

ALTERNATIVE SPLICING REGULATION IN PLANTS

EDITED BY: Ezequiel Petrillo, Maria Kalyna, Craig G. Simpson, Shih-Long Tu
and Kranthi Kiran Mandadi
PUBLISHED IN: Frontiers in Plant Science





frontiers

Frontiers eBook Copyright Statement

The copyright in the text of individual articles in this eBook is the property of their respective authors or their respective institutions or funders. The copyright in graphics and images within each article may be subject to copyright of other parties. In both cases this is subject to a license granted to Frontiers.

The compilation of articles constituting this eBook is the property of Frontiers.

Each article within this eBook, and the eBook itself, are published under the most recent version of the Creative Commons CC-BY licence.

The version current at the date of publication of this eBook is CC-BY 4.0. If the CC-BY licence is updated, the licence granted by Frontiers is automatically updated to the new version.

When exercising any right under the CC-BY licence, Frontiers must be attributed as the original publisher of the article or eBook, as applicable.

Authors have the responsibility of ensuring that any graphics or other materials which are the property of others may be included in the CC-BY licence, but this should be checked before relying on the CC-BY licence to reproduce those materials. Any copyright notices relating to those materials must be complied with.

Copyright and source acknowledgement notices may not be removed and must be displayed in any copy, derivative work or partial copy which includes the elements in question.

All copyright, and all rights therein, are protected by national and international copyright laws. The above represents a summary only. For further information please read Frontiers' Conditions for Website Use and Copyright Statement, and the applicable CC-BY licence.

ISSN 1664-8714

ISBN 978-2-88963-974-8

DOI 10.3389/978-2-88963-974-8

About Frontiers

Frontiers is more than just an open-access publisher of scholarly articles: it is a pioneering approach to the world of academia, radically improving the way scholarly research is managed. The grand vision of Frontiers is a world where all people have an equal opportunity to seek, share and generate knowledge. Frontiers provides immediate and permanent online open access to all its publications, but this alone is not enough to realize our grand goals.

Frontiers Journal Series

The Frontiers Journal Series is a multi-tier and interdisciplinary set of open-access, online journals, promising a paradigm shift from the current review, selection and dissemination processes in academic publishing. All Frontiers journals are driven by researchers for researchers; therefore, they constitute a service to the scholarly community. At the same time, the Frontiers Journal Series operates on a revolutionary invention, the tiered publishing system, initially addressing specific communities of scholars, and gradually climbing up to broader public understanding, thus serving the interests of the lay society, too.

Dedication to Quality

Each Frontiers article is a landmark of the highest quality, thanks to genuinely collaborative interactions between authors and review editors, who include some of the world's best academicians. Research must be certified by peers before entering a stream of knowledge that may eventually reach the public - and shape society; therefore, Frontiers only applies the most rigorous and unbiased reviews.

Frontiers revolutionizes research publishing by freely delivering the most outstanding research, evaluated with no bias from both the academic and social point of view. By applying the most advanced information technologies, Frontiers is catapulting scholarly publishing into a new generation.

What are Frontiers Research Topics?

Frontiers Research Topics are very popular trademarks of the Frontiers Journals Series: they are collections of at least ten articles, all centered on a particular subject. With their unique mix of varied contributions from Original Research to Review Articles, Frontiers Research Topics unify the most influential researchers, the latest key findings and historical advances in a hot research area! Find out more on how to host your own Frontiers Research Topic or contribute to one as an author by contacting the Frontiers Editorial Office: researchtopics@frontiersin.org

ALTERNATIVE SPLICING REGULATION IN PLANTS

Topic Editors:

Ezequiel Petrillo, CONICET Institute of Physiology, Molecular Biology and Neurosciences (IFIBYNE), Argentina

Maria Kalyna, University of Natural Resources and Life Sciences Vienna, Austria

Craig G. Simpson, The James Hutton Institute, United Kingdom

Shih-Long Tu, Institute of Plant and Microbial Biology, Academia Sinica, Taiwan

Kranthi Kiran Mandadi, Texas A&M University, United States

Citation: Petrillo, E., Kalyna, M., Simpson, C. G., Tu, S.-L., Mandadi, K. K., eds. (2020). Alternative Splicing Regulation in Plants. Lausanne: Frontiers Media SA. doi: 10.3389/978-2-88963-974-8

Table of Contents

- 05 Editorial: Alternative Splicing Regulation in Plants**
Ezequiel Petrillo, Maria Kalyna, Kranthi K. Mandadi, Shih-Long Tu and Craig G. Simpson
- 08 Alternative Splicing and Transcription Elongation in Plants**
Micaela A. Godoy Herz and Alberto R. Kornblihtt
- 13 Alternative RNA Splicing Expands the Developmental Plasticity of Flowering Transition**
Young-Joon Park, June-Hee Lee, Jae Young Kim and Chung-Mo Park
- 20 Expanding Alternative Splicing Identification by Integrating Multiple Sources of Transcription Data in Tomato**
Sarah Clark, Feng Yu, Lianfeng Gu and Xiang Jia Min
- 32 New Era in Plant Alternative Splicing Analysis Enabled by Advances in High-Throughput Sequencing (HTS) Technologies**
Renesh Bedre, Sonia Irigoyen, Ezequiel Petrillo and Kranthi K. Mandadi
- 37 Evolution of Alternative Splicing in Eudicots**
Zhihao Ling, Thomas Brockmüller, Ian T. Baldwin and Shuqing Xu
- 55 Alternative Splicing and Protein Diversity: Plants Versus Animals**
Saurabh Chaudhary, Waqas Khokhar, Ibteham Jabre, Anireddy S. N. Reddy, Lee J. Byrne, Cornelia M. Wilson and Naeem H. Syed
- 69 A Role for Pre-mRNA-PROCESSING PROTEIN 40C in the Control of Growth, Development, and Stress Tolerance in Arabidopsis thaliana**
Carlos Esteban Hernando, Mariano García Hourquet, María José de Leone, Daniel Careno, Javier Iserete, Santiago Mora Garcia and Marcelo Javier Yanovsky
- 83 Alternative Splicing Regulation During Light-Induced Germination of Arabidopsis thaliana Seeds**
Rocío Soledad Tognacca, Lucas Servi, Carlos Esteban Hernando, Maite Saura-Sanchez, Marcelo Javier Yanovsky, Ezequiel Petrillo and Javier Francisco Botto
- 95 Quantitative Proteomics Reveals a Role for SERINE/ARGININE-Rich 45 in Regulating RNA Metabolism and Modulating Transcriptional Suppression via the ASAP Complex in Arabidopsis thaliana**
Samuel L. Chen, Timothy J. Rooney, Anna R. Hu, Hunter S. Beard, Wesley M. Garrett, Leann M. Mangalath, Jordan J. Powers, Bret Cooper and Xiao-Ning Zhang
- 109 To Splice or to Transcribe: SKIP-Mediated Environmental Fitness and Development in Plants**
Ying Cao and Ligeng Ma
- 116 Genome-Wide Identification of Splicing Quantitative Trait Loci (sQTLs) in Diverse Ecotypes of Arabidopsis thaliana**
Waqas Khokhar, Musa A. Hassan, Anireddy S. N. Reddy, Saurabh Chaudhary, Ibteham Jabre, Lee J. Byrne and Naeem H. Syed

- 129** *Alternative Splicing of Circadian Clock Genes Correlates With Temperature in Field-Grown Sugarcane*
Luíza L. B. Dantas, Cristiane P. G. Calixto, Maira M. Dourado, Monalisa S. Carneiro, John W. S. Brown and Carlos T. Hotta
- 144** *Thermo-Sensitive Alternative Splicing of FLOWERING LOCUS M is Modulated by Cyclin-Dependent Kinase G2*
Candida Nibau, Marçal Gallemí, Despoina Dadarou, John H. Doonan and Nicola Cavallari
- 158** *Alternative Splicing and DNA Damage Response in Plants*
Barbara Anna Nimeth, Stefan Riegler and Maria Kalyna
- 167** *Current Challenges in Studying Alternative Splicing in Plants: The Case of Physcomitrella patens SR Proteins*
José Pedro Melo, Maria Kalyna and Paula Duque



Editorial: Alternative Splicing Regulation in Plants

Ezequiel Petrillo^{1,2*†}, Maria Kalyna^{3*†}, Kranthi K. Mandadi^{4,5*†}, Shih-Long Tu^{6*†} and Craig G. Simpson^{7*†}

OPEN ACCESS

Edited by:

Zhong-Nan Yang,
Shanghai Normal University, China

Reviewed by:

Xin-Qi Gao,
Shandong Agricultural
University, China

*Correspondence:

Ezequiel Petrillo
petry@fbmc.fcen.uba.ar
Maria Kalyna
mariya.kalyna@boku.ac.at
Kranthi K. Mandadi
kkmandadi@tamu.edu
Shih-Long Tu
tsl@gate.sinica.edu.tw
Craig G. Simpson
craig.simpson@hutton.ac.uk

†ORCID:

Ezequiel Petrillo
orcid.org/0000-0002-0025-3708
Maria Kalyna
orcid.org/0000-0003-4702-7625
Kranthi K. Mandadi
orcid.org/0000-0003-2986-4016
Shih-Long Tu
orcid.org/0000-0001-9436-278X
Craig G. Simpson
orcid.org/0000-0002-1723-1492

Specialty section:

This article was submitted to
Plant Physiology,
a section of the journal
Frontiers in Plant Science

Received: 21 April 2020

Accepted: 04 June 2020

Published: 09 July 2020

Citation:

Petrillo E, Kalyna M, Mandadi KK,
Tu S-L and Simpson CG (2020)
Editorial: Alternative Splicing
Regulation in Plants.
Front. Plant Sci. 11:913.
doi: 10.3389/fpls.2020.00913

¹ Departamento de Fisiología, Biología Molecular y Celular, Facultad de Ciencias Exactas y Naturales, Universidad de Buenos Aires, Buenos Aires, Argentina, ² CONICET-Universidad de Buenos Aires, Instituto de Fisiología, Biología Molecular y Neurociencias, Buenos Aires, Argentina, ³ Department of Applied Genetics and Cell Biology, BOKU—University of Natural Resources and Life Sciences, Vienna, Austria, ⁴ Texas A&M AgriLife Research and Extension Center, Texas A&M University, Weslaco, TX, United States, ⁵ Department of Plant Pathology and Microbiology, Texas A&M University, College Station, TX, United States, ⁶ Institute of Plant and Microbial Biology, Academia Sinica, Taipei, Taiwan, ⁷ Cell and Molecular Sciences, James Hutton Institute, Dundee, United Kingdom

Keywords: splicing factor, development, stress, adaptation, evolution, environment, flowering, germination

Editorial on the Research Topic

Alternative Splicing Regulation in Plants

Research on alternative splicing (AS) in plants has bloomed during the past decade, largely fueled by the advance of high-throughput sequencing (HTS) technologies and pioneering papers demonstrating an unexpectedly high frequency of AS in plants (Filichkin et al., 2010; Lu et al., 2010; Marquez et al., 2012; Syed et al., 2012). The formation and regulation of alternative transcripts from individual transcribed genes by alternative splice site selection pervades all aspects of a eukaryote's development and adaptive response to its changing environment. This is particularly relevant to sessile plant species that must be able to rapidly respond to abiotic, biotic, diurnal, and seasonal fluctuations. The mechanism of pre-mRNA splicing and the process of splice site selection has existed since its divergence from metazoans (Fedorov et al., 2002) and is regulated by splicing factors that are components of the assembling spliceosome. Many knowledge gaps remain to be addressed, not only to define the AS prevalence in different plant species and its impact on various biological processes, but also to understand its mechanistic basis with the aim of manipulating crops for important traits required for food security. Here, we share with the plant biology community a Research Topic that aims to showcase current findings, emerging questions, and technical advances, in the field of AS in plants.

The basic mechanisms underlying AS regulation in different eukaryotes are quite conserved. Core sequence elements are found at the 5' splice site, 3' splice site and branchpoint located upstream of the 3' splice site (Wahl et al., 2009; Meyer et al., 2015). Splicing enhancers and silencers positioned in introns and exons further define selected splice sites. These sequence elements are relatively short and variable such that multiple alternative sequences exist and, along with variability in the expression of splicing factors, lead to AS. However, there are splicing differences between plants and animals that highlight peculiarities in gene and chromatin architecture, transcription, and splicing machineries. Chaudhary et al. review differences in AS mechanisms and its regulation in plants and humans. In animals, it is widely accepted that splicing and transcription are coupled, and emerging evidence indicates that it is conserved in plants (Chaudhary et al.). The co-transcriptional behavior of splicing means that the chromatin environment and the RNA Polymerase II (RNAPII) processivity have a strong influence on splicing outcomes. In plants, light/dark transitions modulate the RNAPII elongation rate, which in turn controls AS, demonstrating a coordination between AS, transcription and plant growth

(Godoy Herz and Kornblihtt). Cao and Ma summarized the role of a highly conserved SKI-INTERACTING PROTEIN (SKIP), which functions both as a transcription factor and as a general splicing factor. It is not only required for precise and efficient splicing of pre-mRNAs on a genome-wide scale, but also delays flowering time by promoting transcription of the flowering repressor *FLC* (Cao and Ma). SKIP also interacts with ELF7, an RNAPII-associated factor 1 complex (Paf1c) component, that regulates transcription elongation. Cao and Ma further outlined SR45 (RNPS1 human ortholog) as a multifunctional splicing factor, while Chen et al. described a proteomic study of the *sr45-1* mutant that identified this factor as a component of the apoptosis and splicing-associated protein complex (ASAP), which is known to regulate RNA metabolism at multiple levels by recruiting histone deacetylases to the *FLC* locus. The *sr45-1* mutant also showed a significant reduction in Sin3-associated protein 18 (SAP18), a component of ASAP, which induces transcriptional silencing in mammalian cells. Lastly, yeast PRE-mRNA-PROCESSING PROTEIN 40 (PRP40) has a role in the early steps of spliceosome complex formation, and the human homolog was initially discovered as a transcriptional modulator and later linked to pre-mRNA splicing. Hernando et al. characterized a mutant of the *Arabidopsis* PRP40C, and found the factor linked the regulation of gene expression and pre-mRNA splicing to modulate plant growth, development, and stress responses in *Arabidopsis*. These multifunctional splicing factors show the complex interplay between splicing and transcription that underlie functional coupling between these processes.

Plants adapt to changing conditions and acquire tolerances to daily, seasonal, and chronic stresses. Proteomics of *sr45-1* revealed increased amounts of enzymes involved in glucosinolate biosynthesis, which are important for disease resistance (Chen et al.). Nimeth et al. uncovered the hidden potential of AS in the DNA damage response in plants by reporting that about 80% of the DNA repair genes are alternatively spliced based on the evidence in the *Arabidopsis* reference transcript dataset (AtRTD2). Hernando et al. found contrasting sensitivity of the PRP40C mutants to salt stress and their enhanced tolerance to *Pseudomonas* infection. They identified over 600 transcripts enriched for genes related to biotic and abiotic stress responses. In both cases, increased proportions of intron retention events were identified, indicating a possible mechanism for regulating expression of stress response genes and a role in fine tuning transcriptome functionality (Hernando et al.). Many intron retention transcripts are constrained in the nucleus avoiding degradation by nonsense-mediated mRNA decay. Some may be translated to produce truncated proteins that could modulate the function of the fully spliced protein (Chaudhary et al.). Precise control of plant developmental transitions, ranging from seed germination to flowering, is essential for plant propagation and reproductive success. Many transitions are initiated in response to changes in temperature and light. For instance, light perception plays key roles in the transition from seed dormancy to germination that involves red and far-red photoreceptors. Tognacca et al. showed that a pulse of red light changes AS of several genes, mostly involved in

splicing regulation, light signaling or dormancy/germination, supporting an important role of AS at germination. Flowering time is regulated by a complex network of factors that integrate environmental and developmental cues. Park et al. reviewed the established roles of alternatively spliced genes that are essential for flowering time. For example, alternatively spliced *CONSTANS* (*CO*) produces CO α and CO β protein isoforms, which form non-DNA-binding heterodimers during photoperiodic flowering and protein turnover (Park et al.). Alternatively spliced *FLOWERING LOCUS M* (*FLM*) encodes a temperature sensitive, MADS box transcription factor functioning as a floral repressor. The protein isoform FLM- β is the functional floral repressor, while FLM- δ has a reduced DNA-binding capability and inhibits FLM- β . Nibau et al. showed that the cyclin-dependent kinase G2 (CDKG2), together with its cognate cyclin, CYCLIN L1 (CYCL1) affects the AS of *FLM*, balancing the levels of the FLM- β and FLM- δ isoforms across a temperature range. Thermosensory and developmental signal induction of AS is therefore important for fine-tuning the initiation of flowering. These genes produce apparently functional protein isoforms translated from alternatively spliced transcripts. However, the full extent to which alternatively spliced transcripts are translated and contribute to protein diversity is still far from clear. CO α and CO β protein levels change during the day but alternative transcript levels remain constant. A significant reduction in SAP18 protein with no significant decrease in *SAP18* RNA is observed in the *sr45-1* mutant (Chen et al.). There is a poor correlation between alternatively spliced transcripts and detectable proteins from proteomic experiments. How this disparity occurs remains to be determined. It is possible that the transcript abundances, transcript stability, transcript retention in the nucleus and other processes will ultimately impact on the proteomic outputs. Mis-interpretation of transcript translation by disregarding authentic start codons and premature termination codons may also lead to false interpretation of proteomic outputs. Alternatively, the proteomic technologies, as opposed to PCR or HTS approaches, may not be sensitive enough to detect the low-abundant proteins arising from AS variants (Brown et al., 2015; Zhang et al., 2017).

Genetic variability at core splicing elements, splicing regulatory sequences or in genes encoding trans-acting splicing factors can substantially affect the abundance of transcript isoforms. These changes impact phenotypic diversity and terrestrial adaptation. Khokhar et al. used 666 diverse natural inbred *Arabidopsis* ecotypes, demographically sourced along the east-west axis of Eurasia. They performed splicing quantitative trait loci (sQTL) analysis to reveal the functional impact of genetic variations on AS patterns of genes related to stress response, flowering and the circadian clock (Khokhar et al.). Dantas et al. analyzed the AS events of circadian clock genes in the C4 sugarcane under different field conditions. The authors found striking differences in the seasonal AS patterns in these genes that could be a result of fluctuating temperatures between winter and summer (Dantas et al.). The strong association between AS and environmental stimuli, including temperature, diurnal, and seasonal changes, underscores that AS is intricately involved in myriad of adaptation processes. SR proteins are a

highly conserved family of AS regulators. Melo et al. performed an *in silico* analysis that identified 16 *Physcomitrella patens* SR genes belonging to the six canonical plant SR protein subfamilies. The number and size of SR subfamilies changes greatly from aquatic green algae to vascular plants. The authors suggest a role for SR proteins in the adaptation to new land habitats and perhaps even speciation (Melo et al.). Ling et al. discuss the evolution of AS in plants compared to that in vertebrates. They performed a comparative analysis of the transcriptomes of both closely and distantly related plants and found a low level of AS conservation among different species with the exception of AS events that generate premature termination codons (Ling et al.). Clark et al. analyzed genome-wide AS in tomato by integrating mRNA, EST, and RNA-seq reads from 27 published projects. They found an ~65% frequency of AS, similar to the frequency detected for *Arabidopsis* (Clark et al.). Lastly, Bedre et al. discuss the value, limitations and future developments of HTS technologies needed to overcome limitations imposed by low coverage of particular genomes, high ploidy levels and sequencing error rates (Bedre et al.). Melo et al. further bolster these conclusions by identifying discrepancies in major databases when characterizing SR genes in *Physcomitrella* that resulted from imperfect gene annotation curation, sometimes lacking support by expression data (Melo et al.). Nevertheless, new HTS and bioinformatics approaches will likely spur further in-depth

identification of AS patterns on a genome-wide scale in diverse plant species that coupled with the fundamental mechanistic studies will likely answer key questions related to the role of AS in adaptation and evolution of plants on this planet.

AUTHOR CONTRIBUTIONS

All authors listed have made a substantial, direct and intellectual contribution to the work, and approved it for publication.

FUNDING

This work was supported in part by funds from the Austrian Science Fund (FWF) (P26333 to MK), the Agencia Nacional de Promoción Científica y Tecnológica of Argentina (PICT-2016-4366, PICT-2017-1343 to EP), Academia Sinica (to S-LT), USDA-NIFA-AFRI (2016-67013-24738), and Texas A&M AgriLife Research Insect Vectored Diseases Seed Grant (114190-96210) to KM, and the Scottish Government Rural and Environment Science and Analytical Services division (RESAS) to CS.

ACKNOWLEDGMENTS

We genuinely acknowledge the contributions of every author, reviewer and editor that made this Research Topic possible.

REFERENCES

- Brown, J. W. S., Simpson, C. G., Marquez, Y., Gadd, G. M., Barta, A., and Kalyna, M. (2015). Lost in translation: pitfalls in deciphering plant alternative splicing transcripts. *Plant Cell* 27, 2083–2087. doi: 10.1105/tpc.15.00572
- Fedorov, A., Merican, A. F., and Gilbert, W. (2002). Large-scale comparison of intron positions among animal, plant, and fungal genes. *Proc. Natl. Acad. Sci. U.S.A.* 99, 16128–16133. doi: 10.1073/pnas.242624899
- Filichkin, S. A., Priest, H. D., Givan, S. A., Shen, R., Bryant, D. W., Fox, S. E., et al. (2010). Genome-wide mapping of alternative splicing in *Arabidopsis thaliana*. *Genome Res.* 20, 45–58. doi: 10.1101/gr.093302.109
- Lu, T., Lu, G., Fan, D., Zhu, C., Li, W., Zhao, Q., et al. (2010). Function annotation of the rice transcriptome at single-nucleotide resolution by RNA-seq. *Genome Res.* 20, 1238–1249. doi: 10.1101/gr.106120.110
- Marquez, Y., Brown, J. W. S., Simpson, C. G., Barta, A., and Kalyna, M. (2012). Transcriptome survey reveals increased complexity of the alternative splicing landscape in *Arabidopsis*. *Genome Res.* 22, 1184–1195. doi: 10.1101/gr.134106.111
- 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. doi: 10.3390/biom5031717
- Syed, N. H., Kalyna, M., Marquez, Y., Barta, A., and Brown, J. W. S. (2012). Alternative splicing in plants—coming of age. *Trends Plant Sci.* 17, 616–623. doi: 10.1016/j.tplants.2012.06.001
- Wahl, M. C., Will, C. L., and Luhmann, R. (2009). The spliceosome: design principles of a dynamic RNP machine. *Cell* 136, 701–718. doi: 10.1016/j.cell.2009.02.009
- Zhang, R., Calixto, C. P. G., Marquez, Y., Venhuizen, P., Tzioutziou, N. A., Guo, W., et al. (2017). A high quality *Arabidopsis* transcriptome for accurate transcript-level analysis of alternative splicing. *Nucleic Acids Res.* 45, 5061–5073. doi: 10.1093/nar/gkx267

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2020 Petrillo, Kalyna, Mandadi, Tu and Simpson. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



Alternative Splicing and Transcription Elongation in Plants

Micaela A. Godoy Herz^{1,2} and Alberto R. Kornblihtt^{1,2*}

¹Departamento de Fisiología, Biología Molecular y Celular, Facultad de Ciencias Exactas y Naturales, Universidad de Buenos Aires (UBA), Buenos Aires, Argentina, ²Instituto de Fisiología, Biología Molecular y Neurociencias (IFIBYNE), CONICET-UBA, Buenos Aires, Argentina

OPEN ACCESS

Edited by:

Kranthi Kiran Mandadi,
Texas A&M University, United States

Reviewed by:

Klaus Grasser,
University of Regensburg,
Germany
Yu Qing-Bo,
Shanghai Normal University,
China

*Correspondence:

Alberto R. Kornblihtt
ark@fbmc.fcen.uba.ar

Specialty section:

This article was submitted to
Plant Physiology,
a section of the journal
Frontiers in Plant Science

Received: 31 January 2019

Accepted: 26 February 2019

Published: 26 March 2019

Citation:

Godoy Herz MA and Kornblihtt AR
(2019) Alternative Splicing and
Transcription Elongation in Plants.
Front. Plant Sci. 10:309.
doi: 10.3389/fpls.2019.00309

Alternative splicing and transcription elongation by RNA polymerase II (RNAPII) are two processes which are tightly connected. Splicing is a co-transcriptional process, and different experimental approaches show that splicing is coupled to transcription in *Drosophila*, yeast and mammals. However, little is known about coupling of transcription and alternative splicing in plants. The kinetic coupling explains how changes in RNAPII elongation rate influence alternative splicing choices. Recent work in *Arabidopsis* shows that expression of a dominant negative transcription elongation factor, TFIIIS, enhances exon inclusion. Furthermore, the *Arabidopsis* transcription elongation complex has been recently described, providing new information about elongation factors that interact with elongating RNAPII. Light regulates alternative splicing in plants through a chloroplast retrograde signaling. We have recently shown that light promotes RNAPII elongation in the affected genes, while in darkness elongation is lower. These changes in transcription are consistent with elongation causing the observed changes in alternative splicing. Altogether, these findings provide evidence that coupling between transcription and alternative splicing is an important layer of gene expression regulation in plants.

Keywords: alternative splicing, transcription, plants, RNAPII, light

TRANSCRIPTION AND ALTERNATIVE SPLICING

Transcription in eukaryotes is functionally coupled to mRNA maturation (Saldi et al., 2017), which includes capping, splicing, and polyadenylation. Alternative splicing and transcription elongation by RNA polymerase II (RNAPII) are two processes which are tightly connected. Splicing is a co-transcriptional process, and different experimental approaches show that splicing is coupled to transcription in *Drosophila*, yeast, and mammals. The first evidence that splicing occurs co-transcriptionally was observed by early electron microscopy in nascent transcripts containing splicing loops in *Drosophila* embryos (Beyer and Osheim, 1988). Tilgner and collaborators have shown in a genome-wide deep-sequencing analysis of nascent transcripts that, in human genes, splicing associated to chromatin is mostly co-transcriptional. To investigate this, the authors developed a method that allows to assess the degree of splicing completion around internal exons. They found that in polyadenylated RNAs from cytosolic fractions splicing is almost fully completed. On the other hand, in nuclear RNAs, the majority of human introns are excised while they are still associated with chromatin (Tilgner et al., 2012). This supports the idea that splicing occurs during transcription.

More recently, a work performed in yeast demonstrates that the process of splicing coincides with intron exit from RNAPII: with high-resolution sequencing techniques, nascent RNAs from 87 yeast genes were analyzed (about one third of intron containing genes in this organism). An adaptor was ligated to each 3' end: this enables to establish the exact position of RNAPII in every nascent RNA molecule. The authors found that, when RNAPII has transcribed about 26–27 nucleotides downstream of the 3' splice site, the splicing catalysis begins: this work is a high-resolution evidence that splicing *in vivo* is co-transcriptional (Carrillo Oesterreich et al., 2016).

Furthermore, multiple evidences for coupling between transcription and alternative splicing have been obtained working with mammalian culture cells (Cramer et al., 1997, 1999; Kadener et al., 2001; Nogués et al., 2002; de la Mata et al., 2003). It is important to point out that we refer to a functional coupling, that is, properties of the splicing reactions itself are affected by the transcriptional process, not only because they occur at the same time and space: coupling occurs if the splicing reactions depend on transcription and transcription depends on splicing (Lazarev and Manley, 2007). However, little is known about coupling of transcription and alternative splicing in plants.

Two mechanisms, which are not mutually exclusive, have been proposed to explain the nature of coupling between transcription and alternative splicing: recruitment coupling and kinetic coupling. The first mechanism involves recruitment of splicing factors by the transcription machinery. The kinetic coupling explains how changes in RNAPII elongation rate influence alternative splicing choices (de la Mata et al., 2003; Dujardin et al., 2014; Fong et al., 2014).

The recruitment coupling focuses on recruitment of splicing factors by the transcription machinery: RNAPII largest subunit contains a carboxy-terminal domain (CTD). In animals and plants, the CTD is composed of a number of repeats of a consensus heptad YSPTSPST that can be subject to different post-translational modifications in their residues. The number of repeats varies among species: the human CTD has 52 repeats, while in plants the CTD has 34 repeats (Koiwa et al., 2004). Modifications of the CTD regulate the affinity for factors involved in capping, 3' end processing and alternative splicing (McCracken et al., 1997). One of them is phosphorylation, and the results that phosphorylation elicits depend on the involved amino acids. For example, phosphorylation of serine 5 is associated with recruitment of enzymes involved in 5' capping, while phosphorylation of serine 2 participates in 3' end processing (Kim et al., 2004). The CTD phosphorylation patterns determinate the recruitment of splicing factors to transcription sites (Moore and Proudfoot, 2009; Perales and Bentley, 2009).

The kinetic coupling explains how changes in RNAPII elongation rate influence alternative splicing. The first direct evidence of kinetic coupling *in vivo* was achieved working with a RNAPII mutant in human culture cells, which encodes a point mutation that produces a slow transcription rate. Transcription by the slow RNAPII mutant favors exon inclusion in a reporter minigene compared to wild-type RNAPII (de la Mata et al., 2003). In this reporter minigene (that encodes fibronectin exon EDI), there are two 3' splice sites: the upstream

3' splice site is weak, while the downstream one is strong (this means that it is more adjusted to the consensus sequence compared to the weak one). If RNAPII elongation rate is fast, both sites are presented to the splicing machinery at the same time, and the strong one is recognized by the splicing machinery more efficiently: this produces exon skipping. On the other hand, if RNAPII transcription rate is slow, the splicing machinery will recognize the first, weak 3' splice site, and afterwards, the strong 3' splice site. This, in turn, leads to exon inclusion. This, however, does not mean that the first intron will necessarily be eliminated before the second one: once "commitment" to include the exon is achieved, the order of intron removal becomes irrelevant (de la Mata et al., 2010). Interesting information has been obtained from single-molecule imaging analyses of constitutive and alternative splicing to identify the intracellular sites of splicing (Vargas et al., 2011). This work demonstrates that, although catalysis of constitutive splicing is co-transcriptional, catalysis of alternative splicing occurs post-transcriptionally in a group of studied events, which does not rule out that recruitment of splicing factors needed for those alternative events takes place co-transcriptionally.

It is important to point out that in most cases, a slow RNAPII produces exon inclusion. However, in some examples, a slow RNAPII promotes exon skipping by allowing the recruitment of negative factors to the splice sites (Dujardin et al., 2014; Fong et al., 2014). In both cases, transcription elongation regulates alternative splicing.

ELONGATION FACTORS: TRANSCRIPTION FACTOR II S

Changes in RNAPII elongation rate can be caused by different factors: histone chaperones and nucleosome remodelers can facilitate RNAPII movement through chromatin; some DNA sequences may be more difficult to transcribe than others owing to their DNA topology; and histone marks can tighten or loosen DNA binding around nucleosomes (for a detailed review on alternative splicing and chromatin modifications, see Luco et al., 2011). Furthermore, the balance between pausing and activating transcription elongation factors modulates the level of pausing (Jonkers and Lis, 2015).

Transcription factor II S (TFIIS) is an elongation factor required for RNAPII processivity that stimulates RNAPII to reassume elongation after pausing (Fish and Kane, 2002). Recently, a dominant negative TFIIS *Arabidopsis* plant was constructed by replacing two key amino acids responsible for its stimulatory activity by alanines (Dolata et al., 2015). TFIIS mutant plants show a range of developmental and morphological defects. Changes in alternative splicing were analyzed in TFIIS mutant plants: these plants show some alternative splicing defects. The authors found that, out of 284 events analyzed, 62 were altered. In 45 of them, the observed changes correspond to what would be predicted by the kinetic model (preferential selection of upstream 5' splice sites and enhanced exon inclusion in the case of exon skipping alternative splicing events) (Dolata et al., 2015).

PLANT TRANSCRIPTION ELONGATION COMPLEX

The *Arabidopsis* transcription elongation complex has been recently described, providing new information about elongation factors that interact with elongating RNAPII. Antosz et al. performed reciprocal tagging in combination with affinity purification and mass spectrometry and demonstrated which transcription elongation factors copurify with elongating RNAPII (using an antibody against Ser2-phosphorylated carboxy-terminal domain repeats of RNAPII). They have shown that RNAPII interacts with TFIIS, PAF1-C, FACT, SPT4/5, and SPT6, among other transcription elongation factors. These results are similar to what has been previously observed in yeast (Krogan et al., 2002), suggesting that the elongation factors that associate with elongating RNAPII both in *Arabidopsis* and in yeast are conserved. In addition, the elongation factors mentioned above also copurified with splicing factors and with some spliceosomal complexes, like U1, U2, U5, among others (Antosz et al., 2017). This suggests that an interplay between transcription elongation and splicing regulation is to be expected in plants.

LIGHT REGULATION OF ALTERNATIVE SPLICING THROUGH TRANSCRIPTION ELONGATION

Our laboratory has shown that light regulates plant alternative splicing (Petrillo et al., 2014). *Arabidopsis* seedlings were exposed to light and dark conditions and alternative splicing was studied

in a group of alternative splicing events. Light initiates a chloroplast retrograde signal that regulates nuclear alternative splicing of a subset of transcripts, which encode proteins involved in RNA processing (Petrillo et al., 2014). This light effect depends on functional chloroplasts, since the use of drugs that block chloroplast photosynthetic transport chain inhibits the effect of light on alternative splicing.

In a recent publication (Godoy Herz et al., 2019), we show that the light control of alternative splicing responds to the kinetic coupling mechanism. Light promotes transcription elongation, while in darkness RNAPII elongation is lower. We show by different experimental approaches that light changes RNAPII elongation, including RNAPII chromatin immunoprecipitation and a modified nascent RNA single molecule intron tracking (SMIT) method, originally described by Carrillo Oesterreich et al. (2016). Briefly, 3' ends of nascent transcripts were sequenced after treating plants with light and darkness. As a result, the detected 3' ends correspond to RNAPII positions. Comparison of the densities of 3' ends along different genes show changes in elongation or processivity in light and dark. This is consistent with elongation causing the observed changes in alternative splicing. Furthermore, the light control on alternative splicing is abolished in the TFIIS mutant plants previously described (Dolata et al., 2015): TFIIS mutant plants do not respond to light signaling on a group of studied alternative splicing events. Thus, light regulates alternative splicing through control of transcriptional elongation. A scheme that summarizes these findings is shown in **Figure 1**.

The mechanism that explains how light promotes RNAPII elongation remains unknown: as a first attempt to study a

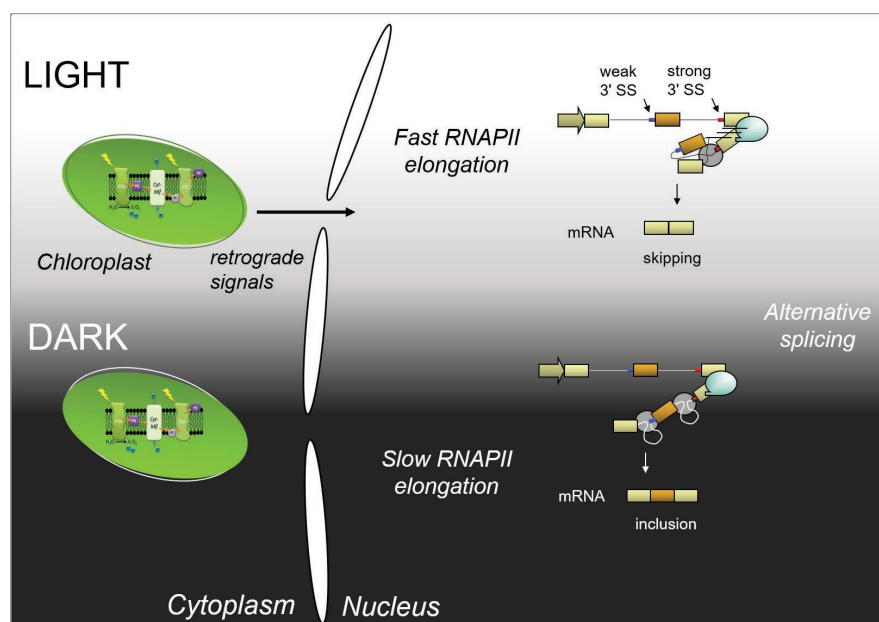


FIGURE 1 | Scheme of light regulation of alternative splicing through the control of transcriptional elongation (adapted from Godoy Herz et al., 2019). Light initiates a chloroplast retrograde signal. In light, RNAPII elongation is fast; both splicing sites (SS) are presented to the splicing machinery at the same time, and the strong one is recognized by the splicing machinery more efficiently: this produces exon skipping. In darkness, RNAPII elongation is slow; the splicing machinery will recognize the first, weak 3' splice site, and afterwards, the strong 3' splice site. This, in turn, leads to exon inclusion.

possible mechanism mRNA levels of different elongation factors were measured. Although mRNA levels of TFIIS do not change in light and dark, the mRNA levels of the PAF1-C subunits CDC73 and ELF7 increase in light-treated plants. This opens the possibility that an increase in RNAPII elongation is caused by an increase in elongation factor expression (Godoy Herz et al., 2019).

