks indicate statistically significant differences (*t-test*, P<0.05). SR45.1T218A and SR45.1T218A+ SR45.1S219A showed increased sensitivity to heat stress as compared to SR45.1 and SR45.1S219A.

we compared the transcriptomes of the SR45 mutant (*sr45)* and WT using high-throughput RNA-Seq. Previous studies with *Arabidopsis* SR45 have indicated that it is a spliceosomal protein that interacts with U1-70K, AFC2 kinase, U2AF35, SC35-like SR (SR33), RSZ21, SR34 and SR34a, and plays multiple roles in plant development (Ali et al., 2007; Day et al., 2012; Golovkin and Reddy, 1998, 1999; Xing et al., 2015; Zhang et al., 2014). The single-base resolution and quantitative nature of RNA-Seq allowed us to identify DE and DS genes. Our RNA-Seq analysis identified 1,345 DE genes and 927 DS genes as SR45-regulated with an FDR of 5%. Given the known molecular functions of SR45, the DE genes are most likely indirect targets of SR45 while the DS genes are potentially its direct and indirect targets (see below). Analysis of DS events indicates that SR45 most likely influences the A3′ss within the CDS. However, in some transcripts, changes in A5′ss within CDS, IR within 3′ UTR, and A5′ss within 3′ UTR were also observed, suggesting its effect on other AS events also. A prior study has indicated that SR45 binds *in vitro* to an intronic region near 5′ss and recruits spliceosomal proteins U1-70K and U2AF^{35} to both A5′ss and A3′ss, respectively (Day et al., 2012).

However, SR45 has also been shown to play a role in regulating the splicing of several other splicing factors (Ali et al., 2007). Therefore, some of these DS events might be indirect effects of SR45, mediated by other splicing regulators directly affected by SR45. Nonetheless, the differentiation of DS from DE genes sets up a framework for further dissecting the pathways connecting SR45-regulated AS events to its biological functions.

The high changes in gene expression due to loss of SR45 may result from altering splicing patterns of certain transcription factors. In our previous work, we have shown that SR45 binds and regulates splicing of the heat stress transcription factor *HSFA2* (AT2G26150); loss of SR45 leads to altered splicing of *HSFA2* and affects the expression level of at least eight downstream

target genes (Albaqami, 2013). In agreement with this result, *HSFA2* (AT2G26150) was one of the DS genes in *sr45*, and validated by RT-PCR (Figure 20).

The altered patterns of gene expression in *sr45* might also be due to the fact that *Arabidopsis* SR45 is involved in RNA-mediated DNA methylation and gene silencing (Ausin et al., 2012). Furthermore, SR45 is involved in regulating the splicing and expression of non-coding RNAs. It is known from previous studies that pre-miRNA AS regulatory mechanisms control miRNA biogenesis (Hirsch et al., 2006; Yan et al., 2012). Several miRNA genes in plants are located in the intronic regions of protein-coding genes, and their expression and processing co-occur with that of their host genes (Brown et al., 2008; Yan et al., 2012; Yang et al., 2012). Our RNA-Seq data revealed that expression levels of six microRNA genes were altered in the mutant, and RT-PCR analysis of mir156C validated that its splicing is altered in *sr45* (Figure 18A). Significantly, at least one target gene for each of the four micro RNAs was also differentially expressed between WT and *sr45*. These results indicate that SR45 likely plays a direct or indirect role in transcription and/or microRNA processing. In addition to these findings, the expression of *Arabidopsis* intronic mir400 has been shown to be down-regulated in a splicing regulator (SR30) mutant (Yan et al., 2012). Here, we suggest that splicing factors including SR45 likely regulate the processing of pre-miRNA and intronic miRNAs in plants. Further investigation of the interactions between those factors and microRNAs will be very helpful in increasing our understanding of AS regulation and miRNA biogenesis in *Arabidopsis*.

Genes identified as DE or DS in the RNA-Seq data were analyzed for enrichment of GO biological process terms. Enriched terms included “response to abscisic acid stimulus,” “response to salt stress,” “response to heat stress,” “carbohydrate metabolic process,” and “regulation of root development”; supporting the validity of this analysis as these terms are

consistent with the previously reported functions of SR45 in ABA/sugar signaling, abiotic stress responses, and root development (Albaqami, 2013; Ali et al., 2007; Carvalho et al., 2010; Zhang and Mount, 2009). While GO terms involved in flowering time were not enriched among misregulated genes, the expression of *FLC*, a flowering time control gene, which is also a known target of SR45, was dramatically increased in the SR45 mutant, consistent with a previous report (Ali et al., 2007). The lack of observed enrichment for any GO terms directly related to flowering suggests that flowering control might is likely due to its effect on one or more key genes rather than a large number of genes involved in flowering. In total, several dozen GO terms were enriched, indicating that SR45 functions in multiple biological processes. However, most of those GO terms were related to hormonal and stress responses (Table 10), suggesting a possible intrinsic relationship among them. It is thus tempting to speculate that SR45 might be a central regulator in the signaling hierarchy of hormonal and stress response pathways. That the biological process "regulation of transcription, DNA-dependent" and the molecular function "sequence-specific DNA binding transcription factor activity" were among the most enriched GO terms might even add another layer of detail to this speculation. That is, it is through AS or altered transcription of transcription factors that SR45 reprograms those pathways in response to a variety of internal and external stimuli.

As a case study for the potential role of SR45 in stress responses, we investigated SR45's function in thermotolerance. It is noteworthy that SR45 is alternatively spliced, generating two isoforms (SR45.1 and SR45.2), and these two isoforms have been shown to function differently in *Arabidopsis* development as well as in response to abiotic stresses (Albaqami, 2013; Zhang and Mount, 2009). Consistent with our previous work, we found that SR45.1, but not SR45.2, was a positive regulator of thermotolerance (Figure 28) at different developmental stages in

Arabidopsis (Albaqami, 2013). In addition, the expression of *SR45.1* under heat stress was significantly higher than that of *SR45.2* (Figure 27).

It is known from our previous studies that SR45.1 likely mediates thermotolerance by regulating the AS of certain transcription factors involved in heat stress response (Albaqami, 2013). For example, *Arabidopsis HSFA2*, a key transcription factor in heat stress response, is alternately spliced under heat stress to generate three isoforms, two of which are known to function positively to enhance thermotolerance (Liu et al., 2013). Intriguingly, it has been shown that splicing patterns and expression levels of *HSFA2* were altered in *sr45* at both normal and high temperatures, and SR45.1 binds to an alternatively spliced *HSFA2*-intron *in vitro* (Albaqami, 2013). Furthermore, expression of several downstream targets of HSFA2 are regulated by SR45.1 (Albaqami, 2013). These observations and the RNA-Seq analysis results highlight the role of SR45.1 in thermotolerance.

In an attempt to understand how SR45.1 confers thermotolerance while SR45.2 does not, we investigated the differences between these two isoforms. SR45.1 is distinguished from SR45.2 by eight amino acids (TSPQRKTG). *Arabidopsis* phosphoproteome analysis and previous studies revealed that two of these additional residues, T218 and S219, are potential phosphorylation sites (Reiland et al., 2009; Zhang and Mount, 2009 and PhosPhAt). It is also noteworthy to mention that other phosphorylation sites are predicted in both isoforms, and a large-scale *Arabidopsis* phosphoproteome assay has revealed that SR45 is extensively phosphorylated at different residues (Figure 29) (Reiland et al., 2009 and PhosPhAt). Similarly, when we examined the phosphorylation status of both isoforms under heat stress using Phos-tag™ SDS-PAGE, we found that both of them were phosphorylated; however, it was not possible to resolve any phosphorylation differences between isoforms using this method. Thus,

given the different heat stress responses of both overexpression lines, the overall change of phosphorylation of both isoforms under heat stress added another layer of regulation concerning SR45's role in heat shock response.

It is known that the phosphorylation/dephosphorylation cycle of SR proteins plays a fundamental role in splice site selection and spliceosome assembly (Zhou and Fu, 2013). Furthermore, SR proteins are known to function in either a cooperative or antagonistic fashion (Bradley et al., 2015), but it is not known if this behavior is specific to certain splice variants. Therefore, we cannot exclude the possibility that the phosphorylation of SR45.2 under heat stress interferes with *Arabidopsis* heat stress tolerance through antagonistic heat-induced alternative splicing events. Future investigation is needed to determine whether SR45.2 might be a negative regulator of thermotolerance in *Arabidopsis*.

Nevertheless, the different responses of these splice variants of SR45 to heat stress, and the conservation of phosphorylation sites within SR45.1's unique amino acids among different organisms (Figure 30), made us wonder if phosphorylation of these residues influences the specific role of SR45.1 in thermotolerance. Therefore, we used genetically modified *SR45.1* with alanine substitution mutations at potential phosphorylation sites T218 and S219, which are located in the alternatively spliced region. Heat stress responses of these mutants confirmed that T218 is likely phosphorylated and that this phosphorylation, but not at S219, is required for thermotolerance in *Arabidopsis* (Figures 32 and 33). Further studies using the long form with "phosphomimetic" mutation at T218 and/or phophoproteomics analysis of SR45 under heat stress are needed to further confirm that phosphorylation of T218 is critical for its role in thermotolerance. From our findings, we suggest that the unique predicted phosphorylation site at T218 of SR45.1 contributes to the different capabilities of these two isoforms in regulating

thermotolerance. Although mutation of S219 alone did not result in a heat stress response phenotype, it may also be involved in different biological roles. Taken together, these findings and previous phosphorylation assays strongly suggest the likelihood that phosphorylation of SR45.1 mediates its biological role in response to heat shock in *Arabidopsis*. The phosphorylation status of the long isoform may influence its function in several different ways. First, it may affect the localization of SR45. It has been shown that subnuclear localization of SR45 is altered by heat stress (Ali et al., 2003). Second, it may alter its ability to interact with other spliceosomal proteins since SR45 is known to interact with U1-70K, AFC2 kinase, U2AF35, SC35-like SR (SR33), RSZ21, SR34 and SR34a. Finally, the interaction of SR45.1 with *cis*-elements in pre-mRNAs may be regulated by phosphorylation. An important are for further investigation is to identify RNA targets of SR45 splice isoform. Further investigations will add more insight into how protein phosphorylation of SR45 influences its splicing activities.

Conclusion

Arabidopsis SR45, one of the SR-like proteins, has been well-characterized as an essential pre-mRNA splicing regulator that functions in plant development and environmental responses. However, how SR45's biological roles are mediated by SR45-regulated alternative splicing events remains to be discovered. Therefore, to address this question and to gain insights into the mode of action of SR45, we compared the transcriptomes of the *sr45* mutant and WT using high throughput sequencing. This analysis has allowed us to identify thousands of DE and DS genes in the mutant. Furthermore, the distribution of AS events along the gene bodies of DS genes suggests that SR45 likely regulates the selection of 3' ss within the CDS region. The top over-represented GO biological process terms in DE and DS genes related to hormonal signaling and

response to abiotic stresses. Remarkably, supporting the validity of these GO enrichment results, the enriched terms include previously characterized functions of SR45.

GO terms suggested that SR45 plays a role in heat stress response; hence, we used that as a case study to investigate the role of SR45 in abiotic response. However, SR45 itself is alternatively spliced to generate two functional proteins, SR45.1 and SR45.2, in which SR45.1 has an additional seven amino acids. We included both isoforms in my investigation to gain insight into how alternative splicing regulates SR45 functions. Heat shock responses of WT, *sr45*, and complemented lines of both isoforms indicated that SR45 has a vital role in thermotolerance in *Arabidopsis*. However, the heat stress responses of plants expressing specific isoforms revealed that only SR45.1, not SR45.2, has a positive role in thermotolerance; this is likely through modulating AS and expression of several heat stress responsive genes. Furthermore, we found that post-translation phosphorylation modifications likely differentiate the function of SR45.1 from SR45.2, where the SR45.1-specific phosphorylation residue T218 impacts its biological role in *Arabidopsis* heat shock tolerance.

Finally, this research established the direct and indirect targets of *Arabidopsis* SR45; however, further investigation to identify direct targets of SR45 isoforms under heat shock will be needed to gain more insights into how SR45.1 confers thermotolerance. Global gene expression analysis with complemented lines under normal and high temperature will be a very suitable first step in this investigation. Mass spectrometry analysis of immunopurified SR45 isoforms will reveal the phosphorylation differences between them under heat stress. It will additionally be important to investigate how the mutation of a single amino acid in SR45.1, T218, specifically affects the expression and AS of heat stress responsive genes that were shown to be misregulated by SR45 mutation.

CHAPTER II

Development of a Plant-derived *In Vitro* Pre-mRNA Splicing System

SUMMARY

Precursor messenger RNA (pre-mRNA) splicing is an essential post-transcriptional process common to all eukaryotes. Evidence indicates that plants may have unique pre-mRNA splicing regulatory mechanisms; however, these mechanisms are poorly understood and have not received attention equivalent to those of animals and yeast. *In vitro* splicing systems using nuclear or cytoplasmic extracts from mammalian cells, yeast, and *Drosophila* have provided a wealth of mechanistic insights into eukaryotic pre-mRNA splicing. A corresponding plant-derived *in vitro* splicing system has long been awaited; therefore, we present here an effort toward developing such a system from plants. We show that nuclear extract (NE) derived from dark-grown (etiolated) *Arabidopsis* seedlings is capable of converting a truncated *LIGHT-HARVESTING CHLOROPHYLL B-BINDING PROTEIN 3* (*LHCB3*) pre-mRNA substrate with a single intron into the expected size of mRNA. Based on several lines of evidence, we suggest that this is an accurate *in vitro* splicing reaction. Supporting evidence include: i) generation of an RNA product that correspond to the size of expected mRNA, ii) generation of RNA species that migrate within the size range expected for splicing intermediates, iii) indications from a junction-mapping assay using S1 nuclease that the two exons are linked together, iv) remarkable similarities between plant and non-plant *in vitro* splicing assay reaction conditions, such as requirements for ATP and Mg^{+2}, and v) finally, more importantly, mutations in conserved donor and acceptor sites abolished the production of the putative spliced product. Unlike mammalian *in vitro* splicing assays, the optimal incubation temperature for splicing with plant extract was

lower, within the optimal growth temperature range of *Arabidopsis* seedlings, 24^{o}C. Collectively, our results suggest that *Arabidopsis* NE is capable of splicing pre-mRNA substrate and that the *in vitro* reaction conditions are similar to those found with non-plant extracts. This is the first step toward developing a plant-derived *in vitro* pre-mRNA splicing assay. Further confirmation of these results with additional approaches and optimization of this assay would lead to development of a robust *in vitro* splicing assay and opens new avenues to investigate spliceosome assembly and composition, splicing regulatory mechanisms specific to plants, and thereby enhances the overall understanding of post-transcriptional gene regulatory mechanisms in eukaryotes.

INTRODUCTION

Shortly after the discovery of introns in 1977, different groups developed the mammalian cell-free system (*in vitro*), using nuclear or cytoplasmic extracts, which were competent for pre-mRNA splicing (Hernandez and Keller, 1983; Krainer et al., 1984; Padgett et al., 1984). Subsequently, the preparation of efficient splicing extracts and the *in vitro* splicing assay have been adapted to other organisms such as budding yeast (*Saccharomyces cerevisiae*) and fruit flies (*Drosophila melanogaster*) (Lin et al., 1985; Rio, 1988). The development of mammalian, yeast, and *Drosophila in vitro* systems to study pre-mRNA splicing has provided essential insights into spliceosome assembly, its composition and splicing mechanisms in non-plant systems. The characterization of the splicing two-step trans-esterification reaction, pre-mRNA splicing intermediates, and the formation of final mature mRNA and lariat structure of intron have all been revealed by *in vitro* splicing studies (Padgett et al., 1984; Ruskin et al., 1984). Furthermore, *in vitro* splicing combined with immunodepletion has long been used to determine the roles of spliceosomal components, such as small nuclear ribonucleic proteins (snRNP) (Black

et al., 1985; Krainer and Maniatis, 1985). Several other splicing regulators were characterized by their ability to promote the *in vitro* splicing assay (Krainer et al., 1990; Screaton et al., 1995). Furthermore, the formation of spliceosomal complexes and their stepwise assembly pathway have been deduced *in vitro* by native (nondenaturing) agarose/polyacrylamyde gel electrophoresis (Das and Reed, 1999; Konarska and Sharp, 1986; Wahl et al., 2009). Additionally, many *in vitro* biochemical splicing studies have allowed purification of spliceosomes and provided a wealth of information on the spliceosome's composition, its structures, and the structure conformational dynamics of spliceosomal complexes (Bertram et al., 2017; Galej et al., 2016; Matera and Wang, 2014; Papasaikas and Valcarcel, 2016; Wahl et al., 2009; Yan et al., 2015; Yan et al., 2017a). The identification of additional splicing regulatory *cis*-elements, such as splicing enhancers or silencers and their cognate factors, has also been expedited by studies using *in vitro* biochemical assays (Ge and Manley, 1990; Liu et al., 1998; Mayeda and Krainer, 1992; Reed and Maniatis, 1986; Schaal and Maniatis, 1999; Tian and Maniatis, 1992). The *in vitro* splicing assay has also been a valuable technique for dissecting abnormal splicing events which cause human genetic disease, and in developing new therapeutic approaches for human disease (Hua et al., 2008; Krainer et al., 1984). Finally, the *in vitro* splicing assay has been used to evaluate molecules with splicing inhibitory functions, such as spliceostatin A (Kaida et al., 2007). These are just several examples of how *in vitro* splicing assays contribute important information regarding this fundamental gene expression regulatory mechanism. The *in vitro* splicing systems have been, and will continue to be, indispensable tools for studying the mechanism of splicing.

As in animals, pre-mRNAs from a majority (over 80%) of plant genes contain non-coding sequences and are processed to generate mature mRNAs via a splicing mechanism where

noncoding sequences (introns) are removed and coding sequences (exons) are ligated together (Reddy, 2007). Pre-mRNA splicing is a post-transcriptional regulatory mechanism that is not only fundamental for controlling gene expression. Recent studies indicate the developmental and environmental cues can reprogram gene expression in plants by regulating post-transcriptional processes, especially pre-mRNA splicing (Lee and Kang, 2016; Lorkovic, 2009; Mazzucotelli et al., 2008; Palusa et al., 2007). However, the molecular mechanism of basic splicing in plants remains completely unknown. Furthermore, the composition of the plant spliceosome and its assembly intermediates are currently undefined (Barta et al., 2012; Reddy et al., 2013). Thus, the study of pre-mRNA splicing in plants requires innovative approaches, which will greatly empower this field of research.