We found that treating *Arabidopsis* seedling with the histone deacetylase inhibitor trichostatin A (TSA) mimics the effect of light on alternative splicing, which we interpreted as a consequence of higher RNAPII elongation due to chromatin relaxation caused by histone hyperacetylation. However, histone acetylation does not participate in the mechanism involved, as has been demonstrated by chromatin immunoprecipitation and Western blot experiments. A role for chromatin modifications on alternative splicing has been intensively studied in mammalian cells (Kadener et al., 2001; Nogués et al., 2002; Alló et al., 2009; Schor et al., 2013; Fiszbein et al., 2016). Chromatin participation on the regulation of alternative splicing in plants remains an interesting field to investigate.

PERSPECTIVES

The results summarized in **Figure 1** point at alternative splicing regulation by transcription elongation as a mechanism to respond to an environmental stimulus. Furthermore, they provide evidence that coupling between transcription and alternative splicing is

important for a whole organism (plants) to respond to environmental cues (light). It is interesting to point out that plants are a model organism suitable for this kind of experimental approaches and that there is still plenty to learn about gene expression regulation from this kingdom. More efforts will be needed to understand the mechanisms involved in light-mediated RNAPII elongation: How does light affect transcription elongation? How do changes in transcription elongation modulate alternative splicing? Nevertheless, the experimental system described here provides a mean to investigate the mechanism behind the proposed regulation of alternative splicing by transcription elongation in whole organisms.

AUTHOR CONTRIBUTIONS

MGH and AK wrote the manuscript.

FUNDING

This work was supported by grants from the Agencia Nacional de Promoción Científica y Tecnológica of Argentina (PICT-2014 2582 and PICT-2015-0341), the Universidad de Buenos Aires (UBACYT 20020130100152BA), and the Howard Hughes Medical Institute. AK is a career investigator and MGH received a fellowship from the Consejo Nacional de Investigaciones Científicas y Técnicas of Argentina (CONICET).

REFERENCES

- Alló, M., Buggiano, V., Fededa, J. P., Petrillo, E., Schor, I., de la Mata, M., et al. (2009). Control of alternative splicing through siRNA-mediated transcriptional gene silencing. *Nat. Struct. Mol. Biol.* 16, 717–725.
- Antosz, W., Pfab, A., Ehrnsberger, H. F., Holzinger, P., Köllen, K., Mortensen, S. A., et al. (2017). The composition of the Arabidopsis RNA polymerase II transcript elongation complex reveals the interplay between elongation and mRNA processing factors. *Plant Cell* 29, 854–870. doi: 10.1105/tpc.16.00735
- Beyer, A. L., and Osheim, Y. N. (1988). Splice site selection, rate of splicing, and alternative splicing on nascent transcripts. *Genes Dev.* 2, 754–765. doi: 10.1101/gad.2.6.754
- Carrillo Oesterreich, F., Herzel, L., Straube, K., Hujer, K., Howard, J., and Neugebauer, K. M. (2016). Splicing of nascent RNA coincides with intron exit from RNA polymerase II. *Cell* 165, 372–381. doi: 10.1016/j.cell.2016.02.045
- Cramer, P., Cáceres, J. F., Cazalla, D., Kadener, S., Muro, A. F., Baralle, F. E., et al. (1999). Coupling of transcription with alternative splicing. *Mol. Cell* 4, 251–258. doi: 10.1016/S1097-2765(00)80372-X
- Cramer, P., Pesce, C. G., Baralle, F. E., and Kornblihtt, A. R. (1997). Functional association between promoter structure and transcript alternative splicing. *PNAS* 94, 11456–11460.
- de la Mata, M., Alonso, C. R., Kadener, S., Fededa, J. P., Blaustein, M., Pelisch, F., et al. (2003). A slow RNA polymerase II affects alternative splicing in vivo. *Mol. Cell* 12, 525–532. doi: 10.1016/j.molcel.2003.08.001
- de la Mata, M., Lafaille, C., and Kornblihtt, A. R. (2010). First come, first served revisited: factors affecting the same alternative splicing event have different effects on the relative rates of intron removal. *RNA* 16, 904–912. doi: 10.1261/rna.1993510
- Dolata, J., Guo, Y., Kolowierz, A., Smolinski, D., Brzyzek, G., Jarmolowski, A., et al. (2015). NTR1 is required for transcription elongation checkpoints at alternative exons in Arabidopsis. *EMBO J.* 34, 544–558. doi: 10.15252/emboj.201489478
- Dujardin, G., Lafaille, C., de la Mata, M., Marasco, L. E., Muñoz, M. J., Jossic-Corcos, C. L., et al. (2014). How slow RNA polymerase II elongation favors alternative exon skipping. *Mol. Cell* 54, 683–690. doi: 10.1016/j.molcel.2014.03.044
- Fish, R. N., and Kane, C. M. (2002). Promoting elongation with transcript cleavage stimulatory factors. *Biochim. Biophys. Acta. Gene Struct. Expr.* 1577, 287–307. doi: 10.1016/S0167-4781(02)00459-1
- Fiszbein, A., Giono, L. E., Quaglino, A., Berardino, B. G., Sigaut, L., von Bilderling, C., et al. (2016). Alternative splicing of G9a regulates neuronal differentiation. *Cell Rep.* 14, 2797–2808. doi: 10.1016/j.celrep.2016.02.063
- Fong, N., Kim, H., Zhou, Y., Ji, X., Qiu, J., Saldi, T., et al. (2014). Pre-mRNA splicing is facilitated by an optimal RNA polymerase II elongation rate. *Genes Dev.* 28, 2663–2676. doi: 10.1101/gad.252106.114
- Godoy Herz, M. A., Kubaczka, M. G., Brzyzek, G., Servi, L., Krzyszton, M., Simpson, C., et al. (2019). Light regulates plant alternative splicing through the control of transcriptional elongation. *Mol. Cell* 73, 1066–1074. doi: 10.1016/j.molcel.2018.12.005
- Jonkers, I., and Lis, J. T. (2015). Getting up to speed with transcription elongation by RNA polymerase II. *Nat. Rev. Mol. Cell Biol.* 16, 167–177. doi: 10.1038/nrm3953
- Kadener, S., Cramer, P., Nogués, G., Cazalla, D., de la Mata, M., Fededa, J. P., et al. (2001). Antagonistic effects of T-Ag and VP16 reveal a role for RNA Pol II elongation on alternative splicing. *EMBO J.* 20, 5759–5768. doi: 10.1093/emboj/20.20.5759
- Kim, M., Krogan, N. J., Vasiljeva, L., Rando, O. J., Nedeá, E., Greenblatt, J. F., et al. (2004). The yeast Rat1 exonuclease promotes transcription termination by RNA polymerase II. *Nature* 432, 517–522.
- Koiwa, H., Hausmann, S., Bang, W. Y., Ueda, A., Kondo, N., Hiraguri, A., et al. (2004). Arabidopsis C-terminal domain phosphatase-like 1 and 2 are essential Ser-5-specific C-terminal domain phosphatases. *PNAS*, 14539–14544. doi: 10.1073/pnas.0403174101

- Krogan, N. J., Kim, M., Ahn, S. H., Zhong, G., Kobor, M. S., Cagney, G., et al. (2002). RNA polymerase II elongation factors of *Saccharomyces cerevisiae*: a targeted proteomics approach. *Mol. Cell. Biol.* 22, 6979–6992. doi: 10.1128/MCB.22.20.6979-6992.2002
- Lazarev, D., and Manley, J. M. (2007). Concurrent splicing and transcription are not sufficient to enhance splicing efficiency. *RNA* 13, 1546–1557. doi: 10.1261/rna.595907
- Luco, R. F., Allo, M., Schor, I. E., Kornblihtt, A. R., and Misteli, T. (2011). Epigenetics in alternative pre-mRNA splicing. *Cell* 144, 16–26. doi: 10.1016/j.cell.2010.11.056
- McCracken, S., Fong, N., Yankulov, K., Ballantyne, S., Pan, G., Greenblatt, J., et al. (1997). The C-terminal domain of RNA polymerase II couples mRNA processing to transcription. *Nature* 385, 357–361. doi: 10.1038/385357a0
- Moore, M. J., and Proudfoot, N. J. (2009). Pre-mRNA processing reaches back to transcription and ahead to translation. *Cell* 136, 688–700. doi: 10.1016/j.cell.2009.02.001
- Nogués, G., Kadener, S., Cramer, P., Bentley, D., and Kornblihtt, A. R. (2002). Transcriptional activators differ in their abilities to control alternative splicing. *J. Biol. Chem.* 277, 43110–43114. doi: 10.1074/jbc.M208418200
- Perales, R., and Bentley, D. (2009). ‘Cotranscriptionality’: the transcription elongation complex as a nexus for nuclear transactions. *Mol. Cell* 36, 178–191. doi: 10.1016/j.molcel.2009.09.018
- Petrillo, E., Godoy Herz, M. A., Fuchs, A., Reifer, D., Fuller, J., Yanovsky, M. J., et al. (2014). A chloroplast retrograde signal regulates nuclear alternative splicing. *Science* 344, 427–430. doi: 10.1126/science.1250322
- Saldi, T., Cortazar, M., Ryan, S., and Bentley, D. (2017). Coupling of RNA polymerase II transcription elongation with pre-mRNA splicing. *J. Mol. Biol.* 428, 69–81.
- Schor, I. E., Fiszbein, A., Petrillo, E., and Kornblihtt, A. R. (2013). Intragenic epigenetic changes modulate NCAM alternative splicing in neuronal differentiation. *EMBO J.* 32, 2264–2274. doi: 10.1038/emboj.2013.167
- Tilgner, H., Knowles, D. G., Johnson, R., Davis, C. A., Chakraborty, S., Djebali, S., et al. (2012). Deep sequencing of subcellular RNA fractions shows splicing to be predominantly co-transcriptional in the human genome but inefficient for lncRNAs. *Genome Res.* 22, 1616–1625. doi: 10.1101/gr.134445.111
- Vargas, D. Y., Shah, K., Batish, M., Levandoski, M., Sinha, S., Marras, S. A. E., et al. (2011). Single-molecule imaging of transcriptionally coupled and uncoupled splicing. *Cell* 147, 1054–1065. doi: 10.1016/j.cell.2011.10.024

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2019 Godoy Herz and Kornblihtt. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



Alternative RNA Splicing Expands the Developmental Plasticity of Flowering Transition

Young-Joon Park¹, June-Hee Lee¹, Jae Young Kim¹ and Chung-Mo Park^{1,2*}

¹ Department of Chemistry, Seoul National University, Seoul, South Korea, ² Plant Genomics and Breeding Institute, Seoul National University, Seoul, South Korea

OPEN ACCESS

Edited by:

Maria Kalyna,
Universität für Bodenkultur Wien,
Austria

Reviewed by:

Sureshkumar Balasubramanian,
Monash University, Australia
John William Slessor Brown,
University of Dundee, United Kingdom

*Correspondence:

Chung-Mo Park
cmpark@snu.ac.kr

Specialty section:

This article was submitted to
Plant Physiology,
a section of the journal
Frontiers in Plant Science

Received: 08 March 2019

Accepted: 25 April 2019

Published: 08 May 2019

Citation:

Park Y-J, Lee J-H, Kim JY and
Park C-M (2019) Alternative RNA
Splicing Expands the Developmental
Plasticity of Flowering Transition.
Front. Plant Sci. 10:606.
doi: 10.3389/fpls.2019.00606

Precise control of the developmental phase transitions, which ranges from seed germination to flowering induction and senescence, is essential for propagation and reproductive success in plants. Flowering induction represents the vegetative-to-reproductive phase transition. An extensive array of genes controlling the flowering transition has been identified, and signaling pathways that incorporate endogenous and environmental cues into the developmental phase transition have been explored in various plant species. Notably, recent accumulating evidence indicate that multiple transcripts are often produced from many of the flowering time genes via alternative RNA splicing, which is known to diversify the transcriptomes and proteomes in eukaryotes. It is particularly interesting that some alternatively spliced protein isoforms, including CO β and FT2 β , function differentially from or even act as competitive inhibitors of the corresponding functional proteins by forming non-functional heterodimers. The alternative splicing events of the flowering time genes are modulated by developmental and environmental signals. It is thus necessary to elucidate molecular schemes controlling alternative splicing and functional characterization of splice protein variants for understanding how genetic diversity and developmental plasticity of the flowering transition are achieved in optimizing the time of flowering under changing climates. In this review, we present current knowledge on the alternative splicing-driven control of flowering time. In addition, we discuss physiological and biochemical importance of the alternative splicing events that occur during the flowering transition as a molecular means of enhancing plant adaptation capabilities.

Keywords: alternative splicing, flowering, photoperiod, temperature, developmental aging

INTRODUCTION

Plants coordinately incorporate both exogenous and endogenous signals to fine-tune the timing of flowering transition under changing environments, among which the effects of light and temperature have been most extensively studied. Therefore, plants have evolved versatile mechanisms to accurately monitor seasonal changes in photoperiod and temperatures (Duncan et al., 2015; Blackman, 2017). Multiple flowering time genes are differentially affected by various environmental conditions. Endogenous cues, including plant aging signals and circadian rhythms, also affect the timing of flowering transition (Wang, 2014; Shim et al., 2017). It is now evident

that the flowering transition is tightly regulated through a complex network of flowering genetic pathways, each monitoring distinct internal and external changes.

The flowering time genes are regulated through diverse molecular and biochemical mechanisms, such as transcriptional, post-transcriptional, and protein-level controls (Liu et al., 2008; Wang et al., 2016). They are also modulated by epigenetic mechanisms (Kim et al., 2004). Accumulating evidence in recent years indicate that alternative splicing, a versatile molecular process that produces multiple transcripts from a single gene and thus is capable of expanding the transcriptomes and proteomes, plays a critical role in flowering time control (Eckardt, 2002). Notably, the alternatively spliced protein isoforms either promote or suppress the corresponding functional proteins, depending on developmental and environmental conditions (Seo et al., 2012). Therefore, we believe that unraveling the functional roles of alternative splicing events would further expand the functional repertoire of the previously identified flowering time genes, especially in response to fluctuating external conditions.

In this review, we summarize recent advances in understanding the functional roles of alternative splicing events during flowering transition. Physiological and mechanistic relevance of the alternative splicing events are also discussed in terms of the developmental plasticity of flowering time control.

ALTERNATIVE SPLICING OF PHOTOPERIODIC FLOWERING GENES

Daylength information is a central determinant of photoperiodic flowering, and genes and molecular mechanisms underlying the photoperiod-dependent flowering induction have been characterized in many plant species (Sawa et al., 2007; Song et al., 2014; Lee et al., 2017). The floral activator CONSTANS (CO) plays a crucial role in photoperiodic flowering (Shim et al., 2017). It has been reported that FLAVIN-BINDING, KELCH REPEAT, F BOX 1 (FKF1) and GIGANTEA (GI) proteins interact with each other under long days (LDs), while this interaction does not occur under short days (SDs) (Sawa et al., 2007). The FKF1-GI complex suppresses the function of CYCLING DOF FACTOR 1 (CDF1), which acts as a CO repressor, thereby inducing the transcription of CO mainly under LDs.

In addition to the transcriptional control of CO, the CO function is also regulated at the protein level. It is known that CO proteins undergo post-translational modifications (Liu et al., 2008; Jung et al., 2012; Lazaro et al., 2012, 2015). On the other hand, the E3 ubiquitin ligase HIGH EXPRESSION OF OSMOTICALLY RESPONSIVE GENES 1 (HOS1) polyubiquitinates CO, leading to a controlled degradation of CO proteins in a red light-dependent manner (Lazaro et al., 2012, 2015). Meanwhile, in the dark, the E3 ubiquitin ligase CONSTITUTIVE PHOTOMORPHOGENIC 1 (COP1) targets the CO proteins (Liu et al., 2008). Notably, FKF1 conveys blue light information into the ubiquitin-proteasome system to enhance the CO protein stability, thus triggering the onset of photoperiodic flowering (Song et al., 2012; Lee et al., 2017).

A previous study has shown that CO undergoes alternative splicing, producing two protein isoforms: the full-size, physiologically functional CO α , which is equivalent to the well-characterized CO flowering promoter, and the C-terminally truncated CO β (Gil et al., 2017). All the previous functional studies on CO have been performed with the CO α protein isoform (Liu et al., 2008; Jung et al., 2012; Lazaro et al., 2012, 2015; Song et al., 2012, 2014; Wang et al., 2014, 2016), and, as a result, the potential importance of its alternative splicing in photoperiodic flowering has been elusive until recently. The CO α form contains both the B-box (BBX) and CCT (for CONSTANS, CONSTANS-LIKE, TOC1) domains, whereas the C-terminally truncated CO β form lacks the CCT domain.

In accordance with the notion that CO is a flowering activator, overexpression of CO α accelerates flowering induction (Gil et al., 2017). Notably, transgenic plants overexpressing CO β exhibited late flowering, which is phenotypically similar to what is observed in CO-deficient mutants. In addition, the promotive effects of CO α on flowering induction were compromised when CO β was co-expressed with CO α , suggesting that CO β acts as a competitive inhibitor of CO α . Transcription factors act typically as dimers to enhance their DNA binding affinity and specificity (Seo et al., 2011). It has been found that CO β attenuates the CO α function by forming non-functional heterodimers, which have a significantly reduced DNA binding capability compared to that of the CO α -CO α homodimer (Figure 1A; Gil et al., 2017).

A critical question is how photoperiodic information is functionally linked with the alternative splicing event of CO. It has been observed that while the absolute level of CO α transcripts is much higher than that of CO β transcripts, the relative ratio between the two RNA isoforms is unchanged during photoperiodic flowering (Gil et al., 2017). Interestingly, CO β protein is resistant to the ubiquitin-proteasome degradation. Meanwhile, the protein stability of CO α is modulated in a complicated manner by a group of E3 ubiquitin ligases. CO β enhances the interaction of CO α with HOS1 and COP1, while CO β suppressed the interaction of CO α with FKF1, leading to a further destabilization of CO α . Together, these observations indicate that CO is not a passive substrate of the E3 ubiquitin ligases. Instead, CO acts as a proactive regulator of its own protein accumulation by modulating its interactions with multiple E3 ubiquitin ligases in a coordinated manner during the induction of photoperiodic flowering (Figure 1B).

CO belongs to the BBX transcription factor family, which consists of 32 members in *Arabidopsis*. It has been reported that other BBX transcription factors, which are structurally similar to either CO α or CO β , are also functionally linked with flowering time control (Wang, 2014), further supporting the functional relevance of alternative splicing for photoperiodic flowering. It is also interesting that the alternative splicing event of CO is not confined to *Arabidopsis*. A putative *Brachypodium* CO also undergoes alternative splicing, producing two protein isoforms: the full-size CO isoform and the C-terminally truncated CO isoform (Gil et al., 2017). Both *Arabidopsis* and *Brachypodium*

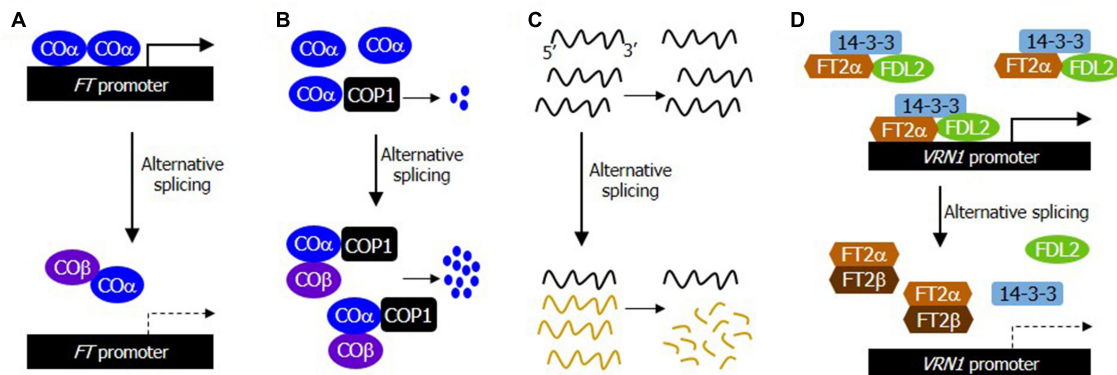


FIGURE 1 | Alternative RNA splicing regulates the timing of flowering transition through diverse regulatory mechanisms. In response to developmental and environmental cues, numerous flowering time genes undergo alternative splicing. The alternatively spliced RNA isoforms are regulated either at the RNA level through the NMD pathway or translated to multiple protein products. The alternatively spliced protein variants modulate the stability or function of different variants by forming heterodimers. They also interact differentially with their interacting partners in a competitive manner. **(A)** Attenuation of the DNA binding affinity of CO α by forming non-DNA-binding, CO α -CO β heterodimers in photoperiodic flowering. While the CO α -CO α homodimers are able to efficiently bind to DNA, the CO α -CO β heterodimers are excluded from DNA binding. **(B)** Facilitation of the interaction of CO α with the COP1 E3 ubiquitin ligase by CO β in photoperiodic flowering. While CO α monomers are poorly targeted by COP1, CO β facilitates the interaction of CO α with COP1, resulting in an elevated ubiquitin-mediated degradation of the CO α proteins. **(C)** Differential stabilization of alternatively spliced transcripts against NMD pathway in thermosensory flowering. The alternatively spliced transcripts are degraded through the NMD pathway. **(D)** Attenuation of the FT2 α /FDL2/14-3-3 florigenic complex formation by FT2 β in aging-induced flowering time control. The alternatively spliced protein isoform FT2 β inhibits the formation of the FT2 α -containing florigenic complexes by forming FT2 α -FT2 β heterodimers in Brachypodium.

are LD plants, flowering early during LDs. It will be interesting to examine whether the CO alternative splicing is a conserved molecular event in all LD plants.

ALTERNATIVE SPLICING EVENTS DURING THERMOSENSORY FLOWERING

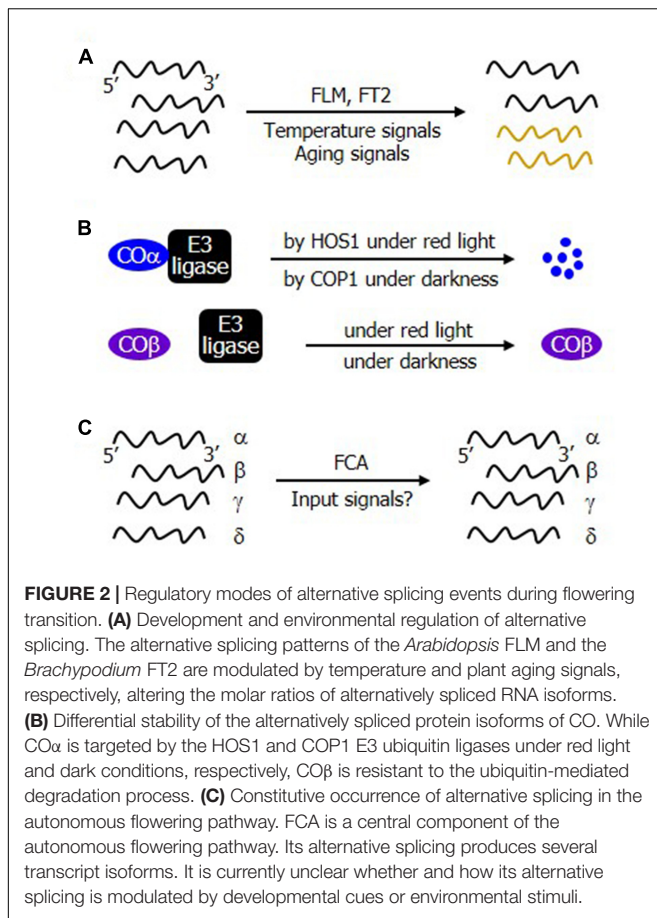
Global warming, a gradual increase of the average global temperature, is widely considered as a serious environmental concern in recent decades. It is well-known that even small changes in ambient temperatures profoundly affect the growth patterning and the timing of developmental transitions in plants, and thus studies on genes and associated molecular mechanisms underlying plant temperature adaptation attracts particular attention in recent years (Quint et al., 2016; Park et al., 2017).

It has been documented for a long time that plants are capable of coping with extreme temperature stress, such as heat and freezing (Ding et al., 2015; Han et al., 2019). Numerous genes and stress adaptation mechanisms have been functionally characterized (Chinnusamy et al., 2007; Ohama et al., 2017). On the other hand, plants often encounter mild temperature changes rather than temperature extremes in natural habitats. In response to changes in ambient temperatures, plants exhibit multiple distinct phenotypes, such as stem elongation, elevation of leaf hyponasty, and acceleration of flowering initiation, which are collectively termed thermomorphogenesis (Koini et al., 2009; Quint et al., 2016; Park et al., 2019). It is known that the thermomorphogenic process is distinct from temperature stress responses and these two thermal responses are regulated by

different sets of genes and regulatory mechanisms (Quint et al., 2016). Among the pleiotropic thermomorphogenic phenotypes, the thermal control of flowering initiation has been extensively studied because of its direct association with reproductive success and crop productivity in temperate areas (Lee et al., 2013).

FLOWERING LOCUS M (FLM) is a MADS box transcription factor functioning as a floral repressor (Sureshkumar et al., 2016). It has been observed that temperature-responsive flowering is nearly diminished in FLM-deficient mutants (Balasubramanian et al., 2006), showing that FLM is involved in thermosensory flowering. A critical question is how temperature signals modulate the FLM function in controlling thermosensory flowering. Interestingly, FLM undergoes alternative splicing, producing multiple FLM transcripts (Sureshkumar et al., 2016). In addition, its alternative splicing pattern is altered in response to temperature changes, supporting that the alternative splicing process of FLM is a critical constituent of temperature-sensitive timing of flowering. A question is how the temperature-mediated production of multiple transcripts is associated with the timing of thermosensory flowering.

Eukaryotes have evolved a molecular surveillance system to remove any potential defects in gene expression by eliminating non-functional or damaged mRNA transcripts, a molecular machinery often termed nonsense-mediated mRNA decay (NMD) (Karousis et al., 2016). In this sense, it is apparent that gene expression is regulated by the NMD-mediated mRNA degradation as well as mRNA transcription. Interestingly, alternatively spliced FLM transcripts are more rapidly degraded by the NMD pathway at warm temperatures, while they are relatively stable at low ambient temperatures (Figure 1C; Sureshkumar et al., 2016). Consistently, temperature-responsive flowering is compromised in *Arabidopsis* mutants that are



defective in the NMD pathway. The differential sensitivity of the alternatively spliced *FLM* transcripts to the NMD pathway at different temperatures illustrates a pivotal role of the NMD-mediated surveillance system in thermosensory flowering.

In addition to the NMD-mediated degradation of the alternatively spliced transcripts, alternative splicing might provide an additional molecular mechanism that regulates *FLM* function. *FLM* physically interacts with another MADS box transcription factor, SHORT VEGETATIVE PHASE (SVP), which also functions as a floral repressor (Lee et al., 2013). Both *flm* and *svp* mutants are insensitive to changes in ambient temperatures, showing that SVP and *FLM* are tightly linked with thermosensory flowering.

It is known that warm temperatures reduce the binding of the SVP transcription factor to the promoter of *FLOWERING LOCUS T* (*FT*) gene (Lee et al., 2007). Interestingly, the temperature-sensitive binding of SVP to DNA is abolished in *flm* mutant backgrounds (Lee et al., 2013), indicating that *FLM* facilitates the DNA binding ability of SVP. *FLM* undergoes alternative splicing, producing multiple protein isoforms, such as *FLM* β and *FLM* δ . *FLM* β is the functional floral repressor, whereas *FLM* δ is one of the alternatively spliced protein variants and has a reduced DNA-binding capability (Lee et al., 2013; Posé et al., 2013). SVP interact efficiently with both *FLM* β and *FLM* δ . However, the SVP- *FLM* δ complex does not

efficiently bind to the target promoter sequence, suggesting a mutually competitive inhibition between the *FLM* protein isoforms. While overexpression of *FLM* δ promotes flowering, this model has been proven to be inappropriate for explaining temperature-responsive flowering in *Arabidopsis* (Sureshkumar et al., 2016; Capovilla et al., 2017; Lutz et al., 2017). It still remains to be elucidated whether and how *FLM* δ contributes to flowering transition.

Most studies on alternative splicing utilize gene overexpressing systems to maximize the effects of individual alternatively spliced RNA or protein variants. However, the phenotypes of the resultant transgenic plants do not necessarily provide any information as to the endogenous effects of alternative splicing. Care should also be taken when interpreting the phenotypes of transgenic plants using endogenous promoters because of the positional effects of the gene insertion into plant genomes. An alternative approach is the CRISPR/Cas9-mediated genome editing system, a powerful technology for studying the effects of alternative splicing in that it is readily applicable to introducing a mutation into a specific splice site so that the patterns of an alternative splicing event are precisely engineered. Indeed, the genome editing system has been applied successfully to generate plants that produce only *FLM* β but not *FLM* δ and *vice versa* (Capovilla et al., 2017).

Genomic *FLM* gene-engineered plants, which lack *FLM* β production, flower earlier than wild-type plants, but not earlier than *FLM*-defective mutants. In addition, plants lacking *FLM* δ production flower later than wild-type plants, but not later than *FLM*-overexpressing plants (Capovilla et al., 2017). Therefore, it is likely that the negative regulatory effect of *FLM* δ is not dominant in wild-type plants. Collectively, these observations indicate that alternative splicing of *FLM* is a critical molecular device for the *FLM*-mediated control of thermoresponsive flowering.

In addition to *FLM*, multiple regulators of flowering timing undergo alternative splicing. It has been reported that trimethylated histone H3 at lysine 36 (*H3K36me3*), a marker of active gene transcription, is enriched in genes undergoing alternative splicing in mammals (Zhou et al., 2014). Chromatin immunoprecipitation assays using *Arabidopsis* plants exposed to different ambient temperatures have shown that *H3K36me3* is enriched in the genomic sequence regions harboring flowering time genes and the *H3K36me3*-enriched regions are broader at warm temperatures (Pajoro et al., 2017). These observations support that temperature-responsive epigenetic control is intimately linked with the effects of ambient temperatures on alternative splicing.

H3K36me3 is directed by histone methyltransferases, SET DOMAIN GROUP 8 (SDG8) and SDG26, in *Arabidopsis* (Xu et al., 2008). Notably, the thermo-responsive alternative splicing of flowering time genes is disturbed in SDG-deficient mutants (Pajoro et al., 2017), further supporting the notion that temperature-responsive epigenetic control by *H3K36me3* is tightly associated with alternative splicing events. Consistent with this observation, the flowering of *sdg* mutants are less sensitive to ambient temperatures. Taken together, these observations indicate that temperature-induced epigenetic modifications,

such as H3K36me3, mediate the thermo-responsive alternative splicing of flowering time genes.

An ultimate question in the field is how plant temperature-sensing mechanisms affect the alternative splicing events of flowering time regulators. There have been studies aimed to identify such thermosensors in plants (Jung et al., 2016; Legris et al., 2016). The best characterized is the red/far-red light-sensing phytochrome photoreceptors, which also function as thermosensors in *Arabidopsis* (Jung et al., 2016; Legris et al., 2016). It is known that photoconversion of the physiologically activated Pfr form to the inactive Pr form is accelerated at warm temperature (Jung et al., 2016; Legris et al., 2016). Notably, the phytochromes have been implicated in the red light-dependent alternative splicing process (Shikata et al., 2014). It will be interesting to examine whether the phytochromes or any putative thermosensors are responsible for the alternative splicing of genes involved in flowering time during thermosensory flowering.

ALTERNATIVE SPLICING DURING DEVELOPMENTAL CONTROL OF FLOWERING

In the juvenile vegetative phase, plants are recalcitrant to floral activating signals, necessitating that plants must spend sufficient time in the vegetative phase to acquire reproductive competence. It is well-known that the mutually interacting microRNA156 (miR156)-miR172 pathway acts to provide developmental aging signals during flowering transition (Wang, 2014). Thus, miRNA-mediated degradation of target transcripts and their translational inhibition are regarded as a major molecular device for transmitting developmental aging signals.

It has been recently reported that alternative splicing plays an important role in the developmental control of flowering initiation in *Brachypodium distachyon*, a representative monocot model for studies on bioenergy grasses and cereal crops in the field (Brkljacic et al., 2011). In response to inductive photoperiodic signals, the FT florigen is produced in the leaves and transported to the shoot apical meristems (SAM) to induce flowering in *Arabidopsis* and other plant species (Jaeger and Wigge, 2007). FT interacts with 14-3-3 and FD proteins in SAM to promote flowering transition (Taoka et al., 2011). FT2 is a potential homolog of the *Arabidopsis* FT protein in *Brachypodium* (Qin et al., 2017). Interestingly, FT2 undergoes alternative splicing, producing the functional FT2 α protein and the alternatively spliced FT2 β protein (Qin et al., 2017).

Protein domain analysis revealed that the N-terminal region of FT2 harboring the phosphatidylethanolamine-binding protein (PEBP) domain is eliminated in the FT2 β protein isoform, suggesting that FT2 β lacks any mechanistic functions conferred by the PEBP domain. Notably, the alternatively spliced FT2 β isoform is unable to interact with 14-3-3 and FD-LIKE 2 (FDL2) proteins, while FT2 α is still able to interact with FT2 α in *Brachypodium* (Qin et al., 2017). Extensive biochemical studies have shown

that FT2 β acts as a competitive inhibitor by attenuating the binding capability of FT2 α with 14-3-3 and FDL2 proteins (Figure 1D). Consistent with this biochemical observations, the expression of *VERNALIZATION1* (*VRN1*) gene is significantly elevated in FT2 β -specific knock-down plants (Qin et al., 2017).

What regulates the alternative splicing of FT2? It has been previously reported that the expression of FT2 gene gradually increases as plant ages (Qin et al., 2017). The levels of both the FT2 α and FT2 β transcripts increase throughout developmental transitions. However, the FT2 β transcripts are more abundant in young plants, while the FT2 α transcripts are more abundant in old plants (Qin et al., 2017). These observations indicate that the alternative splicing patterns of FT2 is developmentally programmed to incorporate endogenous cues into flowering genetic pathways. The miR156-miR172 pathway is widely conserved in plants. It is worthy of examining whether and how miRNA-mediated developmental signals are linked with the alternative splicing event of FT2 in *Brachypodium*.

CONCLUSION AND PERSPECTIVES

Alternative splicing is wide spread in both plants and animals. In plants, it is involved in a variety of plant adaptation processes in response to aging and environmental stimuli. Many flowering time genes undergoes alternative splicing, and plants utilize this molecular device to precisely control the onset of flowering under fluctuating environments.

It is notable that the proven and predicted mechanistic functions of alternatively spliced variants are quite diverse in flowering time control (Figure 1). For example, the alternatively spliced CO β variant interacts with the CO α transcription factor to constitute non-DNA-binding heterodimers during photoperiodic flowering in *Arabidopsis*, while FT2 β interferes with the protein-protein interactions between FT2 α and FDL2 proteins during aging signal-induced flowering in *Brachypodium*. In addition, the CO β splice variant controls the accessibility of E3 ubiquitin ligases to CO α . The stability of alternatively spliced transcripts are also targeted by the NMD pathway at the RNA level during thermosensory flowering. These observations indicate that alternative splicing provides a versatile regulatory system to incorporate multiple developmental and environmental signals into flowering genetic pathways to achieve fine-tuning of the time of flowering induction and maximal productivity.

It is also interesting that the patterns of alternative splicing are differentially regulated during flowering transition. For example, the relative ratio of the alternatively spliced transcripts of FLM in *Arabidopsis* and FT2 in *Brachypodium* is influenced by temperature and developmental aging signals, respectively (Figure 2A). Meanwhile, the ratio of the protein levels between CO α and CO β changes during the day, while their transcript levels are unchanged (Figure 2B). Furthermore, the ratio of the alternatively spliced transcripts of a gene encoding the floral activator FLOWERING CONTROL LOCUS A (FCA), which

functions via the autonomous flowering genetic pathway, is unchanged (**Figure 2C**; Macknight et al., 2002). It is evident that the regulatory modes of alternative splicing is quite diverse in plants.

Alternative splicing events are modulated by differential actions of splicing factors on primary transcripts. For example, SKI-INTERACTING PROTEIN (SKIP), which functions as a splicing factor, directly binds to the pre-mRNA of *SERRATED LEAVES AND EARLY FLOWERING* (*SEF*) to suppress its undesirable alternative splicing (Cui et al., 2017). Since *SEF* activates the transcription of *FLOWERING LOCUS C* (*FLC*), its transcript level is significantly low in SKIP-deficient mutants supporting that splicing factors play crucial roles during floral transition. In addition, it is possible that the activities of splicing factors are modulated by both external and internal cues through multiple molecular mechanisms, such as transcriptional and post-translational modifications and the formation of spliceosome complex (Xiao and Manley, 1998; Stankovic et al., 2016; Shang et al., 2017). It would be interesting to examine the functional relevance of the diversified patterns of alternative splicing in the developmental plasticity of flowering timing.

REFERENCES

- Balasubramanian, S., Sureshkumar, S., Lempe, J., and Weigel, D. (2006). Potent induction of *Arabidopsis thaliana* flowering by elevated growth temperature. *PLoS Genet.* 2:e106. doi: 10.1371/journal.pgen.0020106
- Blackman, B. K. (2017). Changing responses to changing seasons: natural variation in the plasticity of flowering time. *Plant Physiol.* 173, 16–26. doi: 10.1104/pp.16.01683
- Brkljacic, J., Grotewold, E., Scholl, R., Mockler, T., Garvin, D. F., Vain, P., et al. (2011). Brachypodium as a model for the grasses: today and the future. *Plant Physiol.* 157, 3–13. doi: 10.1104/pp.111.179531
- Capovilla, G., Symeonidi, E., Wu, R., and Schmid, M. (2017). Contribution of major FLM isoforms to temperature-dependent flowering in *Arabidopsis thaliana*. *J. Exp. Bot.* 68, 5117–5127. doi: 10.1093/jxb/erx328
- Chinnusamy, V., Zhu, J., and Zhu, J. K. (2007). Cold stress regulation of gene expression in plants. *Trends Plant Sci.* 12, 444–451. doi: 10.1016/j.tplants.2007.07.002
- Cui, Z., Tong, A., Huo, Y., Yan, Z., Yang, W., Yang, X., et al. (2017). SKIP controls flowering time via the alternative splicing of *SEF* pre-mRNA in *Arabidopsis*. *BMC Biol.* 15:80. doi: 10.1186/s12915-017-0422-2
- Ding, Y., Li, H., Zhang, X., Xie, Q., Gong, Z., and Yang, S. (2015). OST1 kinase modulates freezing tolerance by enhancing ICE1 stability in *Arabidopsis*. *Dev. Cell* 32, 278–289. doi: 10.1016/j.devcel.2014.12.023
- Duncan, S., Holm, S., Questa, J., Irwin, J., Grant, A., and Dean, C. (2015). Seasonal shift in timing of vernalization as an adaptation to extreme winter. *Elife* 23:4. doi: 10.7554/eLife.06620
- Eckardt, N. A. (2002). Alternative splicing and the control of flowering time. *Plant Cell* 14, 743–747. doi: 10.1105/tpc.140410
- Gil, K. E., Park, M. J., Lee, H. J., Park, Y. J., Han, S. H., Kwon, Y. J., et al. (2017). Alternative splicing provides a proactive mechanism for the diurnal constans dynamics in *Arabidopsis* photoperiodic flowering. *Plant J.* 89, 128–140. doi: 10.1111/tjp.13351
- Han, S. H., Park, Y. J., and Park, C. M. (2019). Light primes the thermally induced detoxification of reactive oxygen species during development of thermotolerance in *Arabidopsis*. *Plant Cell Physiol.* 60, 230–241. doi: 10.1093/pcp/pcy206

AUTHOR CONTRIBUTIONS

C-MP and Y-JP designed the concept and organization of the manuscript. C-MP and Y-JP wrote the manuscript with helps of J-HL and JK.

FUNDING

This work was supported by the Leaping Research Program (NRF-2018R1A2A1A19020840) provided by the National Research Foundation of Korea (NRF) and the Next-Generation BioGreen 21 Program (PJ013134) provided by the Rural Development Administration of Korea. Y-JP was partially supported by the Global Ph.D. Fellowship Program through NRF (NRF-2016H1A2A1906534).

ACKNOWLEDGMENTS

We apologize to researchers whose work has not been included in this manuscript owing to space limit.

- Jaeger, K. E., and Wigge, P. A. (2007). FT protein acts as a long-range signal in *Arabidopsis*. *Curr. Biol.* 17, 1050–1054. doi: 10.1016/j.cub.2007.05.008
- Jung, J. H., Domijan, M., Klose, C., Biswas, S., Ezer, D., Gao, M., et al. (2016). Phytochromes function as thermosensors in *Arabidopsis*. *Science* 354, 886–889. doi: 10.1126/science.aaf6005
- Jung, J. H., Seo, P. J., and Park, C. M. (2012). The E3 ubiquitin ligase HOS1 regulates *Arabidopsis* flowering by mediating CONSTANS degradation under cold stress. *J. Biol. Chem.* 287, 43277–43287. doi: 10.1074/jbc.M112.394338
- Karousis, E. D., Nasif, S., and Mühlemann, O. (2016). Nonsense-mediated mRNA decay: novel mechanistic insights and biological impact. *Wiley Interdiscip. Rev. RNA* 7, 661–682. doi: 10.1002/wrna.1357
- Kim, H. J., Hyun, Y., Park, J. Y., Park, M. J., Park, M. K., Kim, M. D., et al. (2004). A genetic link between cold responses and flowering time through FVE in *Arabidopsis thaliana*. *Nat. Genet.* 36, 167–171. doi: 10.1038/ng1298
- Koini, M. A., Alvey, L., Allen, T., Tilley, C. A., Harberd, N. P., and Whitelam, G. C. (2009). High temperature-mediated adaptations in plant architecture require the bHLH transcription factor PIF4. *Curr. Biol.* 19, 408–413. doi: 10.1016/j.cub.2009.01.046
- Lazaro, A., Mouriz, A., Piñeiro, M., and Jarrillo, J. A. (2015). Red light-mediated degradation of CONSTANS by the E3 ubiquitin ligase HOS1 regulates photoperiodic flowering in *Arabidopsis*. *Plant Cell* 27, 2437–2454. doi: 10.1105/tpc.15.00529
- Lazaro, A., Valverde, F., Piñeiro, M., and Jarrillo, J. A. (2012). The *Arabidopsis* E3 ubiquitin ligase HOS1 negatively regulates constans abundance in the photoperiodic control of flowering. *Plant Cell* 24, 982–999. doi: 10.1105/tpc.110.081885
- Lee, B. D., Kim, M. R., Kang, M. Y., Cha, J. Y., Han, S. H., Nawkar, G. M., et al. (2017). The F-box protein FKF1 inhibits dimerization of COP1 in the control of photoperiodic flowering. *Nat. Commun.* 8:2259. doi: 10.1038/s41467-017-02476-2
- Lee, J. H., Ryu, H. S., Chung, K. S., Posé, D., Kim, S., Schmid, M., et al. (2013). Regulation of temperature-responsive flowering by MADS-box transcription factor repressors. *Science* 342, 628–632. doi: 10.1126/science.1241097
- Lee, J. H., Yoo, S. J., Park, S. H., Hwang, I., Lee, J. S., and Ahn, J. H. (2007). Role of SVP in the control of flowering time by ambient temperature in *Arabidopsis*. *Genes Dev.* 21, 397–402. doi: 10.1101/gad.151847

- Legris, M., Klose, C., Burgie, E. S., Rojas, C. C., Neme, M., Hiltbrunner, A., et al. (2016). Phytochrome B integrates light and temperature signals in *Arabidopsis*. *Science* 354, 897–900. doi: 10.1126/science.aaf5656
- Liu, L. J., Zhang, Y. C., Li, Q. H., Sang, Y., Mao, J., Lian, H. L., et al. (2008). COP1-mediated ubiquitination of constans is implicated in cryptochrome regulation of flowering in *Arabidopsis*. *Plant Cell* 20, 292–306. doi: 10.1105/tpc.107.057281
- Lutz, U., Nussbaumer, T., Spannagl, M., Diener, J., Mayer, K. F., and Schwechheimer, C. (2017). Natural haplotypes of FLM non-coding sequences fine-tune flowering time in ambient spring temperatures in *Arabidopsis*. *Elife* 6:e22114. doi: 10.7554/eLife.22114
- Macknight, R., Duroux, M., Laurie, R., Dijkwel, P., Simpson, G., and Dean, C. (2002). Functional significance of the alternative transcript processing of the *Arabidopsis* floral promoter FCA. *Plant Cell* 14, 877–888. doi: 10.1105/tpc.010456
- Ohama, N., Sato, H., Shinozaki, K., and Yamaguchi-Shinozaki, K. (2017). Transcriptional regulatory network of plant heat stress response. *Trends Plant Sci.* 22, 53–65. doi: 10.1016/j.tplants.2016.08.015
- Pajoro, A., Severing, E., Angenent, G. C., and Immink, R. G. H. (2017). Histone H3 lysine 36 methylation affects temperature-induced alternative splicing and flowering in plants. *Genome Biol.* 18:102. doi: 10.1186/s13059-017-1235-x
- Park, Y. J., Lee, H. J., Gil, K. E., Kim, J. Y., Lee, J. H., Lee, H., et al. (2019). Developmental programming of thermonastic leaf movement. *Plant Physiol.* doi: 10.1104/pp.19.00139 [Epub ahead of print].
- Park, Y. J., Lee, H. J., Ha, J. H., Kim, J. Y., and Park, C. M. (2017). COP1 conveys warm temperature information to hypocotyl thermomorphogenesis. *New Phytol.* 215, 269–280. doi: 10.1111/nph.14581
- Posé, D., Verhage, L., Ott, F., Yant, L., Mathieu, J., Angenent, G. C., et al. (2013). Temperature-dependent regulation of flowering by antagonistic FLM variants. *Nature* 503, 414–417. doi: 10.1038/nature12633
- Qin, Z., Wu, J., Geng, S., Feng, N., Chen, F., Kong, X., et al. (2017). Regulation of FT splicing by an endogenous cue in temperate grasses. *Nat. Commun.* 8:14320. doi: 10.1038/ncomms14320
- Quint, M., Delker, C., Franklin, K. A., Wigge, P. A., Halliday, K. J., and van Zanten, M. (2016). Molecular and genetic control of plant thermomorphogenesis. *Nat. Plants* 2:15190. doi: 10.1038/nplants.2015.190
- Sawa, M., Nusinow, D. A., Kay, S. A., and Imaizumi, T. (2007). FKF1 and GIGANTEA complex formation is required for day-length measurement in *Arabidopsis*. *Science* 318, 261–265. doi: 10.1126/science.1146994
- Seo, P. J., Kim, M. J., Ryu, J. Y., Jeong, E. Y., and Park, C. M. (2011). Two splice variants of the IDD14 transcription factor competitively form nonfunctional heterodimers which may regulate starch metabolism. *Nat. Commun.* 2:303. doi: 10.1038/ncomms1303
- Seo, P. J., Park, M. J., Lim, M. H., Kim, S. G., Lee, M., Baldwin, I. T., et al. (2012). A self-regulatory circuit of CIRCADIAN CLOCK-ASSOCIATED1 underlies the circadian clock regulation of temperature responses in *Arabidopsis*. *Plant Cell* 24, 2427–2442. doi: 10.1105/tpc.112.098723
- Shang, X., Cao, Y., and Ma, L. (2017). Alternative splicing in plant genes: a means of regulating the environmental fitness of plants. *Int. J. Mol. Sci.* 18:E432. doi: 10.3390/ijms18020432
- Shikata, H., Hanada, K., Ushijima, T., Nakashima, M., Suzuki, Y., and Matsushita, T. (2014). Phytochrome controls alternative splicing to mediate light responses in *Arabidopsis*. *Proc. Natl. Acad. Sci. U.S.A.* 111, 18781–18786. doi: 10.1073/pnas.1407147112
- Shim, J. S., Kubota, A., and Imaizumi, T. (2017). Circadian clock and photoperiodic flowering in *Arabidopsis*: constans is a hub for signal integration. *Plant Physiol.* 173, 5–15. doi: 10.1104/pp.16.01327
- Song, Y. H., Estrada, D. A., Johnson, R. S., Kim, S. K., Lee, S. Y., MacCoss, M. J., et al. (2014). Distinct roles of FKF1, gigantea, and zeitlupe proteins in the regulation of constans stability in *Arabidopsis* photoperiodic flowering. *Proc. Natl. Acad. Sci. U.S.A.* 111, 17672–17677. doi: 10.1073/pnas.1415375111
- Song, Y. H., Smith, R. W., To, B. J., Millar, A. J., and Imaizumi, T. (2012). FKF1 conveys timing information for constans stabilization in photoperiodic flowering. *Science* 336, 1045–1049. doi: 10.1126/science.1219644
- Stankovic, N., Schloesser, M., Joris, M., Sauvage, E., Hanikenne, M., and Motte, P. (2016). Dynamic distribution and interaction of the *Arabidopsis* SRSF1 subfamily splicing factors. *Plant Physiol.* 170, 1000–1013. doi: 10.1104/pp.15.01338
- Sureshkumar, S., Dent, C., Seleznev, A., Tasset, C., and Balasubramanian, S. (2016). Nonsense-mediated mRNA decay modulates FLM-dependent thermosensory flowering response in *Arabidopsis*. *Nat. Plants* 2:16055. doi: 10.1038/nplants.2016.55
- Taoka, K., Ohki, I., Tsuji, H., Furuita, K., Hayashi, K., Yanase, T., et al. (2011). 14-3-3 proteins act as intracellular receptors for rice Hd3a florigen. *Nature* 476, 332–335. doi: 10.1038/nature10272
- Wang, C. Q., Guthrie, C., Sarmast, M. K., and Dehesh, K. (2014). BBX19 interacts with constans to repress flowering locus T transcription, defining a flowering time checkpoint in *Arabidopsis*. *Plant Cell* 26, 3589–3602. doi: 10.1105/tpc.114.130252
- Wang, H., Pan, J., Li, Y., Lou, D., Hu, Y., and Yu, D. (2016). The della-constans transcription factor cascade integrates gibberellic acid and photoperiod signaling to regulate flowering. *Plant Physiol.* 172, 479–488. doi: 10.1104/pp.16.00891
- Wang, J. W. (2014). Regulation of flowering time by the miR156-mediated age pathway. *J. Exp. Bot.* 65, 4723–4730. doi: 10.1093/jxb/eru246
- Xiao, S. H., and Manley, J. L. (1998). Phosphorylation-dephosphorylation differentially affects activities of splicing factor ASF/SF2. *EMBO J.* 17, 6359–6367. doi: 10.1093/emboj/17.21.6359
- Xu, L., Zhao, Z., Dong, A., Soubigou-Taconnat, L., Renou, J. P., Steinmetz, A., et al. (2008). Di- and tri- but not monomethylation on histone H3 lysine 36 marks active transcription of genes involved in flowering time regulation and other processes in *Arabidopsis thaliana*. *Mol. Cell Biol.* 28, 1348–1360. doi: 10.1128/MCB.01607-07
- Zhou, H. L., Luo, G., Wise, J. A., and Lou, H. (2014). Regulation of alternative splicing by local histone modifications: potential roles for RNA-guided mechanisms. *Nucleic Acids Res.* 42, 701–713. doi: 10.1093/nar/gkt875

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2019 Park, Lee, Kim and Park. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



Expanding Alternative Splicing Identification by Integrating Multiple Sources of Transcription Data in Tomato

Sarah Clark¹, Feng Yu², Lianfeng Gu³ and Xiang Jia Min^{1*}

¹ Department of Biological Sciences, Youngstown State University, Youngstown, OH, United States, ² Department of Computer Science and Information Systems, Youngstown State University, Youngstown, OH, United States, ³ Basic Forestry and Proteomics Center, College of Forestry, Fujian Agriculture and Forestry University, Fuzhou, China

OPEN ACCESS

Edited by:

Kranthi Kiran Mandadi,
Texas A&M University, United States

Reviewed by:

Chaoling Wei,
Anhui Agricultural University, China
Daqi Fu,
China Agricultural University, China

*Correspondence:

Xiang Jia Min
xmin@ysu.edu

Specialty section:

This article was submitted to
Plant Physiology,
a section of the journal
Frontiers in Plant Science

Received: 20 February 2019

Accepted: 08 May 2019

Published: 28 May 2019

Citation:

Clark S, Yu F, Gu L and Min XJ
(2019) Expanding Alternative Splicing
Identification by Integrating Multiple
Sources of Transcription Data
in Tomato. *Front. Plant Sci.* 10:689.
doi: 10.3389/fpls.2019.00689

Tomato (*Solanum lycopersicum*) is an important vegetable and fruit crop. Its genome was completely sequenced and there are also a large amount of available expressed sequence tags (ESTs) and short reads generated by RNA sequencing (RNA-seq) technologies. Mapping transcripts including mRNA sequences, ESTs, and RNA-seq reads to the genome allows identifying pre-mRNA alternative splicing (AS), a post-transcriptional process generating two or more RNA isoforms from one pre-mRNA transcript. We comprehensively analyzed the AS landscape in tomato by integrating genome mapping information of all available mRNA and ESTs with mapping information of RNA-seq reads which were collected from 27 published projects. A total of 369,911 AS events were identified from 34,419 genomic loci involving 161,913 transcripts. Within the basic AS events, intron retention is the prevalent type (18.9%), followed by alternative acceptor site (12.9%) and alternative donor site (7.3%), with exon skipping as the least type (6.0%). Complex AS types having two or more basic event accounted for 54.9% of total AS events. Within 35,768 annotated protein-coding gene models, 23,233 gene models were found having pre-mRNAs generating AS isoform transcripts. Thus the estimated AS rate was 65.0% in tomato. The list of identified AS genes with their corresponding transcript isoforms serves as a catalog for further detailed examination of gene functions in tomato biology. The post-transcriptional information is also expected to be useful in improving the predicted gene models in tomato. The sequence and annotation information can be accessed at plant alternative splicing database (<http://proteomics.ysu.edu/altsplice>).

Keywords: alternative splicing, gene expression, tomato, mRNA, plant, *Solanum lycopersicum*, transcriptome

INTRODUCTION

Understanding the transcriptome diversity and gene expression dynamics is critical for developing methods for further improving the quantity and quality of plant products. Tomato (*Solanum lycopersicum*), one of the important fruit and vegetable crops, has its genome being completely sequenced. The complete genome of the inbred tomato cultivar “Heinz 1706” approximately has a size of 900 megabases (Mb) with a total of 34,727 protein-coding genes predicted

(Tomato Genome Consortium, 2012). Recently updated release of the tomato genome assembly (SL3.0, ITAG3.2 release) contains 35,768 gene models¹. Since the release of the tomato complete genome sequences, large amounts of RNA-seq data in tomato have been generated in a number of research projects from tissues at variable different developmental stages, growing under different conditions, or challenged with different pathogens, as well as comparative transcriptome analysis with wild tomato plants (Kumar and Khurana, 2014). Mapping RNA-seq data, generated in domesticated and wild tomato plants under various conditions or treatments, to the released complete genome sequences as a reference genome has significantly contributed to our understanding the transcriptome complexity and regulations including differential gene expressions and alternative splicing (AS) in tomato (Wang K. et al., 2016; Dai et al., 2017).

A gene can be transcribed to form two or more RNA transcripts using the process of AS in intron containing eukaryotic organisms (Reddy et al., 2013), thus, significantly increasing the diversities of mRNAs and proteins in the organism. AS commonly occurs in eukaryotes including protists, fungi, plants, and animals (McGuire et al., 2008). Four basic types of AS including exon skipping (ExonS), alternative donor site (AltD), alternative acceptor (AltA) site, and intron retention (IntronR) were commonly found (Wang and Brendel, 2006; Sablok et al., 2011). Various complex types can be formed by combination of basic events (Sablok et al., 2011). While these basic types can be found in all kingdoms of eukaryotes, ExonS is most prevalent event in animals including humans and IntronR is the dominant event in plants (McGuire et al., 2008), suggesting that the splicing mechanisms may be different in animals and plants. Numerous experimental results showed AS plays important roles in many biological processes in plants such as photosynthesis, defense responses, flowering timing, responses to stresses, etc., (Reddy et al., 2013; Staiger and Brown, 2013). AS isoforms may or may not be functional. The functional isoforms may encode distinct functional proteins and the non-functional isoforms are degraded by a process known as nonsense-mediated decay (NMD) (Lewis et al., 2003; Xu et al., 2014).

Alternative splicing has been examined in a number of plant species including the model species, *Arabidopsis thaliana*, crop plants including rice, maize, sorghum, etc., (Zhang C. et al., 2016). Due to the differences in the amounts of available gene expression data in different plant species, the estimated AS rates vary tremendously from below 10% to ~70% in intron-containing genes (Zhang C. et al., 2016; Min, 2017). For example, due to relatively large amounts of transcription data available in *Arabidopsis*, it was estimated that ~60–70% of multi-exon genes undergoing AS (Filichkin et al., 2010; Marquez et al., 2012; Yu et al., 2016; Zhang R. et al., 2017). Other well analyzed plant species were rice (*Oryza sativa*) (Wang and Brendel, 2006; Min et al., 2015; Wei et al., 2017; Chae et al., 2017; Zhang et al., 2018), maize (*Zea mays*) (Thatcher et al., 2014; Min et al., 2015; Thatcher et al., 2016; Mei et al., 2017; Min, 2017), and sorghum (*Sorghum bicolor*) (Panahi et al., 2014; Min et al., 2015; Abdel-Ghany et al., 2016); fruit plants such as grape (*Vitis vinifera*) (Vitulo et al., 2014;

Sablok et al., 2017), and fiber plants such as cotton (*Gossypium raimondii*; *G. barbadense*; *G. davidsonii*, and *G. hirsutum*) (Li et al., 2014; Min, 2018; Wang et al., 2018a; Zhu G. et al., 2018). However, the above mentioned species are just few examples of AS analysis in plants, not an exhaustive list of all plants AS work. Further, genome-wide conserved AS events in flowering plant species as well as in monocot species have also been analyzed (Chamala et al., 2015; Mei et al., 2017).

Transcriptome analysis in plants have been carried out intensively using recently developed RNA-sequencing (RNA-seq) technology. A number of well-designed experiments in genome-wide transcriptome analysis for identifying differentially expressed genes and/or AS in tomato, a model plant specifically for fruit development, have been reported in recent few years. These studies include comparative transcriptome analysis of domesticated tomato for identifying differentially expressed genes in different tissues (Lopez-Casado et al., 2012; Koenig et al., 2013; Zouine et al., 2014; Sun and Xiao, 2015; Sundaresan et al., 2016; Wang K. et al., 2016; Zhang Y. et al., 2017), diurnal transcriptome changes (Higashi et al., 2016), global transcriptome profiles of tomato leaf responses to exogenous ABA or cytokinin (Wang et al., 2013; Shi et al., 2013), root transcriptome regulations in response to different plant hormone cytokinin and auxin (Gupta et al., 2013), transcriptome profiles with a focus on fruit development or in different fruit tissues (Ye et al., 2015; Zhang S. et al., 2016; Dai et al., 2017; Ezura et al., 2017) and fruit chilling tolerance (Cruz-Mendivil et al., 2015). RNA-seq data were also collected for analysis of differential gene expressions in response to tomato yellow leaf curl virus (TYLCV) infection in the TYLCV-resistant (R) breeding line and TYLCV-susceptible breeding line (Chen et al., 2013), in response to tobacco rattle virus (TRV) (Zheng et al., 2017), *Pseudomonas syringae* (Yang et al., 2015; Worley et al., 2016), *Xanthomonas perforans* (Du et al., 2015), *Cladosporium fulvum* (Xue et al., 2016), *Colletotrichum gloeosporioides* (Alkan et al., 2015), *Verticillium dahlia* (Tan et al., 2015), *Meloidogyne incognita* (root-knot nematode) (Shukla et al., 2018) and in arbuscular mycorrhiza inoculated and control plants (Zouari et al., 2014). Transcriptome analysis was also carried out with mutants including high pigment mutant (Tang et al., 2013) and SIEIN2-silenced tomato (ethylene insensitive 2) mutant which had a non-ripening phenotype (Wang R.H. et al., 2016).

The aforementioned RNA-seq projects generated large amounts of RNA-seq data in tomato, which provide an unprecedented opportunity for integrating these transcriptome data with publicly available expressed sequence tag (EST) and mRNA sequences for identifying alternatively spliced genes in tomato. However, the integrated study for AS events in tomato has not been reported. Thus, the aim of the current work is to maximize AS identification and generate a comprehensive catalog of alternatively spliced genes and AS events in tomato by integrating ESTs, mRNAs, and RNA-seq data available in public databases. The identified alternatively spliced genes and AS events with detailed annotation information in the work are expected to provide a solid resource for tomato researchers for further detailed functional analysis of these genes in tomato growth and development including fruit production.

¹https://solgenomics.net/organism/Solanum_lycopersicum/genome

MATERIALS AND METHODS

Genome, EST, and mRNA Sequence Datasets

Tomato genome sequences (version SL3.0) and associated annotation files (ITAG3.20) were downloaded from the International Tomato Genome Sequencing Project (see text footnote 1). Using “*Solanum lycopersicum*” as “organism” we downloaded 300,665 EST sequences and 53,613 mRNA sequences of tomato from EST and nucleotide database at the National Center for Biotechnology Information (NCBI).

We used a procedure well implemented in our previous analysis for cleaning the data (Min et al., 2015; Min, 2018). The procedure used EMBOSS trimmest tool for trimming the polyA or polyT end (Rice et al., 2000), BLASTN search against UniVec and *E. coli* database for removal of vector and *E. coli* contaminants, and BLASTN search against the plant repeat database for removal of the repetitive sequences including transposable elements. A total of 350,141 cleaned EST and mRNA sequences were obtained and combined with 39,095 transcript sequences generated by Scarano et al. (2017) and 250,676 transcript sequences generated by Wang K. et al. (2016). Thus a total of 639,912 sequences were used for assembling using CAP3 (Huang and Madan, 1999). A total of 452,672 putative unique transcripts including 27,791 contigs and 424,881 singlets were obtained for mapping to tomato genome sequences.

Mapping Assembled Transcripts to the Genome

We used ASFinder to map the assembled transcripts tomato genome sequences (Min, 2013). We applied the threshold values as reported previously (Walters et al., 2013). Mapped transcripts having an intron size > 100 kb were removed for AS identification in order to avoid chimeric transcripts.

RNA-Seq Data Mapping to the Genome

We downloaded tomato RNA-seq sequence data from the NCBI SRA database² using SRA Toolkit. The RNA-seq data were retrieved from 27 published papers, which were listed in **Supplementary Table S1**. In total, 2,543 Gbs RNA-seq data were downloaded. The data from each publication were processed individually. The RNA-seq reads were mapped to tomato genome sequences using TopHat (v2.2.6) with default parameters (Kim et al., 2013). Then the transcript alignment file together with the ITAG3.20 annotation was used as input for Cufflinks (v2.2.1)³ (Trapnell et al., 2010). The GTF (Gene Transfer Format) files generated from each RNA-seq dataset after Cufflinks were merged using Cuffcompare script within the Cufflinks package (Trapnell et al., 2010). The GTF file generated from merged RNA-seq GTF files then was further merged using Cuffcompare script with the GTF file that was generated by the ASFinder for mapping the assembled ESTs and mRNA

(transcripts) sequences to the genome to generate a final GTF file for AS analysis. AStalavista was used for AS event classification (Foissac and Sammeth, 2007).

Transcript Functional Annotation

The sequences of the transcripts were retrieved using gtf_to_fasta tool in the tophat package (Kim et al., 2013), based on the GTF file generated by Cuffcompare program after merging the EST and mRNA mapping GTF file and RNA-seq mapping GTF file. They were functionally annotated using a procedure we reported previously (Min, 2018). The annotation information contains protein coding regions (ORF) prediction, assessment of full-length transcript coverage, protein family, and comparison with sequences of predicted gene models. Gene Ontology (GO) information was extracted also using a procedure reported previously (Min, 2018). Transcripts not having BLASTX hit against UniProtKB-SwissProt database were further used for non-coding RNA (ncRNA) identification by using BLASTN search against the non-coding RNA central database (version 10)⁴ with a cutoff *E*-value of 1e-5.

Transposable Element Analysis in Introns

Intron sequences were retrieved using an in-house script. Transposable elements in the introns were identified using BLASTN searching against the RepBase (Bao et al., 2015)⁵ using a cutoff *E*-value of 1e-5.

Internal Exon and Intron Length Distribution and Exon/Intron Junctions

The lengths of internal exons and introns in all transcripts and sizes of DNA fragments involved in AS were analyzed. The exon and intron junction sequences were extracted from genes not undergoing AS (non-AS genes) and genes undergoing AS (AS-genes). The exon/intron boundary sequence logo was created using the weblogo server⁶ (Crooks et al., 2004).

Availability of Data

The assembled transcripts and AS events identified in this study along with the predicted gene models, along with the data reported previously in our group including *Brachypodium distachyon* (Sablok et al., 2011; Walters et al., 2013), pineapple (Wai et al., 2016), and sacred lotus (*Nelumbo nucifera*) (VanBuren et al., 2013), are available from plant alternative splicing database⁷ (Walters et al., 2013; VanBuren et al., 2013; Min et al., 2015; Wai et al., 2016, 2018; Sablok et al., 2017; Min, 2017, 2018). BLAST search is also available for searching the transcripts. The datasets for database construction and the supplementary data are publicly available at: <http://proteomics.ysu.edu/publication/data/Tomato/>.

²<https://www.ncbi.nlm.nih.gov/sra/docs/sradownload/>

³<http://cole-trapnell-lab.github.io/cufflinks/>

⁴<https://rnacentral.org/>

⁵<https://www.girinst.org/repbase/>

⁶<http://weblogo.threeplusone.com/>

⁷<http://proteomics.ysu.edu/altsplice>

RESULTS

Features of Assembled Transcripts

In this work we integrated genome mapping data from available ESTs and mRNAs in the public nucleotide database with RNA-seq data downloaded from the NCBI SRA database that were obtained from 27 previous publications (see **Supplementary Table S1**). A total of 533,707 putative unique transcripts with an average length of 1,350 bp were obtained based on the final GTF file which was generated by merging the mapping GTF file of assembled EST and mRNA sequences and the mapping GTF file of RNA-seq data to tomato genome sequences (**Table 1**). The basic features of the transcripts in the dataset were summarized (**Table 1**).

The mapped transcripts were clustered to a total of 260,681 genomic loci (**Table 1**), which were significantly higher than the number of protein coding genes (35,768 gene models, 35,768 protein sequences and 35,768 cDNAs) annotated in ITAG3.20. Using ungapped BLASTN search with 100% identity and a minimum length of 80 bp in a high score aligned segment, a total of 260,365 transcripts accounting for 48.8% of total transcripts matched with cDNA sequences of gene models. A total of 34,522 (96.5%) unique loci, represented by cDNA sequences, have matched at least one transcript. Assembled transcripts were annotated functionally using BLASTX search

against the UniProt-SwissProt database. Among them, a total of 226,881 (42.5%) has a BLASTX hit and 182,325 transcripts were predicted to contain a complete ORF region, i.e., a full-length ORF. Within the assembled transcripts, a total of 518,307 ORFs were predicted with an average length of 215 amino acids and 176,324 ORFs were mapped to a protein family (Pfam) using rpsBLAST (**Table 1**). Transcripts which were not able to match with a predicted cDNA transcript sequence were likely novel transcripts. Whether the novel transcripts identified in the projects represent the transcription noise, i.e., without any biological significance, or play certain biological roles remain to be examined in future study.

AS Events Identification

Tomato genome has 12 chromosomes (Tomato Genome Consortium, 2012). Chromosome zero (Chr0) is a segment of chromosome that has not been assigned to a specific chromosome (**Table 2**). There were 369,911 AS events identified from 34,419 genomic loci involving 161,913 transcripts. Although there are variations in the total number of AS events among different chromosomes, the general AS event distribution patterns are consistent across chromosomes (**Table 2**). Among the four basic AS types, IntronR is the prevalent type of AS event (18.9%), followed by AltA (12.9%) and AltD (7.3%), and ExonS as the least type (6.0%). Various complex types can be formed in transcript isoforms by combination of basic events and 54.9% of AS events were complex types (2).

Among 369,911 transcripts generated from pre-mRNAs having AS, 150,131 matched to cDNAs of 23,233 unique annotated gene models (See **Supplementary Table S2**). We noticed that there were some gene models undergoing AS having only one corresponding transcript shown in the table because other transcripts involved in AS did not have a sufficient overlap region with the gene model transcript (**Supplementary Table S2**). However, the complete information of isoform transcripts can be obtained from the database and the genome browser. Based on the mapping analysis, there were 34,522 unique loci (gene models) having EST/mRNA or RNA-seq mapped, i.e., these

TABLE 1 | Basic features of the assembled unique transcripts in tomato plants.

Total unique transcripts	533707
Average transcript length (bp)	1350
Total genomic loci with at least one transcript	260681
Transcripts matching with gene model cDNAs	260365
Unique gene model cDNAs matching with transcripts	34522
Transcripts having a BLASTX match against Swiss-Prot database	226881
Total predicted ORFs from assembled transcripts	518307
Average length of predicted ORFs (amino acids)	215
Predicted full-length ORFs	182325
Predicted ORFs having a PFAM match	176234

TABLE 2 | Summary of alternative splicing events in each chromosome of tomato plants.