In spite of the absence of an *in vitro* plant-splicing system, efforts have been made to identify distinct plant splicing-related mechanisms. Bioinformatics analyses using sequence similarity have identified the core components of the plant spliceosome, including five snRNAs and several orthologs of known spliceosomal proteins (Koncz et al., 2012; Lorković et al., 2000; Reddy et al., 2013; Ru et al., 2008; Wang and Brendel, 2004). Likewise, the highly conserved sequences at the 5' splice site (5'ss), 3' splice site (3'ss), polypyrimidine tract and branch point sequence (BPS) are similar between plants and animals (Reddy, 2007). These similarities are strong signs for comparable basic mechanisms of intron processing across eukaryotic systems, but there are also numerous indications of plant-specific splicing regulatory mechanisms. For example, animal introns cannot be processed in plant systems (Barta et al., 1986), the average sizes of plant introns are shorter than their mammalian counterparts (Reddy et al., 2013), and analysis for proteins similar to mammalian spliceosomal proteins indicates that there is almost twice the number of plant splicing factors compared to human splicing factors (Koncz et al., 2012; Reddy

et al., 2013). Furthermore, comparative analysis of alternative splicing (AS) events between plants and animals has revealed that intron retention is the predominant mode of AS in plants (Ner-Gaon et al., 2004; Syed et al., 2012), whereas exon skipping is the predominant mode in animals (Kim et al., 2007). Although the pre-mRNA splicing mechanisms in plants are poorly understood, based on the reasons cited above it is likely that the mechanisms of recognition of introns and exons may differ between plant and animals.

The development of *in vitro* systems to study RNA-related mechanisms in plants is limited and challenging (Sugiura, 1997). An *in vitro* pre-mRNA splicing system to uncover plant splicing regulatory mechanisms has long been awaited (Barta et al., 2012; Reddy et al., 2013). Therefore, despite difficulties inherent to plant cells, here we describe our efforts to develop an *in vitro* system for plant pre-mRNA splicing using nuclear extracts (NE) prepared from *Arabidopsis* seedlings. We present a detailed procedure for the preparation of the NE and a subsequent *in vitro* splicing reaction using a plant pre-mRNA substrate containing a single intron from *LIGHT-HARVESTING CHLOROPHYLL B-BINDING PROTEIN 3* (*LHCB3*). We show that plant NE is capable of converting *LHCB3* pre-mRNA substrate to the size of expected mRNA. Based on several lines of experimental evidence, we show that *Arabidopsis* NE can process a plant pre-mRNA to produce a putative spliced product. These evidences include i) generation of an RNA product that correspond to the size of expected mRNA, ii) generation of RNA species that that have sizes of predicted splicing intermediate, iii) indication of exon-exon junction using S1 mapping assay, iv) remarkable similarities between plat and non-plant *in vitro* splicing assay reaction conditions, such as requirements for ATP and Mg^{+2}, and v) finally, changing conserved 5' and 3' splice site sequences affected the *in vitro* generation of putative splicing product as well as the putative splicing intermediates. Together, our results suggest that *Arabidopsis* NE is

capable of splicing pre-mRNA substrate and that the *in vitro* reaction conditions are similar to those found with non-plant extracts. This is the first step toward establishing a plant-derived *in vitro* pre-mRNA splicing assay. Further confirmation of these results with additional approaches and optimization of this assay would lead to development a robust *in vitro* splicing and opens new avenues to investigate spliceosome assembly and composition, splicing regulatory mechanisms specific to plants, and thereby enhances the overall understanding of post-transcriptional gene regulatory mechanisms in eukaryotes.

MATERIAL AND METHODS

Arabidopsis Nuclear Extract Preparation

The nuclear extract preparation method presented here is a modification of protocols described previously (Folta and Kaufman, 2006; Kataoka and Dreyfuss, 2008; Xing et al., 2015). For plant material preparation, seeds (50 mg) of *Arabidopsis thaliana* ecotype Columbia-0 (Col-0) were surface-sterilized with 70% ethanol followed by 15% bleach and stratified for 3 days at 4 °C (to break dormancy). Then, seeds were placed into 100 mL of Murashige and Skoog (MS) medium (1x MS basal salt, 1 mL/L MS vitamin solution, and 1% sucrose, pH 5.7) in a 250 mL flask and moved to a growth chamber. Seedlings were grown in dark at 2°C for 4 days, with flasks on a shaker at 150 rpm. Four-day-old seedlings were harvested, rinsed three times with Nanopure water, and excessive water was removed using a few layers of Kimwipes. Afterwards, the seedlings were weighed, directly frozen in liquid N_2, and stored at -80°C.

For nuclear protein preparation, 5 g of etiolated seedlings were ground into a fine powder in liquid nitrogen. Subsequently, the sample was homogenized in 25 mL of Honda buffer (1.25% Ficoll 400, 2.5% Dextran T40, 0.44M sucrose, 10mM $MgCl_2$, 0.5% Triton X-100, 20mM HEPES KOH, pH 7.4, 5 mM DTT, 1 mM PMSF, and 1% protease inhibitor cocktail [Sigma-

Aldrich, St. Louis, MO; catalog number: P9599]) for 15 min on ice, with gentle mixing every minute. The homogenate was filtered through two layers of Miracloth into a 50 mL Corex tube. Then the residue left behind on the Miracloth was washed with 25ml ice-cold Honda buffer, and then the filtration step was repeated and collected into the same Corex tube. The filtrate (total 50 ml) was centrifuged at 2000 *g* for 15 min at 4°C. The supernatant was discarded and the pellet resuspended in 15 mL ice-cold Honda buffer, then incubated on ice for 15 min with gentle mixing every minute. It is not recommended to pipet up and down when mixing, as this can disrupt the nuclei; instead, a camel-hair brush can be used to resuspend the nuclei pellet. After complete resuspension, the sample was centrifuged at 1500 *g* for 15 min at 4 °C. This washing step was repeated two times. The pellet was then resuspended in 15 mL ice-cold washing buffer (20 mM HEPES KOH, pH 7.9, 100 mM KCl, 0.2 mM EDTA, 10% (v/v) glycerol, 1mM DTT, 1 mM PMSF, and 1% protease inhibitor cocktail [Sigma-Aldrich, St. Louis, MO; catalog number: P9599]) for 15 min on ice, mixed gently every minute. It is also not recommended to mix by pipetting here. After complete resuspension, the sample was centrifuged at 1500 *g* for 15 min at 4°C. The nuclei pellet was then resuspended in 0.5X ice-cold nuclei swelling buffer (50 mM Tris–HCl (pH 7.9), 10 mM 2-mercaptoethanol, 20% glycerol, 5 mM $MgCl_2$, 0.44 M sucrose, 1 mM PMSF, and 1% protease inhibitor cocktail [Sigma-Aldrich, St. Louis, MO; catalog number: P9599]) and transferred to a 1.5 ml microcentrifuge tube, then incubated at 4°C for 30 min with gentle rocking. The extract was then centrifuged for 30 min at maximum speed (16,000 *g*) at 4°C and the supernatant removed to a new 1.5 mL microcentrifuge tube. Then, the NE was distributed into 50 ml aliquots, flash-frozen in liquid nitrogen and stored at -80^{o}C for *in vitro* splicing assay.

DNA Templates and *In Vitro* Pre-mRNA Synthesis

For *in vitro* pre-mRNA synthesis, DNA templates were amplified by PCR from *Arabidopsis* genomic DNA, using a gene-specific forward primer plus SP6 promoter sequence and reverse primer plus an adaptor sequence. PCR products of the correct size were then gel-purified using the Thermo Scientific GeneJET Gel Extraction Kit (Thermo Fisher Scientific, Waltham, MA; catalog number: K0691). Purified DNA templates were quantified using a NanoDrop 1000, and approximately 0.250 µg of amplified DNA template was used for *in vitro* transcription. To generate mutations at conserved splicing sites, DNA template sequences with desired sequences were synthesized at Integrated DNA Technologies, Inc. (Coralville, IA; https://www.idtdna.com).

Preparation of P^{32}-labelled pre-mRNA substrates was conducted using an *in vitro* transcription system as described previously in (Palusa and Reddy, 2013). The pre-mRNA substrates were internally labeled with 45 µCi of [a-^{32}P] UTP (800 Ci mmol^{-1}, PerkinElmer, Waltham, MA) using SP6 RNA polymerase (Fermentas, Thermo Fisher Scientific, Waltham, MA; www.fermentas.com) in the presence of 500 µM ATP and CTP, 50 µM GTP and UTP, 50 µM cap analog (^{7m}GpppG), and 20 U RNase inhibitor. *In vitro* synthesized P^{32}-labelled pre-mRNAs with the correct size were gel-purified using TNS solution (25 mM Tris-HCl (pH 7.5), 400 mM NaCl, 0.1% SDS) for overnight at room temperature. Radioactive pre-mRNAs were measured using a liquid scintillation counter (Tri-Carb Liquid Scintillation Counter, PerkinElmer, Waltham, MA); 25,000 CPM (~20 fmol) of P^{32}-labelled pre-mRNA substrates was used for *in vitro* splicing reactions, unless otherwise specified (see figure legends).

In Vitro Splicing

Unless otherwise specified (see figure legends), *in vitro* splicing reactions (25 µl) contained 1 mM ATP, 20 mM creatine phosphate (CP), 10 U RNase inhibitor, 1 mM DTT, 72.5 mM KOAc, 25,000 CPM (~20 fmol) P^{32}-labelled pre-mRNA, and 50% NE. Reactions were incubated at 30°C for the times indicated in figure legends. Subsequently, 175 µl of proteinase K master mix (1× proteinase K buffer, 0.25 mg/ml glycogen, 0.25 mg/ml proteinase K, and sterile water) was added, and the solution incubated at 37^{o}C for 20 min. Afterward, RNA was purified by adding an equal volume of phenol:chloroform, precipitating with 2.5 volumes of 100% ice-cold ethanol, and air-drying the isolated pellet for 5 min (Movassat et al., 2014). Finally, RNA samples were dissolved with formamide/EDTA stop dye (formamide with 0.1% bromophenol blue, 0.1% xylene cyanol, and 2 mM EDTA).

Visualization of Splicing Products

Purified RNA from *in vitro* splicing reactions was analyzed by fractionation on a 6% polyacrylamide-urea gel as described previously (Movassat et al., 2014). RNA samples were heated at 95^{o}C for 3 min, and loaded onto a pre-run gel. The gel was run at 200 V for 3 hours. Subsequently, the gel was transferred onto Whatman paper and dried for 2 hours using a Bio-Rad Gel Dryer at 80°C with suction. The gel was then exposed to a phosphor-imaging screen overnight, and imaged using a STORM 840 imager (Molecular Dynamics, GE Healthcare, Little Chalfont, United Kingdom; www.gehealthcare.com).

S1 Nuclease Protection Assay

For the S1 nuclease protection assay, *in vitro* spliced RNA (putative spliced RNA) was gel-purified as described above. To enhance the intensity of P^{32}-labelled RNA, putative spliced product from at least three samples were pooled together. 100 µM of DNA oligo (50 nt) with

sequence complementary to the exon/exon junction was hybridized to the purified RNA in S1 hybridization buffer (80% formamide, 40mM PIPES (pH 6.4), 500 mM NaCl, 1 mM EDTA) for 2 hours at room temperature after denaturing at 95^{o}C for 5 min (Dupuy and Sonenshein, 1998). Subsequently, total hybridized P^{32}-labelled RNA/oligo DNA was digested with S1 nuclease (100U) (Promega, Madison, WI; catalog number: M5761) in 1x S1 nuclease buffer (provided with the enzyme) for 1 hour at 37^{o}C. The undigested P^{32}-labelled RNA was purified by phenol:chloroform extraction and visualized as described above.

RESULTS

Arabidopsis NE Processed *LHCB3* Pre-mRNA to Produce an Expected Size mRNA

Generally, *in vitro* splicing reactions are conducted using NE prepared from HeLa cells; however, there are reports of NEs from other cells, such as *Drosophila* Kc cells, also being used (Hicks et al., 2005). The quality of NE is a vital factor for a successful and efficient i*n vitro* splicing assay (Hicks et al., 2005). Therefore, we aimed first to develop a method for the preparation of NE from *Arabidopsis* etiolated seedlings to use in *in vitro* splicing assays (Figure 34). This method is adapted from different protocols (Folta and Kaufman, 2006; Kataoka and Dreyfuss, 2008; Xing et al., 2015), and described in the Materials and Methods.

To test this *Arabidopsis* NE for *in vitro* splicing activity, several pre-mRNA substrates from *Arabidopsis* genes were designed. Each pre-mRNA contained a single intron flanked with two truncated exons. In addition, most of these substrates were selected based on their expression pattern throughout different developmental stages and under a variety of environmental conditions (Czechowski et al., 2005). These substrates included *UBC9, UBC10, PP2AA3, GAPC2, AP2M, PEX4, TIP41-like family protein, MON1, RHIP1,* and *PTB1* (Figure 35 and 36). In addition to these substrates, we included light harvesting complex B3 (*LHCB3*), which is a

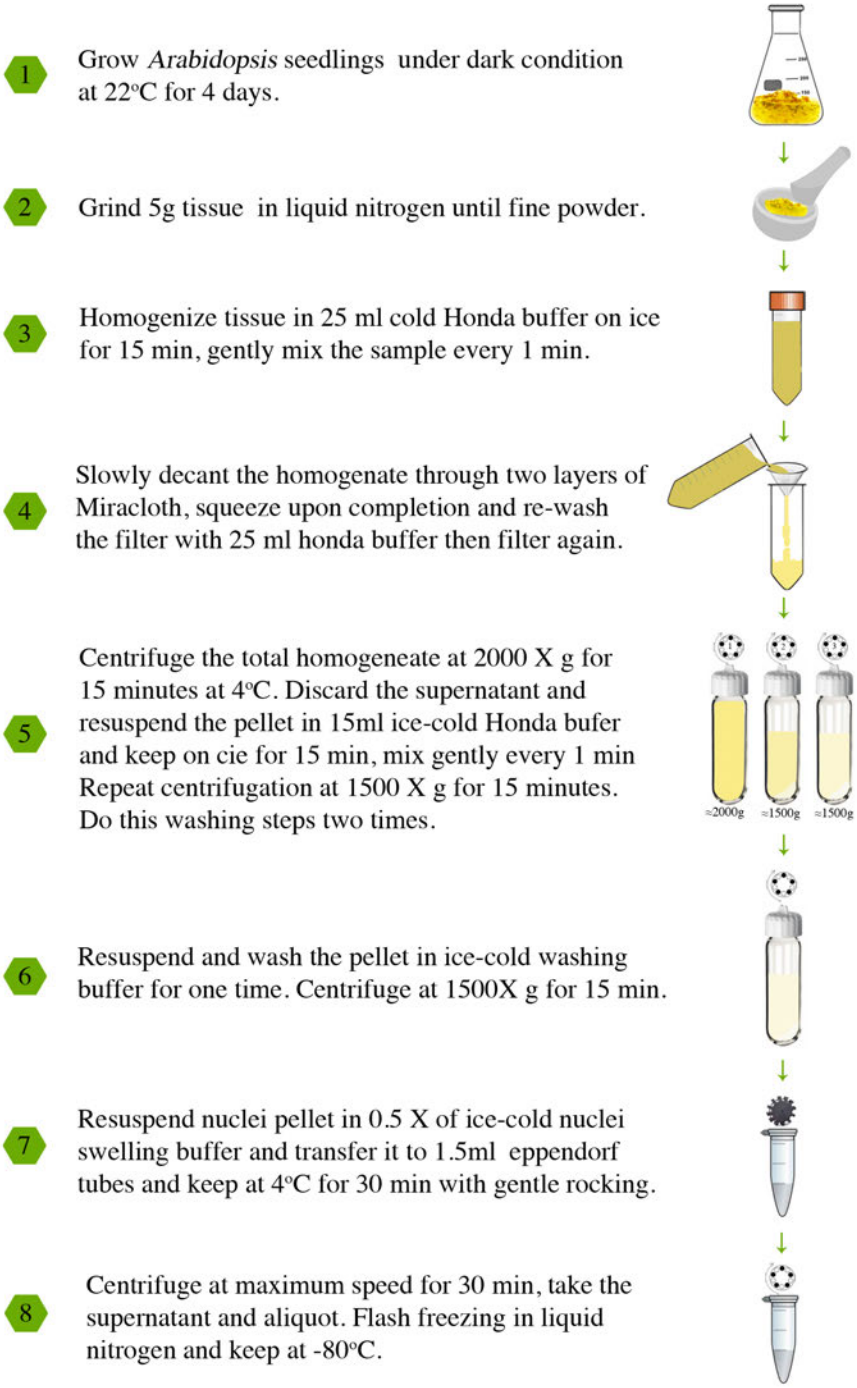


Figure 34: Schematic diagram showing the steps in NE preparation from four-day-old *Arabidopsis thaliana* etiolated seedlings. See material and methods for details.

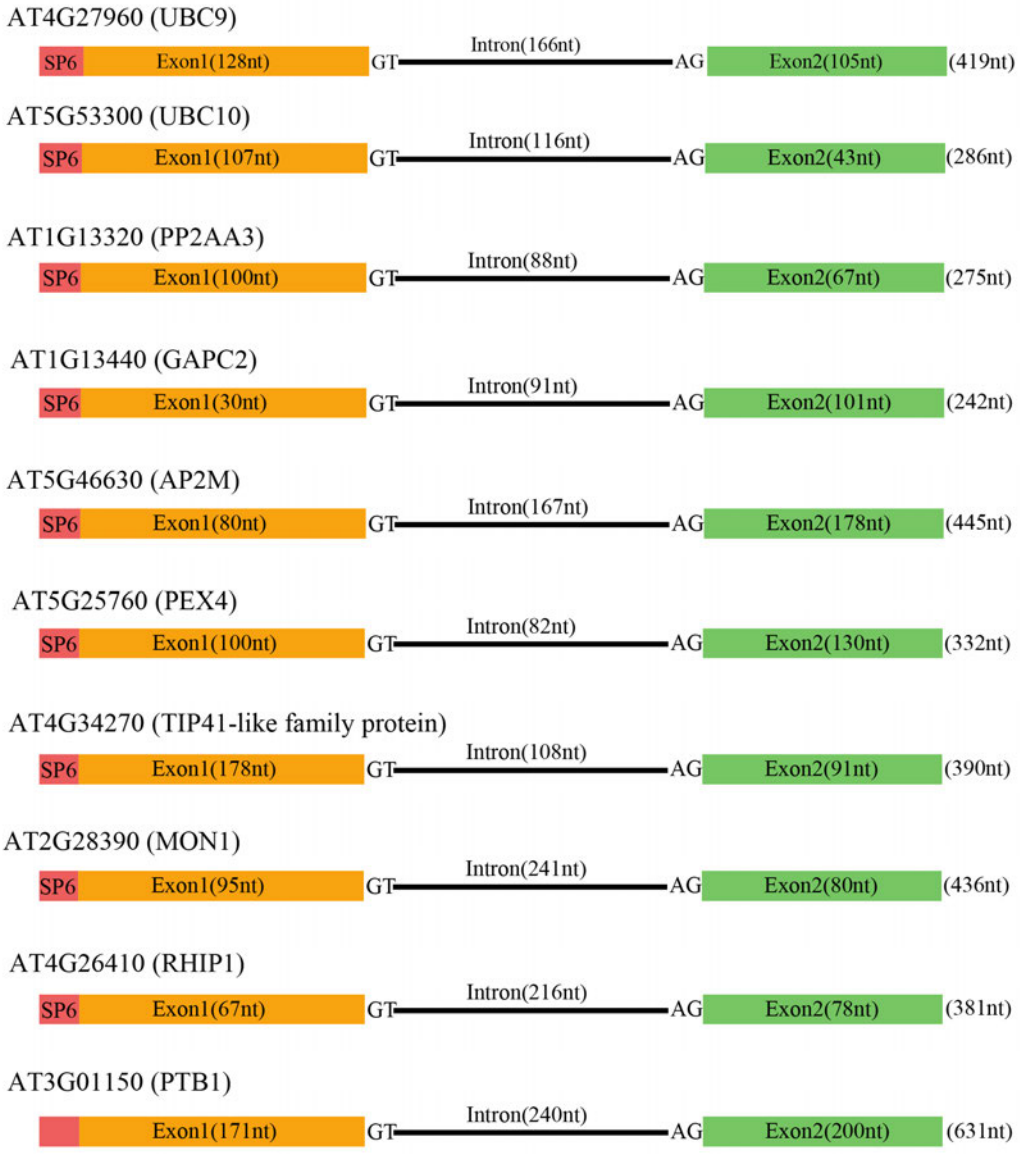


Figure 35: Schematic diagrams of *Arabidopsis* pre-mRNA substrates that were tested in *in vitro* splicing assay with *Arabidopsis* NE. Gene names and ID numbers are displayed at the top of each substrate. In each substrate, intron is shown as a black line, first exon is depicted as an orange box, and second exon as a green box. SP6 promoter is indicated as a red box. Numbers within the boxes or at the top of the introns indicate the length (in nucleotides - nt) of each region. The numbers listed on the right indicate the full length of each substrate. Diagrams are not drawn to scale.