	AltA (%)		AltD (%)		ExonS (%)		IntronR (%)		Others (%)		Total
Chr0	188	11.8	108	6.8	55	3.4	451	28.3	794	49.7	1596
Chr1	6410	13.5	3592	7.6	2923	6.2	9269	19.5	25276	53.2	47470
Chr2	5396	10.1	3156	5.9	2379	4.5	8324	15.6	34184	64.0	53439
Chr3	5168	13.2	2856	7.3	2274	5.8	7432	19.0	21476	54.8	39206
Chr4	3988	14.0	2228	7.8	1851	6.5	5567	19.5	14926	52.3	28560
Chr5	3026	13.4	1818	8.0	1434	6.3	4488	19.8	11883	52.5	22649
Chr6	4077	14.1	2173	7.5	1758	6.1	5938	20.5	15071	51.9	29017
Chr7	3379	12.3	1855	6.7	1661	6.0	5106	18.6	15506	56.4	27507
Chr8	3355	13.3	1839	7.3	1537	6.1	4979	19.8	13426	53.4	25136
Chr9	3281	12.8	1995	7.8	1600	6.3	4896	19.2	13786	53.9	25558
Chr10	2883	13.6	1572	7.4	1378	6.5	4509	21.3	10811	51.1	21153
Chr11	3254	13.0	1876	7.5	1630	6.5	4383	17.5	13973	55.6	25116
Chr12	3316	14.1	1905	8.1	1681	7.2	4693	20.0	11909	50.7	23504
Total	47721	12.9	26973	7.3	22161	6.0	70035	18.9	203021	54.9	369911

genes were supported with transcription data. Thus, using only expressed genes the AS rate in tomato was estimated to be 67.3%. However, when all gene models were used, the estimated AS rate was 65.0%.

Li et al. (2014) reported that transposons were enriched in the retained introns in cotton plants. While only 2.9% of all introns contained transposable elements (TEs), 43% of the retained introns were found to have TEs in the AS transcripts, suggesting TE-insertion may result in IntronR during pre-mRNA splicing in cotton (Li et al., 2014). We retrieved 68,241 retained introns from our datasets with a length > 50 bp and found only 812 TEs (1.2%). While in the whole set (138,127) of introns of the predicted gene models, using the same cutoff value, 3,105 TEs (2.3%) were found. Thus, the TE enrichment phenomenon reported by Li et al. (2014) was not found in tomato.

To facilitate identifying project specific AS events which may aid in elucidating the biological significances of AS isoforms, we analyzed AS events in each individual projects (Table 3). Because we used pooled data from different projects with variable data sizes and treatment conditions, it is difficult to directly compare

the results from these projects (Supplementary Table S1). However, the overall trend in AS type frequency distribution is consistent with the final combined data (Tables 2, 3), that is, IntronR is the most prevalent type of AS, followed by AltA and AltD, with ExonR as the least frequent type. The RNA-seq mapping information including expression levels, AS event analysis, and information for sequence identifier mapping to the final assembled transcripts are available for downloading at <http://proteomics.yzu.edu/publication/data/Tomato/projects/>.

Functional Annotation of Transcripts

All transcripts were functionally annotated including BALSTX search against the UniProtKB-SwissProt database and predicting the ORF regions and completeness of ORFs (see section “Transcript Functional Annotation”). The predicted proteins were further annotated for the protein family (Pfam) analysis. To provide an overview of the protein family distribution in tomato proteome and proteins encoded by genes generating AS isoforms, we used the protein sequences of the gene models for Pfam analysis. Among 35,768 protein sequences of gene models a total of 22,322 entries had Pfam matches with a total of 3,319 unique Pfam. Among a total of 23,233 protein sequences generated from AS genes, a total 16,531 had Pfam matches with a total of 3,114 unique Pfam. The top Pfam in the whole tomato proteome and proteins encoded by genes undergoing AS were listed in Table 4. The numbers of proteins in each Pfam varied significantly from 1 member in some families to 631 members in Pkinase (pfam00069). In average 74.1% of Pfam members were alternatively spliced with varying proportions in different protein families (Table 4). In considering the varying Pfam size, number of exons per gene and gene expression levels as well as functional differences of these protein coding genes, such a difference in AS rates in the genes belonging to different protein families is expected. Comparing with our previous Pfam analysis of proteins generated from AS genes in cereal plants and fruit plants, these Pfams found in tomato AS genes were also well conserved in other plant species (Min et al., 2015; Sablok et al., 2017).

The isoform transcripts generated by alternative pre-mRNA splicing may or may not be functional. Thus the impact of AS on the functionalities of isoforms was assessed using Pfam annotation information of the predicted proteins. Within a total of 369,911 pairs generating AS events, 114,729 (31.0%) pairs did not have Pfam annotation, 164,594 (44.5%) pairs had same Pfam annotation; 64,107 (17.3%) pairs had one isoform with Pfam annotation and one isoform did not have Pfam annotation, suggesting either no protein sequences predicted or a loss of protein functionality; and 26,481 (7.2%) pairs had different Pfam annotation, suggesting a functional domain change in the protein sequences resulting from AS. Thus, the comparative Pfam analysis of the proteins encoded by the isoform transcripts generated by AS revealed that about 24.5% of them may cause functional loss or change in the protein isoforms. Our previous analysis showed that the resulting functional domain loss or change by AS were 19.6% in maize, 20.9% in cotton, and 24.9% in pineapple, respectively (Wai et al., 2016; Min, 2017, 2018). The translation frame changes in AS isoforms is the main reason for protein domain loss or change.

TABLE 3 | Summary of alternative splicing events identified in EST and mRNA assembly dataset and each RNA-seq dataset in tomato plants.

Data ^a	AltA	AltD	ExonS	IntronR	Others
ESTs and mRNAs	13738	7779	9355	17255	32448
RNA-seq projects					
Alkan (2014)	2664	1841	1367	3966	3057
Chen (2013)	1425	1068	711	2238	1528
Cruz-Mendivil (2015)	122	130	88	329	448
Dai (2017)	2075	1440	992	2574	1920
Du (2015)	2557	1873	1273	3851	2878
Ezura (2017)	254	251	88	3698	1658
Gupta (2013)	2619	1909	1375	4723	3437
Higashi (2016)	1773	1256	967	7274	2916
Koenig (2013)	1791	1376	1005	2774	1998
Lopez-Casado (2011)	565	521	353	995	889
Shi (2013)	1773	1431	972	4806	2810
Shukla (2017)	18581	9519	7032	20625	29217
Sun (2015)	5309	3255	1876	5925	4530
Sundaresan (2016)	1857	1286	920	2461	1802
Tan (2015)	1268	954	656	1863	1287
Tang (2013)	1353	1057	816	1813	1401
Wang (2013)	5324	3022	1945	5945	4683
Wang (2016)	3603	2264	1857	5079	4818
Worley (2016)	1921	1308	1078	3769	2402
Xue (2017)	6455	3602	3246	8886	9113
Yang (2015)	2093	1356	1257	3259	2041
Ye (2015)	1389	1127	762	3678	2112
Zhang (2016)	6109	3399	2727	7825	6880
Zhang (2017)	2493	1623	1385	3121	2396
Zheng (2017)	412	376	267	972	683
Zouari (2014)	2679	1820	1436	3878	3091
Zouine (2014)	2853	1935	1357	3278	2453

^aThe last name of the first author with year of the publication is used the project ID. The details of the reference for each project can be found in the Supplementary Table S1.

TABLE 4 | Protein families in gene models and alternatively spliced genes in tomato plants*.

Pfam ID	Total	AS genes	%	Pfam abbreviation	Pfam description
pfam00069	631	503	79.7	Pkinase	Protein kinase domain
pfam07714	467	388	83.1	Pkinase_Tyr	Protein tyrosine kinase
pfam00067	337	216	64.1	p450	Cytochrome P450
pfam13041	316	247	78.2	PPR_2	PPR repeat family
pfam13639	223	140	62.8	zf-RING_2	Ring finger domain
pfam00931	217	151	69.6	NB-ARC	NB-ARC domain
pfam00076	179	162	90.5	RRM_1	RNA recognition motif
pfam00249	178	111	62.4	Myb_DNA-binding	Myb-like DNA-binding domain
pfam00201	169	86	50.9	UDPGT	UDP-glucuronosyl and UDP-glucosyl transferase
pfam10536	167	101	60.5	PMD	Plant mobile domain
pfam03171	147	87	59.2	2OG-Fe(II)_Oxy	2OG-Fe(II) oxygenase superfamily
pfam00847	137	62	45.3	AP2	AP2 domain
pfam02519	130	35	26.9	Auxin_inducible	Auxin responsive protein
pfam00141	123	81	65.9	peroxidase	Peroxidase
pfam05699	118	28	23.7	Dimer_Tnp_hAT	hAT family C-terminal dimerisation
pfam00319	109	34	31.2	SRF-TF	SRF-type transcription factor
pfam02458	99	51	51.5	Transferase	Transferase family
pfam14432	95	68	71.6	DYW_deaminase	DYW family of nucleic acid deaminases
pfam00481	91	79	86.8	PP2C	Protein phosphatase 2C
pfam00854	86	75	87.2	PTR2	POT family
pfam01095	85	42	49.4	Pectinesterase	Pectinesterase
pfam00561	84	70	83.3	Abhydrolase_1	alpha/beta hydrolase fold
pfam00657	83	65	78.3	Lipase_GDSL	GDSL-like Lipase/Acylhydrolase
pfam00010	82	70	85.4	HLH	Helix-loop-helix DNA-binding domain
pfam03106	82	54	65.9	WRKY	WRKY DNA-binding domain
pfam01554	77	66	85.7	MatE	MatE
pfam13839	75	63	84.0	PC-Esterase	GDSL/SGNH-like Acyl-Esterase family found
pfam00004	74	64	86.5	AAA	ATPase family associated with various cellular
pfam02362	74	49	66.2	B3	B3 DNA binding domain
pfam00071	73	58	79.5	Ras	Ras family
pfam00083	73	53	72.6	Sugar_tr	Sugar (and other) transporter
pfam05695	72	13	18.1	DUF825	Plant protein of unknown function (DUF825)
pfam00082	71	41	57.7	Peptidase_S8	Subtilase family
pfam13499	71	40	56.3	EF-hand_7	EF-hand domain pair
Others	17227	13078	75.9		
Total	22322	16531	74.1		

A total of 342,073 transcripts not having a BLASTX hit against the UniProt Swiss-Prot database were further used to search the ncRNA database using BLASTN. The ncRNA database was obtained from the RNACentral (release 10) with a total of 11,963,117 ncRNA sequences. With a cutoff *E*-value of 1e-5, we identified a total of 136,643 transcripts sharing similarities with known ncRNAs. The list of ncRNAs can be downloaded at <http://proteomics.ysu.edu/publication/data/Tomato/>.

Gene Ontology (GO) Analysis

The gene ontology⁸ is “a community-based bioinformatics resource that supplies information about gene product function using ontologies to represent biological knowledge”

⁸<http://www.geneontology.org>

(Gene Ontology Consortium, 2014). GO classification has three major categories including the biological processes, molecular functions, and cellular components. We used the protein sequences of gene models to search the UniProtKB/Swiss-Prot database, then retrieved GO IDs based on the UniProt ID mapping table. Among 35,768 protein sequences predicted from gene models, 25,048 of them had a BLASTP hit with UniProtKB/Swiss-Protein dataset. Among them 18,031 were from protein sequences of genes with pre-mRNAs undergoing AS. We then retrieved 155,246 and 114,667 GO IDs for the whole set and AS gene set, respectively. The GO IDs were further mapped to each category using Slim Viewer with plant specific GO terms (McCarthy et al., 2006). The top GO terms in each category were presented in **Figure 1**. The percentage of each category was calculated based on the total counts in each category.

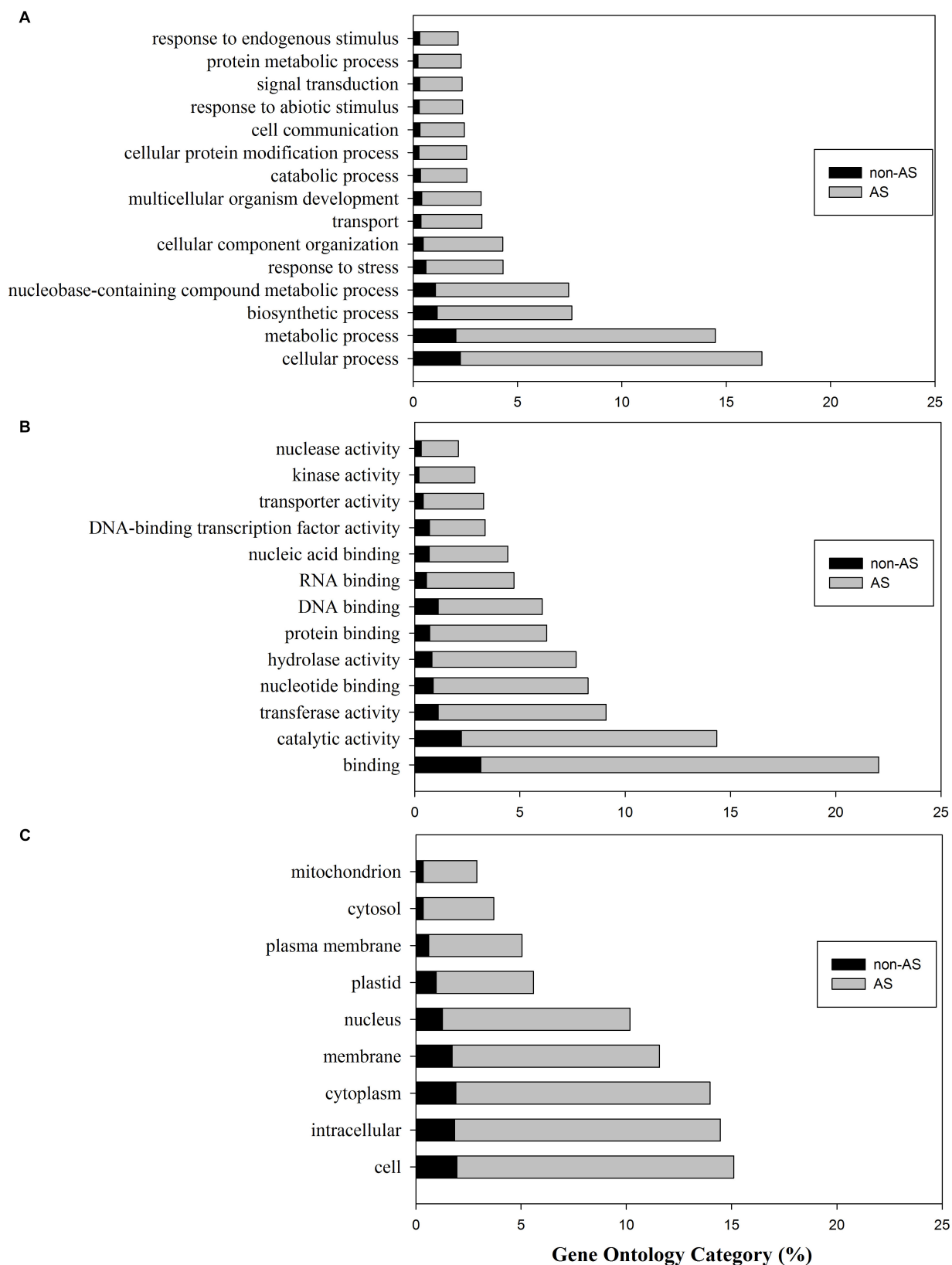
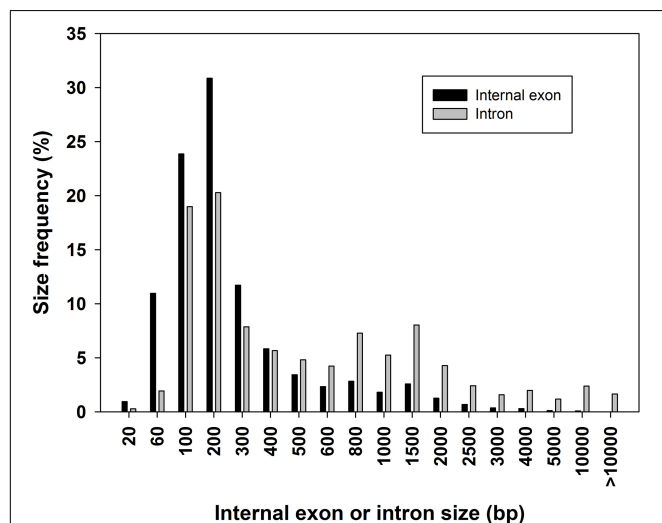


FIGURE 1 | Gene ontology (GO) classification of tomato genes with pre-mRNAs not undergoing alternative splicing (non-AS genes) and genes with pre-mRNA undergoing alternative splicing (AS-genes). **(A)** Biological process; **(B)** Molecular function; **(C)** Cellular components.

TABLE 5 | Summary of internal exon length and intron length of all transcripts, and DNA fragment sizes (bp) involved in alternative splicing events in tomato.

	Sample size	Length range (bp)	Mean (bp)	Standard deviation (bp)
Internal exons	215952	1–88397	282	547
Introns	282296	5–313176	1352	7609
Retained introns	70035	6–19337	366	710
Alternative acceptor sites	47721	1–12110	145	451
Alternative donor sites	26973	1–11238	212	493
Skipped exons	22161	2–25108	214	446

**FIGURE 2 |** Distribution of internal exon size and intron size in tomato genes. Bin size are right inclusive (e.g., bin 100 comprises sequences of lengths 1–100 bp).

In the whole tomato proteome set, the mapped GO term counts were 47,795 in biological processes; 25,989 in molecular functions, and 33,517 in cellular components. The comparative analysis showed that 86.5% of genes involved biological processes may undergo alternative splicing (Figure 1A). The top biological processes include cellular process, metabolic process, biosynthetic process, nucleobase-containing compound metabolic process, response to stress, cellular component organization, etc., (Figure 1A). In particular, 153 genes (76.5%) from a total of 200 genes involved in the secondary metabolic process (GO:0019748) were found undergoing AS. These genes are involved in the biosynthesis of lycopene, carotenoid, abscisic acid, flavonoid, anthocyanins, etc., (Bovy et al., 2007; Liu et al., 2015). AS analyses involved in secondary metabolism pathways including the flavonoid pathway in tea plant have been carried out, suggesting AS may play important roles in regulation of flavonoids biosynthesis (Zhu J. et al., 2018; Qiao et al., 2019).

Similarly, the majority (86.2%) of tomato gene products in the category of molecular functions were also alternatively spliced (Figure 1B). The top categories of molecular functions include binding, catalytic activity, transferase, nucleotide binding, hydrolase activity, protein binding, etc., (Figure 1B). Cellular

components analysis also revealed that ~86.5% of tomato genes with GO cellular component annotation having pre-mRNAs were alternatively spliced (Figure 1C).

Features of Exons, Introns and Exon-Intron Junctions

The lengths of internal exons and introns, as well as the DNA fragment sizes involved in AS events were calculated based on the RNA to genome mapping information (Table 5 and Figure 2). A total of 215,952 internal exons and 282,296 introns were extracted. The sizes of internal exons varied from 1 to 88,397 bp with a mean value of 282 bp; intron sizes varied from 5 to 313,176 bp with a mean value of 1,352 bp (Table 5). Among internal exons, 84.2% of them had a size of ≤ 400 bp and 94.6% were ≤ 1000 bp (Figure 2). In contrast, 55.0% of introns were ≤ 400 bp and 76.5% were ≤ 1000 bp. There were 0.01% of internal exons and 1.44% of introns ≥ 10 kb. In addition, there were 350 introns with a size > 100 kb. As we removed alignments in the mapping of EST/mRNA assembled data, these introns were clearly from the RNA-seq mapping. We manually checked some of the transcripts having long introns and found that these extremely long introns were likely due to the mapping of the fused transcripts generated from different genes. The fused transcripts in plants often are ignored as transcription noise. However, it is known that fusion of transcripts in human is related to cancer development (Mertens et al., 2015; Kumar et al., 2016). Similar to what we found in *B. distachyon* and in fruit plants (Walters et al., 2013; Sablok et al., 2017), DNA fragments involved in AS events were relatively shorter than the average size of the internal exons or introns in tomato (Table 5). Thus, it is reasonable to conclude that small exons tend to be skipped and small introns tend to be retained.

There were a total 46,114 introns extracted from genes not having AS (non-AS genes) and 236,182 introns from genes having AS (AS genes). Two nucleotides from each end of introns were extracted from both non-AS genes and AS genes. The majority of introns (90.6% in average) in both AS genes (90.9%) and non-AS genes (89.0%) had a canonical splicing junction site of 5'-GT..AG-3' (Table 6). The minor types of splicing sites in introns included 5'-GC..AG-3', 5'-GC..AT-3', 5'-AT..AC-3', and many others types (Table 6). A chi-square test showed there was no significant difference in the frequencies of the types of splicing sites between AS genes and non-AS genes (Table 6).

TABLE 6 | The usage of different splicing sites at both ends of the introns in tomato.

Types	Total	%	Non-AS gene	%	AS gene	%
5'-GT..AG-3'	255669	90.6	41053	89.0	214616	90.9
5'-GC..AG-3'	7571	2.7	1511	3.3	6060	2.6
5'-GC..AT-3'	3060	1.1	712	1.5	2348	1.0
5'-AT..AC-3'	2751	1.0	844	1.8	1907	0.8
5'-CT..AC-3'	2067	0.7	232	0.5	1835	0.8
5'-GT..AT-3'	2036	0.7	554	1.2	1482	0.6
Others	9142	3.2	1208	2.6	7934	3.4
Total	282296		46114		236182	

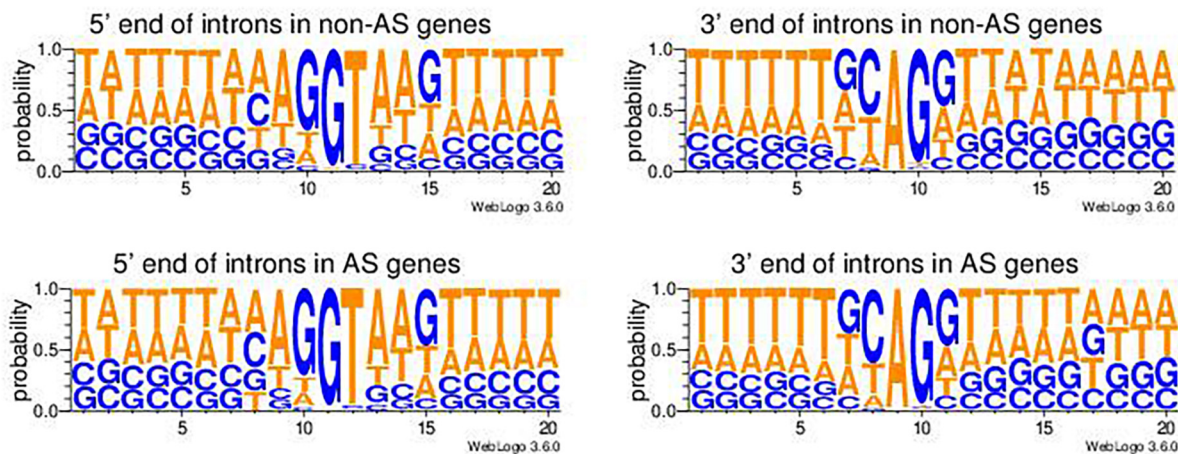


FIGURE 3 | Pictograms of nucleotide probabilities at each position of the exon-intron junctions in genes not undergoing alternative splicing (non-AS genes) and genes undergoing alternative splicing (AS genes). The 5'-end intronic nucleotides are from position 11 to 20 in the left panel pictograms and the 3'-end intronic nucleotides are from position 1 to 10 in the right panel pictograms.

The pictograms showed that the only noticeable difference in nucleotide usage probabilities in the junction sites at the 5'-end of introns between non-AS genes and AS-genes was position 8 (**Figure 3**, left panels). However, there were noticeable differences in the nucleotide probabilities at the 3'-end of introns but within the 5'-end of the next exonic region of, i.e., at position 14, 16, and 17 (**Figure 3**, right panels). Whether there is any biological significance remains to be examined.

DISCUSSION

In this work a much higher number of transcripts, than any previous report, were identified in tomato, as we integrated all currently available EST/mRNA sequences with RNA-seq data generated from 27 RNA-seq projects covering a broad range of biological samples with plants grown under various conditions (see **Supplementary Table S1**). The human ENCODE project reported the genome is pervasively transcribed (~76% of the full genome transcribed), and “many novel non-protein-coding transcripts have been identified, with many of these overlapping protein-coding loci and others located in regions of the genome previously thought to be transcriptionally silent” (Encode Project Consortium, 2007; Pennisi, 2012). Thus the current set of transcripts represents the most comprehensive and complete set of transcripts identified in tomato by now.

The basic type distribution patterns of AS events in tomato are consistent with findings in other plant species (Wang and Brendel, 2006; Walters et al., 2013; VanBuren et al., 2013; Thatcher et al., 2014; Min et al., 2015; Sablok et al., 2017). However, the proportion of the complex type is related to the transcriptome sampling size and thus the completeness and the average length of the transcripts (Min et al., 2015; Min, 2017, 2018; Sablok et al., 2017). Long transcripts have more exons covered and thus are able to detect AS isoforms having more than one type of AS event in their sequences.

The estimated AS rate in tomato was estimated ~65.0% in the analysis. This AS rate in tomato is nearly reached to the maximal rate, though such a value may never be able to obtain due to the dynamic nature of transcriptomes. At least this rate is comparable with the rate reported in Arabidopsis (~60%) and in maize (55%) (Marquez et al., 2012; Mei et al., 2017; Min, 2017). Obtaining such a high rate is clearly due to relatively large number of RNA-seq data were used in current analysis. Using RNA-seq data to detect AS isoforms has been widely accepted as a suitable approach by the research community (Reddy et al., 2012; Sablok et al., 2013), and some of the identified isoforms have been experimentally validated using RT-PCR (Thatcher et al., 2014; Wang K. et al., 2016; Xu et al., 2016). As a meta-analysis in this work, we did not perform any validation on the identified isoforms generated by AS genes. In considering the dynamic nature of AS in responding to the changing environmental conditions and developmental regulations, however, experimental validations are needed in each specific experiment. Pacific BioSciences (PacBio) single-molecule real-time (SMRT) long-read isoform sequencing (Iso-Seq) and Nanopore sequencing from Oxford Nanopore Technologies (ONT) were two tools revolutionizing the way AS are identified (Zhao et al., 2019). The long reads sequencing technologies could avoid the error-prone step during transcripts assembly for RNA-seq reads from Illumina sequencing platform. In future, it will be interesting to validate these AS events based on RNA-seq short reads by using long reads from PacBio or ONT technologies. The list of potential isoforms identified in the work provides a foundation for designing experiments for exploring the biological significances of these AS events.

We also identified a large number of ncRNAs including miRNA and long ncRNAs. The ncRNAs play important regulatory roles in plant biology (Borges and Martienssen, 2015; Chekanova, 2015). A number of studies have reported the biological significances of ncRNAs in tomato plants (Wang et al., 2015, 2018b; Zhu et al., 2015). The list of ncRNAs compiled in the

work will aid in further elucidating the roles of ncRNA playing in tomato biology.

DATA AVAILABILITY

Publicly available datasets were analyzed in this study. This data can be found here: <http://proteomics.yzu.edu/publication/data/Tomato/>.

AUTHOR CONTRIBUTIONS

XM designed the experiments and performed the functional annotation. SC collected and processed the data for genome mapping. XM, FY, and LG analyzed the data. XM and LG wrote the manuscript.

REFERENCES

- Abdel-Ghany, S. E., Hamilton, M., Jacobi, J. L., Ngam, P., Devitt, N., Schilkey, F., et al. (2016). A survey of the sorghum transcriptome using single-molecule long reads. *Nat. Commun.* 7:11706. doi: 10.1038/ncomms11706
- Alkan, N., Friedlander, G., Ment, D., Prusky, D., and Fluhr, R. (2015). Simultaneous transcriptome analysis of *Colletotrichum gloeosporioides* and tomato fruit pathosystem reveals novel fungal pathogenicity and fruit defense strategies. *New Phytol.* 205, 801–815. doi: 10.1111/nph.13087
- Bao, W., Kojima, K. K., and Kohany, O. (2015). Repbase Update, a database of repetitive elements in eukaryotic genomes. *Mob. DNA* 6:11. doi: 10.1186/s13100-015-0041-9
- Borges, F., and Martienssen, R. A. (2015). The expanding world of small RNAs in plants. *Nat. Rev. Mol. Cell Biol.* 16:727. doi: 10.1038/nrm4085
- Bovy, A., Schijlen, E., and Hall, R. D. (2007). Metabolic engineering of flavonoids in tomato (*Solanum lycopersicum*): the potential for metabolomics. *Metabolomics* 3:399.
- Chae, S., Kim, J. S., Jun, K. M., Lee, S. B., Kim, M. S., Nahm, B. H., et al. (2017). Analysis of genes with alternatively spliced transcripts in the leaf, root, panicle and seed of rice using a long oligomer microarray and RNA-seq. *Mol. Cells* 40, 714–730. doi: 10.14348/molcells.2017.2297
- Chamala, S., Feng, G., Chavarro, C., and Barbazuk, W. B. (2015). Genome-wide identification of evolutionarily conserved alternative splicing events in flowering plants. *Front. Bioeng. Biotechnol.* 3:33. doi: 10.3389/fbioe.2015.00033
- Chekanova, J. A. (2015). Long non-coding RNAs and their functions in plants. *Curr. Opin. Plant Biol.* 27, 207–216. doi: 10.1016/j.pbi.2015.08.003
- Chen, T., Lv, Y., Zhao, T., Li, N., Yang, Y., Yu, W., et al. (2013). Comparative transcriptome profiling of a resistant vs. susceptible tomato (*Solanum lycopersicum*) cultivar in response to infection by tomato yellow leaf curl virus. *PLoS One* 8:e80816. doi: 10.1371/journal.pone.0080816
- Crooks, G. E., Hon, G., Chandonia, J. M., and Brenner, S. E. (2004). WebLogo: a sequence logo generator. *Genome Res.* 14, 1188–1190. doi: 10.1101/gr.849004
- Cruz-Mendivil, A., López-Valenzuela, J. A., Calderón-Vázquez, C. L., Vega-García, M. O., Reyes-Moreno, C., and Valdez-Ortiz, A. (2015). Transcriptional changes associated with chilling tolerance and susceptibility in 'Micro-Tom' tomato fruit using RNA-Seq. *Postharvest Biol. Tec.* 99, 141–151. doi: 10.1016/j.postharvbio.2014.08.009
- Dai, Q., Geng, L., Lu, M., Jin, W., Nan, X., He, P. A., et al. (2017). Comparative transcriptome analysis of the different tissues between the cultivated and wild tomato. *PLoS One* 12:e0172411. doi: 10.1371/journal.pone.0172411
- Du, H., Wang, Y., Yang, J., and Yang, W. (2015). Comparative transcriptome analysis of resistant and susceptible tomato lines in response to infection by *Xanthomonas perforans* race T3. *Front. Plant Sci.* 6:1173. doi: 10.3389/fpls.2015.01173

FUNDING

The work was funded by Youngstown State University (YSU) YSU Research Council grant (Project #13-18) to XM. The College of Graduate Studies of YSU supported the article processing charges. The work was also supported by a Research Professorship award and reassigned time for scholarship from College of Science, Technology, Engineering, and Mathematics and Department of Biological Sciences at YSU to XM.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fpls.2019.00689/full#supplementary-material>

- Encode Project Consortium (2007). Identification and analysis of functional elements in 1% of the human genome by the ENCODE pilot project. *Nature* 447, 799–816. doi: 10.1038/nature05874
- Ezura, K., Ji-Seong, K., Mori, K., Suzuki, Y., Kuhara, S., Ariizumi, T., et al. (2017). Genome-wide identification of pistil-specific genes expressed during fruit set initiation in tomato (*Solanum lycopersicum*). *PLoS One* 12:e0180003. doi: 10.1371/journal.pone.0180003
- Filichkin, S. A., Priest, H. D., Givan, S. A., Shen, R., Bryant, D. W., Fox, S. E., et al. (2010). Genome-wide mapping of alternative splicing in *Arabidopsis thaliana*. *Genome Res.* 20, 45–58. doi: 10.1101/gr.093302.109
- Foissac, S., and Sammeth, M. (2007). ASTALAVISTA: dynamic and flexible analysis of alternative splicing events in custom gene datasets. *Nucleic Acids Res.* 2007, W297–W299. doi: 10.1093/nar/gkm311
- Gene Ontology Consortium (2014). Gene ontology consortium: going forward. *Nucleic Acids Res.* 43, D1049–D1056. doi: 10.1093/nar/gku1179
- Gupta, S., Shi, X., Lindquist, I. E., Devitt, N., Mudge, J., and Rashotte, A. M. (2013). Transcriptome profiling of cytokinin and auxin regulation in tomato root. *J. Exp. Bot.* 64, 695–704. doi: 10.1093/jxb/ers365
- Higashi, T., Tanigaki, Y., Takayama, K., Nagano, A. J., Honjo, M. N., and Fukuda, H. (2016). Detection of diurnal variation of tomato transcriptome through the molecular timetable method in a sunlight-type plant factory. *Front. Plant Sci.* 7:87. doi: 10.3389/fpls.2016.00087
- Huang, X., and Madan, A. (1999). CAP3: A DNA sequence assembly program. *Genome Res.* 9, 868–877.
- Kim, D., Pertea, G., Trapnell, C., Pimentel, H., Kelley, R., and Salzberg, S. L. (2013). TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. *Genome Biol.* 14:R36. doi: 10.1186/gb-2013-14-4-r36
- Koenig, D., Jiménez-Gómez, J. M., Kimura, S., Fulop, D., Chitwood, D. H., Headland, L. R., et al. (2013). Comparative transcriptomics reveals patterns of selection in domesticated and wild tomato. *Proc. Natl. Acad. Sci. U.S.A.* 110, E2655–E2662. doi: 10.1073/pnas.1309606110
- Kumar, R., and Khurana, A. (2014). Functional genomics of tomato: opportunities and challenges in post-genome NGS era. *J. Biosci.* 39, 917–929.
- Kumar, S., Razzaq, S. K., Vo, A. D., Gautam, M., and Li, H. (2016). Identifying fusion transcripts using next generation sequencing. *Wiley Interdiscip. Rev. RNA* 7, 811–823. doi: 10.1002/wrna.1382
- Lewis, B. P., Green, R. E., and Brenner, S. E. (2003). Evidence for the widespread coupling of alternative splicing and nonsense-mediated mRNA decay in humans. *Proc. Natl. Acad. Sci. U.S.A.* 100, 189–192.
- Li, Q., Xiao, G., and Zhu, Y. X. (2014). Single-nucleotide resolution mapping of the *Gossypium raimondii* transcriptome reveals a new mechanism for alternative splicing of introns. *Mol. Plant* 7, 829–840. doi: 10.1093/mp/sst175
- Liu, L., Shao, Z., Zhang, M., and Wang, Q. (2015). Regulation of carotenoid metabolism in tomato. *Mol. Plant* 8, 28–39.

- Lopez-Casado, G., Covey, P. A., Bedinger, P. A., Mueller, L. A., Thannhauser, T. W., Zhang, S., et al. (2012). Enabling proteomic studies with RNA-Seq: the proteome of tomato pollen as a test case. *Proteomics* 12, 761–774. doi: 10.1002/pmic.201100164
- Marquez, Y., Brown, J. W., Simpson, C. G., Barta, A., and Kalyna, M. (2012). Transcriptome survey reveals increased complexity of the alternative splicing landscape in *Arabidopsis*. *Genome Res.* 22, 1184–1195. doi: 10.1101/gr.134106.111
- McCarthy, F. M., Wang, N., Magee, G. B., Nanduri, B., Lawrence, M. L., Camon, E. B., et al. (2006). AgBase: a functional genomics resource for agriculture. *BMC Genomics* 7:229. doi: 10.1186/1471-2164-7-229
- McGuire, A. M., Pearson, M. D., Neafsey, D. E., and Galagan, J. E. (2008). Cross-kingdom patterns of alternative splicing and splice recognition. *Genome Biol.* 9:R50. doi: 10.1186/gb-2008-9-3-r50
- Mei, W., Boatwright, L., Feng, G., Schnable, J. C., and Barbazuk, W. B. (2017). Evolutionarily conserved alternative splicing across monocots. *Genetics* 207, 465–480. doi: 10.1534/genetics.117.300189
- Mertens, F., Johansson, B., Fioretos, T., and Mitelman, F. (2015). The emerging complexity of gene fusions in cancer. *Nat. Rev. Cancer* 15, 371–381. doi: 10.1038/nrc3947
- Min, X. J. (2013). ASFinder: a tool for genome-wide identification of alternatively spliced transcripts from EST-derived sequences. *Int. J. Bioinform. Res. Appl.* 9, 221–226. doi: 10.1504/IJBRA.2013.053603
- Min, X. J. (2017). Comprehensive cataloging and analysis of alternative splicing in maize. *Comp. Mol. Biol.* 7, 1–11. doi: 10.5376/cmb.2017.07.0001
- Min, X. J. (2018). A survey of alternative splicing in allotetraploid cotton (*Gossypium hirsutum* L.). *Comp. Mol. Biol.* 8, 1–13. doi: 10.5376/cmb.2018.08.0001
- Min, X. J., Powell, B., Braessler, J., Meinken, J., Yu, F., and Sablok, G. (2015). Genome-wide cataloging and analysis of alternatively spliced genes in cereal crops. *BMC Genomics* 16:721. doi: 10.1186/s12864-015-1914-5
- Panahi, B., Abbaszadeh, B., Taghizadehgan, M., and Ebrahimie, E. (2014). Genome-wide survey of alternative splicing in *Sorghum bicolor*. *Physiol. Mol. Biol. Plants* 20, 323–329. doi: 10.1007/s12298-014-0245-3
- Pennisi, E. (2012). ENCODE project writes eulogy for junk DNA. *Science* 337, 1159–1161. doi: 10.1126/science.337.6099.1159
- Qiao, D., Yang, C., Chen, J., Guo, Y., Li, Y., Niu, S., et al. (2019). Comprehensive identification of the full-length transcripts and alternative splicing related to the secondary metabolism pathways in the tea plant (*Camellia sinensis*). *Sci. Rep.* 9:2709. doi: 10.1038/s41598-019-39286-z
- Reddy, A. S., Marquez, Y., Kalyna, M., and Barta, A. (2013). Complexity of the alternative splicing landscape in plants. *Plant Cell* 25, 3657–3683. doi: 10.1105/tpc.113.117523
- Reddy, A. S., Rogers, M. F., Richardson, D. N., Hamilton, M., and Ben-Hur, A. (2012). Deciphering the plant splicing code: experimental and computational approaches for predicting alternative splicing and splicing regulatory elements. *Front. Plant Sci.* 7:18. doi: 10.3389/fpls.2012.00018
- Rice, P., Longden, I., and Bleasby, A. (2000). EMBOSS: the european molecular biology open software suite. *Trends Genet.* 16, 276–277.
- Sablok, G., Gupta, P. K., Baek, J. M., Vazquez, F., and Min, X. J. (2011). Genome-wide survey of alternative splicing in the grass *Brachypodium distachyon*: an emerging model biosystem for plant functional genomics. *Biotechnol. Lett.* 33, 629–636. doi: 10.1007/s10529-010-0475-6
- Sablok, G., Harikrishnam, J. A., and Min, X. J. (2013). Next generation sequencing for better understanding alternative splicing: way ahead for model and non-model plants. *Transcriptomics* 1:e103.
- Sablok, G., Powell, B., Braessler, J., Yu, F., and Min, X. J. (2017). Comparative landscape of alternative splicing in fruit plants. *Curr. Plant Biol.* 9, 29–36. doi: 10.1016/j.cpb.2017.06.001
- Scarano, D., Rao, R., and Corrado, G. (2017). In Silico identification and annotation of non-coding RNAs by RNA-seq and De Novo assembly of the transcriptome of tomato fruits. *PLoS One* 12:e0171504. doi: 10.1371/journal.pone.0171504
- Shi, X., Gupta, S., Lindquist, I. E., Cameron, C. T., Mudge, J., and Rashotte, A. M. (2013). Transcriptome analysis of cytokinin response in tomato leaves. *PLoS One* 8:e55090. doi: 10.1371/journal.pone.0055090
- Shukla, N., Yadav, R., Kaur, P., Rasmussen, S., Goel, S., Agarwal, M., et al. (2018). Transcriptome analysis of root-knot nematode (*Meloidogyne incognita*)-infected tomato (*Solanum lycopersicum*) roots reveals complex gene expression profiles and metabolic networks of both host and nematode during susceptible and resistance responses. *Mol. Plant Pathol.* 19, 615–633. doi: 10.1111/mpp.12547
- Staiger, D., and Brown, J. W. (2013). Alternative splicing at the intersection of biological timing, development, and stress responses. *Plant Cell* 25, 3640–3656. doi: 10.1105/tpc.113.113803
- Sun, Y., and Xiao, H. (2015). Identification of alternative splicing events by RNA sequencing in early growth tomato fruits. *BMC Genomics* 16:948. doi: 10.1186/s12864-015-2128-6
- Sundaresan, S., Philosoph-Hadas, S., Riov, J., Mugasimangalam, R., Kuravadi, N. A., Kochanek, B., et al. (2016). De novo transcriptome sequencing and development of abscission zone-specific microarray as a new molecular tool for analysis of tomato organ abscission. *Front. Plant Sci.* 6:1258. doi: 10.3389/fpls.2015.01258
- Tan, G., Liu, K., Kang, J., Xu, K., Zhang, Y., Hu, L., et al. (2015). Transcriptome analysis of the compatible interaction of tomato with *Verticillium dahliae* using RNA-sequencing. *Front. Plant Sci.* 6:428. doi: 10.3389/fpls.2015.00428
- Tang, X., Tang, Z., Huang, S., Liu, J., Liu, J., Shi, W., et al. (2013). Whole transcriptome sequencing reveals genes involved in plastid/chloroplast division and development are regulated by the HP1/DBP1 at an early stage of tomato fruit development. *Planta* 238, 923–936. doi: 10.1007/s00425-013-1942-9
- Thatcher, S. R., Danilevskaia, O. N., Meng, X., Beatty, M., Zastrow-Hayes, G., Harris, C., et al. (2016). Genome-wide analysis of alternative splicing during development and drought stress in maize. *Plant Physiol.* 170, 586–599. doi: 10.1104/pp.15.01267
- Thatcher, S. R., Zhou, W., Leonard, A., Wang, B. B., Beatty, M., Zastrow-Hayes, G., et al. (2014). Genome-wide analysis of alternative splicing in *Zea mays*: landscape and genetic regulation. *Plant Cell* 26, 3472–3487. doi: 10.1105/tpc.114.130773
- Tomato Genome Consortium (2012). The tomato genome sequence provides insights into fleshy fruit evolution. *Nature* 485, 635–641. doi: 10.1038/nature11119
- Trapnell, C., Williams, B. A., Pertea, G., Mortazavi, A., Kwan, G., Van Baren, M. J., et al. (2010). Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. *Nat. Biotechnol.* 28, 511–515. doi: 10.1038/nbt.1621
- VanBuren, R., Walters, B., Ming, R., and Min, X. J. (2013). Analysis of expressed sequence tags and alternative splicing genes in sacred lotus (*Nelumbo nucifera* Gaertn.). *Plant Omics* 6, 311–317.
- Vitulo, N., Forcato, C., Carpinelli, E. C., Telatin, A., Campagna, D., D'Angelo, M., et al. (2014). A deep survey of alternative splicing in grape reveals changes in the splicing machinery related to tissue, stress condition and genotype. *BMC Plant Biol.* 14:99. doi: 10.1186/1471-2229-14-99
- Wai, C. M., Powell, B., Ming, R., and Min, X. J. (2016). Analysis of alternative splicing landscape in pineapple (*Ananas comosus*). *Trop. Plant Biol.* 9, 150–160. doi: 10.1007/s12042-016-9168-1
- Wai, C. M., Powell, B., Ming, R., and Min, X. J. (2018). “Analysis of transcriptome and alternative splicing landscape in pineapple,” in *Genetics and Genomics of Pineapple. Plant Genetics and Genomics: Crops and Models*, Vol. 22, ed. R. Ming (Cham: Springer).
- Walters, B. R., Lum, G. E., Sablok, G. A., and Min, X. J. (2013). Genome-wide landscape of alternative splicing events in *Brachypodium distachyon*. *DNA Res.* 20, 163–171. doi: 10.1093/dnares/dss041
- Wang, B., and Brendel, V. (2006). Genome wide comparative analysis of alternative splicing in plants. *Proc. Natl. Acad. Sci. U.S.A.* 103, 7175–7180. doi: 10.1073/pnas.0602039103
- Wang, J., Yu, W., Yang, Y., Li, X., Chen, T., Liu, T., et al. (2015). Genome-wide analysis of tomato long non-coding RNAs and identification as endogenous target mimic for microRNA in response to TYLCV infection. *Sci. Rep.* 5:16946. doi: 10.1038/srep16946
- Wang, K., Jiao, Z., Xu, M., Wang, Y., Li, R., Cui, X., et al. (2016). Landscape and fruit developmental regulation of alternative splicing in tomato by genome-wide analysis. *Hortic. Plant J.* 2, 338–350. doi: 10.1016/j.hpj.2017.01.007
- Wang, R. H., Yuan, X. Y., Meng, L. H., Zhu, B. Z., Zhu, H. L., Luo, Y. B., et al. (2016). Transcriptome analysis provides a preliminary regulation route of the ethylene signal transduction component, SIEIN2, during tomato ripening. *PLoS One* 11:e0168287. doi: 10.1371/journal.pone.0168287

- Wang, M., Wang, P., Liang, F., Ye, Z., Li, J., Shen, C., et al. (2018a). A global survey of alternative splicing in allopolyploid cotton: landscape, complexity and regulation. *New Phytol.* 217, 163–178. doi: 10.1111/nph.14762
- Wang, M., Zhao, W., Gao, L., and Zhao, L. (2018b). Genome-wide profiling of long non-coding RNAs from tomato and a comparison with mRNAs associated with the regulation of fruit ripening. *BMC Plant Biol.* 18:75. doi: 10.1186/s12870-018-1300-y
- Wang, Y., Tao, X., Tang, X. M., Xiao, L., Sun, J. L., Yan, X. F., et al. (2013). Comparative transcriptome analysis of tomato (*Solanum lycopersicum*) in response to exogenous abscisic acid. *BMC Genomics* 14:841. doi: 10.1186/1471-2164-14-841
- Wei, H., Lou, Q., Xu, K., Yan, M., Xia, H., Ma, X., et al. (2017). Alternative splicing complexity contributes to genetic improvement of drought resistance in the rice maintainer HuHan2B. *Sci. Rep.* 7:11686. doi: 10.1038/s41598-017-12020-3
- Worley, J. N., Pombo, M. A., Zheng, Y., Dunham, D. M., Myers, C. R., Fei, Z., et al. (2016). A novel method of transcriptome interpretation reveals a quantitative suppressive effect on tomato immune signaling by two domains in a single pathogen effector protein. *BMC Genomics* 17:229. doi: 10.1186/s12864-016-2534-4
- Xu, P., Kong, Y., Song, D., Huang, C., Li, X., and Li, L. (2014). Conservation and functional influence of alternative splicing in wood formation of *Populus* and *Eucalyptus*. *BMC Genomics* 15:780. doi: 10.1186/1471-2164-15-780
- Xu, Z., Luo, H., Ji, A., Zhang, X., Song, J., and Chen, S. (2016). Global identification of the full-length transcripts and alternative splicing related to phenolic acid biosynthetic genes in *Salvia miltiorrhiza*. *Front. Plant Sci.* 7:100. doi: 10.3389/fpls.2016.00100
- Xue, D. Q., Chen, X. L., Zhang, H., Chai, X. F., Jiang, J. B., Xu, X. Y., et al. (2016). Transcriptome analysis of the Cf-12-mediated resistance response to *Cladosporium fulvum* in tomato. *Front. Plant Sci.* 7:2012. doi: 10.3389/fpls.2016.02012
- Yang, Y. X., Wang, M. M., Yin, Y. L., Onac, E., Zhou, G. F., Peng, S., et al. (2015). RNA-seq analysis reveals the role of red light in resistance against *Pseudomonas syringae* pv. *tomato* DC3000 in tomato plants. *BMC Genomics* 16:120. doi: 10.1186/s12864-015-1228-7
- Ye, J., Hu, T., Yang, C., Li, H., Yang, M., Ijaz, R., et al. (2015). Transcriptome profiling of tomato fruit development reveals transcription factors associated with ascorbic acid, carotenoid and flavonoid biosynthesis. *PLoS One* 10:e0130885. doi: 10.1371/journal.pone.0130885
- Yu, H., Tian, C., Yu, Y., and Jiao, Y. (2016). Transcriptome survey of the contribution of alternative splicing to proteome diversity in *Arabidopsis thaliana*. *Mol. Plant* 9, 749–752. doi: 10.1016/j.molp.2015.12.018
- Zhang, C., Yang, H., and Yang, H. (2016). Evolutionary character of alternative splicing in plants. *Bioinform. Biol. Insights* 9(Suppl. 1), 47–52. doi: 10.4137/BBI.S33716
- Zhang, S., Xu, M., Qiu, Z., Wang, K., Du, Y., Gu, L., et al. (2016). Spatiotemporal transcriptome provides insights into early fruit development of tomato (*Solanum lycopersicum*). *Sci. Rep.* 6:23173. doi: 10.1038/srep23173
- Zhang, G., Sun, M., Wang, J., Lei, M., Li, C., Zhao, D., et al. (2018). PacBio full-length cDNA sequencing integrated with RNA-seq reads drastically improves the discovery of splicing transcripts in rice. *Plant J.* 97, 296–305. doi: 10.1111/tj.14120
- Zhang, R., Calixto, C. P., Marquez, Y., Venhuizen, P., Tzioutziou, N. A., Guo, W., et al. (2017). A high quality Arabidopsis transcriptome for accurate transcript-level analysis of alternative splicing. *Nucleic Acids Res.* 45, 5061–5073. doi: 10.1093/nar/gkx267
- Zhang, Y., Zhao, G., Li, Y., Zhang, J., Shi, M., Muhammad, T., et al. (2017). Transcriptome profiling of tomato uncovers an involvement of cytochrome P450s and peroxidases in stigma color formation. *Front. Plant Sci.* 8:897. doi: 10.3389/fpls.2017.00897
- Zhao, L., Zhang, H., Kohnen, M., Prasad, K., Gu, L., and Reddy, A. S. (2019). Analysis of transcriptome and epitranscriptome in plants using PacBio Iso-Seq and Nanopore-based direct RNA sequencing. *Front. Genet.* 10:253. doi: 10.3389/fgene.2019.00253
- Zheng, Y., Ding, B., Fei, Z., and Wang, Y. (2017). Comprehensive transcriptome analyses reveal tomato plant responses to tobacco rattle virus-based gene silencing vectors. *Sci. Rep.* 7:9771. doi: 10.1038/s41598-017-10143-1
- Zhu, B., Yang, Y., Li, R., Fu, D., Wen, L., Luo, Y., et al. (2015). RNA sequencing and functional analysis implicate the regulatory role of long non-coding RNAs in tomato fruit ripening. *J. Exp. Bot.* 66, 4483–4495. doi: 10.1093/jxb/erv203
- Zhu, G., Li, W., Zhang, F., and Guo, W. (2018). RNA-seq analysis reveals alternative splicing under salt stress in cotton, *Gossypium davidsonii*. *BMC Genomics* 19:73. doi: 10.1186/s12864-018-4449-8
- Zhu, J., Wang, X., Xu, Q., Zhao, S., Tai, Y., and Wei, C. (2018). Global dissection of alternative splicing uncovers transcriptional diversity in tissues and associates with the flavonoid pathway in tea plant (*Camellia sinensis*). *BMC Plant Biol.* 18:266. doi: 10.1186/s12870-018-1497-9
- Zouari, I., Salvioli, A., Chialva, M., Novero, M., Miozzi, L., Tenore, G. C., et al. (2014). From root to fruit: RNA-Seq analysis shows that arbuscular mycorrhizal symbiosis may affect tomato fruit metabolism. *BMC Genomics* 15:221. doi: 10.1186/1471-2164-15-221
- Zouine, M., Fu, Y., Chateigner-Boutin, A. L., Mila, I., Frasse, P., Wang, H., et al. (2014). Characterization of the tomato ARF gene family uncovers a multi-levels post-transcriptional regulation including alternative splicing. *PLoS One* 9:e84203. doi: 10.1371/journal.pone.0084203

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2019 Clark, Yu, Gu and Min. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



New Era in Plant Alternative Splicing Analysis Enabled by Advances in High-Throughput Sequencing (HTS) Technologies

Renesh Bedre¹, Sonia Irigoyen¹, Ezequiel Petrillo^{2,3} and Kranthi K. Mandadi^{1,4*}

¹ Texas A&M AgriLife Research and Extension Center, Texas A&M University, Weslaco, TX, United States, ² Departamento de Fisiología, Biología Molecular y Celular, Facultad de Ciencias Exactas y Naturales, Universidad de Buenos Aires, Buenos Aires, Argentina, ³ CONICET-Universidad de Buenos Aires, Instituto de Fisiología, Biología Molecular y Neurociencias (IFIBYNE), Buenos Aires, Argentina, ⁴ Department of Plant Pathology and Microbiology, Texas A&M University, College Station, TX, United States

Keywords: alternative splicing, high-throughput sequencing, bioinformatics, RNA-seq, PCR, non-sense-mediated decay

OPEN ACCESS

Edited by:

Laigeng Li,
Institute of Plant Physiology and
Ecology, Shanghai Institutes for
Biological Sciences (CAS), China

Reviewed by:

Peng Xu,
University of Alabama at Birmingham,
United States
Xiangjia Min,
Youngstown State University,
United States
Dominika Lewandowska,
James Hutton Institute,
United Kingdom

*Correspondence:

Kranthi K. Mandadi
kmandadi@tamu.edu

Specialty section:

This article was submitted to
Plant Physiology,
a section of the journal
Frontiers in Plant Science

Received: 21 March 2019

Accepted: 17 May 2019

Published: 04 June 2019

Citation:

Bedre R, Irigoyen S, Petrillo E and
Mandadi KK (2019) New Era in Plant
Alternative Splicing Analysis Enabled
by Advances in High-Throughput
Sequencing (HTS) Technologies.
Front. Plant Sci. 10:740.
doi: 10.3389/fpls.2019.00740

Alternative splicing (AS) is a crucial posttranscriptional mechanism of gene expression which promotes transcriptome and proteome diversity. At the molecular level, splicing and AS involves recognition and elimination of intronic regions of a precursor messenger RNA (pre-mRNA) and joining of exonic regions to generate the mature mRNA. AS generates more than one mRNA transcript (transcripts) differing in coding and/or untranslated regions (UTRs). AS can be classified into four major types including the exon skipping (ES), intron retention (IR), alternative donor (AD), and alternative acceptor (AA), of which IR is the most prevalent event in plants (Mandadi and Scholthof, 2015). In addition to these AS types, a subfamily of IR called exitrons, which has dual features of introns and protein-coding exons were first reported in *Arabidopsis thaliana* (*Arabidopsis*) and later also found in humans (Marquez et al., 2015). These spliced transcripts influence multiple biological processes such as growth, development and response to biotic and abiotic stresses in plants (Filichkin et al., 2015; Mandadi and Scholthof, 2015; Wang et al., 2018a).

FUNCTIONAL RELEVANCE OF AS

AS can produce aberrant or unstable transcripts with premature termination codons (PTCs). The PTC-containing transcripts are often targeted to degradation by a conserved cytoplasmic RNA degradation mechanism called NMD (non-sense-mediated mRNA decay). NMD mechanisms ensure that there is a balance or homeostasis in the functional vs. non-functional transcripts (Kalyna et al., 2012). In contrast to mammals, where NMD targets are degraded by a suppressor with morphogenetic effect on genitalia (SMG7) endonucleolytic pathway, plants NMD primarily occurs via. SMG7 exonucleolytic pathway (Shaul, 2015). AS can also produce stable transcripts, which encode proteins with altered functional domains, subcellular localization, and/or biological functions (Reddy et al., 2013; Shang et al., 2017). In humans, ~15% of genetic diseases are a result of aberrant splicing (Staiger and Brown, 2013). In plants, several studies have showed that AS has biologically significant implications in growth, development, stress-responses, and/or adaptation. For instance, AS in a MADS-box transcription factor gene, *SHORT VEGETATIVE PHASE* (SVP), results in multiple transcripts (SVP1 and SVP3), which encode proteins with altered interaction domains (Severing et al., 2012). Overexpression of SVP1, but not SVP3, resulted in repression of flowering (Severing et al., 2012). In rice, AS occurs in the *DEHYDRATION-RESPONSIVE ELEMENT BINDING PROTEIN 2* (*DREB2B*) gene, but only when subjected to drought and heat stress, and results in the production of an alternative transcript which encodes the full-length

functional protein that confers tolerance to the stresses (Matsukura et al., 2010). Similarly, in tobacco, the classical resistance gene (*N*) against *Tobacco mosaic virus* (TMV) is alternatively spliced, resulting in production of two forms—a short and a long transcript (Dinesh-Kumar and Baker, 2000). Functional analysis revealed that both transcripts are required in certain ratio to confer full resistance to TMV (Dinesh-Kumar and Baker, 2000). Recently, by employing high-throughput sequencing (HTS), Mandadi and Scholthof (2015), identified ~670 intron-containing genes in *Brachypodium* that were aberrantly spliced in response to viral infection. Several of these genes encoded resistance proteins, transcription factors, and splicing factors (Mandadi and Scholthof, 2015). Together, these studies suggest that many AS events, if not all, have biologically-significant implications in plant growth, development and response to stresses.

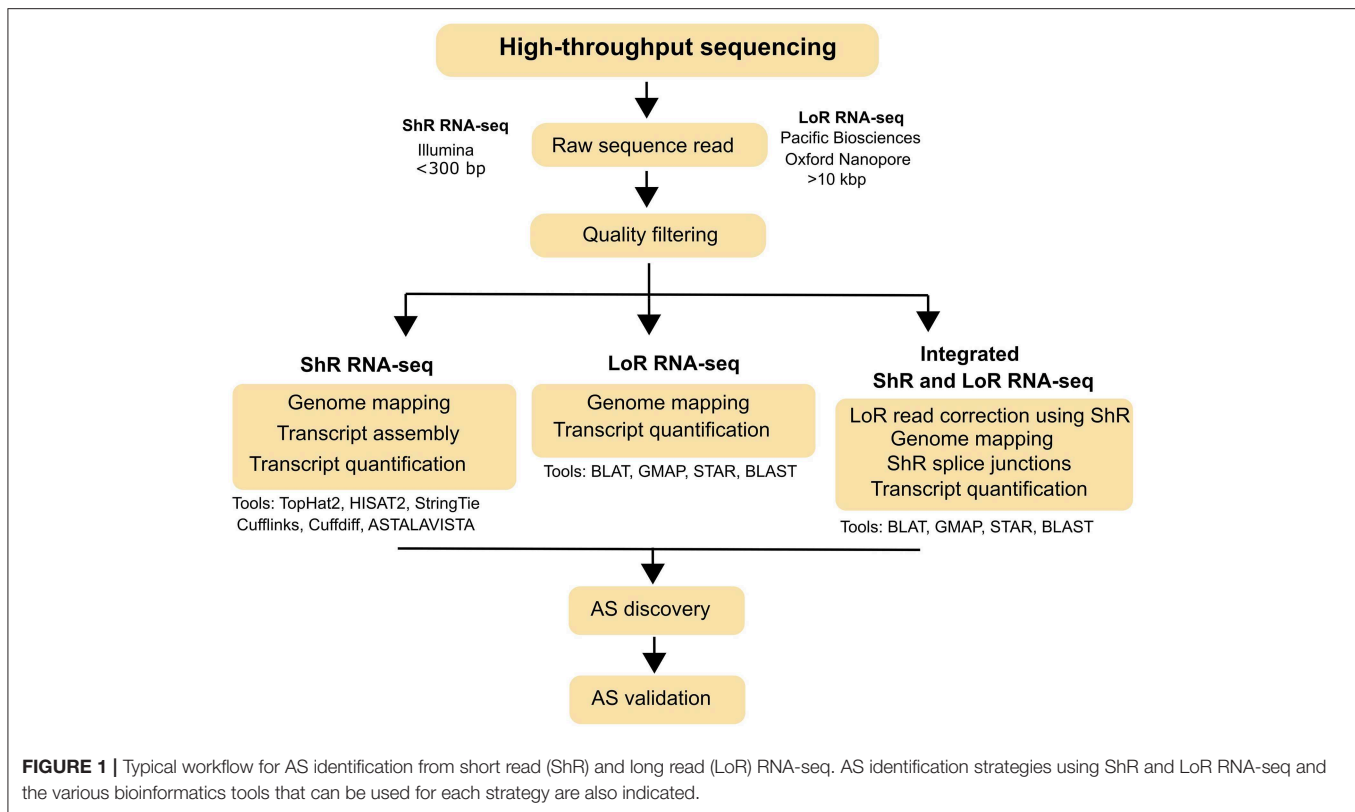
HIGH-THROUGHPUT SEQUENCING (HTS) FOR AS ANALYSIS

Historically, our knowledge of plant alternative splicing and how it affects biological processes was primarily gleaned from studies of few plant species (e.g., *Arabidopsis*, rice; Modrek and Lee, 2002). However, with the rapid developments in HTS (a.k.a. next- and third-generation sequencing) technologies, particularly long-read, single-molecule real-time sequencing (SMRT) and direct RNA-sequencing platforms, the field is rapidly changing. Several existing and emerging next-generation sequencing (NGS) platforms, and bioinformatics tools are useful for genome-wide queries of AS in diverse plant species (Filichkin et al., 2010; Mandadi and Scholthof, 2015; Thatcher et al., 2016). HTS-based genome analysis studies estimated that ~33–70% of plant genes undergo AS, suggesting a broader influence of AS in shaping the functional transcriptome and proteome landscapes of plants (Pan et al., 2008; Chamala et al., 2015; Filichkin et al., 2015; Mandadi and Scholthof, 2015; Wang et al., 2018a). The seemingly lower number of genes undergoing AS in plants when compared to humans (~95%) could be due to lack of enough studies or in-depth annotations of the plant genomes. In early reports dating back to 2004, the AS rates in the model plant *Arabidopsis* was reported at a meager ~11.6%, when the AS rates in humans was ~42% (Iida et al., 2004). Efforts by several groups over the years, and with the advent of HTS technologies, the AS rates in *Arabidopsis* and humans ascended comparably to ~60 and ~95%, respectively (Wang and Brendel, 2006; Zhang et al., 2015; Laloum et al., 2018). Hence, we presume that with recent advances in HTS technologies, the AS frequencies in plant species would likely increase further. Alternatively, it is quite possible that differential gene structure/number, spliceosome composition, as well as variations in the types of tissues sampled, and detection methods could contribute to the observed lower AS rate in plants when compared to humans.

HTS-based short read (Illumina) and long read (Pacific Biosciences and Oxford Nanopore) sequencing technologies

have revolutionized the field of DNA and RNA sequencing. Specifically, short read (<300 bp) RNA-sequencing (ShR RNA-seq), which integrates qualitative (gene discovery) and quantitative (gene quantification) assays, became a popular tool for genome-wide AS identification in plants as well as in other organisms. Because ShR RNA-seq provides high sequencing depth, a low error rate (<1%) and relatively-lower cost, it has been extensively used to characterize and quantify spliced transcripts in well-annotated plant genomes such as *Arabidopsis* (Calixto et al., 2018), *Oryza sativa* (Zhang and Xiao, 2018), *Brachypodium distachyon* (Mandadi and Scholthof, 2015), *Zea mays* (Thatcher et al., 2016), and *Glycine max* (Shen et al., 2014). Further, discovery of AS in plants was improved by the continuous development of open-source bioinformatics tools and pipelines. Identifying spliced transcripts from ShR-RNA-seq involves, mapping of high quality reads to reference genomes (HISAT2, TopHat2), transcript assembly (StringTie, Cufflinks, Trinity), AS events analysis (ASTALAVISTA), and transcript quantification (Cuffdiff, DESeq2) (Figure 1; Haas et al., 2013; Trapnell et al., 2013; Love et al., 2014; Foissac and Sammeth, 2015; Perte et al., 2016; Irigoyen et al., 2018). Among these tools, HISAT2 and StringTie analysis pipeline (new Tuxedo package; Perte et al., 2016) perform much faster, requires less memory and generates more accurate results over the TopHat2 and Cufflinks analysis pipeline (original Tuxedo package; Trapnell et al., 2012).

Despite the advances in ShR RNA-seq for AS discovery, this technology has limited scope in polyploids plant species (e.g., sugarcane, cotton) without reference genomes or those lacking comprehensive transcript-level annotations. To overcome this limitation, long read RNA-sequencing (LoR RNA-seq) technologies such as Pacific Biosciences (SMRT sequencing) and Oxford Nanopore (MinION) which has the ability to sequence full-length transcripts and direct RNA sequencing, have been used to study complex AS landscapes in polyploids (Liu et al., 2017; Wang et al., 2018a). LoR RNA-seq using SMRT (Iso-Seq method) and MinION can generate exceptionally long reads (>10 Kbp), which could cover most of the full-length eukaryotic transcripts and thus eliminate the need for transcript assembly (Liu et al., 2017; Ardui et al., 2018). Bioinformatics steps to identify AS from LoR RNA-seq involves mapping of high quality reads to reference genome (GMAP, BLAT) and, identifications of alternative transcripts and AS events from the alignments (Figure 1; Kent, 2002; Wu and Watanabe, 2005). For species where a reference genome is not available, the self-BLAST based pipeline on LoR RNA-seq can be used to detect AS based on the INDELs (Figure 1; Liu et al., 2017). Even though LoR RNA-seq is exceptional in resolving transcript structures, it has some limitations when compared to ShR RNA-seq including lower sequencing depth, high error-rate (up to ~15%), and being poorly suited for transcript quantifications (Ardui et al., 2018; Clark et al., 2018). These limitations in LoR RNA-seq could be mitigated by employing an integrated strategy involving both LoR and ShR RNA-seq reads (Figure 1; Koren et al., 2012). However, error correction methods are highly dependent on the availability of reference genomes and transcript annotations (Liu et al., 2017).



VALIDATION OF AS

Before embarking on functional analyses, it is recommended to validate the AS events using conventional molecular biology techniques. Reverse transcription followed by polymerase chain reaction (RT-PCR), cloning and Sanger-based sequencing are widely used approaches to confirm the presence and the sequence of the alternatively spliced transcripts (Simpson et al., 2008, 2016; Mandadi and Scholthof, 2015). Semi-quantitative RT-PCR (SqRT-PCR) and quantitative RT-PCR (qRT-PCR) can be used to quantify the alternatively spliced transcripts to understand AS regulation in different conditions (Harvey and Cheng, 2016). Both methods require cDNA synthesis as a first step, using either oligo-dT, which amplifies several mRNAs from the same sample, or specific oligos to generate cDNAs from specific transcripts. SqRT-PCR estimates relative amounts of the different templates in a sample and can be used to compare changes across different conditions/treatments. Additionally, to quantify the relative expression of the transcripts, quantitative PCR (qPCR) could be performed. In qPCR-based assays, transcript-specific primers are designed to quantify the expression levels. Subsequent to a PCR-based analysis, the various transcripts can be resolved in an agarose gel and purified for further cloning and Sanger-based sequencing to validate the sequence. Although less frequently employed, other molecular-biology methods that do not require PCR such as the Northern blotting and RNase protection assays, can also be utilized to validate alternatively spliced transcripts, particularly when the size differences between

transcripts allow a clear distinction. Furthermore, large-scale proteomics experiments (e.g., mass spectrometry) can be used to identify and study the proteins resulting from the alternatively-spliced transcripts and to support the *in-silico* protein sequence predictions (Tress et al., 2017).

Lastly, the AS analysis combined with functional genetic experiments will ultimately allow understanding of the biological significance of the various alternatively spliced transcripts. These methods typically involve selective overexpression or knockdown of the transcripts (and the encoded proteins) using stable or transient plant transformations, followed by evaluation of the trait of interest. Biochemical experiments to decipher the encoded protein localization, protein-protein, and protein-RNA interactions can also provide mechanistic insights into the functions of the alternatively spliced transcripts (Dinesh-Kumar and Baker, 2000; Severing et al., 2012; Staiger and Brown, 2013; Szakonyi and Duque, 2018; Wang et al., 2018b).

UPCOMING RESEARCH AND CONCLUSIONS

In the past few years, HTS technologies has unraveled the breadth of AS that is occurring in plants (Shen et al., 2014; Mandadi and Scholthof, 2015; Thatcher et al., 2016; Calixto et al., 2018; Zhang and Xiao, 2018). The availability of vast amounts of omics data (currently >1 Petabases), largely based on ShR RNA-seq and gene-level analysis, within the publicly available repositories

such as the NCBI SRA database (Leinonen et al., 2011) offer new opportunities to data mine and uncover AS landscapes among diverse plants and conditions. Such studies will allow global determination of conserved AS landscapes, patterns, and phenomenon occurring among the diverse evolutionary lineages of plants. Furthermore, combining the ShR RNA-seq data with LoR RNA-seq will allow discovery of the low-abundant transcripts, and/or resolve inadequacies that exist in reconstructing transcripts and complex transcript structures. The HTS will also advance our knowledge of AS landscapes and processes in complex polyploid plant genomes, which has largely remained understudied when compared to diploid plant genomes. Lastly, despite being well-positioned to study AS in plants at an unprecedented scale using HTS technologies, the challenge ahead lies in deciphering the biological relevance and molecular function of the various alternatively spliced transcripts

and the encoded proteins. Thus, we suggest that the AS research community place equal emphasis in the AS validation using reverse-genetic approaches.

AUTHOR CONTRIBUTIONS

RB, SI, EP, and KM designed the study and prepared the manuscript for submission. All authors have read and approved the manuscript.

FUNDING

This study was supported in part by funds from USDA-NIFA-AFRI (2016-67013-24738), and Texas A&M AgriLife Research Insect Vectored Diseases Seed Grant (114190-96210) to KM.