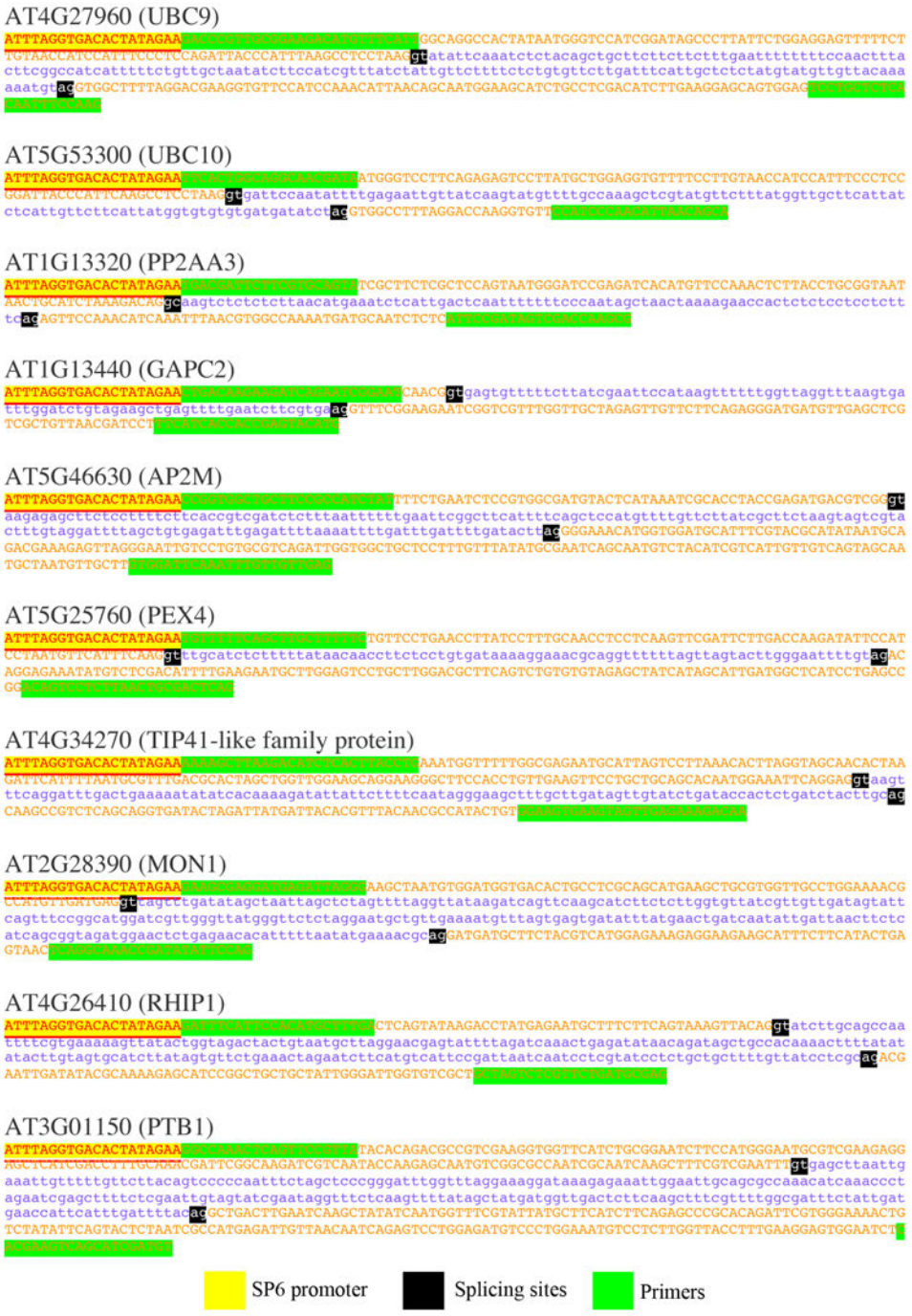


Figure 36: Sequences of all *Arabidopsis* pre-mRNA substrates presented in Fig. 35. Sequences were extracted from The *Arabidopsis* Information Resource (TAIR). Exonic sequences are shown in upper case letters, while the intronic sequence are shown in lower case. SP6 promoter sequence is highlighted in yellow; primers are highlighted in green and conserved splicing sites are highlighted in black.

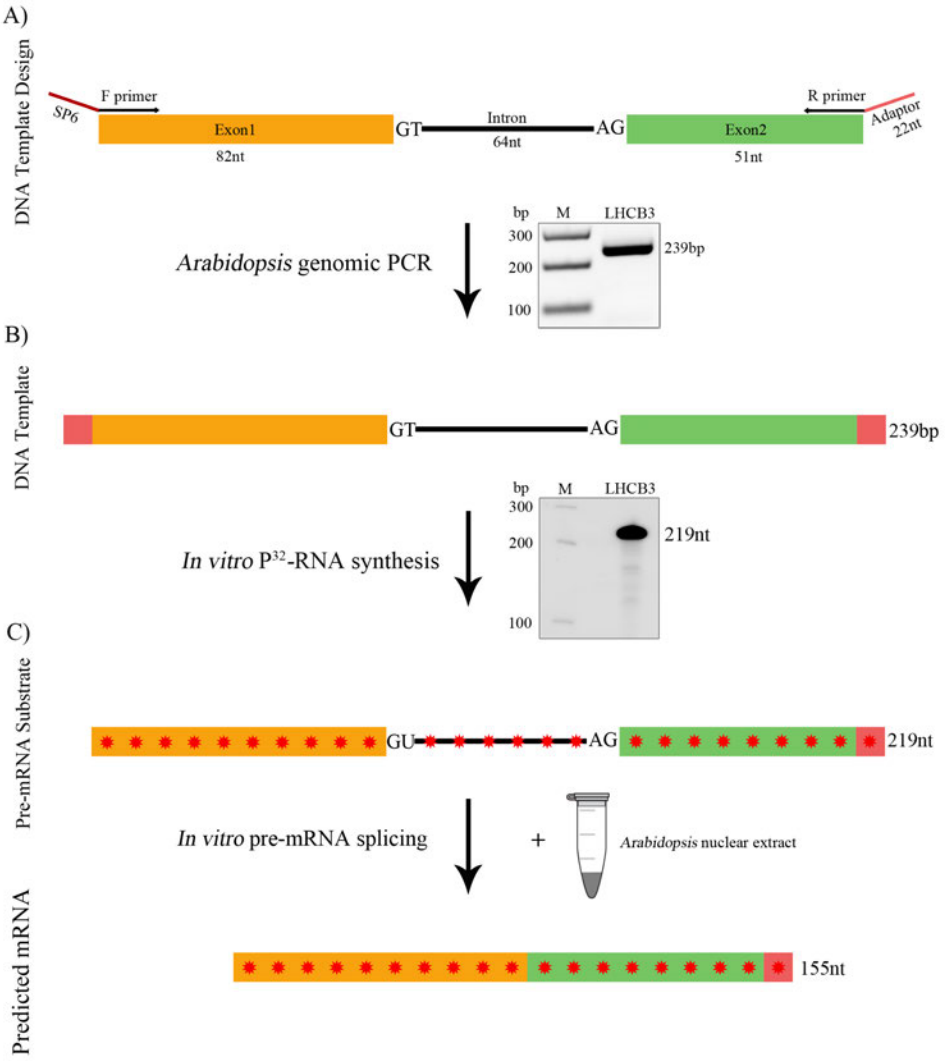


Figure 37: Preparation of *LHCB3* ^{32}P-labeled pre-mRNA substrate used for *in vitro* splicing assay. (A) Top, A schematic representation of a region of *Arabidopsis LHCB3* (AT5G54270) gene used to prepare DNA template for pre-mRNA substrate. A portion of the third and fourth exons (Orange and green boxes, respectively, labeled as exon 1 and exon2) and second intron (black line, labeled as intron) was used. F primer, forward primer with SP6 promoter sequence (red line), R primer reverse primer with an adaptor sequence (red line). Bottom, PCR fragment amplified with F and R primers using *Arabidopsis* genomic DNA. The PCR product was gel purified and used for *in vitro* transcription reaction. (B) Top, schematic of DNA template that was used to synthesize ^{32}P-labeled *LHCB3* pre-mRNA substrate. Bottom, A representative autoradiogram of *in vitro* ^{32}P-labeled *LHCB3* pre-mRNA substrate. (C) Top, Schematic representation of labeled *LHCB3* pre-mRNA substrate used for *in vitro* splicing assay. Bottom, predicted mRNA after *in vitro* splicing of pre-mRNA substrate. Sizes of intron, exons, pre-mRNA, and predicted mRNA are indicated. Red asterisks indicate ^{32}P-nucleotides in RNA.

A) DNA template sequence (239nt):

ATTTAGGTGACACTATAGAAAGAATTCCCTGGCGATTATGGTTGGGACACCGCCGGTTTATCCGCAG

ACCCTGAAGCCTTTGCCAAAAACAGAGCTCTTGAGgttagtctctaaatctcactcactcaatg

taaaaacagagtaaataactaatcatcatttgcagGTGATCCATGGGAGATGGGCAATGTTGGG

AGCTTTTGGTTGCATAACCCCTCTACACGACGCTCTTCCGATAA

B) Pre-mRNA substrate sequence (219nt):

AGAAUUCCCUGGCGAUUAUGGUUGGGACACCGCCGGUUUAUCCGCAGACCCUGAAGCCUUUGCC

AAAAACAGAGCUCUUGAGguuagucucuaaaucucacucacucaauguaaaaacagaguaaaua

acuaaucaucauuugcagGUGAUCCAUGGGAGAUGGGCAAUGUUGGGAGCUUUUGGUUGCAUAA

CCCCUCUACACGACGCUCUUCCGAUAA

C) Predicted mRNA sequence:

AGAAUUCCCUGGCGAUUAUGGUUGGGACACCGCCGGUUUAUCCGCAGACCCUGAAGCCUUUGCC

AAAAACAGAGCUCUUGAGGUGAUCCAUGGGAGAUGGGCAAUGUUGGGAGCUUUUGGUUGCAUAA

CCCCUCUACACGACGCUCUUCCGAUAA

SP6 promoter
Splicing sites
Primers
Adapter sequence

Figure 38: Sequences of DNA template (A), pre-mRNA substrate (B), and predicted mRNA (C). *Arabidopsis LHCB3* (AT5G54270) sequence was extracted from The *Arabidopsis* Information Resource (TAIR). Exonic sequences are shown in upper case letters, while the intron sequence is shown in lower case. SP6 promoter sequence is highlighted in yellow; primers are highlighted in green with either SP6 or adapter sequences: conserved splicing sites (GT – AG) are highlighted in black, and adaptor sequence is highlighted in blue.

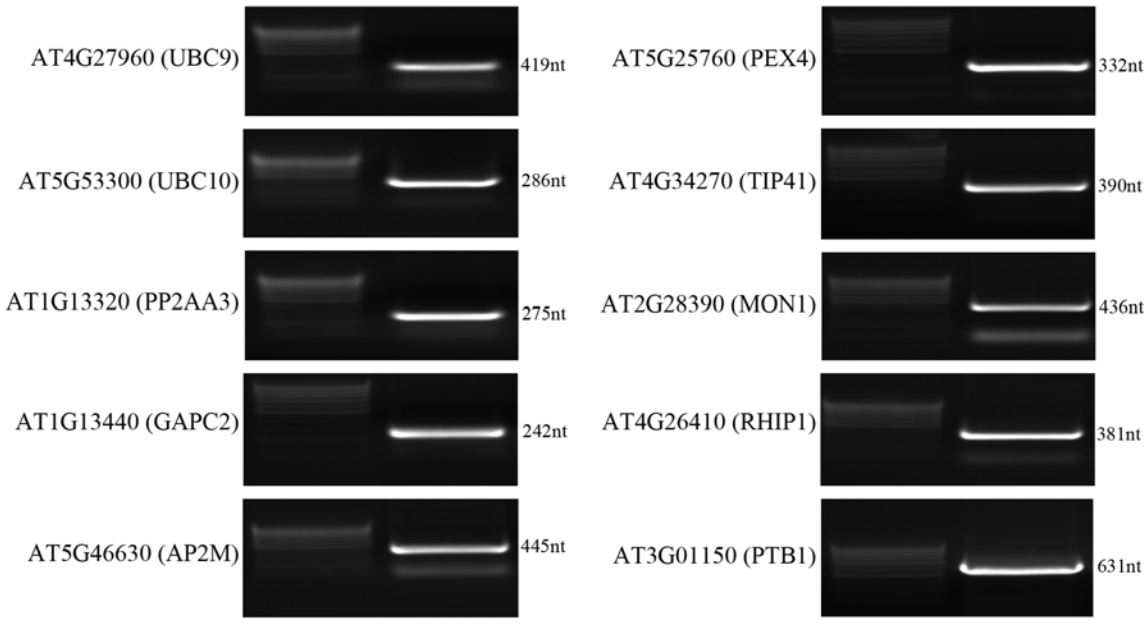


Figure 39: Preparation of DNA template for all *Arabidopsis* pre-mRNA substrates described in Figure 35. PCR reactions were performed with *Arabidopsis* genomic DNA using the primers shown in Figure 36. The numbers listed on the right of each gel indicate the length of the expected size PCR product.

plant-specific gene related to the photosynthesis system (Figure 37A and 38). *Arabidopsis* genomic DNA was used to generate DNA templates for *in vitro* pre-mRNA synthesis, and an SP6 promoter sequence was appended to the forward primer for all substrates (Figure 37A and 39). An *in vitro* transcription system using SP6 polymerase was used to radiolabel the substrates (Figure 37B and 40). Initially, all P^{32}-labelled pre-mRNA substrates were tested for *in vitro* splicing activity using *Arabidopsis* NE. Interestingly, only two of them, *LHCB3* and *UBC10,* produced labeled RNA bands other than the template size upon incubation in the *Arabidopsis* NE (Figure 41 and 42). The other pre-mRNA substrates that were tested did not exhibit any RNA products above background (Figure 43).

The *LHCB3* pre-mRNA substrate carried an 82 nt 5' exon, the 64 nt intron, and 51 nt of 3' exon plus a 22 nt adaptor sequence for a total length of 219 nt (Figure 37B). The P^{32}-labelled *LHCB3* pre-mRNA was incubated at 30^{o}C in *Arabidopsis* NE for 0, 90, or 180 min. The reaction products were isolated and analyzed by 6% denaturing polyacrylamide gel (Figure 41). We did not observe any P^{32}-labelled RNA products at 0 min incubation time; however, at 90 and 180 min there were a variety of P^{32}-labelled species smaller than the pre-mRNA substrate. The predicted spliced mRNA is 155 nt long (Figure 37C). Certainly, a 155 nt RNA species, the expected size of spliced mRNA, was observed after 90 min, and this product was accumulated to a higher level after 180 min. We also generated a P^{32}-labelled marker mRNA from an *LHCB3* cDNA template (M^{*}) for comparison with the *in vitro* product. In addition, we observed several other RNA species, one of the size of the intron plus the second exon (~137nt), and one with size of only the intron (~ 65nt), which may correspond to the pre-mRNA splicing intermediates. This finding suggests the possible *in vitro* splicing of *LHCB3* pre-mRNA using *Arabidopsis* NE. However, a variety of other RNA species were also observed at both 90 and 180 min incubation

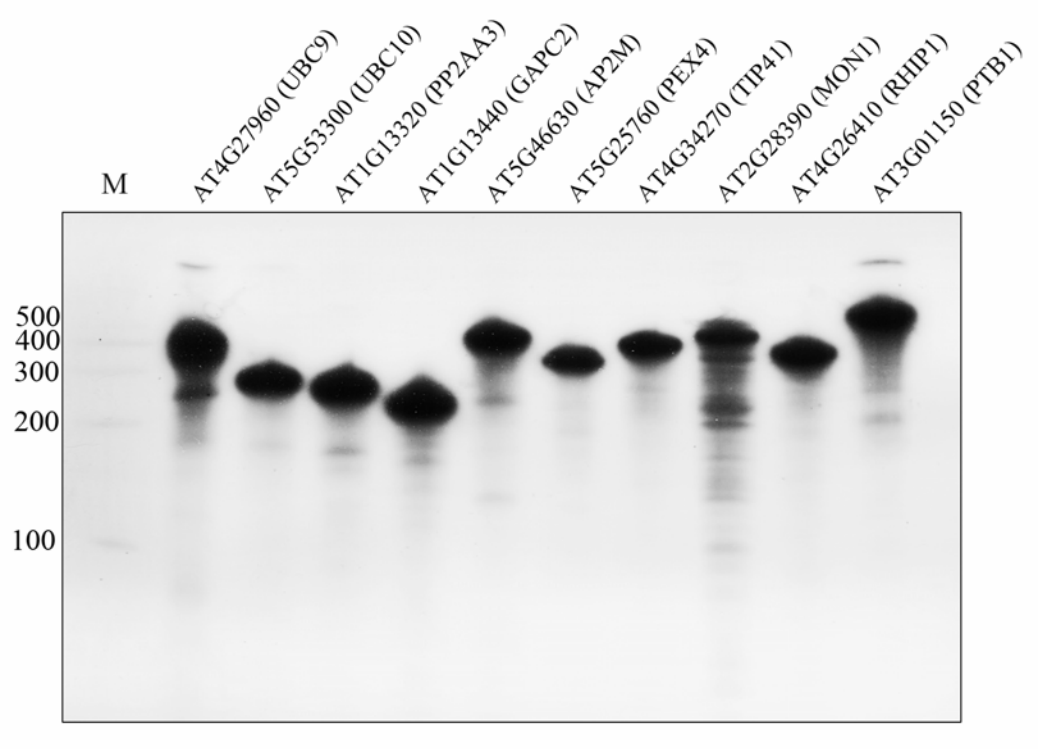


Figure 40: An autoradiogram showing the size of ten different ^{32}P-labelled pre-mRNA substrates prepared from the DNA templates shown in Figure 39. Radioactive pre-mRNA substrates were synthesized *in vitro* from DNA templates using SP6 RNA polymerase as described in materials and methods. ^{32}P-pre-mRNAs corresponding to the expected size were gel purified and used in *in vitro* splicing reaction.

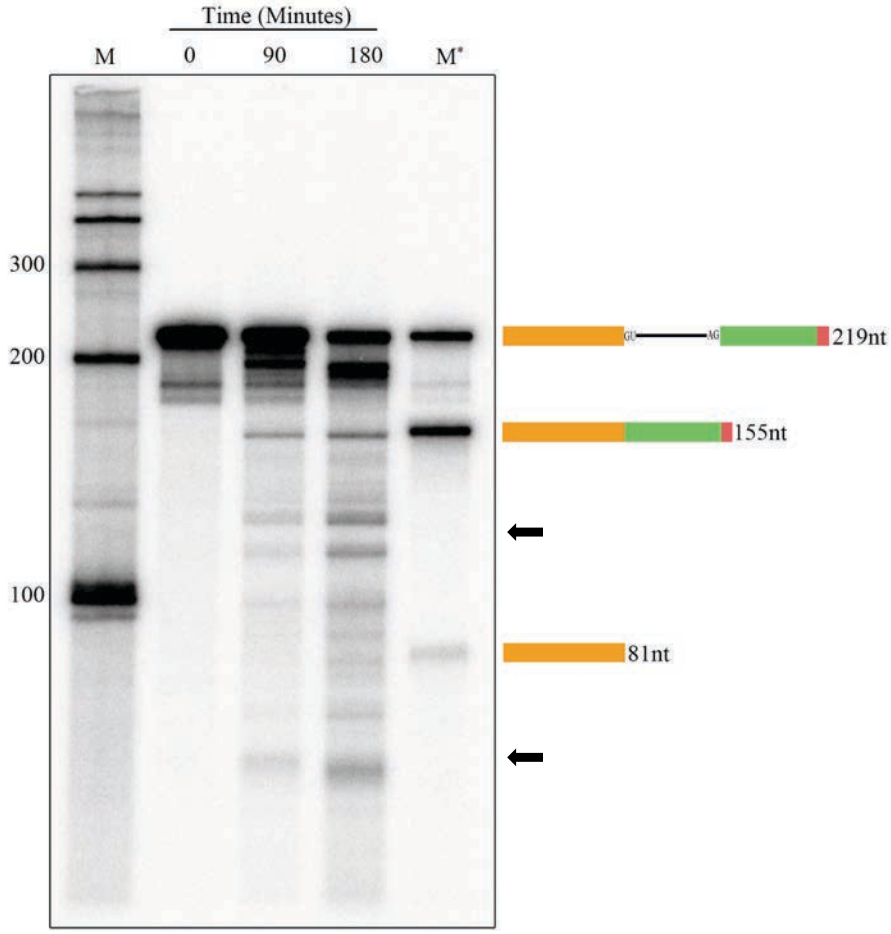


Figure 41: *In vitro* splicing assay with the *Arabidopsis LHCB3* pre-mRNA substrate. Radioactive *LHCB3* pre-mRNA substrate was synthesized *in vitro* with a DNA template using SP6 RNA polymerase (see Fig. 37B) as described in materials and methods. ^{32}P-labeled *LHCB3* pre-mRNA substrate (25,000 cpm) was incubated with nuclear extract from *Arabidopsis* etiolated seedlings at 30^{o}C as described in materials and methods. Samples were withdrawn at intervals (0, 90 and 180 min), ^{32}P-RNA was extracted and analyzed by electrophoresis on a 6% polyacrylamide gel containing 7 M urea, The gel was dried and exposed to a phosphor-imaging screen. M indicates ^{32}P-labeled RNA markers synthesized *in vitro* using RNA Century^{TM}-Plus Marker Templates (Applied Biosystems, AM7782). M* lane contains ^{32}P-labeled *LHCB3* pre-mRNA, spliced mRNA, and exon1. Schematic diagrams on the right show pre-mRNA, spliced mRNA and exon 1 and their sizes. One of the ^{32}P-RNA products formed in *in vitro* splicing assay corresponds to the size of spliced ^{32}P-mRNA marker, suggesting that it could be a putative spliced product. The arrowheads on the right side indicate the potential splicing intermediates. Other ^{32}P-RNA products could be another pre-mRNA splicing intermediates and/or degradation products.

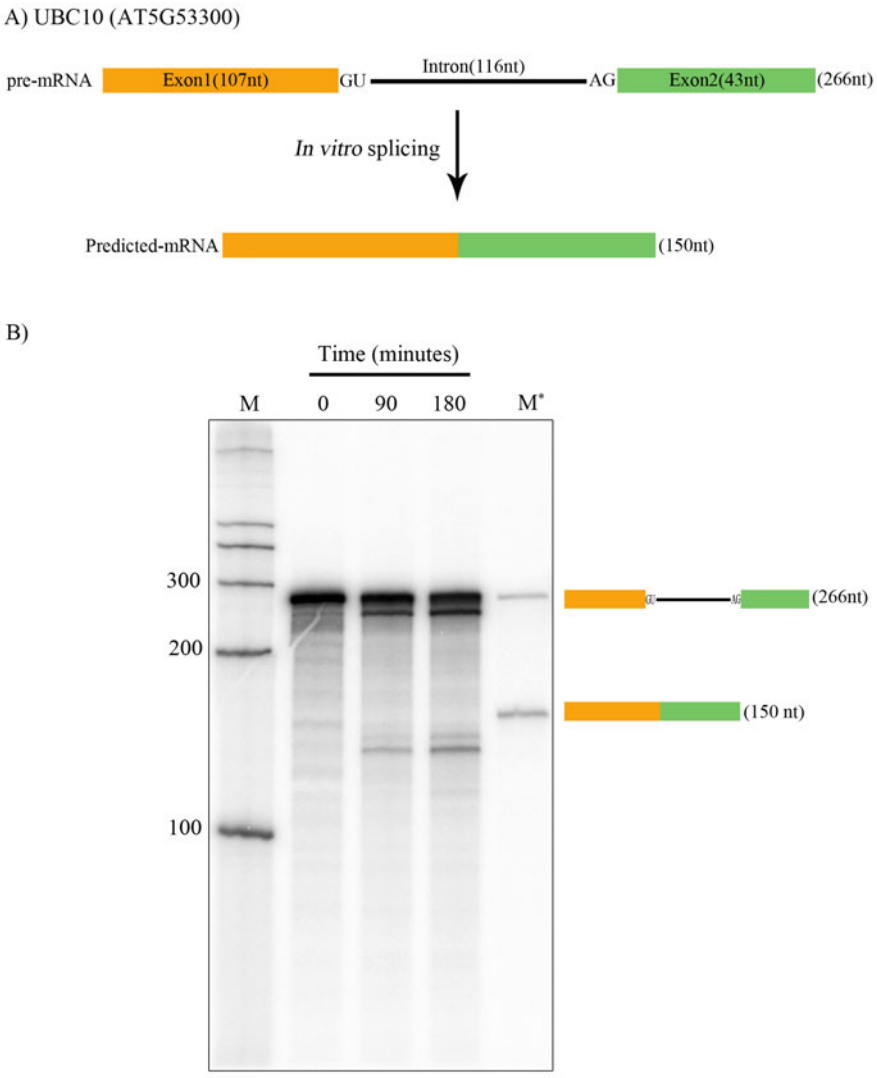


Figure 42: Time course of *in vitro* splicing of *Arabidopsis* ^{32}P labeled *UBC10* pre-mRNA substrate. (A) Top, Schematic representation of *UBC10* pre-mRNA substrate used for *in vitro* splicing assay. Bottom, Predicted mRNA after *in vitro* splicing. Sizes of intron, exons, pre-mRNA and predicted mRNA are shown. (B) *In vitro* splicing of *UBC10* ^{32}P pre-mRNA substrate. ^{32}P-labeled pre-mRNA (25,000cpm) was incubated at 30^{o}C in the *Arabidopsis* nuclear extract under *in vitro* splicing conditions as described above. Samples were withdrawn at 0, 90 and 180 min. ^{32}P-RNA was then isolated and analyzed by electrophoresis on a 6% polyacrylamide-7 M urea gel, followed by autoradiography. M indicates ^{32}P-RNA markers. M* lane contains ^{32}P-labeled RNA markers for *UBC10* pre-mRNA and spliced mRNA that are synthesized *in vitro* with DNA template using SP6 polymerase. Panel on the right shows schematic diagrams of pre-mRNA and the spliced product and their sizes.

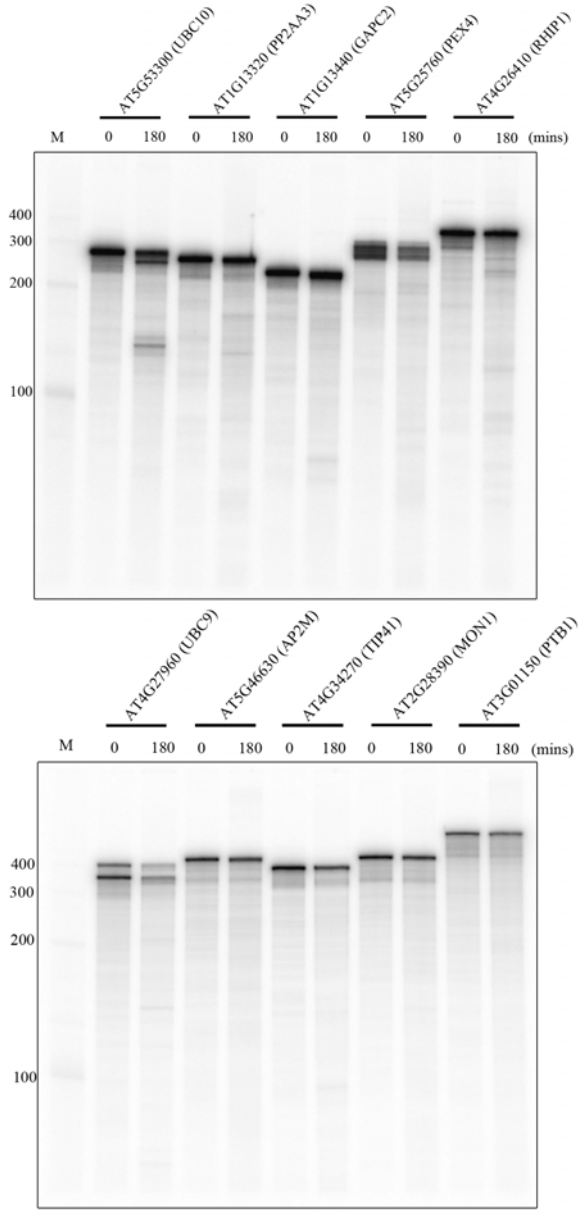


Figure 43: *In vitro* splicing assay with different pre-mRNA substrates. Labeled pre-mRNAs substrates were incubated individually in the *Arabidopsis* nuclear extract at 30^{o}C. *In vitro* splicing reaction conditions were as described above. Samples were withdrawn at time "0" and after 3 hrs (180 mins). ^{32}P-RNA was recovered and analyzed by electrophoresis on a 6% polyacrylamide-7 M urea gel, followed by autoradiography. (Top gel) Splicing of *UBC10, PP2AA3, GAPC2, PEX4, and RHIP1,* (Bottom gel) Splicing of *UBC9, AP2M, TIP41, MON1, and PTB1.* M indicates ^{32}P-RNA marker.

times, which may be other splicing intermediates, aberrant splicing or degradation products.

The *UBC10* pre-mRNA substrate contains a 107 nt 5' exon, the 116 nt intron, and 43 nt of 3' exon for a total length of 266 nt (Figure 42A). A timecourse *in vitro* splicing reaction using *Arabidopsis* NE and the P^{32}-labelled *UBC10* pre-mRNA was conducted and processed as described above (Figure 42B). Based on the known sequence of the *UBC10* pre-mRNA substrate, the expected size of the spliced mRNA is 150 nt (Figure 42A); P^{32}-labelled *UBC10* mRNA M^* was also included here as a marker. However, the *in vitro* splicing reaction revealed unexpected results. None of the RNA species produced from the *in vitro* splicing reaction matched the size of the predicted *UBC10* mRNA. That said, there was an RNA species close to the size of the predicted mRNA; this discrepancy may be a result of inaccurate processing *in vitro* compared to what happens *in vivo*.

Taken together, these findings suggest that the NE derived from *Arabidopsis* etiolated seedlings may support splicing activity *in vitro* of *LHCB* pre-mRNA. However, the variety of RNA species that were produced indicates it can also give rise to false positive results. Despite this uncertainty, these findings provide an opportunity for further investigation and illustrate the potential of developing a plant *in vitro* splicing system. Since only the *LHCB3* pre-mRNA substrate exhibited promising results and generated the exact expected size of mRNA from the *in vitro* reaction (Figure 41), We concentrated our efforts on investigating this further to confirm that splicing of *LHCB* pre-mRNA. For further experiments, we used only *LHCB3* pre-mRNA substrate and considered the RNA band of correct size as the putative *in vitro* spliced product.

Heating of *Arabidopsis* NE Inactivated *In Vitro* Splicing Reaction

Pre-mRNA splicing is mediated via the spliceosome, which is a large and dynamic machine containing five snRNPs and numerous proteins (Wahl et al., 2009). It has been reported that

heat-treated NE prepared from HeLa cells or yeast whole cell extract were unable to form a functional spliceosome and splice pre-mRNA *in vitro* (Lin et al., 1987; Shukla et al., 1990). Therefore, we incubated the *Arabidopsis* NE at 90^{o}C for 3 min, and then tested its splicing activity with P^{32}-labelled *LHCB3* pre-mRNA substrate. While the untreated NE converted the input pre-mRNA to the expected product, heat-treated NE was unable to do so (Figure 44). In agreement with non-plant splicing extracts, these results suggest that the *Arabidopsis* NE contains heat-sensitive components that are required for producing the putative splicing product.

Putative Spliced Product Increased With Increasing NE and Pre-mRNA Concentration

In order to determine the effect of NE concentration on the production of the putative spliced mRNA, a range of NE concentrations from 0 to 50% (v/v) were tested while maintaining consistent volume and chemical composition of the reactions by adding an appropriate amount of NE-containing buffer. Indeed, the production of the putative splicing product increased with increasing NE concentration (Figure 45). In addition, it seems that 30% (v/v) is an optimal NE concentration for this *in vitro* splicing assay. Thus, these results support that the appearance of the spliced mRNA is dependent on the concentration of proteins present in the NE.

Likewise, we examined the effect of input pre-mRNA substrate concentration on formation of the putative spliced mRNA. We found that increasing concentration of *LHCB3* pre-mRNA substrate increased the putative spliced product (Figure 46). This result suggests that the production of putative spliced mRNA depends on the amount of pre-mRNA substrate that is available to be processed *in vitro*.

Characterization of Putative Spliced Product Using S1 Nuclease Protection Assay

To determine whether the P^{32}-labelled *LHCB3* pre-mRNA substrate was accurately spliced *in vitro*, the putative spliced product was analyzed by S1 nuclease protection assay. For this assay,

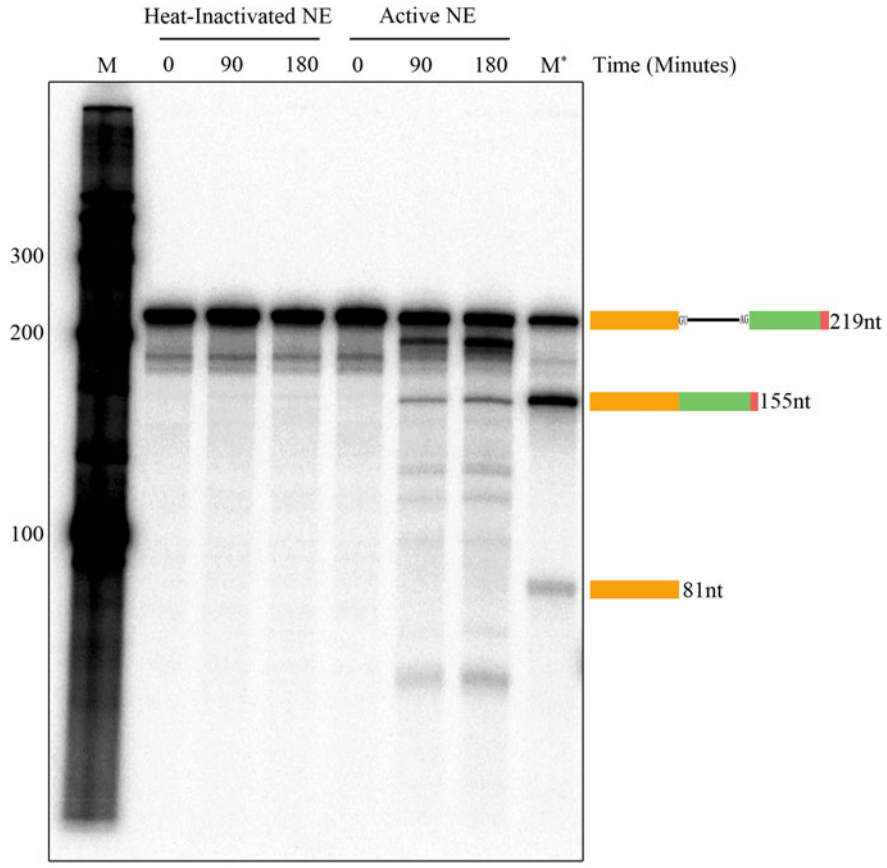


Figure 44: Heat-inactivation of *Arabidopsis* NE abolished the production of a putative spliced product. NE from *Arabidopsis* etiolated seedlings was incubated at 90^{o}C for 3 min or kept on ice (as a control) were used for splicing assays at 30^{o}C with the *LHCB3* ^{32}P-pre-mRNA. Samples were withdrawn at different time points (0, 90 and 180 minutes), ^{32}P-RNA was extracted and analyzed as described above in Fig. 8 legend. RNA markers (M and M*) and schematic diagrams on the right were also described in Fig. 41.