REFERENCES

- Ardui, S., Ameer, A., Vermeesch, J. R., and Hestand, M. S. (2018). Single molecule real-time (SMRT) sequencing comes of age: applications and utilities for medical diagnostics. *Nucleic Acids Res.* 46, 2159–2168. doi: 10.1093/nar/gky066
- Calixto, C. P. G., Guo, W., James, A. B., Tzioutziou, N. A., Entizne, J. C., Panter, P. E., et al. (2018). Rapid and dynamic alternative splicing impacts the arabidopsis cold response transcriptome. *Plant Cell* 30, 1424–1444. doi: 10.1105/tpc.18.00177
- Chamala, S., Feng, G., Chavarro, C., and Barbazuk, W. B. (2015). Genome-wide identification of evolutionarily conserved alternative splicing events in flowering plants. *Front. Bioeng. Biotechnol.* 3:33. doi: 10.3389/fbioe.2015.00033
- Clark, M., Wrzesinski, T., Garcia-Bea, A., Kleinman, J., Hyde, T., Weinberger, D., et al. (2018). Long-read sequencing reveals the splicing profile of the calcium channel gene *CACNA1C* in human brain. *bioRxiv [preprint]*. doi: 10.1101/260562
- Dinesh-Kumar, S. P., and Baker, B. J. (2000). Alternatively spliced *N* resistance gene transcripts: Their possible role in *Tobacco mosaic virus* resistance. *Proc. Natl. Acad. Sci. U.S.A.* 97, 1908–1913. doi: 10.1073/pnas.020367497
- Filichkin, S., Priest, H. D., Megraw, M., and Mockler, T. C. (2015). Alternative splicing in plants: directing traffic at the crossroads of adaptation and environmental stress. *Curr. Opin. Plant Biol.* 24, 125–135. doi: 10.1016/j.pbi.2015.02.008
- Filichkin, S. A., Priest, H. D., Givan, S. A., Shen, R., Bryant, D. W., Fox, S. E., et al. (2010). Genome-wide mapping of alternative splicing in *Arabidopsis thaliana*. *Genome Res.* 20, 45–58. doi: 10.1101/gr.093302.109
- Foissac, S., and Sammeth, M. (2015). Analysis of alternative splicing events in custom gene datasets by AStalavista. *Methods Mol. Biol.* 1269, 379–392. doi: 10.1007/978-1-4939-2291-8_24
- Haas, B. J., Papanicolaou, A., Yassour, M., Grabherr, M., Blood, P. D., Bowden, J., et al. (2013). *De novo* transcript sequence reconstruction from RNA-seq using the Trinity platform for reference generation and analysis. *Nat. Protoc.* 8, 1494–1512. doi: 10.1038/nprot.2013.084
- Harvey, S. E., and Cheng, C. (2016). Methods for characterization of alternative RNA splicing. *Methods Mol. Biol.* 1402, 229–241. doi: 10.1007/978-1-4939-3378-5_18
- Iida, K., Seki, M., Sakurai, T., Satou, M., Akiyama, K., Toyoda, T., et al. (2004). Genome-wide analysis of alternative pre-mRNA splicing in *Arabidopsis thaliana* based on full-length cDNA sequences. *Nucleic Acids Res.* 32, 5096–5103. doi: 10.1093/nar/gkh845
- Irigoyen, S., Bedre, R. H., Scholthof, K. B., and Mandadi, K. K. (2018). Genomic approaches to analyze alternative splicing, a key regulator of transcriptome and proteome diversity in *Brachypodium distachyon*. *Methods Mol. Biol.* 1667, 73–85. doi: 10.1007/978-1-4939-7278-4_7
- Kalyana, M., Simpson, C. G., Syed, N. H., Lewandowska, D., Marquez, Y., Kusenda, B., et al. (2012). Alternative splicing and nonsense-mediated decay modulate expression of important regulatory genes in Arabidopsis. *Nucleic Acids Res.* 40, 2454–2469. doi: 10.1093/nar/gkr932
- Kent, W. J. (2002). BLAT—the BLAST-like alignment tool. *Genome Res.* 12, 656–664. doi: 10.1101/gr.229202
- Koren, S., Schatz, M. C., Walenz, B. P., Martin, J., Howard, J. T., Ganapathy, G., et al. (2012). Hybrid error correction and *de novo* assembly of single-molecule sequencing reads. *Nat. Biotechnol.* 30, 693–700. doi: 10.1038/nbt.2280
- Laloum, T., Martin, G., and Duque, P. (2018). Alternative splicing control of abiotic stress responses. *Trends Plant Sci.* 23, 140–150. doi: 10.1016/j.tplants.2017.09.019
- Leinonen, R., Sugawara, H., Shumway, M., and International Nucleotide Sequence Database, C. (2011). The sequence read archive. *Nucleic Acids Res.* 39, D19–21. doi: 10.1093/nar/gkq1019
- Liu, X., Mei, W., Soltis, P. S., Soltis, D. E., and Barbazuk, W. B. (2017). Detecting alternatively spliced transcript isoforms from single-molecule long-read sequences without a reference genome. *Mol. Ecol. Resour.* 17, 1243–1256. doi: 10.1111/1755-0998.12670
- Love, M. I., Huber, W., and Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* 15:550. doi: 10.1186/s13059-014-0550-8
- Mandadi, K. K., and Scholthof, K.-B. G. (2015). Genome-wide analysis of alternative splicing landscapes modulated during plant-virus interactions in *Brachypodium distachyon*. *Plant Cell* 27, 71–85. doi: 10.1105/tpc.114.133991
- Marquez, Y., Hopfler, M., Ayatollahi, Z., Barta, A., and Kalyana, M. (2015). Unmasking alternative splicing inside protein-coding exons defines exons and their role in proteome plasticity. *Genome Res.* 25, 995–1007. doi: 10.1101/gr.186585.114
- Matsukura, S., Mizoi, J., Yoshida, T., Todaka, D., Ito, Y., Maruyama, K., et al. (2010). Comprehensive analysis of rice DREB2-type genes that encode transcription factors involved in the expression of abiotic stress-responsive genes. *Mol. Genet. Genomics* 283, 185–196. doi: 10.1007/s00438-009-0506-y
- Modrek, B., and Lee, C. (2002). A genomic view of alternative splicing. *Nat. Genet.* 30, 13–19. doi: 10.1038/ng0102-13
- 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. doi: 10.1038/ng.259
- Perteau, M., Kim, D., Perteau, G. M., Leek, J. T., and Salzberg, S. L. (2016). Transcript-level expression analysis of RNA-seq experiments with HISAT, StringTie and Ballgown. *Nat. Protoc.* 11, 1650–1667. doi: 10.1038/nprot.2016.095
- Reddy, A. S. N., Marquez, Y., Kalyana, M., and Barta, A. (2013). Complexity of the alternative splicing landscape in plants. *Plant Cell* 25, 3657–3683. doi: 10.1105/tpc.113.117523
- Severing, E. I., Van Dijk, A. D., Morabito, G., Busscher-Lange, J., Immink, R. G., and Van Ham, R. C. (2012). Predicting the impact of alternative splicing on plant MADS domain protein function. *PLoS ONE* 7:e30524. doi: 10.1371/journal.pone.0030524

- Shang, X., Cao, Y., and Ma, L. (2017). Alternative splicing in plant genes: a means of regulating the environmental fitness of plants. *Int. J. Mol. Sci.* 18:18020432. doi: 10.3390/ijms18020432
- Shaul, O. (2015). Unique aspects of plant nonsense-mediated mRNA decay. *Trends Plant Sci.* 20, 767–779. doi: 10.1016/j.tplants.2015.08.011
- Shen, Y., Zhou, Z., Wang, Z., Li, W., Fang, C., Wu, M., et al. (2014). Global dissection of alternative splicing in paleopolyploid soybean. *Plant Cell* 26, 996–1008. doi: 10.1105/tpc.114.122739
- Simpson, C. G., Fuller, J., Calixto, C. P., Mcnicol, J., Booth, C., Brown, J. W., et al. (2016). Monitoring alternative splicing changes in *Arabidopsis Circadian* clock genes. *Methods Mol. Biol.* 1398, 119–132. doi: 10.1007/978-1-4939-3356-3_11
- Simpson, C. G., Fuller, J., Maronova, M., Kalyna, M., Davidson, D., Mcnicol, J., et al. (2008). Monitoring changes in alternative precursor messenger RNA splicing in multiple gene transcripts. *Plant J.* 53, 1035–1048. doi: 10.1111/j.1365-3113X.2007.03392.x
- Staiger, D., and Brown, J. W. S. (2013). Alternative splicing at the intersection of biological timing, development, and stress responses. *Plant Cell* 25, 3640–3656. doi: 10.1105/tpc.113.113803
- Szakonyi, D., and Duque, P. (2018). Alternative splicing as a regulator of early plant development. *Front. Plant Sci.* 9, 1174–1174. doi: 10.3389/fpls.2018.01174
- Thatcher, S. R., Danilevskaya, O. N., Meng, X., Beatty, M., Zastrow-Hayes, G., Harris, C., et al. (2016). Genome-wide analysis of alternative splicing during development and drought stress in Maize. *Plant Physiol.* 170, 586–599. doi: 10.1104/pp.15.01267
- Trapnell, C., Hendrickson, D. G., Sauvageau, M., Goff, L., Rinn, J. L., and Pachter, L. (2013). Differential analysis of gene regulation at transcript resolution with RNA-seq. *Nat. Biotechnol.* 31:46. doi: 10.1038/nbt.2450
- Trapnell, C., Roberts, A., Goff, L., Pertea, G., Kim, D., Kelley, D. R., et al. (2012). Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and Cufflinks. *Nat. Protocols* 7, 562–578. doi: 10.1038/nprot.2012.016
- Tress, M. L., Abascal, F., and Valencia, A. (2017). Alternative splicing may not be the key to proteome complexity. *Trends Biochem. Sci.* 42, 98–110. doi: 10.1016/j.tibs.2016.08.008
- 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. doi: 10.1073/pnas.0602039103
- Wang, M., Wang, P., Liang, F., Ye, Z., Li, J., Shen, C., et al. (2018a). A global survey of alternative splicing in allopolyploid cotton: landscape, complexity and regulation. *New Phytol.* 217, 163–178. doi: 10.1111/nph.14762
- Wang, Y., Zhang, T., Song, X., Zhang, J., Dang, Z., Pei, X., et al. (2018b). Identification and functional analysis of two alternatively spliced transcripts of *ABSCISIC ACID INSENSITIVE3 (ABI3)* in linseed flax (*Linum usitatissimum* L.). *PLoS ONE* 13:e0191910. doi: 10.1371/journal.pone.0191910
- Wu, T. D., and Watanabe, C. K. (2005). GMAP: a genomic mapping and alignment program for mRNA and EST sequences. *Bioinformatics* 21, 1859–1875. doi: 10.1093/bioinformatics/bti310
- Zhang, C., Yang, H., and Yang, H. (2015). Evolutionary character of alternative splicing in plants. *Bioinform Biol. Insights* 9, 47–52. doi: 10.4137/BBI.S33716
- Zhang, Z., and Xiao, B. (2018). Comparative alternative splicing analysis of two contrasting rice cultivars under drought stress and association of differential splicing genes with drought response QTLs. *Euphytica* 214:73. doi: 10.1007/s10681-018-2152-0

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2019 Bedre, Irigoyen, Petrillo and Mandadi. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



Evolution of Alternative Splicing in Eudicots

Zhihao Ling¹, Thomas Brockmüller¹, Ian T. Baldwin¹ and Shuqing Xu^{2*}

¹ Max Planck Institute for Chemical Ecology, Jena, Germany, ² Plant Adaptation-in-action Group, Institute for Evolution and Biodiversity, University of Münster, Münster, Germany

OPEN ACCESS

Edited by:

Maria Kalyna,
University of Natural Resources
and Life Sciences, Vienna, Austria

Reviewed by:

Yamile Marquez,
Centre for Genomic Regulation
(CRG), Spain
Craig G. Simpson,
The James Hutton Institute,
United Kingdom
Julie Thomas,
University of Arkansas, United States

*Correspondence:

Shuqing Xu
shuqing.xu@uni-muenster.de

Specialty section:

This article was submitted to
Plant Physiology,
a section of the journal
Frontiers in Plant Science

Received: 04 December 2018

Accepted: 13 May 2019

Published: 12 June 2019

Citation:

Ling Z, Brockmüller T, Baldwin IT
and Xu S (2019) Evolution
of Alternative Splicing in Eudicots.
Front. Plant Sci. 10:707.
doi: 10.3389/fpls.2019.00707

Alternative pre-mRNA splicing (AS) is prevalent in plants and is involved in many interactions between plants and environmental stresses. However, the patterns and underlying mechanisms of AS evolution in plants remain unclear. By analyzing the transcriptomes of four eudicot species, we revealed that the divergence of AS is largely due to the gains and losses of AS events among orthologous genes. Furthermore, based on a subset of AS, in which AS can be directly associated with specific transcripts, we found that AS that generates transcripts containing premature termination codons (PTC), are likely more conserved than those that generate non-PTC containing transcripts. This suggests that AS coupled with nonsense-mediated decay (NMD) might play an important role in affecting mRNA levels post-transcriptionally. To understand the mechanisms underlying the divergence of AS, we analyzed the key determinants of AS using a machine learning approach. We found that the presence/absence of alternative splice site (SS) within the junction, the distance between the authentic SS and the nearest alternative SS, the size of exon-exon junctions were the major determinants for both alternative 5' donor site and 3' acceptor site among the studied species, suggesting a relatively conserved AS mechanism. The comparative analysis further demonstrated that variations of the identified AS determinants significantly contributed to the AS divergence among closely related species in both Solanaceae and Brassicaceae taxa. Together, these results provide detailed insights into the evolution of AS in plants.

Keywords: alternative splicing, evolution, transcriptome, splicing code, deep learning, nonsense-mediated decay

INTRODUCTION

Due to their sessile lifestyle, plants have evolved various mechanisms to respond to environmental stresses. Alternative splicing (AS), a mechanism by which different mature RNAs are formed by removing different introns or using different splice sites (SS) from the same pre-mRNA, is known to be important for stress-induced responses in plants (Mastrangelo et al., 2012; Staiger and Brown, 2013). Both biotic and abiotic stresses such as herbivores (Ling et al., 2015), pathogens (Howard et al., 2013), and cold and salt (Ding et al., 2014) can all induce genome-wide changes in AS in plants. The environment-induced AS changes in turn can affect phenotypic traits of plants and may contribute to their adaptations to different stresses (Mastrangelo et al., 2012; Staiger and Brown, 2013). For example, low temperature-induced AS changes of flowering regulator genes affect flowering time and floral development in *Arabidopsis thaliana* (Severing et al., 2012;

Rosloski et al., 2013). The strong association between AS and environmental stimuli suggests that AS is involved in adaptation processes and thus may have evolved rapidly.

Two main functions of AS have been postulated: (i) to expand proteome diversity when different transcript isoforms are translated into different proteins (with different subcellular localization, stability, enzyme activity etc.) (Kazan, 2003; Reddy, 2007; Barbazuk et al., 2008); (ii) to influence gene expression (GE) by generating transcripts harboring premature termination codons (PTC) that are recognized by the nonsense-mediated decay (NMD) machinery and degraded (Chang et al., 2007; Hori and Watanabe, 2007; Kalyna et al., 2012; Kervestin and Jacobson, 2012). For example, different environmental stresses can induce AS events that generate PTC containing (+PTC) transcripts in key splicing regulators and circadian genes (Palusa et al., 2007; Filichkin et al., 2010, 2015). Although initially considered to be transcriptional noise, several AS events that introduce PTCs have been found to be highly conserved in animals (Ni et al., 2007; Lareau and Brenner, 2015) and plants (Iida and Go, 2006; Kalyna et al., 2006; Darracq and Adams, 2013), suggesting that the combination of AS with NMD might play an important role in affecting mRNA levels post-transcriptionally. However, it is unclear whether NMD-coupled AS is more conserved than the AS that generates transcripts without PTC at the genome-wide level.

The evolution of AS in plants, compared to that in vertebrates, remains largely unclear. Studies that compared organ-specific transcriptomes from different vertebrate species spanning ~350 million years of evolution showed that AS complexity differs dramatically among vertebrate lineages, and AS evolved much faster than GE (Barbosa-Morais et al., 2012; Merkin et al., 2012). For example, within 6 million years, the splicing profiles of an organ are more similar to other organs of the same species than the same organ in other species, while the expression profiles of the same organ are similar to the organ in other species (Barbosa-Morais et al., 2012; Merkin et al., 2012). In plants, largely due to the lack of comprehensive transcriptomic data, such comparative analysis remains unavailable. However, several indications suggest that AS in plants and vertebrates may share a similar evolution pattern. For example, only 16.4% of AS between maize and rice, and 5.4% between *Brassica* and *Arabidopsis* are conserved (Severing et al., 2009; Darracq and Adams, 2013). A more recent study further showed that only 2.8% of genes showed conserved AS between two species of mung beans, *Vigna radiata* and *V. angularis* (Satyawan et al., 2016). Furthermore, large changes in AS also exist between different ecotypes of the same species (Ner-Gaon et al., 2007). However, such low conservation of AS found among species could also be due to several other confounding effects. For example, it is also known that the levels of gene expression, which are highly associated with AS, also diverge rapidly in plants (Yang and Wang, 2013). As a consequence, it remains unclear whether the low observed levels of AS conservation are resulted from the rapid expression changes between species. Additionally, AS detection is highly dependent on sequencing depth and the tissue types used for generating transcriptomic data (Xu et al., 2002; Ellis et al., 2012; Ling et al., 2015). Therefore, it is necessary to systematically

control for different confounding effects in order to understand the evolutionary patterns of AS in plants.

From a mechanistic perspective, the divergence of AS among species is affected by factors that affect the exon-intron splicing process, which is mediated by the spliceosome. While the recognition processes of exonic and intronic regions are directed by sequence features of the pre-mRNA in animals, how the spliceosome removes introns and ligates exons is poorly understood in plants. In metazoans, it is known that four crucial signals are required for accurate splicing: (i) 5' SS, which contains a GU dinucleotide at the intron start surrounded by a piece of longer consensus sequences of lower conservation, (ii) 3' SS, which includes an AG dinucleotide at the 3' end surrounded by similar sequences of 5' SS, (iii) a polypyrimidine tract and (iv) a branch site sequence located ~17–40 nt upstream of the 3' SS (Black, 2003; Fu and Ares, 2014). In plants, similar sequence features with a small difference at specific positions were found, except for the requirement of a branch site (Reddy, 2007). In addition, a UA-rich tract in introns has also been found to be important for efficient splicing in plants (Lewandowska et al., 2004; Simpson et al., 2004; Baek et al., 2008). In animals, the regulation of splicing also depends largely on *cis* signals and *trans*-acting splicing factors that can recognize the signals (Barbosa-Morais et al., 2012; Merkin et al., 2012). Among different splicing factors, serine/arginine-rich (SR) proteins are from an important splicing factor family that has been shown to be involved in AS regulation (Lopato et al., 1999; Gao et al., 2004; Wang and Brendel, 2004; Reddy, 2007; Reddy and Shad Ali, 2011). In addition, many splicing regulatory elements (SREs) and RNA-binding proteins (RBPs) have been identified in animals, and the interactions among these SREs in the pre-mRNA and RBPs were found either to promote or suppress the use of particular SS (Licatalosi et al., 2008; Chen and Manley, 2009; Barash et al., 2010). The number of SR proteins genes in plants (on average > 20) is nearly twice of the number found in non-photosynthetic organisms, although the number varies among different species (Iida and Go, 2006; Isshiki et al., 2006; Richardson et al., 2011). To date, more than 1,000 RBPs and 80 SREs have been identified in plants using computational approaches (Lorkovic, 2009; Marondedze et al., 2016), however, only a few of these have been functionally validated (Yoshimura et al., 2002; Pertea et al., 2007; Schoning et al., 2008; Thomas et al., 2012).

In mammals, the emergence of AS originated from constitutive splicing with the fixation of SREs and the creation of alternative competing SS (Koren et al., 2007; Lev-Maor et al., 2007). Distinctive features that distinguish alternatively spliced exons/introns from constitutively spliced exons/introns can be used to accurately predict the specific AS type (Koren et al., 2007; Braunschweig et al., 2014). Furthermore, other factors including secondary and tertiary RNA structures, chromatin remodeling, insertion of transposable elements (TEs) and gene duplication may also be involved in regulating AS (Liu et al., 1995; Sorek et al., 2002; Donahue et al., 2006; Su et al., 2006; Kolasinska-Zwierz et al., 2009; Schwartz et al., 2009; Warf and Berglund, 2010; Lambert et al., 2015). However, the extent to which changes in these factors contributed to the evolutionary

history of AS in vertebrates remains largely unclear. Recently, a study using millions of synthetic mini-genes with degenerated subsequences demonstrated that the likelihood of AS decreases exponentially with increasing distance between constitutive and newly introduced alternative SS (Rosenberg et al., 2015), suggesting that sequence changes between constitutive and alternative SS might contribute to the changes of AS among species. In plants, however, the detailed mechanisms that affect AS remain largely unclear (Reddy et al., 2013). Although it has been proposed that changes in chromatin features such as DNA methylation, histone marks as well as RNA structural features, and SREs are important in regulating AS in plants, experimental evidence is largely lacking (Reddy et al., 2013). A recent study shows that DNA methylation could affect AS in rice (Wang X.T. et al., 2016), indicating changes in DNA methylation can contribute to the variations of AS among species, however, this hypothesis has not been thoroughly tested.

Because AS regulation is a complex process involving many factors, computational modeling is a useful tool to identify key factors and predict the outcome of splicing. While the Bayesian neural network (BNN) method was developed for decoding the splicing code in mammals (Barbosa-Morais et al., 2012), deep learning approaches, which refers to methods that map data through multiple levels of abstraction, have recently been shown to surpass BNN-based approaches (Leung et al., 2014; Mamoshina et al., 2016). Furthermore, deep learning methods are also able to cope with large, heterogeneous and high-dimensional datasets, an issue that is involved in predicting DNA and RBPs (Alipanahi et al., 2015) and AS (Leung et al., 2014; Mamoshina et al., 2016).

Here, we performed a comparative analysis of the transcriptomes of both closely and distantly related plant species to explore the evolutionary history of AS in plants. To further understand the mechanisms underlying the AS evolution in plants, we applied a deep learning approach to investigate the determinants of AS and their effects on AS evolution. Specifically, we aimed to address the following questions in plants: (1) What are the evolutionary patterns of AS? (2) Are the AS events that are coupled with NMD more conserved than regular AS events? (3) What are the important AS determinants? (4) Which AS determinants contributed to the AS divergence between closely related plant species?

MATERIALS AND METHODS

Read Mapping, Transcripts Assembly, and Abundance Estimation

All RNA-seq data of *Nicotiana attenuata* were generated in our lab, while the data of other species were downloaded from the short reads archive¹. The mapping information and SRA IDs of all datasets are listed in **Supplementary Tables S1, S2**. All of the RNA-seq reads were generated from polyA selected libraries. The raw sequence reads were trimmed using AdapterRemoval

(v1.1) (Lindgreen, 2012) with parameters “–collapse –trimns –trimqualities 2 –minlength 36.” The trimmed reads from each species were then aligned to the respective reference genome using Tophat2 (v2.0.6) (Trapnell et al., 2009), with maximum and minimum intron size set to 50,000 and 41 bp, respectively. After our analysis, we noticed that introns in plants are usually larger than 60 bp. However, in our dataset, only less than 0.2% of introns are less than 60 bp. Therefore, including these small introns (between 41 and 60 bp) that might be due to mapping errors should not affect the results. The numbers of uniquely mapped reads and splice junctions mapped reads were then counted using SAMtools (v0.1.19) (Li et al., 2009) by searching “50” in the MAPQ string and “*N*” flag in the CIGAR string of the resulting BAM files. The uniquely mapped reads from each sample were sub-sampled with the same sequencing depth (17 million) using SAMtools (v0.1.19) (Li et al., 2009).

The transcripts of each species were assembled using Cufflinks (v2.2.0) (Trapnell et al., 2012) with the genome annotation as the reference. The open reading frame (ORF) of each transcript was analyzed using TransDecoder from TRINITY (v2.1.0) (Grabherr et al., 2011). To estimate the expression level of genes/transcripts, all trimmed reads were re-mapped to the assembled transcripts using RSEM (v1.2.8) (Grabherr et al., 2011). Transcripts per million (TPM) was calculated for each gene/transcript (Wagner et al., 2012). Only genes with TPM greater than five in at least one sample were considered as an expressed gene.

AS Detection

All AS analysis were based on splicing junctions obtained from the BAM files produced by Tophat2. To remove the false positive junctions that were likely due to non-specific or erroneous alignments, all original junctions were removed if the overhang size was smaller than 13 bp, as suggested in Ling et al. (2015). All filtered junctions were then used for AS identification and annotation using JUNCBASE v0.6 (Graveley et al., 2011). Due to the relatively low sequencing depth of each individual sample of Brassicaceae RNA-seq data (**Supplementary Table S2**), we merged the BAM files of each three replicates together and randomly subsampled 17 million (the lowest depth among all merged samples) unique mapped reads from each merged file to avoid the heterogeneity of sequencing depth. The summary of all detected junctions is shown in **Supplementary Table S3**.

The percent spliced index (PSI) of each AS event, which represents the relative ratio of two different isoforms generated by the AS was calculated in each sample. $PSI = (\text{number of reads of inclusion isoform}) / (\text{number of reads of inclusion isoform} + \text{number of reads to exclusion isoform})$ as suggested in Graveley et al. (2011), Lareau and Brenner (2015). To avoid false-positives, only AS events that supported by at least 10 reads were considered. For alternative 5' donor site (AltD), alternative 5' acceptor site (AltA), and exon skipping (ES), the number of supporting reads was calculated as the sum of reads that support junctions, whereas for intron retention (IR), the total number of supporting reads was calculated as the sum of reads that mapped to both junctions and the intron region.

¹<http://www.ncbi.nlm.nih.gov/sra>

Identification of Conserved Exon–Exon Junctions (EEJs) and AS

We separately extracted the 100 bp sequence from the flanking upstream exon and downstream exon of each junction that has mapped read to support, and combine each side of exon sequence (in total 200 bp sequence) to represent the EEJ. The sequences of all EEJs were compared between species using TBLASTX (v.2.2.25) (Altschul et al., 1990) to find homologous relationships (**Supplementary Figure S1**). A python script was used to filter the TBLASTX results based on the following requirements: (1) The gene pair containing the EEJs must be the one-to-one orthologous gene pair between the two species; (2) the EEJ sequences between two species must be the best reciprocal blast hit based on the bit score; (3) at least 3 amino acid (aa) from both the flanking upstream exon and downstream exon sequence were aligned and (4) alignment coverage ≥ 60 bp, (5) E -value $< 1E-3$.

We only consider an AS event to be conserved if the same type of AS was found on the conserved EEJs between two plant species.

Identification of AS Events That Generate Premature Termination Codons (PTC)

The junctions related to each AS event were mapped back to assembled transcripts; only AS events which were related to junctions that mapped to two unique transcripts (had no structural difference except the AS region) were retained to avoid the situation where the sequence differences of the two transcripts resulted from multiple AS events. The transcript was considered to have a PTC if the stop codon of the longest ORF was at least 50 nucleotides upstream of an exon–exon boundary (the 50 nucleotides rule) (Nagy and Maquat, 1998; Schoenberg and Maquat, 2012; Weischenfeldt et al., 2012). To identify AS events that generate PTC-containing and non-PTC-containing transcripts, we used following criteria: (a) the AS events that can only be mapped to two unique transcripts; (b) the AS region is the only difference between the two transcripts; (c) at least one transcript does not contain PTC, as the AS events that generate two PTC-containing transcripts are likely due to assembly or annotation errors.

One-to-One Orthologous Gene Identifications and Gene Family Size Estimation

One-to-one orthologous gene pairs were predicted based on pairwise sequence similarities between species of the corresponding dataset. First, we calculated the sequence similarities between all protein-coding genes using BLASTP for the selected species and filtered the results based on E -value less than $1E-6$. Second, we selected the groups of genes that represent the best reciprocal hits that are shared among all species from the corresponding dataset.

For calculating the gene family size, we first defined gene families among different species by using a similarity-based approach. To do so, the homolog groups that were identified from our previous work were used, which were predicted from 11 plant species (Xu et al., 2017). In brief, all-vs.-all BLASTP was used to compare the sequence similarity of all protein

coding genes, and the results were filtered based on the following criteria: E -value less than $1E-20$; match length greater than 60 amino acids; sequence coverage greater than 60% and identity greater than 50%. All BLASTP results that remained after filtering were clustered into gene families using the Markov cluster algorithm (mcl). The gene family size for a species is represented by the number of genes of this species within the corresponding gene family.

Correlation and Clustering

For the pairwise comparison of AS, Spearman correlation and binary distance was applied to the PSI data ($0.05 < \text{PSI} < 0.95$ in at least one sample) and binary data (only one-to-one orthologous were used, and all genes that had no AS in all of the four species were excluded), respectively. A non-parametric correlation was selected for PSI level because of its bimodal nature distribution (0 and 100). For the pairwise comparison of gene expression, Pearson correlation was applied to \log_2 (TPM+1) of expressed genes to avoid infinite values.

The R package “pvcluster” was used for clustering of samples with 1,000 bootstrap replications. When we clustered and performed principal component analysis (PCA) of gene expression, the TPM values were normalized by GC% (EDASeq package in R) and TMM (the trimmed mean of M -values).

Identification of Possible Alternative Splice Sites (SS) and Regulatory Sequences

The 5' and 3' SS including 5 bp up and downstream sequences of all EEJs were used as the positive dataset, while the sequences extracted using the same method for all inter-GT (for 5' splice site) and inter-AG (for 3' splice site) within junction regions were used as background dataset. The putative SS motifs (12-mer) of both 5' and 3' SS were separately identified using Homer V3.12 (Heinz et al., 2010) and only motifs present in at least 5% of total positive sequences and P -value $< 1E-20$ were kept. The appearance of putative SS was identified using scanMotifGenomeWide, a Perl script included in the Homer toolkits and only sequence regions with match score > 2 were kept.

Homer was also used to identify the putative regulatory intronic and exonic sequence motifs (6-mer) of AltD, AltA and IR. The 50 bp up and downstream sequence of 5' SS was regarded as exonic and intronic sequence and vice versa for 3' SS. For AltD and AltA, the related sequences of EEJs with AS were used as the positive dataset, while 10,000 related sequences of EEJs without AS by random selection (due to a large number of sequences) were used as background dataset. The enriched motifs in the positive dataset were regarded as splicing enhancers, while the enriched motifs in the negative dataset were considered as splicing silencers. For IR, the related sequences from both SS of EEJs with IR were used as the positive dataset and the same sequences from EEJs without IR were used as background dataset. The conserved motifs between species were identified using compareMotifs, a Perl script included in the Homer toolkits and only one mismatch

was allowed. To identify polypyrimidine tracts, UA-rich tracts and branch site of each EEJ, we used the algorithm and scripts from Schwartz et al. (2008) and Szcześniak et al. (2013). In brief, polypyrimidine tracts and UA-rich tracts, intronic regions of up to 50 bases upstream of the 3' SS were searched using the algorithm that searches for the longest string with the C + U (in the case of polypyrimidine tracts) or A + U (for UA tracts) composition exceeding 85%. Polypyrimidine tracts that end within the last 10 bases of an intron were considered. Moreover, the tracts were required to be at least five bases long. The branch site identification consists three steps. First, the 100 nucleotides (nt) upstream of the 3' SS were used to identify the heptamers that were found in other systems (Bon et al., 2003): NNYTRAY, NNCTYAC, NNRTAAC, and NNCTAAA. Second, each heptamer was scored according to the number of mismatches from the optimal consensus of TACTAAC. Third, branch site containing introns were considered only if the introns in which the most downstream hit also has the best score. Although the last step discarded a relatively large fraction of introns, it reduces the false-positive rate significantly (Schwartz et al., 2008). To estimate the effect of each putative sequence motif, polypyrimidine and UA-tracts, we calculated the AS frequency of EEJs containing or not containing the motif/tract. Then for each motif/tract, the \log_2 odds ratio (effect size) with and without the motif/tract were calculated to quantify to what extent the presence of the motif/tract increases or decreases the AS frequency compare to its absence:

$$\text{Effect Size} = \log_2 \frac{p(\text{AS}|\text{motif}) / (1 - p(\text{AS}|\text{motif}))}{p(\text{AS}|\text{— motif}) / (1 - p(\text{AS}|\text{— motif}))}$$

Deciphering the Splicing Codes and AS Conservation Using a Deep Learning Algorithm

To investigate which sequence determinants contributed to the AS in plants, we constructed multi-layer feed-forward artificial neural networks using H2O's deep learning algorithm ("h2o" package) in R 3.0.2 (R Development Core Team 2013). For each AS type, a matrix was created based on the information of all EEJs that contain the AS (only that single event) and other EEJs within the same gene. The AS status (either AS or constitutive) was considered as output and the features that were known to be associated with splicing recognition and regulation in eukaryotes (Lewandowska et al., 2004; Kandul and Noor, 2009; Rosenberg et al., 2015) (listed in **Supplementary Data Sheet S1**) were used as input for training the model. On average, 25 features were used in the AltA model and 13 features were used in the AltD model. "TanhWithDropout" was used as the activation function and three hidden layers were used. Furthermore, "logloss" was used for model selection and the "Gedeon" method was used to compute the variable importances for input features. To reduce the background noise, we removed the EEJs which were supported by less than five reads on average. In addition, because the number of constitutively spliced EEJs in all cases is much larger than alternatively spliced EEJs, we randomly selected the same number of constitutive spliced EEJs as negative examples

and combined them together with all alternative spliced EEJs as the full dataset (50% precision by chance). To train and test the deep neural networks (DNN), the full dataset was randomly split, which 60% of data were used for training, 20% used for validation and the other 20% were reserved for testing. We trained for a fixed number (10,000) of epochs or stopped the training once the top 10 model were within 1% of improvement, and selected the hyper-parameters that gave the optimal area under the receiver operating characteristic curve (AUC) performance on the validation data. The model was then retrained using these selected hyper-parameters with the full dataset. The AUC value (range between 0 and 1) is an indicator showing the performance of a classification model, which is equivalent to the probability that a randomly chosen positive example is ranked higher than a randomly chosen negative example. A higher AUC value indicates a better performance of the model (high precision and high specificity).

Using a similar approach, we constructed the model for AS conservation. For each AS type, a matrix was created based on the information of all orthologous EEJ pairs between two species that contain the AS in at least one species. To reduce the background noise, any EEJ with multiple AS types, low number of support reads (less than five) or orthologous EEJ pair have different AS types were removed. The conservation levels (conserved, lost or gained in the other species) were used as the output of the model and the difference of features that were known to be important to AS and AS conservation (Su et al., 2006; Kelley et al., 2014; Sierro et al., 2014; Lambert et al., 2015; Rosenberg et al., 2015) (listed in **Supplementary Data Sheet S2**) between two species were used as input to train the model. Yass v1.15 was used to align the SS' flanking sequences (combined 50 bp upstream and downstream sequences of 5'/3' splice site, 100 bp in total) of each orthologous EEJ pair, the similarity was calculated as: (length of alignment – number of gaps – number of mismatches)/(total sequence length). To reduce the bias from different transition types in the dataset (much higher proportion of loss/gain than conserved AS), the data used to train the model was selected as the ratio of 1:1:1 for conserved, lost and gained situations (33.3% precision by chance). Due to small sample size of conserved AS, the model based on the same original data may differ as the randomly selected data of AS lost/gained were different each time. Therefore, the model construction process was repeated 10 times and the models that achieved the highest AUC for the complete dataset were considered. The one-to-one ortholog gene list, genes with/without AS from the four eudicot species are provided in **Supplementary Data Sheet S3**.

RESULTS

Genome-Wide AS Patterns Are Species-Specific in Plants

To provide an overview of AS evolution among different plant families, we studied the genome-wide AS in *A. thaliana*, soybean (*Glycine max*), tomato (*Solanum lycopersicum*), and wild tobacco (*N. attenuata*), from which comparable transcriptomic datasets are available from the same tissues (roots, leaves, and flowers) and

they represent a wide-range of eudicots. The overall distributions of different AS types within each species are consistent with previous studies. In all investigated species, intron retention (IR) and alternative 3' acceptor site (AltA) are the two major AS types (**Supplementary Figure S2**; Aoki et al., 2010; Marquez et al., 2012; Shen et al., 2014; Ling et al., 2015).

To investigate the evolutionary patterns of AS, we compared AS profiles across selected tissues and species. Because sequencing depth is known to strongly affect AS detection, we randomly subsampled 17 million (the lowest depth among all samples) uniquely mapped reads from each sample to standardize for the heterogeneity of sequencing depths. Overall, more than 75% of the splice junctions that were identified from the full dataset can be detected from these randomly selected 17 million uniquely mapped (later referred as 17M) reads (**Supplementary Figure S3A**), indicating the 17M reads is sufficient to reveal the AS evolution pattern among species. In addition, plotting the saturation curve of detected splice junctions with different sequencing depths showed that 17M reads have reached or at least are close to the saturation point (**Supplementary Figure S3B**). Thus, all downstream comparative analyses were based on this subsampled dataset. To investigate the conservation level of AS among different plant species, we focused on only one-to-one orthologous relationships, because relationships among complex one-to-many or many-to-many orthologous relationships are much more difficult to infer. Clustering analyses using PSI (Graveley et al., 2011; Lareau and Brenner, 2015) that measures the quantitative differences of AS among samples showed that different tissues of the same species are more similar to each other than the same tissue from different species (**Figure 1A**). Using the measures of AS that consider the presence or absence of AS (binary) from the genes that are one-to-one orthologous among all compared species, the same species-specific clustering pattern was found (**Figure 1B**). Consistent results were also obtained using all available reads (**Supplementary Figures S3C,D**) or when each type of AS was analyzed separately (**Supplementary Figure S4**).