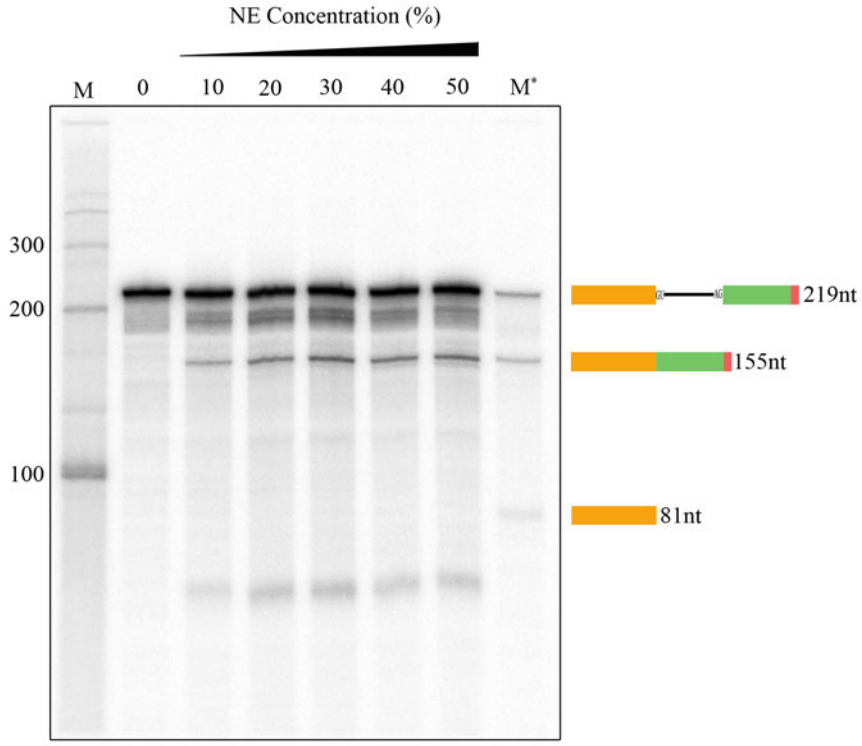


Figure 45: Putative spliced product is increased with increasing nuclear extract concentration. *In vitro* splicing of ^{32}P-labeled *LHCB3* pre-mRNA substrate (25,000 cpm) was carried out at 30^{o}C in 25 ml reaction volume containing different concentrations 0-50% (v/v) of nuclear extract as described in materials and methods. All reactions were stopped after three hours; ^{32}P-RNA was extracted and analyzed by electrophoresis on a 6% polyacrylamide gel containing 7 M urea, followed by autoradiography. RNA markers (M and M*) and schematic diagrams on the right were described in Fig. 41.

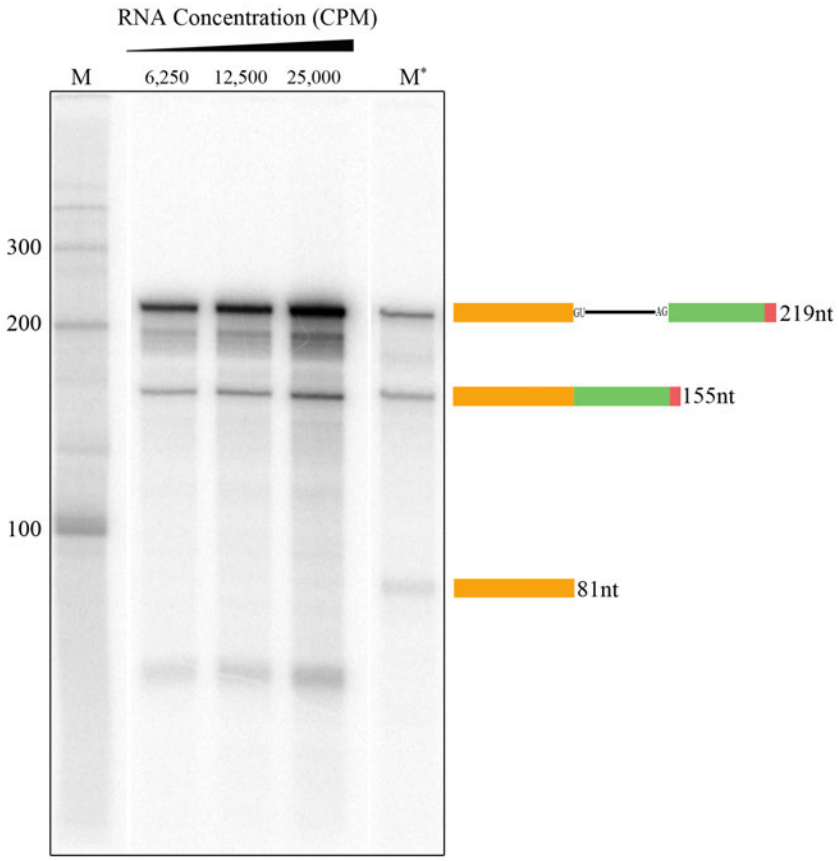


Figure 46: Increasing concentration of ^{32}P-labeled *LHCB3* pre-mRNA substrate in the *in vitro* splicing assay resulted in a proportional increase in putative spliced product. Different concentrations of ^{32}P-pre-mRNA substrate (6,250; 12,500; 25,000 cpm) were incubated at 30^{o}C with 50% (v/v) nuclear extract. Reactions were stopped after three hours; ^{32}P-RNA was isolated and analyzed by electrophoresis as described above. RNA markers (M and M*) and schematic diagrams on the right were described in Fig.41.

we used a DNA oligonucleotide (50nt) probe designed to pair with the exon junction that is predicted to be joined together during *in vitro* splicing (Figure 47A). The *in vitro* putative splice product was gel-purified and at least three samples pooled together to enhance the amount of RNA present. Probes were allowed to hybridize with the purified RNA sample. In addition, the probes were also hybridized with the unspliced P^{32}-labelled *LHCB3* pre-mRNA substrate as a negative control, and with P^{32}-labelled *LHCB3* mRNA as a positive control. Subsequently, the hybridized molecules were digested with single-strand nucleic acid-specific S1 nuclease. The S1-resistant products were then separated on a denaturing polyacrylamide gel and visualized by autoradiography (Figure 47B). The estimated size of the protected sequence is ~50 nt long, and indeed, both the predicted *in vitro* splice product and the positive control generated S1-resistant products with approximately equal sizes of ~50 nt. Conversely, the unspliced pre-mRNA generated only S1-resistant products with a size of 25 nt that correspond to unligated exon regions on each side of the intron. These results suggest that the *LHCB3* pre-mRNA is being accurately spliced *in vitro* to generate the authentic mRNA sequence spanning the *LHCB3* exon1-exon2 junction.

Mutations in Conserved Splice Sites Modulate The Production of Spliced Product

During the pre-mRNA splicing process, the spliceosome assembles around the 5′ss at the beginning of an intron and the 3′ss at the end of that intron (Matera and Wang, 2014). It is well-established that each splice site in plants and animals consists of consensus sequences, and these include almost invariant dinucleotides: GT at the 5′ss and AG at the 3′ss (Reddy, 2007). In addition to these major conserved splice site sequences, there are minor non-canonical splice sites with AT at the 5′ss and AC at the 3′ss (Reddy, 2007). *In vitro* mutation analysis of the conserved 5′ss GT revealed that these mutations could modulate splice site choices and reduce

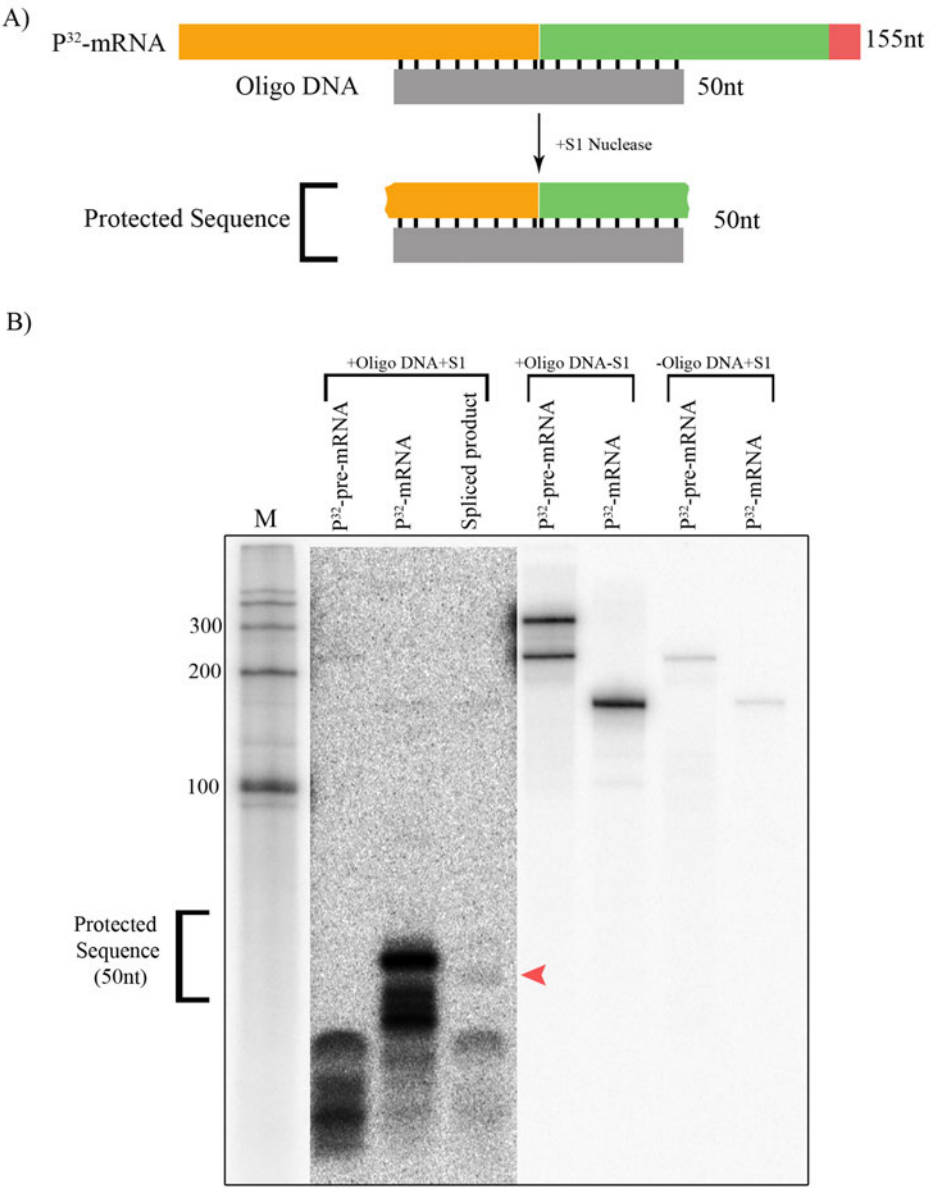


Figure 47: Characterization of putative spliced product using S1 nuclease. (A) Schematic representation of S1 nuclease protection assay. Top, diagram of hybrid formed between spliced RNA and DNA oligonucleotide (50nt) complementary to exons junction. Bottom, diagram of protected sequences after S1 nuclease digestion. (B) Putative spliced ^{32}P-RNA produced in *in vitro* splicing assay was gel purified as described in material and methods. ^{32}P-pre-mRNA (negative control), ^{32}P-mRNA (positive control), and putative spliced product ^{32}P-RNA were hybridized to oligo DNA that is complementary to exons junction. Following hybridization, ^{32}P-RNA-DNA hybrids were digested with S1 nuclease that degraded single stranded nucleic acids. +oligo DNA+S1. ^{32}P-RNA-DNA hybrids with S1 nuclease; +oligo DNA –S1 ^{32}P-RNA-DNA hybrids without S1 nuclease; oligo DNA+S1 ^{32}P-RNA without oligo DNA hybridization and plus S1 nuclease digestion. The size of the protected region (50nts) is indicated. Red arrowhead shows protected exon junction sequence with putative spliced product.

the rate of two exons ligating (Aebi et al., 1987). Therefore, we aimed here to investigate if substitution mutations of consensus splice site sequences in the *LHCB3* pre-mRNA substrate affect the production of its putative *in vitro* spliced product. To this end, we generated different *LHCB3* pre-mRNA substrates carrying several substitution mutations of one or both splice sites (Figure 48A). These mutated substrates include: 5′ss GU to AC (Mutant 1, M1), 3′ss AG to UC (Mutant 2, M2), and both 5′ss GU to AC and 3′ss AG to UC (Mutant 3, M3). In addition, we generated a pre-mRNA substrate in which the major conserved splice site sequences were substituted by minor splice site sequences, changing the 5′ss GU to AU and 3′ss AG to AC (Mutant 4, M4). Since a previous study showed that mutation of the 5′ss GU did not prevent cleavage in 5′ss, but only modulated the production of spliced mRNA (Aebi, M. et al., 1987), we also included another mutated pre-mRNA substrate (Mutant 5, M5). In this substrate, both 5′ss and 3′ss conserved sequences (-3, +5) are changed to less conserved sequences reported in (Reddy, 2007).

In vitro splicing assays using these pre-mRNA substrates revealed that mutations of the conserved splice site sequences modulate the *in vitro* production of putative *LHCB3* spliced mRNA (Figure 48B). In agreement with a previous study (Aebi et al., 1987), M1 did not completely abolish the production of spliced mRNA; however, this mutation clearly reduced the generation of the putative spliced mRNA. M2 enhanced the production of the spliced mRNA, while M3 did not exhibit any clear effect on *in vitro* splicing. Interestingly, M4 was capable of normal *in vitro* generation of the putative spliced mRNA. Mutant M5 completely eliminated the production of the expected product, and instead resulted in a new RNA band of ~100 nt. These findings further demonstrate the authenticity of our *in vitro* splicing assay using *Arabidopsis* NE, and potential for developing a robust plant *in vitro* splicing assay system.

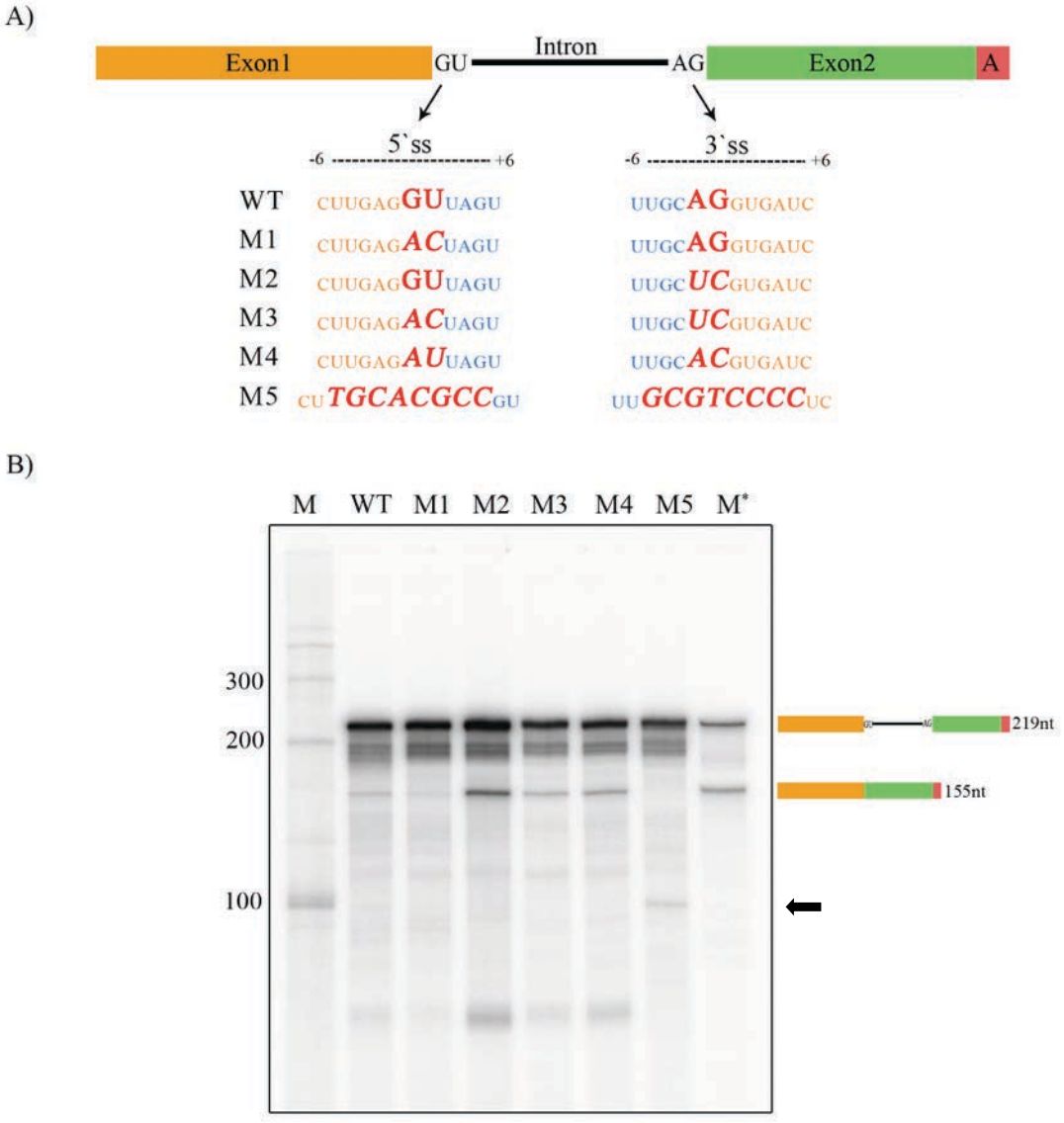


Figure 48: Mutations in conserved splice sites of *LHCB3* pre-mRNA substrate modulated the production of putative spliced product. (A) Diagram shows sequence substitutions of conserved 5' GU and 3' AG splice sites (ss) of *LHCB3* pre-mRNA substrate. DNA templates with different splice site mutations (M1-5) were synthesized (Integrated DNA Technologies, Inc., Coralville, IA) for *in vitro* ^{32}P-labelled RNA synthesis. (B) *In vitro* splicing of ^{32}P-labeled wild type and five mutants (M1 to M5) of *LHCB3* pre-mRNA substrate was carried as described before. Reactions were stopped after three hours. ^{32}P-RNA was isolated and analyzed by electrophoresis as above. The arrowhead on the right side indicates a new RNA band of ~100 nt that results from M5.

Effects of ATP and Mg^{2+} on Pre-mRNA Splicing *In Vitro*

It has been well-established that mammalian, yeast, and *Drosophila in vitro* splicing reactions require exogenous ATP and Mg^{2+} (Krainer et al., 1984; Lin et al., 1985; Rio, 1988). Therefore, we aimed here to investigate and optimize the requirement of these cofactors for the *Arabidopsis*-derived *in vitro* splicing reaction. Notably, the NE preparation method does not involve a dialysis step against the swelling buffer, hence the NE contains endogenous concentrations of ATP and Mg^{2+}. In addition, the 25 ul *in vitro* splicing reaction containing 50% NE has 2.5 mM Mg^{2+}, as the swelling buffer contains 5 mM $MgCl_2$.

The ATP concentration of the *in vitro* splicing reaction was varied while holding the concentrations of remaining components constant (Figure 49). In the absence of ATP, some RNA degradation was observed. In contrast, the addition of ATP maintained input pre-mRNA integrity and also enhanced production of the putative spliced mRNA. There were no obvious differences in effect between the ATP concentrations (1, 2, and 3 mM) tested. Therefore, these results indicate that the addition of exogenous ATP might support splicing *in vitro*, and the optimal concentration is ~1 mM.

In the same manner, the effect of varying Mg^{2+} concentration on the *in vitro* reaction was investigated (Figure 50). We found that 2.5 mM Mg^{2+} was likely an optimal concentration for the *in vitro* splicing reaction. Indeed, increasing Mg^{2+} to 5 mM caused input pre-mRNA instability. On the other hand, *in vitro* splicing activity was reduced by the addition of 2.5 or 5 mM EDTA, an ion-chelating agent, indicating a divalent cation requirement (Figure 50). Taken together, these findings suggest that the plant-derived *in vitro* splicing reaction is like other *in vitro* splicing system requires ATP and Mg^{2+}.

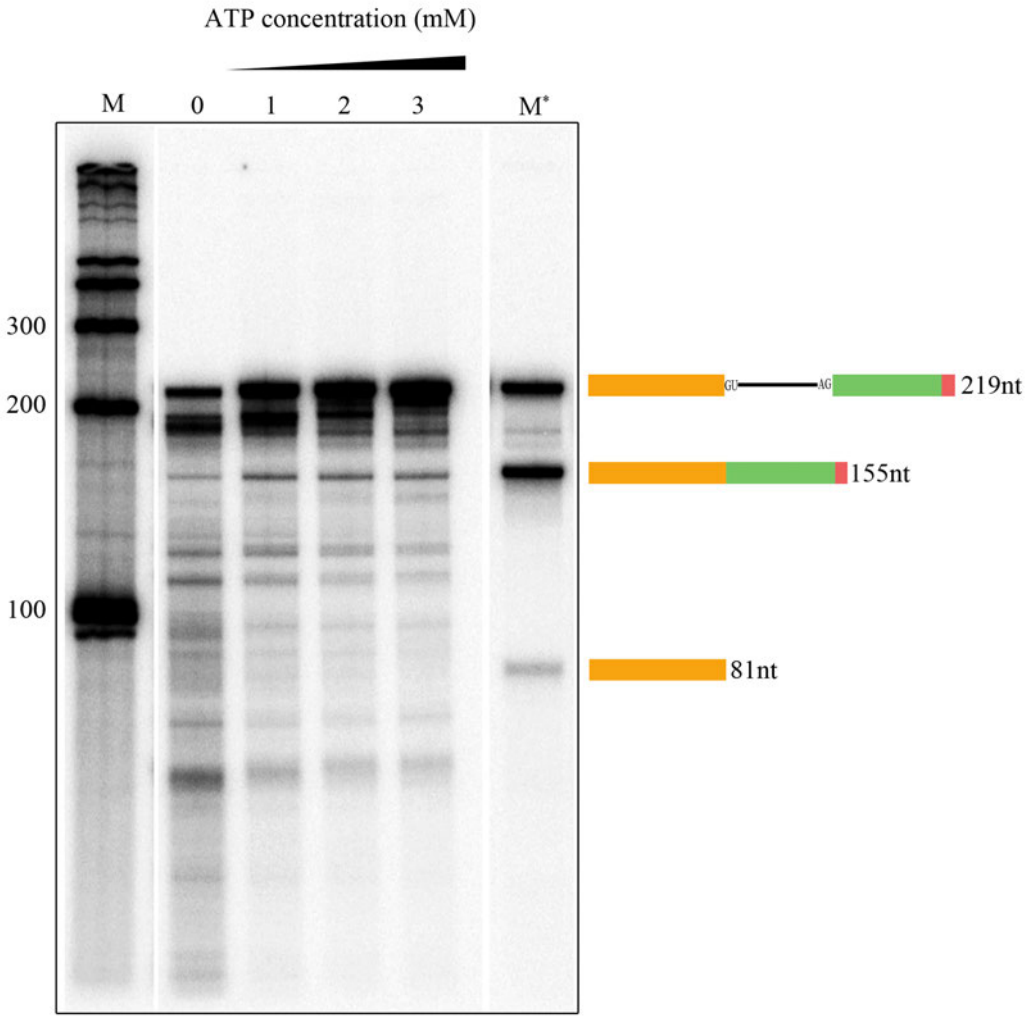


Figure 49: Exogenous ATP to *in vitro* splicing assay increased the amount of putative spliced product from *LHCB3* pre-mRNA. *In vitro* splicing reaction of *LHCB3* ^{32}P-pre-mRNA substrate (25,000cpm) was carried out as described above without (0mM) with increasing concentrations ATP (1, 2, 3mM). Reactions were stopped after three hours; ^{32}P-RNA was recovered and analyzed by electrophoresis on a 6% polyacrylamide gel containing 7 M urea, followed by autoradiography. RNA markers (M and M*) and schematic diagrams on the right were described in Fig.41.

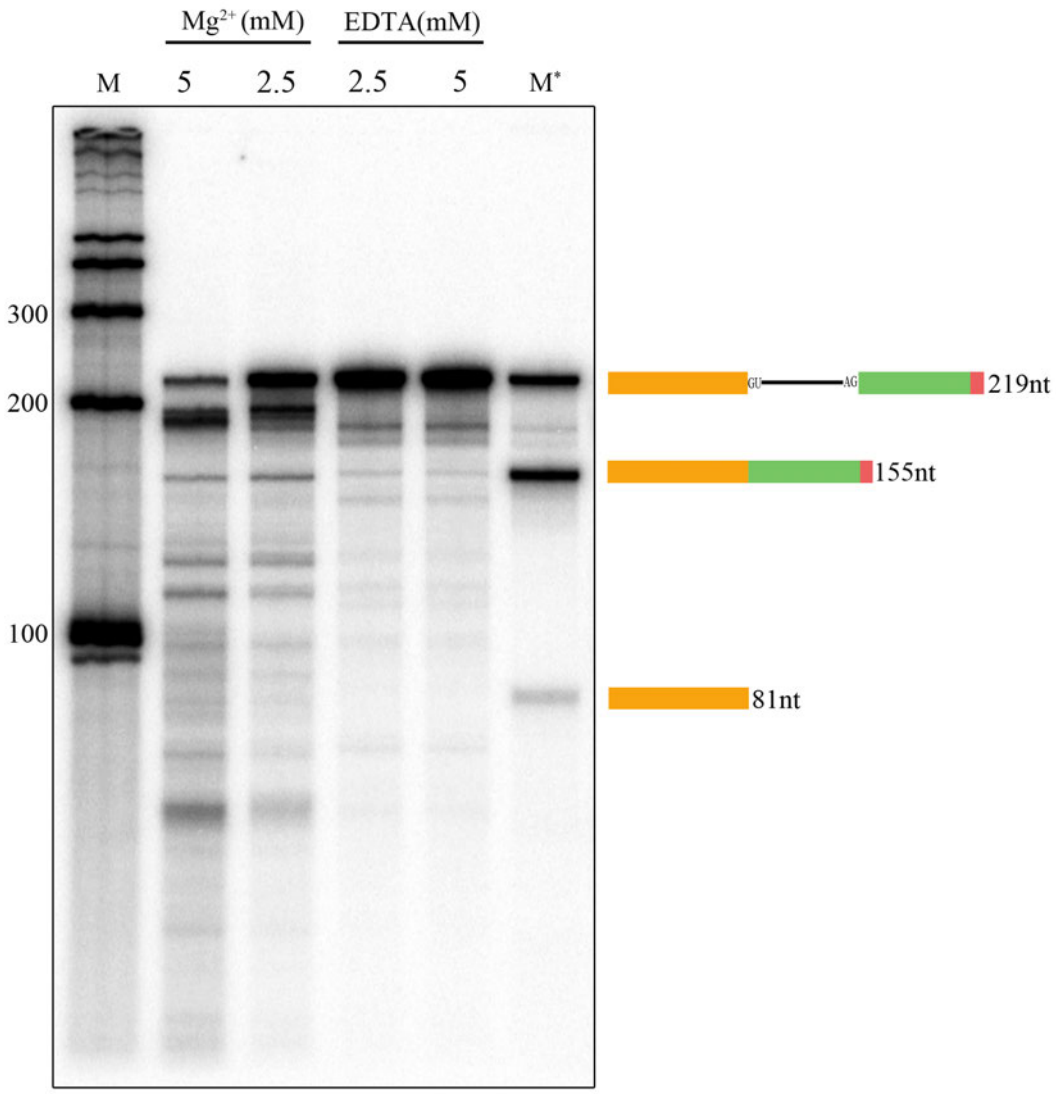


Figure 50: Effect of various concentrations of Mg^{2+} on the production of putative spliced product. *In vitro* splicing reaction of *LHCB3* ^{32}P-pre-mRNA substrate (25,000cpm) was performed as described previously with different concentrations of Mg^{2+} (2.5 and 5 mM), or in the presence of different concentration (2,5 and 5 mM) of EDTA, a divalent cation chelator (EDTA). Reactions were stopped after three hours. ^{32}P-RNA was recovered and analyzed by electrophoresis on a 6% polyacrylamide-7 M urea gel, followed by autoradiography. RNA markers (M and M^*) and schematic diagrams on the right were described in Fig.41.

Incubation Temperature of *In Vitro* Splicing Reaction affects Putative Spliced Product

It has been shown previously that different *in vitro* splicing reactions have narrow optimum temperatures. The optimum temperature for the mammalian *in vitro* splicing assay is 30^{o}C, while for the yeast *in vitro* splicing system it is 25^{o}C (Krainer et al., 1984; Lin et al., 1985). Therefore, to address the effect of incubation temperature on the accumulation of predicted spliced product, reactions were carried out at different temperatures: 24^{o}C, 30^{o}C, 37^{o}C, and 42^{o}C (Figure 51). The results showed that 24^{o}C is likely the optimal incubation temperature for high accumulation of the putative spliced mRNA. The quantity of spliced product is decreased with increasing incubation temperature. In fact, the incubation at 42^{o}C greatly reduced the appearance of spliced mRNA, suggesting sensitivity of the reaction to high temperature, as was shown in the previous experiment (Figure 44). Thus, these findings indicate that the *in vitro* splicing assay using *Arabidopsis* NE has an optimal temperature of 24^{o}C, and further demonstrate that the NE contains heat-labile components required for generating the putative spliced product.

DISCUSSION

In vitro splicing systems derived from mammals, yeast, and *Drosophila* have allowed remarkable progress in illustrating splicing mechanisms in eukaryotes. There is no *in vitro* splicing assay for plant systems. Hence, many aspects of pre-mRNA splicing in plants are unknown. In an effort to develop a plant *in vitro* pre-mRNA splicing system, we report here, for the first time, experimental evidence suggesting that NE derived from *Arabidopsis* etiolated seedlings is capable of splicing plant pre-mRNA *in vitro*. This system should provide a starting point for further development of plant *in vitro* splicing assay to investigate plant-specific pre-mRNA splicing mechanisms.

In this study, we show that *Arabidopsis* NE was able to convert the pre-mRNA of *LHCB3*

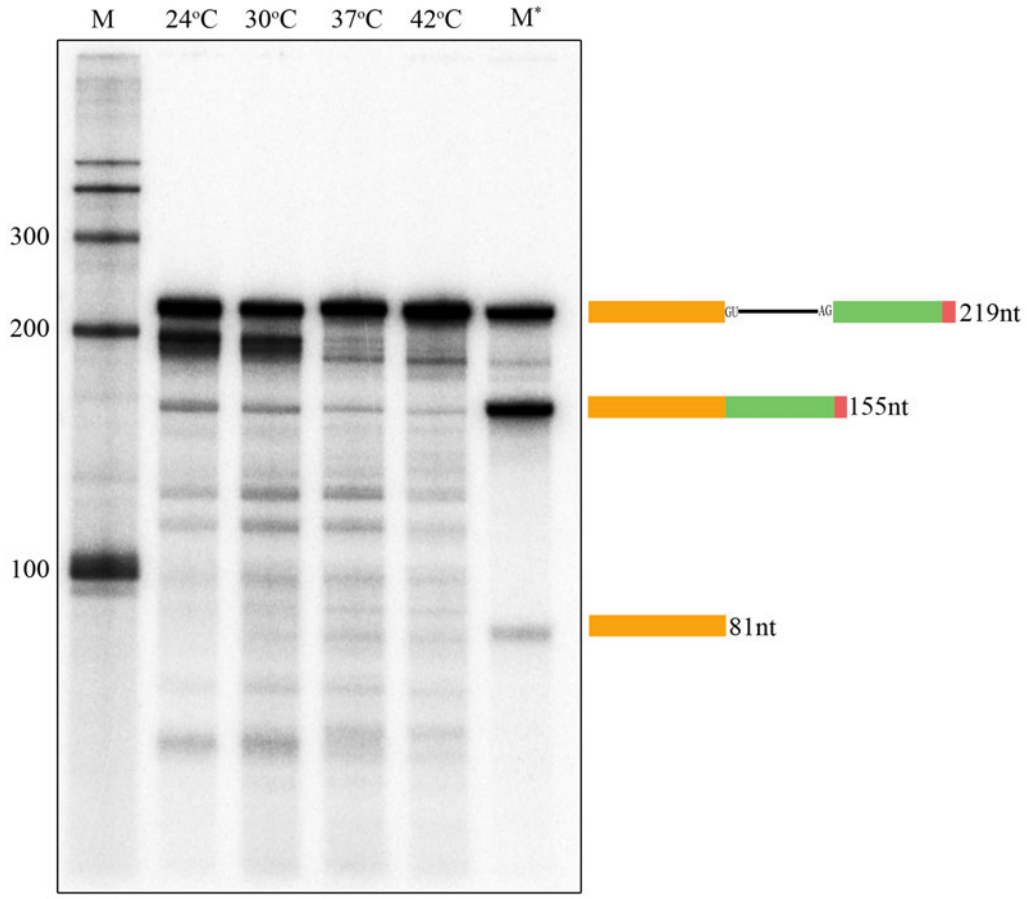


Figure 51: The amount of putative spliced product at different temperatures. *In vitro* splicing of *LHCB3* ^{32}P-pre-mRNA substrate (25,000cpm) was carried out as described earlier at different temperatures (24^{o}C, 30^{o}C, 37^{o}C, 40^{o}C). Reactions were stopped after three hours. ^{32}P-RNA was recovered and analyzed by electrophoresis on a 6% polyacrylamide-7 M urea gel, followed by autoradiography. RNA markers (M and M*) and schematic diagrams on the right were described in Fig. 41.

substrates into an expected size of spliced mRNA. Although additional experiments are necessary to further confirm the identity of the splice product, all the data presented here strongly suggest that the RNA product that corresponds to the spliced mRNA marker is likely a spliced product. Production of the expected size of mRNA upon incubation of pre-mRNA template in the NE, indication that the two exons are linked together according to a junction mapping assay using S1 nuclease, and demonstration that substitution mutations of conserved splice site sequences inhibit the appearance of putative spliced mRNA – all suggest splicing of *LHCB* pre-mRNA in plant NE. In addition, we found this system to be similar to the well-established non-plant *in vitro* pre-mRNA splicing assays in several ways. First, this system is sensitive to high temperature; second, it requires Mg^{2+}; and third, ATP is necessary for the generation of putative spliced mRNA. From these experiments, we conclude that this is a promising progress towards developing an efficient plant *in vitro* splicing system.

The establishment of an *in vitro* splicing system for plants has been long overdue; difficulties in developing one may be because of characteristics inherent to plant cells. Our success in establishing this *in vitro* assay may be attributed to plant materials used for NE preparation, NE preparation method, and choice of pre-mRNA substrate. My NE is sourced from *Arabidopsis* seedlings that were grown under dark conditions, which are actively growing. The actively growing seedlings have high levels of gene expression for which active pre-mRNA splicing machinery is also needed. Furthermore, unlike other preparation methods that include a high salt lysis buffer, we applied a method that uses low denaturing conditions. This may maintain protein integrity and result in protection of the native states of spliceosomal machinery. Moreover, it is known that not every pre-mRNA substrate can be spliced *in vitro*, and each substrate requires optimized *in vitro* splicing conditions (Mayeda and Krainer, 1999a). Thus, the choice of pre-

mRNA substrate and its primary structure is a significant factor for successful *in vitro* splicing systems. Interestingly, none of the pre-mRNA substrates that we tested exhibited encouraging results, except the *LHCB3* pre-mRNA, indicating that this substrate was the most appropriate selection for *in vitro* splicing and for further studies. In animals also, only a few pre-mRNAs such as β-globin, β-tropomyosin, adenovirus, δ-Crystallin, and Simian virus 40 (SV40) pre-mRNA are widely used, suggesting that only few pre-mRNA are efficiently spliced under *in vitro* conditions (Mayeda and Krainer, 1999a; Padgett et al., 1983; Reichert and Moore, 2000).

Convincing evidence also suggests that the plant *in vitro* splicing assay system is quite similar to assays used for mammals, yeast, and *Drosophila*. Similarities involve assay conditions, including concentrations of ATP, Mg^{2+}, and K^{+} (Hicks et al., 2005). In addition, the time course of a splicing reaction (0 to 180 min) is comparable between this *in vitro* splicing assay and other such systems (Krainer et al., 1984; Lin et al., 1985; Rio, 1988). Meanwhile, the optimal incubation temperature for this splicing reaction was unlike other splicing systems; it was within the range of the optimal growth temperature for *Arabidopsis* (23-25^{o}C) (Rivero et al., 2014).

Another point of interest is that *in vitro* splicing of the *LHCB3* pre-mRNA substrate generated a number of RNA species in addition to the putative accurately spliced RNA. Given the sizes of each part of the pre-mRNA substrate, the production of RNA species within the size range of the intron alone (64nt) and the intron plus exon2 (137nt) suggest these species are likely splicing intermediates; this is supported by the well-characterized two-step splicing intermediates in other eukaryotes (Padgett et al., 1986; Ruskin et al., 1984). In addition, the other RNA species that exhibited unexpected mobility on the gel might correspond to lariat-containing RNA intermediates generated during splicing (Padgett et al., 1986; Ruskin et al., 1984). Meanwhile, we cannot rule out the possibility that some of these RNA species might result from degradation

of the pre-mRNA substrate. It would be interesting to further confirm the identity of these RNA species using nuclease protection assay, primer extension, RT-PCR and DNA sequencing. It is also possible to examine each RNA species for the lariat structure using RNase R, which digests only linear but not circular RNAs (Suzuki et al., 2006).