To further investigate the evolutionary patterns of AS among closely related species, we analyzed a recently published transcriptome dataset from three Brassicaceae species (*A. thaliana*, *A. lyrata*, and *Capsella rubella*), each of which has comparable transcriptome data from two tissues (root and shoot) and two treatments (control and cold treated). Using both quantitative (PSI) and qualitative measures (binary) of AS, a similar species-specific clustering pattern was observed (**Figures 1C,D**). Interestingly, within same species and same tissue, samples exposed to cold stress clustered together regarding levels of PSI (**Figure 1C**), a result which is consistent with previous studies that demonstrate that stresses can induce genome-wide AS responses (Li et al., 2013; Ding et al., 2014; Ling et al., 2015).

Species-specific clustering patterns were also reported at the level of GE of one-to-one orthologs among *A. thaliana*, rice and maize (Yang and Wang, 2013). To examine whether species-specific AS clustering results from GE divergences, we compared the divergence patterns of AS and GE among transcriptomes of different species. Comparisons among species from different plant families showed that both GE and

AS cluster in species-specific patterns (**Figures 1A,B** and **Supplementary Figures S5A,B**). However, when species from the same plant family are compared, such as tomato and *N. attenuata* (Solanaceae), the species-specific AS pattern remained (**Figures 1A,B**), but the GE data clustered in a tissue-specific pattern (**Supplementary Figures S5A,B**). This shows that the expression profiles of the same tissues from different species are more similar to each other than the expression patterns from different tissues of the same species, indicating that the observed species-specific AS clustering is not due to GE divergence. A similar pattern was also found in the expression profiles of tissue samples from the three Brassicaceae species, among which the expression profiles of shoots and roots from different species were clearly separated (**Supplementary Figures S5C,D**). The observed difference in species-specific clustering patterns between GE and AS is consistent with the pattern found in animals (Barbosa-Morais et al., 2012; Merkin et al., 2012).

Massive Gains and Losses of AS Among Different Species

Species-specific clustering of AS pattern suggests a low level of AS conservation among species. Overall, among 3,857 one-to-one orthologous genes among the four eudicot species that have AS in at least one species, only ~7% of them have AS in all four species, while ~41% of them have species-specific AS. A similar pattern was also found when using the full dataset (not subsampled). We further investigated the pattern by looking at each exon-exon junction (EEJ) among orthologous groups, and found that more than 87.7% of AS events were species-specific (**Supplementary Figure S6**). Because the rapid change of AS could result from the rapid loss or gain of EEJ between species, we further compared the conservation of EEJs and AS among orthologous genes. Among the four eudicot species, 60% of EEJs are conserved in at least two species, which is much higher than the conservation of AS (~12%). Additional analysis showed that 92% of AS events identified from the conserved EEJs (shared among all four species) are species-specific. A similar analysis using the data from the three Brassicaceae species revealed the similar pattern (**Supplementary Figures S7A,B**). Together, the results from the comparison between divergent species and closely related species consistently suggest that AS are highly variable among plants.

To investigate the transition spectrum of AS at the conserved EEJs between species pairs, we calculated the AS changes among different types of AS. Among the four eudicot species, while the transitions among different AS types are rare, the gain/loss of AS is the most abundant transition type among all three pairwise comparisons (**Figures 2A–F**). For example, while an AltA event was found in *XCT* in *N. attenuata*, which was also confirmed by RT-PCR in our previous work (Ling et al., 2015), no AS was found at its orthologous junction in tomato (**Supplementary Figure S8D**). Among different AS types, AltA and exon skipping (ES) are the most and least conserved AS, respectively. Similar patterns were observed among three closely related species in Brassicaceae (**Supplementary Figures S8A–C**). These results suggest that the species-specific AS pattern is largely not due to the changes of EEJs among species, but rather the species-specific gains and losses of AS.

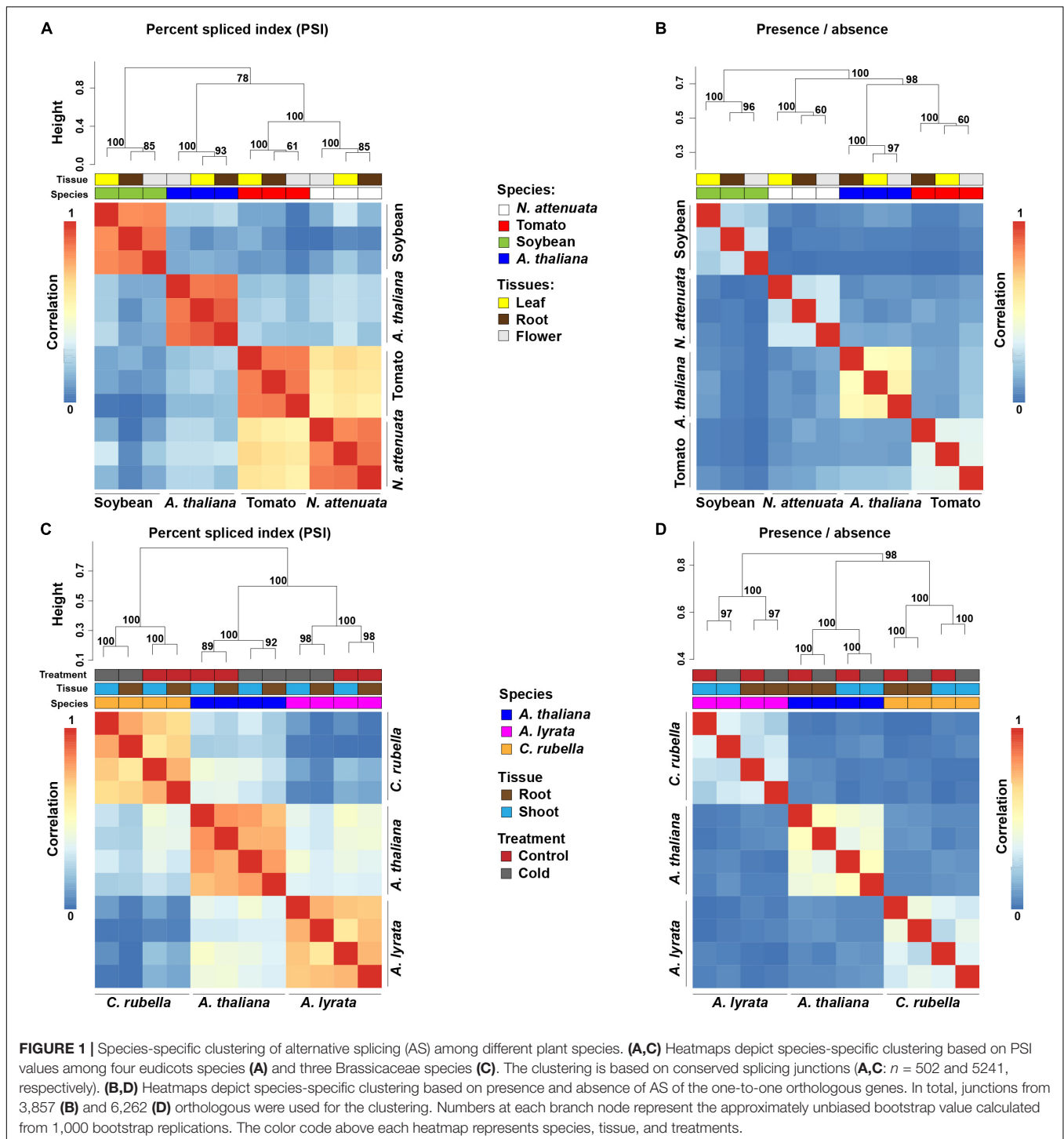


FIGURE 1 | Species-specific clustering of alternative splicing (AS) among different plant species. **(A,C)** Heatmaps depict species-specific clustering based on PSI values among four eudicots species **(A)** and three Brassicaceae species **(C)**. The clustering is based on conserved splicing junctions **(A,C)**: $n = 502$ and 5241 , respectively). **(B,D)** Heatmaps depict species-specific clustering based on presence and absence of AS of the one-to-one orthologous genes. In total, junctions from $3,857$ **(B)** and $6,262$ **(D)** orthologous were used for the clustering. Numbers at each branch node represent the approximately unbiased bootstrap value calculated from 1,000 bootstrap replications. The color code above each heatmap represents species, tissue, and treatments.

The AS Events That Result in PTC-Containing Transcripts Are Likely More Conserved Than Others

Previous studies suggest that many pre-mRNAs underwent unproductive AS, which generates transcripts with in-frame PTCs that are coupled with NMD in plants (Schwartz et al., 2006; Hori and Watanabe, 2007; Kerenyi et al., 2008;

Kalyna et al., 2012; Drechsel et al., 2013). To investigate whether unproductive AS can affect the AS conservation and contribute to the loss/gain of AS among different plant species, we separated the AS into two groups: (1) AS+ PTC and (2) AS- PTC (details see section “Materials and Methods”). Overall, the portion of AS+ PTC ranges from 9 to 15% among the four dicots (**Supplementary Figure S9**), suggesting that only a small

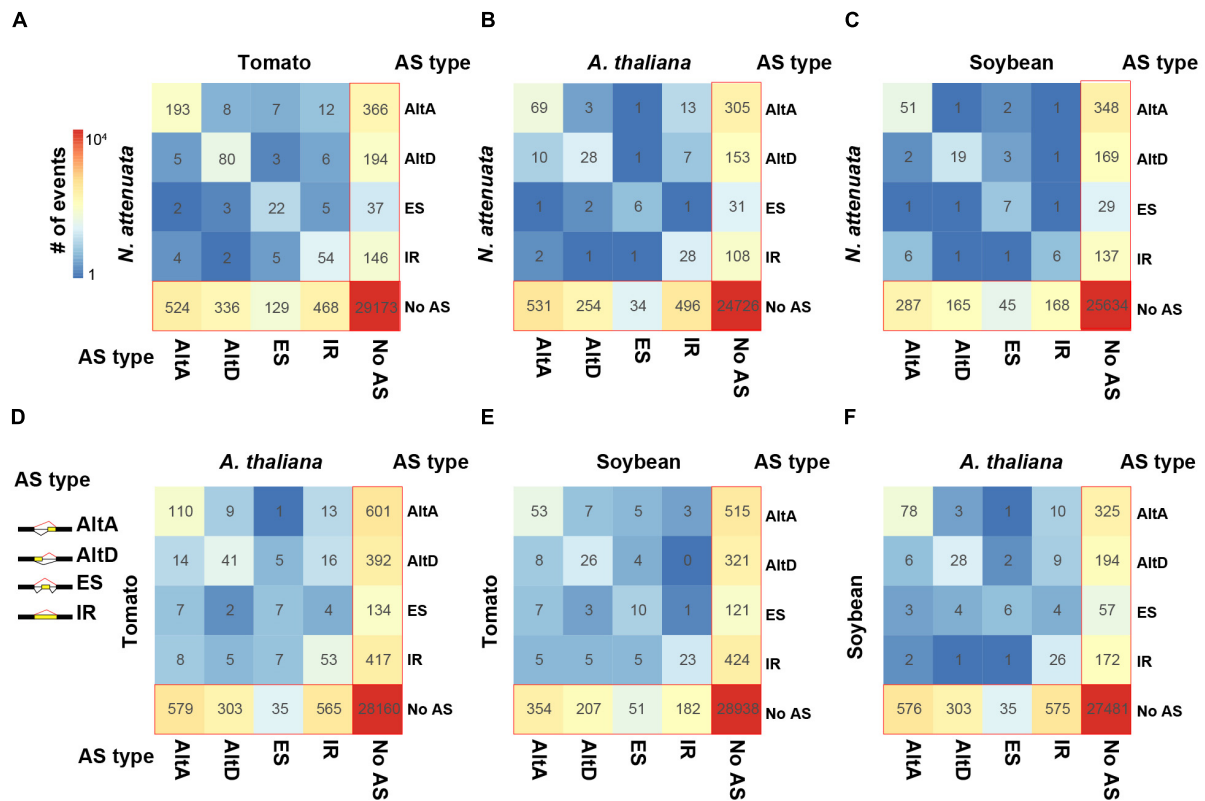


FIGURE 2 | Transition spectrum of AS between each species pairs. (A) *N. attenuata* vs. tomato, (B) *N. attenuata* vs. *A. thaliana*, (C) *N. attenuata* vs. soybean, (D) tomato vs. *A. thaliana*, (E) tomato vs. soybean, and (F) soybean vs. *A. thaliana*. The color of each grid indicates the abundance of the transitions between two AS types. The values are also shown. Only exon-exon junctions (EEJs) with at most one AS event were considered. AltA: alternative 3' acceptor site; AltD: alternative 5' donor site; ES, exon skipping; IR, intron retention.

portion of AS generated PTC-containing transcripts. Comparing the levels of conservation between tomato and *N. attenuata*, we found the AS+PTC is significantly more conserved than AS-PTC ($P < 0.02$, **Figure 3A**). For example, among nine AS+PTC of *N. attenuata* which are both conserved and have PTC information in tomato, eight of them (89%) also generated +PTC transcripts in tomato. Similar patterns were also observed in the three Brassica species (**Supplementary Figures S10A,B**).

To further investigate the level of conservation of AS+PTC, we extended our analysis by adding the transcriptome data of a very ancient plant species, the spreading earthmoss (*Physcomitrella patens*). Our rationale is that if AS+PTC events are more conserved than AS-PTC events, we would expect to see many AS+PTC events from the ultra-conserved AS events. Here, we focused on the 10 most highly conserved AS events found in all four eudicot plants (**Supplementary Figure S6B**) and checked for their presence in moss. In total, we found six AS events that were also present in moss, indicating these AS events might have evolved since land plants and played essential functions in plants. Interestingly, two of these ultra-conserved AS events were from serine/arginine-rich (SR) genes (*RS2Z33*-like and *RS40*-like), which are part of the RNA splicing machinery. The *RS2Z33*-like gene also has AS in rice and *Pinus taeda* (Iida and Go, 2006; Kalyna et al., 2006). Analyzing the protein coding potential of

the transcripts generated by these six ultra-conserved AS events showed that five resulted in +PTC transcripts. For example, the AS events of *RS2Z33*-like and *RS40*-like genes result in +PTC alternative transcripts in all five species and are likely the targets of NMD (**Figure 3B**). To further investigate whether these +PTC transcripts are affected by NMD, we analyzed the available transcriptome data from *A. thaliana* wild-type (WT) and NMD-deficient (*lba1* and *upf3-1* double mutant) plants (Drechsel et al., 2013). Among all five +PTC transcripts in *A. thaliana*, three showed significantly higher expression in NMD-deficient plants ($P < 7e-06$), including *RS2Z33*-like and *RS40*-like genes (**Figure 3B**). Together, these results suggest that AS coupled with PTC is likely more conserved than regular AS and some of these AS+PTC pairs may play essential roles in plants.

Mechanisms Involved in Determining AS Are Overall Conserved Among Different Plant Species

To further understand the mechanisms that contribute to the divergence of AS among species, it is necessary to identify the key features of AS in plants, which are largely unknown (Reddy et al., 2013; Staiger and Brown, 2013). Because splicing is often mediated by SS, we first investigated

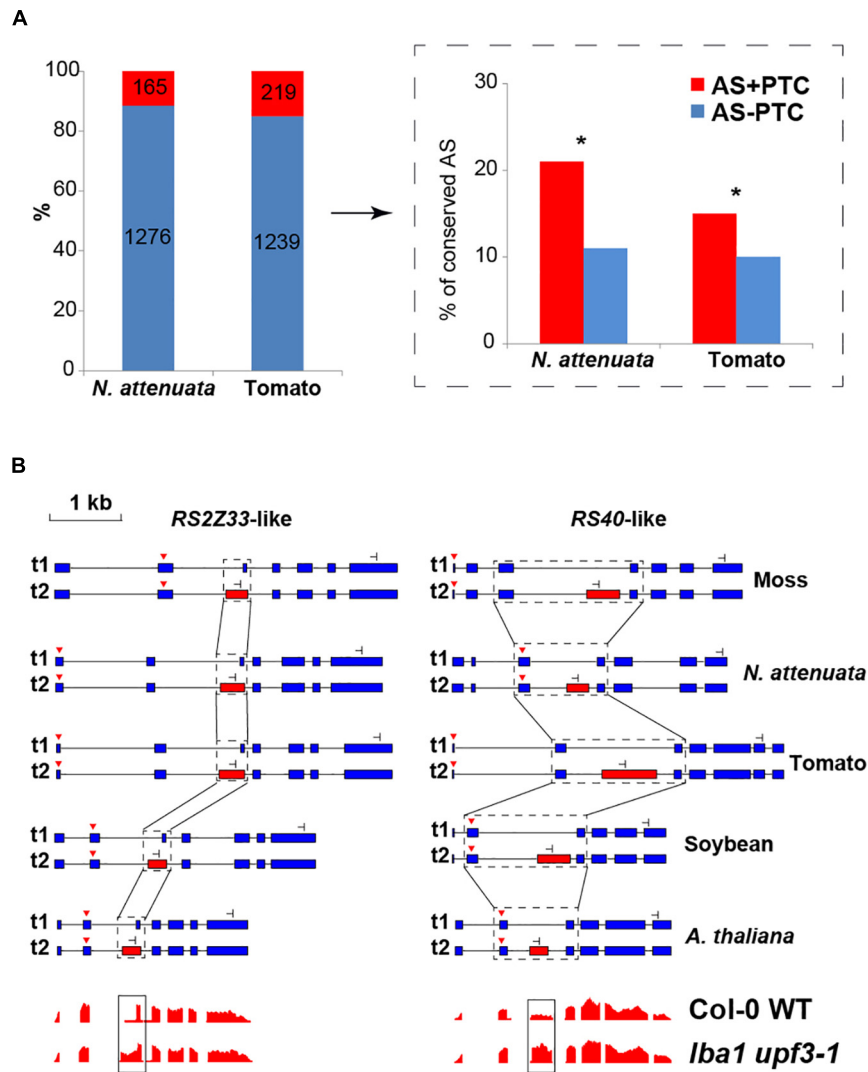


FIGURE 3 | The conservation of AS between AS+PTC and AS-PTC. **(A)** The number and relative portions of AS±PTC in *N. attenuata* and tomato. The insert indicated by the black arrow depicts the percentage of AS+PTC and AS-PTC that are conserved between *N. attenuata* and tomato. Asterisks indicate the significant difference between two AS groups as determined by Fisher's exact test ($P < 0.05$). **(B)** Conserved AS between moss and eudicots in serine/arginine-rich splicing factor *RS2Z33*-like and *RS40*-like gene. The diagrams of the structure of transcripts generated by the AS in all five species, the dominant and minor transcripts are represented by t1 and t2, respectively. Constitutive exons are represented by blue boxes, alternatively spliced exons are represented by red boxes and introns are represented by black solid lines. The black dotted boxes highlight the conserved AS and the start and stop codons are shown as red triangles and stop signs, respectively. The diagrams in the bottom panel showed the relative reads coverage of *AtRS2Z33* and *AtRS40* exons in wild-type plant and *lba1 upf3-1* double mutants. The black box highlights the coverage of the spliced region which is significantly increased in *lba1 upf3-1* double mutants (The diagrams are modified based on the data shown in <http://gbrowse.cbio.mskcc.org/gb/gbrowse/NMD201>).

whether the SS were different between constitutively and alternatively spliced junctions. Comparisons of the SS and their surrounding 12 bp sequences between constitutively and alternatively spliced junctions revealed that their SS are overall very similar (Supplementary Figure S11). Furthermore, we separately identified sequence motifs (12-mer) that are enriched in 5' and 3' SS compared to random sequences and found that these identified motifs are also highly conserved among the studied species (Supplementary Figure S12).

From the mechanistic point of view, the junction size (distance between the 3' SS and 5' SS of the two exons), the presence and positions of alternative SS, which are the additional SS motifs that compete with the authentic splice donors or acceptors can be important for the regulations of different types of AS (Gopal et al., 2005; Kandul and Noor, 2009; Braunschweig et al., 2014; Rosenberg et al., 2015). For the different AS types, we compared these features from both constitutively and alternatively spliced junctions. Because ES events are rare in all species, we only studied

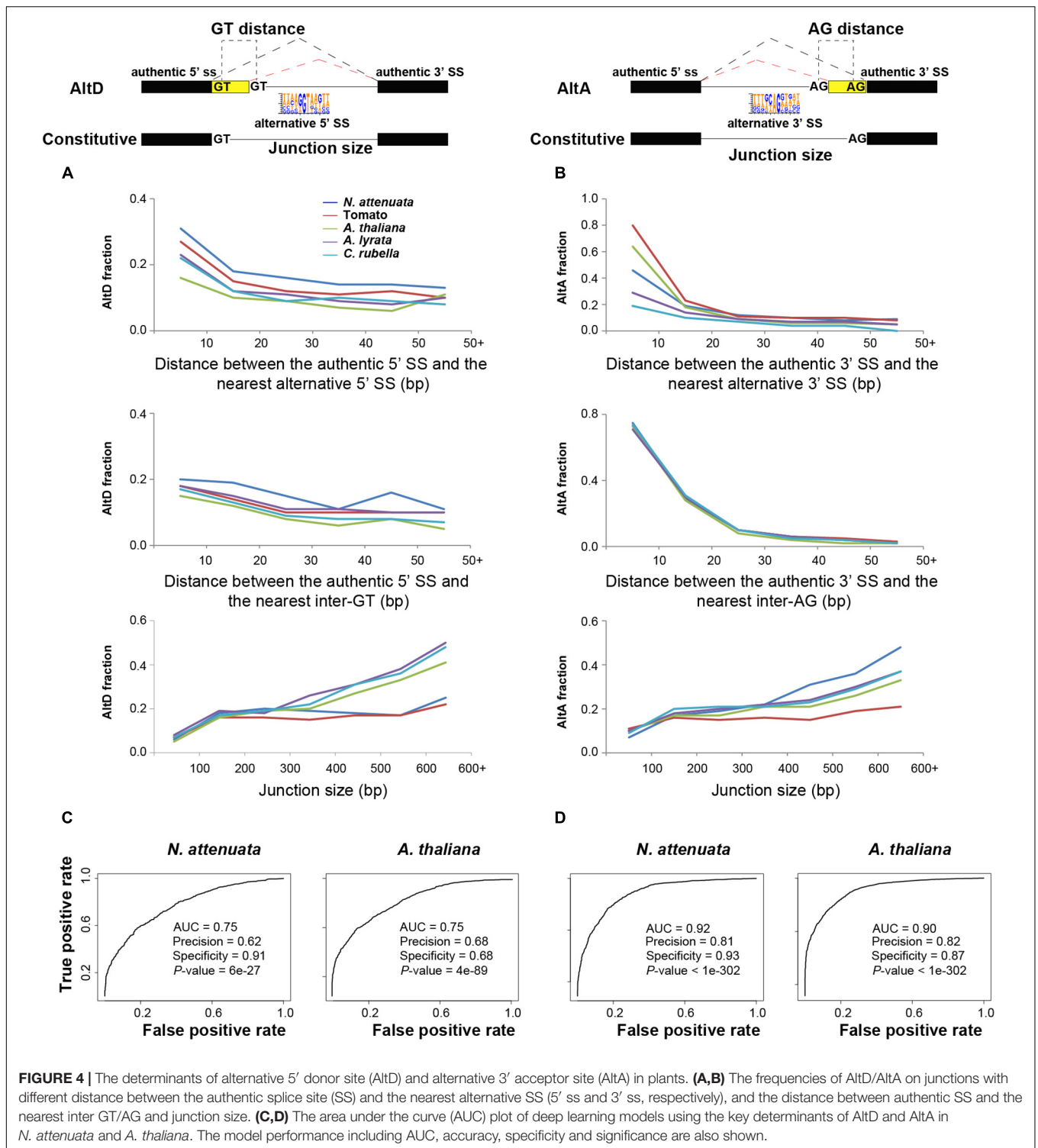


FIGURE 4 | The determinants of alternative 5' donor site (AltD) and alternative 3' acceptor site (AltA) in plants. **(A,B)** The frequencies of AltD/AltA on junctions with different distance between the authentic splice site (SS) and the nearest alternative SS (5' ss and 3' ss, respectively), and the distance between authentic SS and the nearest inter GT/AG and junction size. **(C,D)** The area under the curve (AUC) plot of deep learning models using the key determinants of AltD and AltA in *N. attenuata* and *A. thaliana*. The model performance including AUC, accuracy, specificity and significance are also shown.

the three most abundant AS types AltD, AltA, and IR. As expected, the results showed that for a given junction, the likelihood of both AltD and AltA increases with junction size, while the likelihood of both AltD and AltA decreases with the distance between authentic and alternative SS as well as the distance between authentic SS and the nearest

internal GT/AG (Figures 4A,B). Interestingly, although the likelihood of IR in smaller junctions appears larger than in large junctions, no significant correlation with junction size was found (Supplementary Figure S13A). Both 5' and 3' SS of the junction with IR are significantly weaker than those of the constitutive junction (Supplementary Figure S13B).

Furthermore, the presence/absence of UA-rich tract, polypyrimidine tract and branch site are also known to be associated with 3' SS recognition in eukaryotes (Lewandowska et al., 2004; Fu and Ares, 2014). We compared the frequency of AltA and IR between junctions of the AS gene with and without the presence of UA, polypyrimidine tract and branch site within 100 bp upstream of 3' SS. We found that the frequencies of both AltA and IR are significantly higher in the junctions without UA and polypyrimidine tract than the junctions with them, while the presence of branch site had no significant effect (**Supplementary Figure S14**).

Cis-regulatory elements, including splicing enhancers and silencers located close to SS are also important for the regulations of splicing. To identify these candidate regulatory elements, we performed a *de novo* hexamer motif enrichment analysis by comparing 50 bp sequences from the 5' and 3' sides of both donor and acceptor sites between alternatively spliced and constitutively spliced junctions. The results showed that most of the putative enhancer motifs for alternatively spliced junctions are highly similar to the identified SS. In addition, we also identified several putative silencer motifs (range from 5 to 10 for AltD and 10 to 18 for AltA in the five species), some of which were significantly more enriched in constitutively spliced junctions than alternatively spliced junctions in all species (**Supplementary Figures S15A,B**). However, it is worth noticing that the *cis*-regulatory elements that are located 50 bp further away from the SS or less than 6 bp may have been missed from our analysis.

To evaluate whether these identified features represent the AS determinants, we used a machine learning approach and modeled the different types of AS in each of the studied species. The rationale for this approach is that if the features we identified as representative of the key AS determinants were accurate, we would be able to predict whether an exon-intron junction is constitutively or alternatively spliced based on their quantitative or qualitative information. For this, we combined all of the extracted features mentioned above. In addition, we also extracted information on whether the alternative SS would introduce a frameshift, which may result in premature terminate code (PTC) or different open reading frames (ORFs), the number of reads that support the junction, which represent levels of expression that is known to be associated with AS, as well as the presence and absence of the identified *cis*-motifs. Using this information, our model achieved high precision and specificity for both AltD and AltA in all five species (**Figures 4C,D** and **Supplementary Figures S16A,B**), which suggests that the identified features can provide sufficient information to discriminate AS junctions from constitutively spliced junctions. However, for IR, the extracted features were not predictively useful, as the average model performance measured by area under the receiver operating characteristic curve (AUC) was only 0.54, suggesting low precision and low specificity. This indicates that additional undetected factors have contributed to the determination of IR.

This modeling approach further provides indicative information on the relative importance of each feature to the prediction model. The results showed that for AltD, the distance between the authentic SS and the nearest alternative

5' SS or inter GT, the junction size and presence/absence of 5' additional SS in the intron are among the top important features for the prediction in all species (**Supplementary Data Sheet S1**). In addition, the frame shifts introduced by the nearest alternative 5' SS and nearest GT were also important contributors to the model (**Supplementary Data Sheet S1**). For AltA, the distance to the nearest inter-AG dinucleotide is the top feature for the prediction among all five species. Interestingly, all of the identified putative silencers/enhancers (6-mers motifs) only had a marginal role for the predictions of both AltD and AltA (**Supplementary Data Sheet S1**), the same top important features were presented in models without these motif features. Together, these results showed that the mechanisms regulating AltD and AltA are likely overall conserved among the studied species.

Changes in AS Determinants Contributed to the Divergence of AS in Plants

The relatively conserved AS regulation mechanisms among studied species provide a foundation for investigating the mechanisms that contributed to the divergence of AS among closely related plant species. We hypothesized that the changes in the identified AS determinants among species resulted in the divergence of AS among species. To test this, we associated the changes of the identified AS determinants and AS conservation among closely related species. Because we did not find determinants for IR, we only focused on the evolution of AltA and AltD.

Variation in the distance between authentic SS and alternative SS or inter-GT/AG were negatively associated with AS conservation: the levels of AS conservation decreased with increasing distance in all three pairs of comparisons (**Figures 5A,B**), for both AltD and AltA. In addition, the changes in the reading frame introduced by the alternative SS also significantly decreased the conservations of both AltA and AltD (**Figures 5C,D**). The similar pattern was also found for the distance between authentic SS and the nearest inter-GT/AG (**Figures 5E–H**).

Variation in the *cis*-regulatory elements UA-tract, polypyrimidine tract and branch site significantly reduced the conservation for AltA, but did not affect the conservation of AltD among species (**Supplementary Figures S17A,B** and **Supplementary Data Sheet S2**). This result is consistent with the functional roles of these *cis*-regulatory elements in regulating AltA.

To further systematically analyze different factors that might affect the conservation of AS, we constructed an AS evolution model for each closely related species pair using a deep learning method. In addition to the key AS determinants identified in this study, we also included several other features that were previously hypothesized to be important for AS conservation between species in the model, such as changes in copy numbers (role of gene duplications), transposable element (TE) insertion within the junction, GC-content and sequence similarity of SS. For the AltD, all three models between species pairs achieved significantly better prediction than by chance (highest *P*-value = $3e-44$), with an average precision of 0.63 and specificity of 0.82. In all

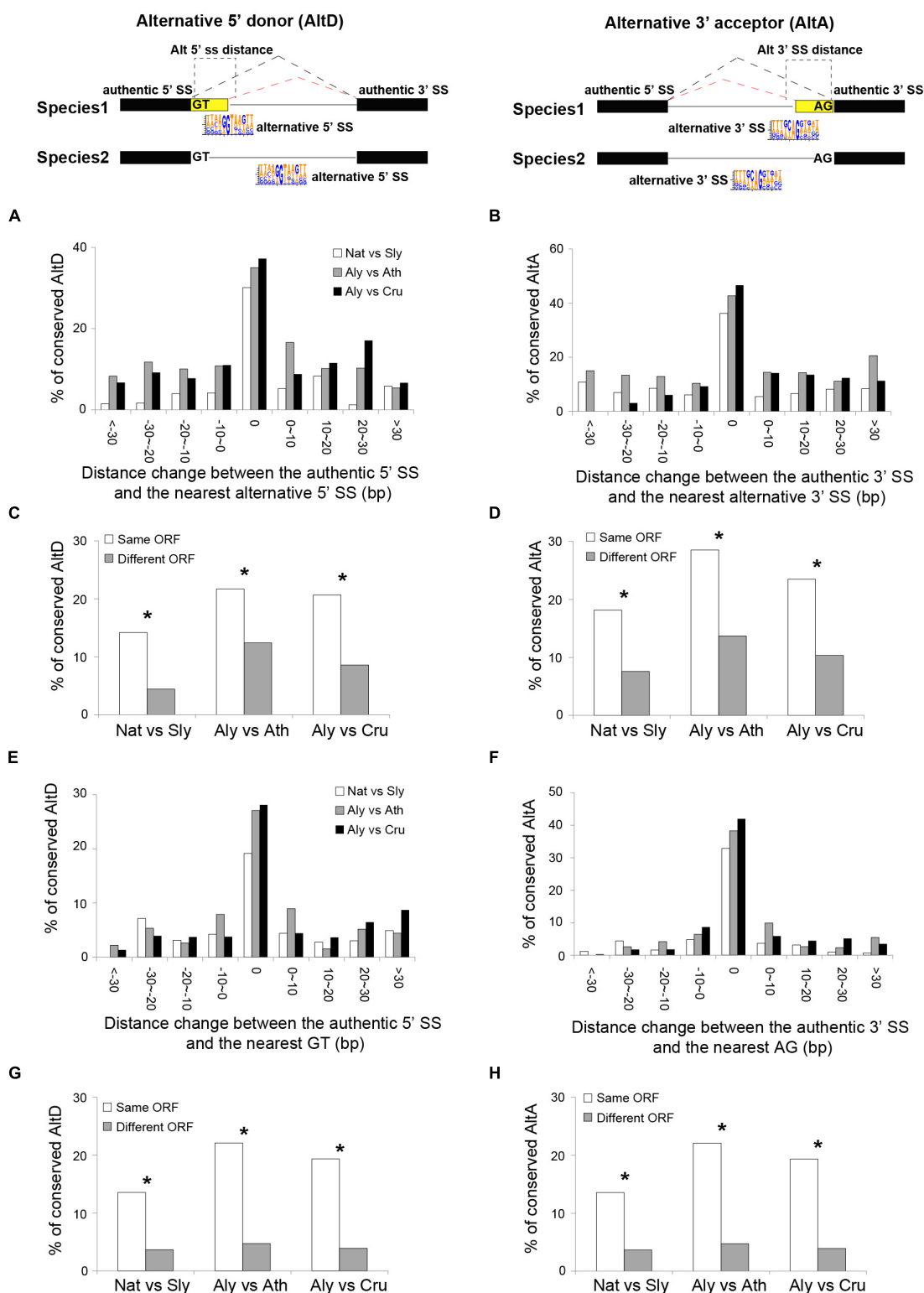


FIGURE 5 | Features that affect the conservation of AltD and AltA between closely related plant species. **(A,B)** The portion of conserved AltD/AltA decreases with the distance between authentic and alternative splice site (SS) between two species. **(C,D)** The percentage of conserved AltD/AltA which the nearest alternative 5'/3' SS generates transcripts with same or different open reading frame (ORF) between species. **(E,F)** The portion of conserved AltD/AltA decreases with changes in the distance between authentic SS and nearest inter-GT/AG sites between two species. **(G,H)** The percentage of conserved AltD/AltA in which that the nearest inter-GT/AG generates transcripts with same or different ORF between species. Nat, *N. attenuata*; Sly, Tomato; Ath, *A. thaliana*; Aly, *A. lyrata*. The asterisks indicate the significance as determined by Fisher's exact test ($P < 0.05$).

three pairwise comparison models, the distance changes between authentic and nearest alternative 5' SS or inter-GT/AG are among the top five important features (**Supplementary Figure S18A** and **Supplementary Data Sheet S2**). For AltA, all three models achieved a precision and specificity (average 0.70 and 0.85, respectively) that was significantly higher than by chance (highest P -value = $3e-145$). In all three models, distance changes between authentic SS and the nearest inter-AG or alternative 3' SS and the changes on *cis*-regulatory elements (UA and polypyrimidine tracts) represent the top five most important features that contributed to the model predictions (**Supplementary Figure S18B** and **Supplementary Data Sheet S2**).

Interestingly, we found TE insertions to also be an important factor that reduced the conservation of both AltD and AltA between *N. attenuata* and tomato but not between any pair of the Brassicaceae species (**Supplementary Figures S18A,B**). This is likely due to the difference of TE abundance between *N. attenuata* (~63%) and tomato (~81%), values which are much higher than the differences between *A. thaliana* (~23%) and *A. lyrata* (28%) (Hu et al., 2011; Tomato Genome Consortium, 2012). Furthermore, we also analyzed the impact of DNA methylation changes between *A. thaliana* and *A. lyrata* using data from (Seymour et al., 2014) and found no significant effects (**Supplementary Figures S18A,B**).