Still another point of interest is provided by the observation that substitution mutations of splice site consensus sequences in the *LHCB3* pre-mRNA modulate the *in vitro* splicing reaction. Some early studies using a point mutation strategy in the mammalian *in vitro* splicing assay have investigated the function of splice site consensus sequences (Aebi et al., 1986; Aebi et al., 1987). Compared with their results, my finding that mutation of the 5'ss causes reduction of splicing efficiency indicates some similarities between these two systems. However, unlike the prior study, we found that mutation of the 3'ss of the *LHCB3* pre-mRNA enhanced *in vitro* splicing efficiency, suggesting some differences in splicing regulation for plants. It would be interesting to further investigate how this mutation affects the splicing rate *in vitro* and whether this mutation has the same effects *in vivo*. Taken together, these findings strongly support the view that a robust plant *in vitro* splicing system could be developed to uncover plant-specific splicing regulatory mechanisms.

It is also noteworthy that the efficiency of splicing in this *in vitro* system was rather low. This low efficiency limited our ability to proceed with more validation experiments such as sequencing of the spliced RNA and splicing intermediates. It is worth mentioning that the efficiency of other initial *in vitro* splicing assays was also low; however, researchers were able to develop highly efficient assays over time (Goldenberg and Hauser, 1983; Kole and Weissman, 1982; Reichert and Moore, 2000). Further investigations will be very helpful in gaining more insights for improving the efficiency of this plant *in vitro* splicing assay. In addition, it is

possible that the concentration of some spliceosomal proteins is less than optimum in our NE, leading to low efficiency. In animals, splicing deficient extracts can be made competent by adding one or more splicing factors such as SR proteins (Krainer et al., 1990; Mayeda and Krainer, 1999b). In future, one could purify one or more SR proteins and supplement the NE or use transgenic lines that are overexpressing one or more SR proteins to prepare NE to enhance splicing efficiency.

Conclusion

In summary, this work represents a first step toward developing a plant-derived *in vitro* pre-mRNA splicing assay. We show, for the first time, that NE derived from *Arabidopsis* etiolated seedlings is capable of splicing *LHCB3* pre-mRNA. This system that we developed will permit further refinement of a plant *in vitro* splicing assay system. Further confirmation of these results with additional approaches and optimization of this assay would lead to development of a robust *in vitro* splicing assay, which would open new avenues to investigate the spliceosome assembly and composition in plants, splicing regulatory mechanisms specific to plants, and thereby enhance the overall understanding of post-transcriptional gene regulatory mechanisms in eukaryotes.

REFERENCES

Aebi, M., Hornig, H., Padgett, R.A., Reiser, J., and Weissmann, C. (1986). Sequence requirements for splicing of higher eukaryotic nuclear pre-mRNA. Cell *47*, 555-565.

Aebi, M., Hornig, H., and Weissmann, C. (1987). 5' cleavage site in eukaryotic pre-mRNA splicing is determined by the overall 5' splice region, not by the conserved 5' GU. Cell *50*, 237-246.

Albaqami, M.M. (2013). Functional analyses of splice variants of the splicing regulator SR45 in abiotic stresses in Arabidopsis (Colorado State University).

Ali, G.S., Golovkin, M., and Reddy, A.S. (2003). Nuclear localization and in vivo dynamics of a plant-specific serine/arginine-rich protein. Plant J *36*, 883-893.

Ali, G.S., Palusa, S.G., Golovkin, M., Prasad, J., Manley, J.L., and Reddy, A.S. (2007). Regulation of plant developmental processes by a novel splicing factor. PLoS One *2*, e471.

Ali, G.S., Prasad, K.V., Hanumappa, M., and Reddy, A.S. (2008). Analyses of in vivo interaction and mobility of two spliceosomal proteins using FRAP and BiFC. PLoS One *3*, e1953.

Ali, G.S., and Reddy, A.S. (2006). ATP, phosphorylation and transcription regulate the mobility of plant splicing factors. J Cell Sci *119*, 3527-3538.

Atambayeva, S.A., Khailenko, V., and Ivashchenko, A. (2008). Intron and exon length variation in arabidopsis, rice, nematode, and human. Molecular biology *42*, 312-320.

Ausin, I., Greenberg, M.V., Li, C.F., and Jacobsen, S.E. (2012). The splicing factor SR45 affects the RNA-directed DNA methylation pathway in Arabidopsis. Epigenetics *7*, 29-33.

Barta, A., Kalyna, M., and Reddy, A.S. (2010). Implementing a rational and consistent nomenclature for serine/arginine-rich protein splicing factors (SR proteins) in plants. The Plant Cell *22*, 2926-2929.

Barta, A., Marquez, Y., and Brown, J.W. (2012). Challenges in plant alternative splicing. Alternative pre-mRNA Splicing: Theory and Protocols, 79-91.

Barta, A., Sommergruber, K., Thompson, D., Hartmuth, K., Matzke, M.A., and Matzke, A.J. (1986). The expression of a nopaline synthase - human growth hormone chimaeric gene in transformed tobacco and sunflower callus tissue. Plant Mol Biol *6*, 347-357.

Berget, S.M., Moore, C., and Sharp, P.A. (1977). Spliced segments at the 5' terminus of adenovirus 2 late mRNA. Proc Natl Acad Sci U S A *74*, 3171-3175.

Bertram, K., Agafonov, D.E., Liu, W.T., Dybkov, O., Will, C.L., Hartmuth, K., Urlaub, H., Kastner, B., Stark, H., and Luhrmann, R. (2017). Cryo-EM structure of a human spliceosome activated for step 2 of splicing. Nature *542*, 318-323.

Bita, C.E., and Gerats, T. (2013). Plant tolerance to high temperature in a changing environment: scientific fundamentals and production of heat stress-tolerant crops. Front Plant Sci *4*, 273.

Black, D.L., Chabot, B., and Steitz, J.A. (1985). U2 as well as U1 small nuclear ribonucleoproteins are involved in premessenger RNA splicing. Cell *42*, 737-750.

Bradley, T., Cook, M.E., and Blanchette, M. (2015). SR proteins control a complex network of RNA-processing events. RNA *21*, 75-92.

Brown, J.W., Marshall, D.F., and Echeverria, M. (2008). Intronic noncoding RNAs and splicing. Trends Plant Sci *13*, 335-342.

Brown, J.W., Simpson, C.G., Thow, G., Clark, G.P., Jennings, S.N., Medina-Escobar, N., Haupt, S., Chapman, S.C., and Oparka, K.J. (2002). Splicing signals and factors in plant intron removal. Biochem Soc Trans *30*, 146-149.

Cao, W., Jamison, S.F., and Garcia-Blanco, M.A. (1997). Both phosphorylation and dephosphorylation of ASF/SF2 are required for pre-mRNA splicing in vitro. RNA *3*, 1456-1467.

Carmona-Saez, P., Chagoyen, M., Tirado, F., Carazo, J.M., and Pascual-Montano, A. (2007). GENECODIS: a web-based tool for finding significant concurrent annotations in gene lists. Genome Biol *8*, R3.

Carvalho, R.F., Carvalho, S.D., and Duque, P. (2010). The plant-specific SR45 protein negatively regulates glucose and ABA signaling during early seedling development in Arabidopsis. Plant Physiol *154*, 772-783.

Carvalho, R.F., Szakonyi, D., Simpson, C.G., Barbosa, I.C., Brown, J.W., Baena-Gonzalez, E., and Duque, P. (2016). The Arabidopsis SR45 Splicing Factor, a Negative Regulator of Sugar Signaling, Modulates SNF1-Related Protein Kinase 1 Stability. Plant Cell *28*, 1910-1925.

Chen, M., and Manley, J.L. (2009). Mechanisms of alternative splicing regulation: insights from molecular and genomics approaches. Nat Rev Mol Cell Biol *10*, 741-754.

Chen, T., Cui, P., Chen, H., Ali, S., Zhang, S., and Xiong, L. (2013). A KH-domain RNA-binding protein interacts with FIERY2/CTD phosphatase-like 1 and splicing factors and is important for pre-mRNA splicing in Arabidopsis. PLoS Genet *9*, e1003875.

Chow, L.T., Gelinas, R.E., Broker, T.R., and Roberts, R.J. (1977). An amazing sequence arrangement at the 5' ends of adenovirus 2 messenger RNA. Cell *12*, 1-8.

Clark, A.J., Archibald, A.L., McClenaghan, M., Simons, J.P., Wallace, R., and Whitelaw, C.B. (1993). Enhancing the efficiency of transgene expression. Philos Trans R Soc Lond B Biol Sci *339*, 225-232.

Cowling, V.H., and Cole, M.D. (2010). Myc Regulation of mRNA Cap Methylation. Genes Cancer *1*, 576-579.

Cruz, T.M., Carvalho, R.F., Richardson, D.N., and Duque, P. (2014). Abscisic acid (ABA) regulation of Arabidopsis SR protein gene expression. Int J Mol Sci *15*, 17541-17564.

Czechowski, T., Stitt, M., Altmann, T., Udvardi, M.K., and Scheible, W.R. (2005). Genome-wide identification and testing of superior reference genes for transcript normalization in Arabidopsis. Plant Physiol *139*, 5-17.

Das, R., and Reed, R. (1999). Resolution of the mammalian E complex and the ATP-dependent spliceosomal complexes on native agarose mini-gels. RNA *5*, 1504-1508.

Day, I.S., Golovkin, M., Palusa, S.G., Link, A., Ali, G.S., Thomas, J., Richardson, D.N., and Reddy, A.S. (2012). Interactions of SR45, an SR-like protein, with spliceosomal proteins and an intronic sequence: insights into regulated splicing. Plant J *71*, 936-947.

de la Fuente van Bentem, S., Anrather, D., Roitinger, E., Djamei, A., Hufnagl, T., Barta, A., Csaszar, E., Dohnal, I., Lecourieux, D., and Hirt, H. (2006). Phosphoproteomics reveals extensive in vivo phosphorylation of Arabidopsis proteins involved in RNA metabolism. Nucleic Acids Res *34*, 3267-3278.

Dupuy, B., and Sonenshein, A.L. (1998). Regulated transcription of Clostridium difficile toxin genes. Mol Microbiol *27*, 107-120.

Edwalds-Gilbert, G. (2010). Regulation of mRNA Splicing by Signal Transduction. Nature Education.

Filichkin, S.A., Priest, H.D., Givan, S.A., Shen, R., Bryant, D.W., Fox, S.E., Wong, W.K., and Mockler, T.C. (2010). Genome-wide mapping of alternative splicing in Arabidopsis thaliana. Genome Res *20*, 45-58.

Floris, M., Mahgoub, H., Lanet, E., Robaglia, C., and Menand, B. (2009). Post-transcriptional regulation of gene expression in plants during abiotic stress. Int J Mol Sci *10*, 3168-3185.

Folta, K.M., and Kaufman, L.S. (2006). Isolation of Arabidopsis nuclei and measurement of gene transcription rates using nuclear run-on assays. Nat Protoc *1*, 3094-3100.

Fu, X.D., and Ares, M., Jr. (2014). Context-dependent control of alternative splicing by RNA-binding proteins. Nat Rev Genet *15*, 689-701.

Galej, W.P., Wilkinson, M.E., Fica, S.M., Oubridge, C., Newman, A.J., and Nagai, K. (2016). Cryo-EM structure of the spliceosome immediately after branching. Nature *537*, 197-201.

Garneau, N.L., Wilusz, J., and Wilusz, C.J. (2007). The highways and byways of mRNA decay. Nat Rev Mol Cell Biol *8*, 113-126.

Ge, H., and Manley, J.L. (1990). A protein factor, ASF, controls cell-specific alternative splicing of SV40 early pre-mRNA in vitro. Cell *62*, 25-34.

Gifford, M.L., Dean, A., Gutierrez, R.A., Coruzzi, G.M., and Birnbaum, K.D. (2008). Cell-specific nitrogen responses mediate developmental plasticity. Proc Natl Acad Sci U S A *105*, 803-808.

Gilbert, W. (1978). Why genes in pieces? Nature *271*, 501.

Glisovic, T., Bachorik, J.L., Yong, J., and Dreyfuss, G. (2008). RNA-binding proteins and post-transcriptional gene regulation. FEBS Lett *582*, 1977-1986.

Goldenberg, C.J., and Hauser, S.D. (1983). Accurate and efficient in vitro splicing of purified precursor RNAs specified by early region 2 of the adenovirus 2 genome. Nucleic Acids Res *11*, 1337-1348.

Golovkin, M., and Reddy, A.S. (1998). The plant U1 small nuclear ribonucleoprotein particle 70K protein interacts with two novel serine/arginine-rich proteins. Plant Cell *10*, 1637-1648.

Golovkin, M., and Reddy, A.S. (1999). An SC35-like protein and a novel serine/arginine-rich protein interact with Arabidopsis U1-70K protein. J Biol Chem *274*, 36428-36438.

Goodwin, S., McPherson, J.D., and McCombie, W.R. (2016). Coming of age: ten years of next-generation sequencing technologies. Nat Rev Genet *17*, 333-351.

Graveley, B.R. (2000). Sorting out the complexity of SR protein functions. RNA *6*, 1197-1211.

Guerra, D., Crosatti, C., Khoshro, H.H., Mastrangelo, A.M., Mica, E., and Mazzucotelli, E. (2015). Post-transcriptional and post-translational regulations of drought and heat response in plants: a spider's web of mechanisms. Front Plant Sci *6*, 57.

Guhaniyogi, J., and Brewer, G. (2001). Regulation of mRNA stability in mammalian cells. Gene *265*, 11-23.

Hernandez, N., and Keller, W. (1983). Splicing of in vitro synthesized messenger RNA precursors in HeLa cell extracts. Cell *35*, 89-99.

Hicks, M.J., Lam, B.J., and Hertel, K.J. (2005). Analyzing mechanisms of alternative pre-mRNA splicing using in vitro splicing assays. Methods *37*, 306-313.

Hirsch, J., Lefort, V., Vankersschaver, M., Boualem, A., Lucas, A., Thermes, C., d'Aubenton-Carafa, Y., and Crespi, M. (2006). Characterization of 43 non-protein-coding mRNA genes in Arabidopsis, including the MIR162a-derived transcripts. Plant Physiol *140*, 1192-1204.

Hoskins, A.A., and Moore, M.J. (2012). The spliceosome: a flexible, reversible macromolecular machine. Trends Biochem Sci *37*, 179-188.

Hsieh, L.C., Lin, S.I., Shih, A.C., Chen, J.W., Lin, W.Y., Tseng, C.Y., Li, W.H., and Chiou, T.J. (2009). Uncovering small RNA-mediated responses to phosphate deficiency in Arabidopsis by deep sequencing. Plant Physiol *151*, 2120-2132.

Hua, Y., Vickers, T.A., Okunola, H.L., Bennett, C.F., and Krainer, A.R. (2008). Antisense masking of an hnRNP A1/A2 intronic splicing silencer corrects SMN2 splicing in transgenic mice. Am J Hum Genet *82*, 834-848.

Irimia, M., and Roy, S.W. (2014). Origin of spliceosomal introns and alternative splicing. Cold Spring Harb Perspect Biol *6*.

Jones-Rhoades, M.W., and Bartel, D.P. (2004). Computational identification of plant microRNAs and their targets, including a stress-induced miRNA. Mol Cell *14*, 787-799.

Kabran, P., Rossignol, T., Gaillardin, C., Nicaud, J.M., and Neuveglise, C. (2012). Alternative splicing regulates targeting of malate dehydrogenase in Yarrowia lipolytica. DNA Res *19*, 231-244.

Kaida, D., Motoyoshi, H., Tashiro, E., Nojima, T., Hagiwara, M., Ishigami, K., Watanabe, H., Kitahara, T., Yoshida, T., Nakajima, H., *et al.* (2007). Spliceostatin A targets SF3b and inhibits both splicing and nuclear retention of pre-mRNA. Nat Chem Biol *3*, 576-583.

Kalyna, M., and Barta, A. (2004). A plethora of plant serine/arginine-rich proteins: redundancy or evolution of novel gene functions? Biochem Soc Trans *32*, 561-564.

Kalyna, M., Lopato, S., and Barta, A. (2003). Ectopic expression of atRSZ33 reveals its function in splicing and causes pleiotropic changes in development. Mol Biol Cell *14*, 3565-3577.

Kalyna, M., Simpson, C.G., Syed, N.H., Lewandowska, D., Marquez, Y., Kusenda, B., Marshall, J., Fuller, J., Cardle, L., McNicol, J., *et al.* (2012). Alternative splicing and nonsense-mediated decay modulate expression of important regulatory genes in Arabidopsis. Nucleic Acids Res *40*, 2454-2469.

Kataoka, N., and Dreyfuss, G. (2008). Preparation of efficient splicing extracts from whole cells, nuclei, and cytoplasmic fractions. Methods Mol Biol *488*, 357-365.

Kim, E., Magen, A., and Ast, G. (2007). Different levels of alternative splicing among eukaryotes. Nucleic Acids Res *35*, 125-131.

Kinoshita, E., Kinoshita-Kikuta, E., and Koike, T. (2009). Separation and detection of large phosphoproteins using Phos-tag SDS-PAGE. Nat Protoc *4*, 1513-1521.

Kole, R., and Weissman, S.M. (1982). Accurate in vitro splicing of human beta-globin RNA. Nucleic Acids Res *10*, 5429-5445.

Konarska, M.M., and Sharp, P.A. (1986). Electrophoretic separation of complexes involved in the splicing of precursors to mRNAs. Cell *46*, 845-855.

Koncz, C., Dejong, F., Villacorta, N., Szakonyi, D., and Koncz, Z. (2012). The spliceosome-activating complex: molecular mechanisms underlying the function of a pleiotropic regulator. Front Plant Sci *3*, 9.

Krainer, A.R., Conway, G.C., and Kozak, D. (1990). Purification and characterization of pre-mRNA splicing factor SF2 from HeLa cells. Genes Dev *4*, 1158-1171.

Krainer, A.R., and Maniatis, T. (1985). Multiple factors including the small nuclear ribonucleoproteins U1 and U2 are necessary for pre-mRNA splicing in vitro. Cell *42*, 725-736.

Krainer, A.R., Maniatis, T., Ruskin, B., and Green, M.R. (1984). Normal and mutant human beta-globin pre-mRNAs are faithfully and efficiently spliced in vitro. Cell *36*, 993-1005.

Laemmli, U.K. (1970). Cleavage of structural proteins during the assembly of the head of bacteriophage T4. Nature *227*, 680-685.

Lee, K., and Kang, H. (2016). Emerging Roles of RNA-Binding Proteins in Plant Growth, Development, and Stress Responses. Mol Cells *39*, 179-185.
Lee, Y., and Rio, D.C. (2015). Mechanisms and Regulation of Alternative Pre-mRNA Splicing. Annu Rev Biochem *84*, 291-323.
Lewis, J.D., and Izaurralde, E. (1997). The role of the cap structure in RNA processing and nuclear export. Eur J Biochem *247*, 461-469.
Lin, R.J., Lustig, A.J., and Abelson, J. (1987). Splicing of yeast nuclear pre-mRNA in vitro requires a functional 40S spliceosome and several extrinsic factors. Genes Dev *1*, 7-18.
Lin, R.J., Newman, A.J., Cheng, S.C., and Abelson, J. (1985). Yeast mRNA splicing in vitro. J Biol Chem *260*, 14780-14792.
Lin, S., and Fu, X.D. (2007). SR proteins and related factors in alternative splicing. Adv Exp Med Biol *623*, 107-122.
Liu, H.X., Zhang, M., and Krainer, A.R. (1998). Identification of functional exonic splicing enhancer motifs recognized by individual SR proteins. Genes Dev *12*, 1998-2012.
Liu, J., Sun, N., Liu, M., Liu, J., Du, B., Wang, X., and Qi, X. (2013). An autoregulatory loop controlling Arabidopsis HsfA2 expression: role of heat shock-induced alternative splicing. Plant Physiol *162*, 512-521.
Long, J.C., and Caceres, J.F. (2009). The SR protein family of splicing factors: master regulators of gene expression. Biochem J *417*, 15-27.
Lopato, S., Kalyna, M., Dorner, S., Kobayashi, R., Krainer, A.R., and Barta, A. (1999). atSRp30, one of two SF2/ASF-like proteins from Arabidopsis thaliana, regulates splicing of specific plant genes. Genes Dev *13*, 987-1001.
Lorkovic, Z.J. (2009). Role of plant RNA-binding proteins in development, stress response and genome organization. Trends Plant Sci *14*, 229-236.
Lorković, Z.J., Kirk, D.A.W., Lambermon, M.H., and Filipowicz, W. (2000). Pre-mRNA splicing in higher plants. Trends in plant science *5*, 160-167.
Lowry, O.H., Rosebrough, N.J., Farr, A.L., and Randall, R.J. (1951). Protein measurement with the Folin phenol reagent. J Biol Chem *193*, 265-275.
Lundmark, M., Korner, C.J., and Nielsen, T.H. (2010). Global analysis of microRNA in Arabidopsis in response to phosphate starvation as studied by locked nucleic acid-based microarrays. Physiol Plant *140*, 57-68.
Marquez, Y., Brown, J.W., Simpson, C., Barta, A., and Kalyna, M. (2012). Transcriptome survey reveals increased complexity of the alternative splicing landscape in Arabidopsis. Genome Res *22*, 1184-1195.
Martinez-Contreras, R., Cloutier, P., Shkreta, L., Fisette, J.F., Revil, T., and Chabot, B. (2007). hnRNP proteins and splicing control. Adv Exp Med Biol *623*, 123-147.
Mastrangelo, A.M., Marone, D., Laido, G., De Leonardis, A.M., and De Vita, P. (2012). Alternative splicing: enhancing ability to cope with stress via transcriptome plasticity. Plant Sci *185-186*, 40-49.
Matera, A.G., and Wang, Z. (2014). A day in the life of the spliceosome. Nat Rev Mol Cell Biol *15*, 108-121.
Matlin, A.J., Clark, F., and Smith, C.W. (2005). Understanding alternative splicing: towards a cellular code. Nat Rev Mol Cell Biol *6*, 386-398.
Mayeda, A., Badolato, J., Kobayashi, R., Zhang, M.Q., Gardiner, E.M., and Krainer, A.R. (1999). Purification and characterization of human RNPS1: a general activator of pre-mRNA splicing. EMBO J *18*, 4560-4570.

Mayeda, A., and Krainer, A.R. (1992). Regulation of alternative pre-mRNA splicing by hnRNP A1 and splicing factor SF2. Cell *68*, 365-375.
Mayeda, A., and Krainer, A.R. (1999a). Mammalian in vitro splicing assays. Methods Mol Biol *118*, 315-321.
Mayeda, A., and Krainer, A.R. (1999b). Preparation of HeLa cell nuclear and cytosolic S100 extracts for in vitro splicing. Methods Mol Biol *118*, 309-314.
Mazzucotelli, E., Mastrangelo, A.M., Crosatti, C., Guerra, D., Stanca, A.M., and Cattivelli, L. (2008). Abiotic stress response in plants: when post-transcriptional and post-translational regulations control transcription. Plant Science *174*, 420-431.
Meyer, K., Koester, T., and Staiger, D. (2015). Pre-mRNA Splicing in Plants: In Vivo Functions of RNA-Binding Proteins Implicated in the Splicing Process. Biomolecules *5*, 1717-1740.
Misteli, T. (1999). RNA splicing: What has phosphorylation got to do with it? Curr Biol *9*, R198-200.
Morello, L., and Breviario, D. (2008). Plant spliceosomal introns: not only cut and paste. Curr Genomics *9*, 227-238.
Movassat, M., Mueller, W.F., and Hertel, K.J. (2014). In vitro assay of pre-mRNA splicing in mammalian nuclear extract. Methods Mol Biol *1126*, 151-160.
Ner-Gaon, H., Halachmi, R., Savaldi-Goldstein, S., Rubin, E., Ophir, R., and Fluhr, R. (2004). Intron retention is a major phenomenon in alternative splicing in Arabidopsis. Plant J *39*, 877-885.
Nilsen, T.W. (2003). The spliceosome: the most complex macromolecular machine in the cell? Bioessays *25*, 1147-1149.
Padgett, R.A., Grabowski, P.J., Konarska, M.M., Seiler, S., and Sharp, P.A. (1986). Splicing of messenger RNA precursors. Annu Rev Biochem *55*, 1119-1150.
Padgett, R.A., Hardy, S.F., and Sharp, P.A. (1983). Splicing of adenovirus RNA in a cell-free transcription system. Proc Natl Acad Sci U S A *80*, 5230-5234.
Padgett, R.A., Konarska, M.M., Grabowski, P.J., Hardy, S.F., and Sharp, P.A. (1984). Lariat RNA's as intermediates and products in the splicing of messenger RNA precursors. Science *225*, 898-903.
Palusa, S.G., Ali, G.S., and Reddy, A.S. (2007). Alternative splicing of pre-mRNAs of Arabidopsis serine/arginine-rich proteins: regulation by hormones and stresses. Plant J *49*, 1091-1107.
Palusa, S.G., and Reddy, A.S. (2010). Extensive coupling of alternative splicing of pre-mRNAs of serine/arginine (SR) genes with nonsense-mediated decay. New Phytol *185*, 83-89.
Palusa, S.G., and Reddy, A.S. (2013). Analysis of RNA-protein Interactions Using Electrophoretic Mobility Shift Assay (Gel Shift Assay). Bio protocol *22(3), 1-10*.
Pan, Q., Shai, O., Lee, L.J., Frey, B.J., and Blencowe, B.J. (2008). Deep surveying of alternative splicing complexity in the human transcriptome by high-throughput sequencing. Nat Genet *40*, 1413-1415.
Papasaikas, P., and Valcarcel, J. (2016). The Spliceosome: The Ultimate RNA Chaperone and Sculptor. Trends Biochem Sci *41*, 33-45.
Proudfoot, N.J. (2011). Ending the message: poly(A) signals then and now. Genes Dev *25*, 1770-1782.
Rearick, D., Prakash, A., McSweeny, A., Shepard, S.S., Fedorova, L., and Fedorov, A. (2011). Critical association of ncRNA with introns. Nucleic Acids Res *39*, 2357-2366.

Reddy, A.S. (2004). Plant serine/arginine-rich proteins and their role in pre-mRNA splicing. Trends Plant Sci *9*, 541-547.
Reddy, A.S. (2007). Alternative splicing of pre-messenger RNAs in plants in the genomic era. Annu Rev Plant Biol *58*, 267-294.
Reddy, A.S., Day, I.S., Gohring, J., and Barta, A. (2012). Localization and dynamics of nuclear speckles in plants. Plant Physiol *158*, 67-77.
Reddy, A.S., Marquez, Y., Kalyna, M., and Barta, A. (2013). Complexity of the alternative splicing landscape in plants. Plant Cell *25*, 3657-3683.
Reed, R., and Maniatis, T. (1986). A role for exon sequences and splice-site proximity in splice-site selection. Cell *46*, 681-690.
Reichert, V., and Moore, M.J. (2000). Better conditions for mammalian in vitro splicing provided by acetate and glutamate as potassium counterions. Nucleic Acids Res *28*, 416-423.
Reiland, S., Messerli, G., Baerenfaller, K., Gerrits, B., Endler, A., Grossmann, J., Gruissem, W., and Baginsky, S. (2009). Large-scale Arabidopsis phosphoproteome profiling reveals novel chloroplast kinase substrates and phosphorylation networks. Plant Physiol *150*, 889-903.
Remy, E., Cabrito, T.R., Baster, P., Batista, R.A., Teixeira, M.C., Friml, J., Sa-Correia, I., and Duque, P. (2013). A major facilitator superfamily transporter plays a dual role in polar auxin transport and drought stress tolerance in Arabidopsis. Plant Cell *25*, 901-926.
Rino, J., and Carmo-Fonseca, M. (2009). The spliceosome: a self-organized macromolecular machine in the nucleus? Trends Cell Biol *19*, 375-384.
Rio, D.C. (1988). Accurate and efficient pre-mRNA splicing in Drosophila cell-free extracts. Proc Natl Acad Sci U S A *85*, 2904-2908.
Ritchie, D.B., Schellenberg, M.J., and MacMillan, A.M. (2009). Spliceosome structure: piece by piece. Biochim Biophys Acta *1789*, 624-633.
Rivero, L., Scholl, R., Holomuzki, N., Crist, D., Grotewold, E., and Brkljacic, J. (2014). Handling Arabidopsis plants: growth, preservation of seeds, transformation, and genetic crosses. Methods Mol Biol *1062*, 3-25.
Rogers, M.F., Thomas, J., Reddy, A.S., and Ben-Hur, A. (2012). SpliceGrapher: detecting patterns of alternative splicing from RNA-Seq data in the context of gene models and EST data. Genome Biol *13*, R4.
Rogozin, I.B., Carmel, L., Csuros, M., and Koonin, E.V. (2012). Origin and evolution of spliceosomal introns. Biol Direct *7*, 11.
Rose, A.B. (2008). Intron-mediated regulation of gene expression. Curr Top Microbiol Immunol *326*, 277-290.
Ru, Y., Wang, B.B., and Brendel, V. (2008). Spliceosomal proteins in plants. Curr Top Microbiol Immunol *326*, 1-15.
Ruskin, B., Krainer, A.R., Maniatis, T., and Green, M.R. (1984). Excision of an intact intron as a novel lariat structure during pre-mRNA splicing in vitro. Cell *38*, 317-331.
Savaldi-Goldstein, S., Aviv, D., Davydov, O., and Fluhr, R. (2003). Alternative splicing modulation by a LAMMER kinase impinges on developmental and transcriptome expression. Plant Cell *15*, 926-938.
Schaal, T.D., and Maniatis, T. (1999). Selection and characterization of pre-mRNA splicing enhancers: identification of novel SR protein-specific enhancer sequences. Mol Cell Biol *19*, 1705-1719.

Screaton, G.R., Caceres, J.F., Mayeda, A., Bell, M.V., Plebanski, M., Jackson, D.G., Bell, J.I., and Krainer, A.R. (1995). Identification and characterization of three members of the human SR family of pre-mRNA splicing factors. EMBO J *14*, 4336-4349.

Shefer, K., Sperling, J., and Sperling, R. (2014). The Supraspliceosome - A Multi-Task Machine for Regulated Pre-mRNA Processing in the Cell Nucleus. Comput Struct Biotechnol J *11*, 113-122.

Shi, Y., and Manley, J.L. (2007). A complex signaling pathway regulates SRp38 phosphorylation and pre-mRNA splicing in response to heat shock. Mol Cell *28*, 79-90.

Shin, C., Feng, Y., and Manley, J.L. (2004). Dephosphorylated SRp38 acts as a splicing repressor in response to heat shock. Nature *427*, 553-558.

Shukla, R.R., Dominski, Z., Zwierzynski, T., and Kole, R. (1990). Inactivation of splicing factors in HeLa cells subjected to heat shock. J Biol Chem *265*, 20377-20383.

Silva-Correia, J., Freitas, S., Tavares, R.M., Lino-Neto, T., and Azevedo, H. (2014). Phenotypic analysis of the Arabidopsis heat stress response during germination and early seedling development. Plant Methods *10*, 7.

Singh, K.B. (1998). Transcriptional regulation in plants: the importance of combinatorial control. Plant Physiol *118*, 1111-1120.

Staiger, D., and Brown, J.W. (2013). Alternative splicing at the intersection of biological timing, development, and stress responses. Plant Cell *25*, 3640-3656.

Stamm, S., Ben-Ari, S., Rafalska, I., Tang, Y., Zhang, Z., Toiber, D., Thanaraj, T.A., and Soreq, H. (2005). Function of alternative splicing. Gene *344*, 1-20.

Stief, A., Altmann, S., Hoffmann, K., Pant, B.D., Scheible, W.R., and Baurle, I. (2014). Arabidopsis miR156 Regulates Tolerance to Recurring Environmental Stress through SPL Transcription Factors. Plant Cell *26*, 1792-1807.

Stoss, O., Schwaiger, F.W., Cooper, T.A., and Stamm, S. (1999). Alternative splicing determines the intracellular localization of the novel nuclear protein Nop30 and its interaction with the splicing factor SRp30c. J Biol Chem *274*, 10951-10962.

Sugiura, M. (1997). Plant in Vitro Transcription Systems. Annu Rev Plant Physiol Plant Mol Biol *48*, 383-398.

Suzuki, H., Zuo, Y., Wang, J., Zhang, M.Q., Malhotra, A., and Mayeda, A. (2006). Characterization of RNase R-digested cellular RNA source that consists of lariat and circular RNAs from pre-mRNA splicing. Nucleic Acids Res *34*, e63.

Syed, N.H., Kalyna, M., Marquez, Y., Barta, A., and Brown, J.W. (2012). Alternative splicing in plants--coming of age. Trends Plant Sci *17*, 616-623.

Tian, M., and Maniatis, T. (1992). Positive control of pre-mRNA splicing in vitro. Science *256*, 237-240.

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

Trapnell, C., Roberts, A., Goff, L., Pertea, G., Kim, D., Kelley, D.R., Pimentel, H., Salzberg, S.L., Rinn, J.L., and Pachter, L. (2012). Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and Cufflinks. Nat Protoc *7*, 562-578.

Turunen, J.J., Niemela, E.H., Verma, B., and Frilander, M.J. (2013). The significant other: splicing by the minor spliceosome. Wiley Interdiscip Rev RNA *4*, 61-76.

Twyffels, L., Gueydan, C., and Kruys, V. (2011). Shuttling SR proteins: more than splicing factors. FEBS J *278*, 3246-3255.

Ule, J., Stefani, G., Mele, A., Ruggiu, M., Wang, X., Taneri, B., Gaasterland, T., Blencowe, B.J., and Darnell, R.B. (2006). An RNA map predicting Nova-dependent splicing regulation. Nature *444*, 580-586.

Wahl, M.C., Will, C.L., and Luhrmann, R. (2009). The spliceosome: design principles of a dynamic RNP machine. Cell *136*, 701-718.

Wang, B., Sun, Y.F., Song, N., Wei, J.P., Wang, X.J., Feng, H., Yin, Z.Y., and Kang, Z.S. (2014). MicroRNAs involving in cold, wounding and salt stresses in Triticum aestivum L. Plant Physiol Biochem *80*, 90-96.

Wang, B.B., and Brendel, V. (2004). The ASRG database: identification and survey of Arabidopsis thaliana genes involved in pre-mRNA splicing. Genome Biol *5*, R102.

Wang, B.B., and Brendel, V. (2006). Genomewide comparative analysis of alternative splicing in plants. Proc Natl Acad Sci U S A *103*, 7175-7180.

Wang, E.T., Sandberg, R., Luo, S., Khrebtukova, I., Zhang, L., Mayr, C., Kingsmore, S.F., Schroth, G.P., and Burge, C.B. (2008). Alternative isoform regulation in human tissue transcriptomes. Nature *456*, 470-476.

Wang, K., Singh, D., Zeng, Z., Coleman, S.J., Huang, Y., Savich, G.L., He, X., Mieczkowski, P., Grimm, S.A., Perou, C.M., *et al.* (2010). MapSplice: accurate mapping of RNA-seq reads for splice junction discovery. Nucleic Acids Res *38*, e178.

Will, C.L., and Luhrmann, R. (2011). Spliceosome structure and function. Cold Spring Harb Perspect Biol *3*.

Wu, S.H. (2014). Gene expression regulation in photomorphogenesis from the perspective of the central dogma. Annu Rev Plant Biol *65*, 311-333.

Xiao, S.H., and Manley, J.L. (1997). Phosphorylation of the ASF/SF2 RS domain affects both protein-protein and protein-RNA interactions and is necessary for splicing. Genes Dev *11*, 334-344.

Xing, D., Wang, Y., Hamilton, M., Ben-Hur, A., and Reddy, A.S. (2015). Transcriptome-Wide Identification of RNA Targets of Arabidopsis SERINE/ARGININE-RICH45 Uncovers the Unexpected Roles of This RNA Binding Protein in RNA Processing. Plant Cell *27*, 3294-3308.

Yan, C., Hang, J., Wan, R., Huang, M., Wong, C.C., and Shi, Y. (2015). Structure of a yeast spliceosome at 3.6-angstrom resolution. Science *349*, 1182-1191.

Yan, C., Wan, R., Bai, R., Huang, G., and Shi, Y. (2017a). Structure of a yeast step II catalytically activated spliceosome. Science *355*, 149-155.

Yan, K., Liu, P., Wu, C.A., Yang, G.D., Xu, R., Guo, Q.H., Huang, J.G., and Zheng, C.C. (2012). Stress-induced alternative splicing provides a mechanism for the regulation of microRNA processing in Arabidopsis thaliana. Mol Cell *48*, 521-531.

Yan, Q., Xia, X., Sun, Z., and Fang, Y. (2017b). Depletion of Arabidopsis SC35 and SC35-like serine/arginine-rich proteins affects the transcription and splicing of a subset of genes. PLoS Genet *13*, e1006663.

Yang, G.D., Yan, K., Wu, B.J., Wang, Y.H., Gao, Y.X., and Zheng, C.C. (2012). Genomewide analysis of intronic microRNAs in rice and Arabidopsis. J Genet *91*, 313-324.

Zhang, X.N., Mo, C., Garrett, W.M., and Cooper, B. (2014). Phosphothreonine 218 is required for the function of SR45.1 in regulating flower petal development in Arabidopsis. Plant Signal Behav *9*, e29134.

Zhang, X.N., and Mount, S.M. (2009). Two alternatively spliced isoforms of the Arabidopsis SR45 protein have distinct roles during normal plant development. Plant Physiol *150,* 1450-1458.

Zhou, Z., and Fu, X.D. (2013). Regulation of splicing by SR proteins and SR protein-specific kinases. Chromosoma *122,* 191-207.