DISCUSSION

Here, we showed that species-specific gain and loss of AS resulted in lineage-specific AS profiles in plants. Between closely related species, AS events that introduce PTCs are likely more conserved than AS events that do not introduce PTC (**Figure 3A**). Consistently, several AS events that generate PTC-containing transcripts were ultra-conserved among highly divergent plants. To understand the mechanisms that resulted in a rapid divergence of AS between closely related species, we identified several key determinants for both alternative donor (AltD) and alternative acceptor (AltA) splicing. We found that the change of these key determinants between species is associated with the gain and loss of AS in plants.

In this analysis, we observed a dominant species-specific pattern of AS among different species (**Figure 1**). Although, the relatively low sequence depth (17 million) or incomplete genome assembly and annotation might be a confounding effect to draw this conclusion. We did several analyses to examine this and found: (i) overall, 17M unique mapped reads can sufficiently detect more than 75% of total splice junctions in all four species (**Supplementary Figure S3A**); (ii) increase of sequencing depth from 17M didn't dramatically increase the number of identified splice junctions, suggesting that 17M reads already reached or at least is close to the saturation point (**Supplementary Figure S3B**); (iii) the same patterns have been observed using all available reads (**Supplementary Figures S3C,D**). Thus, we believe that the main conclusions of the work are not affected by the relatively low sequencing depth or stochasticity from the random sampling. However, further studies using similar datasets with higher number of reads can provide stronger evidence for this.

The species-specific AS clustering pattern was also found among vertebrate species that span ~350 million years of evolution (Barbosa-Morais et al., 2012; Merkin et al., 2012), indicating that this might be universal among eukaryotes. Interestingly, in vertebrates, some tissues, such as brain, testis, heart and muscle still showed a strong tissue-specific splicing signature, despite the dominant species-specific splicing background (Barbosa-Morais et al., 2012; Merkin et al., 2012). Although all three tissues (root, leaves, and flowers) used in our study did not show such strong tissue-specific splicing signatures, some other plant tissues might. For example, the transcriptomes of sexual tissues are substantially different from those of vegetative tissues, and anthers harbor the most diverged specialized metabolomes (Yang and Wang, 2013; Li et al., 2016). Future studies that include transcriptome data of much more fine-scaled tissue samples will provide new insights on this aspect.

AS events that resulted in transcripts with PTC, are coupled with nonsense-mediated decay NMD. They are more conserved than the AS that do not generate PTC-containing transcripts in plants (**Figure 3A** and **Supplementary Figure S10**). Consistently, among six ultra-conserved AS events across different plant species including the spreading earth moss, five produced +PTC transcripts, indicating that AS+PTC might be more important than it was previously thought. Previous studies showed that all human serine/arginine-rich (SR) genes and some SR genes in plants produce AS resulted in +PTC transcripts (Kalyna et al., 2006; Lareau et al., 2007; Filichkin et al., 2010; Palusa and Reddy, 2010). Furthermore, the junction regions that contain AS+PTC in numerous splicing factors (SFs) are ultra-conserved between different kingdoms and the loss of the ancient AS+PTC in paralogs through gene duplications were repeatedly replaced by newly created distinct unproductive splicing (Lewis et al., 2003; Lareau et al., 2007; Lareau and Brenner, 2015). Similar to these previous works, our results are consistent with the hypothesis that the unproductive splicing coupled with NMD can be a functional process that influences the abundance of active proteins at a post-transcriptional level.

One caveat from our analysis on the conservation of AS+PTC events is that we focused on only a subset of AS events, due to methodological challenges of associating AS events with specific transcripts and annotating PTC. Identifying and annotating AS+PTC events from RNA-seq data is computationally challenging, to reduce the false positives, we applied stringent filtering parameters, and only focused on the transcripts that can be uniquely associated with a single AS event. Although the same filtering parameters were used for all of the AS events and the observed pattern is unlikely to be the result of such filtering, it remains unclear whether this pattern represents all AS events. Future studies that combine full-length transcript (Wang B. et al., 2016) and long reads sequencing technologies will reduce the computational complexity and errors involved in associating AS events with transcripts and may provide more robust analysis on the evolution and conservation of AS+PTC in plants.

Among the five investigated plant species, the distance between the 5'/3' nearest alternative SS and the authentic SS is the main determinant that distinguishes AltD/AltA from

constitutive splicing (**Figure 4** and **Supplementary Data Sheet S1**). For a given spliced junction, the likelihood of AS decreases with an increased distance between the authentic and nearest alternative SS (**Figures 4A,B**). Interestingly, similar patterns were also found in mammals, in which, the closer the alternative SS was to the authentic SS, the more likely it was used for AS (Dou et al., 2006; Rosenberg et al., 2015). The frequency of AltA also decreases with the increased distance between the authentic SS and nearest inter-AG dinucleotide. This result is consistent with the pattern found in humans in that only closely located AGs (< 6 nt) can effectively compete with the authentic SS and the distance between branch site and the first downstream AG can affect the 3' SS selection (Chiara et al., 1997; Chua and Reed, 2001). Although, the BS in plants is not well studied and BS was not identified in ~30% of junctions, a similar effect of inter-AG distance on AltA in both plants and mammals indicates that the mechanisms of generating AS, at least for AltA, might be similar between these two kingdoms.

While the deep learning model for AltA achieved high precision and specificity among five species (AUC > 0.9, indicating high precision and high specificity), the models for AltD performed less well, although still performing better than by chance (AUC > 0.75, **Figures 4C,D** and **Supplementary Figure S16**). This indicates that additional determinants that contribute to the regulations of AltD were not detected by our method. It is known that the mechanisms involved in AltD are more complex than in AltA. For example, in both human and mouse, while both the presence and quantity of exon splicing enhancer and exon splicing silencer are important for generating AltD (Koren et al., 2007). While binding sites of splicing factors can also be important, AltA is mainly affected by the competition of closely located AG dinucleotide by a scanning mechanism for the downstream sequence of the branch site polypyrimidine tract (Smith et al., 1989; Smith et al., 1993; Chiara et al., 1997; Chua and Reed, 2001). Furthermore, it is known that NAGNAG (N is any nucleotide), which is a subset of SS for AltA that are separated by three nucleotides, are enriched in genes encoding DNA-binding proteins in both plants and animals (Vogan et al., 1996; Iida et al., 2008; Schindler et al., 2008). These results suggest that splicing regulatory elements (SREs) may play more important roles in the proper selection of alternative SS in AltD than AltA. This may also explain why the junction size contributed more in the AltD model than in the AltA model (**Supplementary Data Sheet S1**), since larger junction size might increase the likelihood of introducing intronic SREs. Although a few candidate sequence motifs were identified using the enrichment analysis, none of them significantly contributed to the model predictions. Two non-exclusive possibilities may partially explain this failure. First, the identified motifs are not involved in affecting splicing processes, although their density was significantly different between constitutively and alternatively spliced junctions. Second, they might be essential for tissue-specific AS, which likely did not contribute to the overall AltD prediction based on all

three tissues. Future studies using different approaches to investigate the alteration of AS by introducing millions of random hexamers into specific regions of a gene junction in a plant then measuring the consequences of splicing, may allow us to more reliably detect splicing regulators of AltD in plants.

Although we found that both junction size and SS for IR junctions were different between the constitutively and alternatively spliced junctions (**Supplementary Figure S13**), the identified features did not improve the AS prediction from that occurring by chance. There are three non-exclusive possible reasons. First, the expression level of IR is usually low and therefore requires high sequencing depth for their detection (**Supplementary Figure S19**). It is possible that the sequencing depth of the transcriptome data used in this study was not sufficient to detect all of the IR junctions. In such case, many true IR junctions may not have been considered as IRs in our analysis, which reduced prediction precision and power. Second, a recent study showed that a subset of IR junctions – exon – has different features from regular IR junctions (Marquez et al., 2015). Thus their determinant might also be different. Third, previous work showed that a large proportion of IRs (76.5%) identified from RNA-seq result from incompletely spliced pre-mature mRNA (Zhang et al., 2015), thus increasing the false positives of IR. Future studies that sequence transcriptomes of different tissues among species using polyribosomal RNA-seq technology (Zhang et al., 2015) in high depth will likely reveal the mechanisms underlying IR regulations in plants.

For both AltA and AltD, their divergence between closely related species was likely due to variations in the key sequence determinants near the SS (**Figure 5** and **Supplementary Figures S17, S18**) and the key sequence determinants such as distance to authentic SS and *cis*-elements (branch site, polypyrimidine tract, UA-rich tract for AltA), which are all located within intronic regions. Intron sequences diverge faster than protein coding regions (Mattick, 1994; Hare and Palumbi, 2003), therefore, the process that likely have contributed to the species-specific gains and losses of AS among different lineages to produce species-specific AS profiles in plants. For example, a decreased distance between alternative SS and authentic SS as a result of a short deletion of the intron sequence could lead to a gain of AS at the junction, and as the consequence it is likely to be shared among different tissues. Consistently, in vertebrates, the mutations that affect intronic SREs were shown to be the main factor that resulted in the dominant species-specific splicing pattern (Merkin et al., 2012). However, it is unclear whether the observed changes of AS were neutral or under selection, because defining the null model of the AS evolution remains challenging. Furthermore, our data cannot exclude the possibility that the species-specific *trans*-factors, such as the SR protein families, which have distinct numbers of homologs among species (**Supplementary Figure S20**; Iida and Go, 2006; Isshiki et al., 2006; Ling et al., 2015), may have also contributed to the divergence of AS among different species (Ast, 2004; Barbosa-Morais et al., 2012). For example, there are 38 SR homologs in soybean, which is much

higher than the number of SR homologs in other plant species (**Supplementary Figure S20**). Such species-specific expansion of certain SR families may contribute to the relative unique AS pattern of soybean (**Figures 1A,B**).

We also investigated other factors that were hypothesized to affect AS evolution, such as gene duplication, DNA methylation and TE insertion (Sorek et al., 2002; Su et al., 2006; Flores et al., 2012). However, with the exception of TE insertions, the effects of which were found to be species-specific, most of the tested factors did not show significant effects on the levels of AS conservation between closely related species (**Supplementary Figure S18** and **Supplementary Data Sheet S2**). The species-specific effects of TE on the AS conservation were likely due to the different abundance of TE insertions in the genomes of different species (Hu et al., 2011; Tomato Genome Consortium, 2012; Slotte et al., 2013; Sierro et al., 2014), suggesting genomic composition of each species might also affect the evolutionary alteration of AS.

CONCLUSION

We found that the divergence of AS profile among species is associated with massive gains and losses of AS in each lineage, while a group of AS that generate PTC-containing transcripts were highly conserved even among very distantly related plants. The alteration of a few key sequence determinants of AltA and AltD, all located in the intron region, likely contributed to the divergence of AS among closely related plant species. These results provide mechanistic insights into the evolution of AS in plants and highlight the role of post-transcriptional regulation of a plant's responses to environmental stresses.

REFERENCES

- Alipanahi, B., Delong, A., Weirauch, M. T., and Frey, B. J. (2015). Predicting the sequence specificities of DNA- and RNA-binding proteins by deep learning. *Nat. Biotechnol.* 33:831. doi: 10.1038/nbt.3300
- Altschul, S. F., Gish, W., Miller, W., Myers, E. W., and Lipman, D. J. (1990). Basic local alignment search tool. *J. Mol. Biol.* 215, 403–410. doi: 10.1016/S0022-2836(05)80360-2
- Aoki, K., Yano, K., Suzuki, A., Kawamura, S., Sakurai, N., Suda, K., et al. (2010). Large-scale analysis of full-length cDNAs from the tomato (*Solanum lycopersicum*) cultivar Micro-Tom, a reference system for the Solanaceae genomics. *BMC Genomics* 11:210. doi: 10.1186/1471-2164-11-210
- Ast, G. (2004). How did alternative splicing evolve? *Nat. Rev. Genet.* 5, 773–782. doi: 10.1038/nrg1451
- Baek, J. M., Han, P., Iandolino, A., and Cook, D. R. (2008). Characterization and comparison of intron structure and alternative splicing between *Medicago truncatula*, *Populus trichocarpa*, *Arabidopsis* and rice. *Plant Mol. Biol.* 67, 499–510. doi: 10.1007/s11103-008-9334-4
- Barash, Y., Calarco, J. A., Gao, W., Pan, Q., Wang, X., Shai, O., et al. (2010). Deciphering the splicing code. *Nature* 465, 53–59. doi: 10.1038/nature09000
- Barbazuk, W. B., Fu, Y., and McGinnis, K. M. (2008). Genome-wide analyses of alternative splicing in plants: opportunities and challenges. *Genome Res.* 18, 1381–1392. doi: 10.1101/gr.053678.106
- Barbosa-Morais, N. L., Irimia, M., Pan, Q., Xiong, H. Y., Gueroussov, S., Lee, L. J., et al. (2012). The evolutionary landscape of alternative splicing in vertebrate species. *Science* 338, 1587–1593. doi: 10.1126/science.1230612

AUTHOR CONTRIBUTIONS

ZL and SX designed the research. ZL, TB, and SX performed the experiments and analyzed the data. ZL, IB, and SX wrote the manuscript.

FUNDING

The work was supported by Max Planck Society (All), Swiss National Science Foundation (Project No. PEBZP3-142886 to SX), Marie Curie Intra-European Fellowship (IEF) (Project No. 328935 to SX), and European Research Council advanced grant ClockworkGreen (Project No. 293926 to IB).

ACKNOWLEDGMENTS

We thank Danell Seymour and Daniel Koenig for providing the methylation data, Michal Szczesniak for providing the Perl scripts for finding UA tracts. We also thank Martin Schäfer for providing comments on the language. We acknowledge support from the Open Access Publication Fund of the University of Münster. An earlier version of this manuscript has been submitted to a preprint server (<https://www.biorxiv.org/content/biorxiv/early/2017/02/13/107938.full.pdf>).

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fpls.2019.00707/full#supplementary-material>

- Black, D. L. (2003). Mechanisms of alternative pre-messenger RNA splicing. *Annu. Rev. Biochem.* 72, 291–336. doi: 10.1146/annurev.biochem.72.121801.161720
- Bon, E., Casaregola, S., Blandin, G., Llorente, B., Neuveglise, C., Munsterkotter, M., et al. (2003). Molecular evolution of eukaryotic genomes: hemiascomycetous yeast spliceosomal introns. *Nucleic Acids Res.* 31, 1121–1135. doi: 10.1093/nar/gkg213
- Braunschweig, U., Barbosa-Morais, N. L., Pan, Q., Nachman, E. N., Alipanahi, B., Gonatopoulos-Pournatzis, T., et al. (2014). Widespread intron retention in mammals functionally tunes transcriptomes. *Genome Res.* 24, 1774–1786. doi: 10.1101/gr.177790.114
- Chang, Y. F., Imam, J. S., and Wilkinson, M. F. (2007). The nonsense-mediated decay RNA surveillance pathway. *Annu. Rev. Biochem.* 76, 51–74. doi: 10.1146/annurev.biochem.76.050106.093909
- 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. doi: 10.1038/nrm2777
- Chiara, M. D., Palandjian, L., Feld Kramer, R., and Reed, R. (1997). Evidence that U5 snRNP recognizes the 3' splice site for catalytic step II in mammals. *EMBO J.* 16, 4746–4759. doi: 10.1093/emboj/16.15.4746
- Chua, K., and Reed, R. (2001). An upstream AG determines whether a downstream AG is selected during catalytic step II of splicing. *Mol. Cell Biol.* 21, 1509–1514. doi: 10.1128/MCB.21.5.1509-1514.2001
- Darracq, A., and Adams, K. L. (2013). Features of evolutionarily conserved alternative splicing events between *Brassica* and *Arabidopsis*. *New Phytol.* 199, 252–263. doi: 10.1111/nph.12238

- Ding, F., Cui, P., Wang, Z., Zhang, S., Ali, S., and Xiong, L. (2014). Genome-wide analysis of alternative splicing of pre-mRNA under salt stress in *Arabidopsis*. *BMC Genomics* 15:431. doi: 10.1186/1471-2164-15-431
- Donahue, C. P., Muratore, C., Wu, J. Y., Kosik, K. S., and Wolfe, M. S. (2006). Stabilization of the tau exon 10 stem loop alters pre-mRNA splicing. *J. Biol. Chem.* 281, 23302–23306. doi: 10.1074/jbc.C600143200
- Dou, Y., Fox-Walsh, K. L., Baldi, P. F., and Hertel, K. J. (2006). Genomic splice-site analysis reveals frequent alternative splicing close to the dominant splice site. *RNA* 12, 2047–2056. doi: 10.1261/rna.151106
- Drechsel, G., Kahles, A., Kesarwani, A. K., Stauffer, E., Behr, J., Drewe, P., et al. (2013). Nonsense-mediated decay of alternative precursor mRNA splicing variants is a major determinant of the *Arabidopsis* steady state transcriptome. *Plant Cell* 25, 3726–3742. doi: 10.1105/tpc.113.115485
- Ellis, J. D., Barrios-Rodiles, M., Colak, R., Irimia, M., Kim, T., Calarco, J. A., et al. (2012). Tissue-specific alternative splicing remodels protein-protein interaction networks. *Mol. Cell* 46, 884–892. doi: 10.1016/j.molcel.2012.05.037
- Filichkin, S. A., Cumbie, J. S., Dharmawardhana, P., Jaiswal, P., Chang, J. H., Palusa, S. G., et al. (2015). Environmental stresses modulate abundance and timing of alternatively spliced circadian transcripts in *Arabidopsis*. *Mol. Plant* 8, 207–227. doi: 10.1016/j.molp.2014.10.011
- Filichkin, S. A., Priest, H. D., Givan, S. A., Shen, R., Bryant, D. W., Fox, S. E., et al. (2010). Genome-wide mapping of alternative splicing in *Arabidopsis thaliana*. *Genome Res.* 20, 45–58. doi: 10.1101/gr.093302.109
- Flores, K., Wolschin, F., Corneveaux, J. J., Allen, A. N., Huentelman, M. J., and Amdam, G. V. (2012). Genome-wide association between DNA methylation and alternative splicing in an invertebrate. *BMC Genomics* 13:480. doi: 10.1186/1471-2164-13-480
- Fu, X. D., and Ares, M. Jr. (2014). Context-dependent control of alternative splicing by RNA-binding proteins. *Nat. Rev. Genet.* 15, 689–701. doi: 10.1038/nrg3778
- Gao, H., Gordon-Kamm, W. J., and Lyznick, L. A. (2004). ASF/SF2-like maize pre-mRNA splicing factors affect splice site utilization and their transcripts are alternatively spliced. *Gene* 339, 25–37. doi: 10.1016/j.gene.2004.06.047
- Gopal, S., Awadalla, S., Gaasterland, T., and Cross, G. A. (2005). A computational investigation of kinetoplastid trans-splicing. *Genome Biol.* 6:R95. doi: 10.1186/gb-2005-6-11-r95
- Grabherr, M. G., Haas, B. J., Yassour, M., Levin, J. Z., Thompson, D. A., Amit, I., et al. (2011). Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nat. Biotechnol.* 29, 644–652. doi: 10.1038/nbt.1883
- Graveley, B. R., Brooks, A. N., Carlson, J. W., Duff, M. O., Landolin, J. M., Yang, L., et al. (2011). The developmental transcriptome of *Drosophila melanogaster*. *Nature* 471, 473–479. doi: 10.1038/nature09715
- Hare, M. P., and Palumbi, S. R. (2003). High intron sequence conservation across three mammalian orders suggests functional constraints. *Mol. Biol. Evol.* 20, 969–978. doi: 10.1093/molbev/msg111
- Heinz, S., Benner, C., Spann, N., Bertolino, E., Lin, Y. C., Laslo, P., et al. (2010). Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. *Mol. Cell* 38, 576–589. doi: 10.1016/j.molcel.2010.05.004
- Hori, K., and Watanabe, Y. (2007). Context analysis of termination codons in mRNA that are recognized by plant NMD. *Plant Cell Physiol.* 48, 1072–1078. doi: 10.1093/pcp/pcm075
- Howard, B. E., Hu, Q., Babaoglu, A. C., Chandra, M., Borghi, M., Tan, X., et al. (2013). High-throughput RNA sequencing of *Pseudomonas*-infected *Arabidopsis* reveals hidden transcriptome complexity and novel splice variants. *PLoS One* 8:e74183. doi: 10.1371/journal.pone.0074183
- Hu, T. T., Pattyn, P., Bakker, E. G., Cao, J., Cheng, J. F., Clark, R. M., et al. (2011). The *Arabidopsis lyrata* genome sequence and the basis of rapid genome size change. *Nat. Genet.* 43, 476–481. doi: 10.1038/ng.807
- Iida, K., and Go, M. (2006). Survey of conserved alternative splicing events of mRNAs encoding SR proteins in land plants. *Mol. Biol. Evol.* 23, 1085–1094. doi: 10.1093/molbev/msj118
- Iida, K., Shionyu, M., and Suso, Y. (2008). Alternative splicing at NAGNAG acceptor sites shares common properties in land plants and mammals. *Mol. Biol. Evol.* 25, 709–718. doi: 10.1093/molbev/msn015
- Isshiki, M., Tsumoto, A., and Shimamoto, K. (2006). The serine/arginine-rich protein family in rice plays important roles in constitutive and alternative splicing of pre-mRNA. *Plant Cell* 18, 146–158. doi: 10.1105/tpc.105.037069
- Kalyna, M., Lopato, S., Voronin, V., and Barta, A. (2006). Evolutionary conservation and regulation of particular alternative splicing events in plant SR proteins. *Nucleic Acids Res.* 34, 4395–4405. doi: 10.1093/nar/gkl570
- Kalyna, M., Simpson, C. G., Syed, N. H., Lewandowska, D., Marquez, Y., Kusenda, B., et al. (2012). Alternative splicing and nonsense-mediated decay modulate expression of important regulatory genes in *Arabidopsis*. *Nucleic Acids Res.* 40, 2454–2469. doi: 10.1093/nar/gkr932
- Kandul, N. P., and Noor, M. A. (2009). Large introns in relation to alternative splicing and gene evolution: a case study of *Drosophila* bruno-3. *BMC Genet.* 10:67. doi: 10.1186/1471-2156-10-67
- Kazan, K. (2003). Alternative splicing and proteome diversity in plants: the tip of the iceberg has just emerged. *Trends Plant Sci.* 8, 468–471. doi: 10.1016/j.tplants.2003.09.001
- Kelley, D. R., Hendrickson, D. G., Tenen, D., and Rinn, J. L. (2014). Transposable elements modulate human RNA abundance and splicing via specific RNA-protein interactions. *Genome Biol.* 15:537. doi: 10.1186/s13059-014-0537-5
- Kerenyi, Z., Merai, Z., Hiripi, L., Benkovics, A., Gyula, P., Lacomme, C., et al. (2008). Inter-kingdom conservation of mechanism of nonsense-mediated mRNA decay. *EMBO J.* 27, 1585–1595. doi: 10.1038/emboj.2008.88
- Kervestin, S., and Jacobson, A. (2012). NMD: a multifaceted response to premature translational termination. *Nat. Rev. Mol. Cell Biol.* 13, 700–712. doi: 10.1038/nrm3454
- Kolasinska-Zwierz, P., Down, T., Latorre, I., Liu, T., Liu, X. S., and Ahinger, J. (2009). Differential chromatin marking of introns and expressed exons by H3K36me3. *Nat. Genet.* 41, 376–381. doi: 10.1038/ng.322
- Koren, E., Lev-Maor, G., and Ast, G. (2007). The emergence of alternative 3' and 5' splice site exons from constitutive exons. *PLoS Comput. Biol.* 3:e95. doi: 10.1371/journal.pcbi.0030095
- Lambert, M. J., Cochran, W. O., Wilde, B. M., Olsen, K. G., and Cooper, C. D. (2015). Evidence for widespread subfunctionalization of splice forms in vertebrate genomes. *Genome Res.* 25, 624–632. doi: 10.1101/gr.184473.114
- Lareau, L. F., and Brenner, S. E. (2015). Regulation of splicing factors by alternative splicing and NMD is conserved between kingdoms yet evolutionarily flexible. *Mol. Biol. Evol.* 32, 1072–1079. doi: 10.1093/molbev/msv002
- Lareau, L. F., Inada, M., Green, R. E., Wengrod, J. C., and Brenner, S. E. (2007). Unproductive splicing of SR genes associated with highly conserved and ultraconserved DNA elements. *Nature* 446, 926–929. doi: 10.1038/nature05676
- Leung, M. K., Xiong, H. Y., Lee, L. J., and Frey, B. J. (2014). Deep learning of the tissue-regulated splicing code. *Bioinformatics* 30, i121–i129. doi: 10.1093/bioinformatics/btu277
- Lev-Maor, G., Goren, A., Sela, N., Kim, E., Keren, H., Doron-Faigenboim, A., et al. (2007). The “alternative” choice of constitutive exons throughout evolution. *PLoS Genet.* 3:e203. doi: 10.1371/journal.pgen.0030203
- Lewandowska, D., Simpson, C. G., Clark, G. P., Jennings, N. S., Barciszewska-Pacak, M., Lin, C. F., et al. (2004). Determinants of plant U12-dependent intron splicing efficiency. *Plant Cell* 16, 1340–1352. doi: 10.1105/tpc.020743
- Lewis, B. P., Green, R. E., and Brenner, S. E. (2003). Evidence for the widespread coupling of alternative splicing and nonsense-mediated mRNA decay in humans. *Proc. Natl. Acad. Sci. U.S.A.* 100, 189–192. doi: 10.1073/pnas.0136770100
- Li, D., Heiling, S., Baldwin, I. T., and Gaquerel, E. (2016). Illuminating a plant's tissue-specific metabolic diversity using computational metabolomics and information theory. *Proc. Natl. Acad. Sci. U.S.A.* 113, E7610–E7618. doi: 10.1073/pnas.1610218113
- Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., et al. (2009). The sequence alignment/map format and SAMtools. *Bioinformatics* 25, 2078–2079. doi: 10.1093/bioinformatics/btp352
- Li, W., Lin, W. D., Ray, P., Lan, P., and Schmidt, W. (2013). Genome-wide detection of condition-sensitive alternative splicing in *Arabidopsis* roots. *Plant Physiol.* 162, 1750–1763. doi: 10.1104/pp.113.217778
- Licatalosi, D. D., Mele, A., Fak, J. J., Ule, J., Kayikci, M., Chi, S. W., et al. (2008). HITS-CLIP yields genome-wide insights into brain alternative RNA processing. *Nature* 456, 464–469. doi: 10.1038/nature07488
- Lindgreen, S. (2012). AdapterRemoval: easy cleaning of next-generation sequencing reads. *BMC Res. Notes* 5:337. doi: 10.1186/1756-0500-5-337
- Ling, Z., Zhou, W., Baldwin, I. T., and Xu, S. (2015). Insect herbivory elicits genome-wide alternative splicing responses in *Nicotiana attenuata*. *Plant J.* 84, 228–243. doi: 10.1111/tj.12997

- Liu, H. X., Goodall, G. J., Kole, R., and Filipowicz, W. (1995). Effects of secondary structure on pre-mRNA splicing: hairpins sequestering the 5' but not the 3' splice site inhibit intron processing in *Nicotiana plumbaginifolia*. *EMBO J.* 14, 377–388. doi: 10.1002/j.1460-2075.1995.tb07012.x
- Lopato, S., Gattoni, R., Fabini, G., Stevenin, J., and Barta, A. (1999). A novel family of plant splicing factors with a Zn knuckle motif: examination of RNA binding and splicing activities. *Plant Mol. Biol.* 39, 761–773.
- Lorkovic, Z. J. (2009). Role of plant RNA-binding proteins in development, stress response and genome organization. *Trends Plant Sci.* 14, 229–236. doi: 10.1016/j.tplants.2009.01.007
- Mamoshina, P., Vieira, A., Putin, E., and Zhavoronkov, A. (2016). Applications of deep learning in biomedicine. *Mol. Pharm.* 13, 1445–1454. doi: 10.1021/acs.molpharmaceut.5b00982
- Marondedze, C., Thomas, L., Serrano, N. L., Lilley, K. S., and Gehring, C. (2016). The RNA-binding protein repertoire of *Arabidopsis thaliana*. *Sci. Rep.* 6:29766. doi: 10.1038/srep29766
- 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. doi: 10.1101/gr.134106.111
- Marquez, Y., Hopfner, M., Ayatollahi, Z., Barta, A., and Kalyna, M. (2015). Unmasking alternative splicing inside Protein-coding exons defines exons and their role in proteome plasticity. *Genome Res.* 25, 995–1007. doi: 10.1101/gr.186585.114
- Mastrangelo, A. M., Marone, D., Laido, G., De Leonardi, A. M., and De Vita, P. (2012). Alternative splicing: enhancing ability to cope with stress via transcriptome plasticity. *Plant Sci.* 18, 40–49. doi: 10.1016/j.plantsci.2011.09.006
- Mattick, J. S. (1994). Introns: evolution and function. *Curr. Opin. Genet. Dev.* 4, 823–831. doi: 10.1016/0959-437x(94)90066-3
- Merkin, J., Russell, C., Chen, P., and Burge, C. B. (2012). Evolutionary dynamics of gene and isoform regulation in Mammalian tissues. *Science* 338, 1593–1599. doi: 10.1126/science.1228186
- Nagy, E., and Maquat, L. E. (1998). A rule for termination-codon position within intron-containing genes: when nonsense affects RNA abundance. *Trends Biochem. Sci.* 23, 198–199. doi: 10.1016/s0968-0004(98)01208-0
- Ner-Gaon, H., Leviatan, N., Rubin, E., and Fluhr, R. (2007). Comparative cross-species alternative splicing in plants. *Plant Physiol.* 144, 1632–1641. doi: 10.1104/pp.107.098640
- Ni, J. Z., Grate, L., Donohue, J. P., Preston, C., Nobida, N., O'Brien, G., et al. (2007). Ultraconserved elements are associated with homeostatic control of splicing regulators by alternative splicing and nonsense-mediated decay. *Gene Dev.* 21, 708–718. doi: 10.1101/gad.1525507
- 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. doi: 10.1111/j.1365-313X.2006.03020.x
- 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. doi: 10.1111/j.1469-8137.2009.03065.x
- Pertea, M., Mount, S. M., and Salzberg, S. L. (2007). A computational survey of candidate exonic splicing enhancer motifs in the model plant *Arabidopsis thaliana*. *BMC Bioinformatics* 8:159. doi: 10.1186/1471-2105-8-159
- R Development Core Team (2013). *R: A Language and Environment for Statistical Computing*. Vienna: R Foundation for Statistical Computing.
- Reddy, A. S. (2007). Alternative splicing of pre-messenger RNAs in plants in the genomic era. *Annu. Rev. Plant Biol.* 58, 267–294. doi: 10.1146/annurev.arplant.58.032806.103754
- Reddy, A. S., Marquez, Y., Kalyna, M., and Barta, A. (2013). Complexity of the alternative splicing landscape in plants. *Plant Cell* 25, 3657–3683. doi: 10.1105/tpc.113.117523
- Reddy, A. S., and Shad Ali, G. (2011). Plant serine/arginine-rich proteins: roles in precursor messenger RNA splicing, plant development, and stress responses. *Wiley Interdiscip. Rev. RNA* 2, 875–889. doi: 10.1002/wrna.98
- Richardson, D. N., Rogers, M. F., Labadorf, A., Ben-Hur, A., Guo, H., Paterson, A. H., et al. (2011). Comparative analysis of serine/arginine-rich proteins across 27 eukaryotes: insights into sub-family classification and extent of alternative splicing. *PLoS One* 6:e24542. doi: 10.1371/journal.pone.0024542
- Rosenberg, A. B., Patwardhan, R. P., Shendure, J., and Seelig, G. (2015). Learning the sequence determinants of alternative splicing from millions of random sequences. *Cell* 163, 698–711. doi: 10.1016/j.cell.2015.09.054
- Rosloski, S. M., Singh, A., Jali, S. S., Balasubramanian, S., Weigel, D., and Grbic, V. (2013). Functional analysis of splice variant expression of MADS AFFECTING FLOWERING 2 of *Arabidopsis thaliana*. *Plant Mol. Biol.* 81, 57–69. doi: 10.1007/s11103-012-9982-2
- Satyawan, D., Kim, M. Y., and Lee, S. H. (2016). Stochastic alternative splicing is prevalent in mungbean (*Vigna radiata*). *Plant Biotechnol. J.* 15, 174–182. doi: 10.1111/pbi.12600
- Schindler, S., Szafranski, K., Hiller, M., Ali, G. S., Palusa, S. G., Backofen, R., et al. (2008). Alternative splicing at NAGNAG acceptors in *Arabidopsis thaliana* SR and SR-related protein-coding genes. *BMC Genomics* 9:159. doi: 10.1186/1471-2164-9-159
- Schoenberg, D. R., and Maquat, L. E. (2012). Regulation of cytoplasmic mRNA decay. *Nat. Rev. Genet.* 13, 246–259. doi: 10.1038/nrg3160
- Schoning, J. C., Streitner, C., Meyer, I. M., Gao, Y., and Staiger, D. (2008). Reciprocal regulation of glycine-rich RNA-binding proteins via an interlocked feedback loop coupling alternative splicing to nonsense-mediated decay in *Arabidopsis*. *Nucleic Acids Res.* 36, 6977–6987. doi: 10.1093/nar/gkn847
- Schwartz, A. M., Komarova, T. V., Skulachev, M. V., Zvereva, A. S., Dorokhov Iu, L., and Atabekov, J. G. (2006). Stability of plant mRNAs depends on the length of the 3'-untranslated region. *Biochemistry* 71, 1377–1384. doi: 10.1134/s0006297906120145
- Schwartz, S., Meshorer, E., and Ast, G. (2009). Chromatin organization marks exon-intron structure. *Nat. Struct. Mol. Biol.* 16, 990–995. doi: 10.1038/nsmb.1659
- Schwartz, S. H., Silva, J., Burstein, D., Pupko, T., Eyra, E., and Ast, G. (2008). Large-scale comparative analysis of splicing signals and their corresponding splicing factors in eukaryotes. *Genome Res.* 18, 88–103. doi: 10.1101/gr.6818908
- Severing, E. I., van Dijk, A. D., Morabito, G., Busscher-Lange, J., Immink, R. G., and van Ham, R. C. (2012). Predicting the impact of alternative splicing on plant MADS domain protein function. *PLoS One* 7:e30524. doi: 10.1371/journal.pone.0030524
- Severing, E. I., van Dijk, A. D. J., Stiekema, W. J., and van Ham, R. C. (2009). Comparative analysis indicates that alternative splicing in plants has a limited role in functional expansion of the proteome. *BMC Genomics* 10:154. doi: 10.1186/1471-2164-10-154
- Seymour, D. K., Koenig, D., Hagmann, J., Becker, C., and Weigel, D. (2014). Evolution of DNA methylation patterns in the Brassicaceae is driven by differences in genome organization. *PLoS Genet.* 10:e1004785. doi: 10.1371/journal.pgen.1004785
- Shen, Y., Zhou, Z., Wang, Z., Li, W., Fang, C., Wu, M., et al. (2014). Global dissection of alternative splicing in paleopolyploid soybean. *Plant Cell* 26, 996–1008. doi: 10.1105/tpc.114.122739
- Sierro, N., Battey, J. N., Ouadi, S., Bakaher, N., Bovet, L., Willig, A., et al. (2014). The tobacco genome sequence and its comparison with those of tomato and potato. *Nat. Commun.* 5:3833. doi: 10.1038/ncomms4833
- Simpson, C. G., Jennings, S. N., Clark, G. P., Thow, G., and Brown, J. W. (2004). Dual functionality of a plant U-rich intronic sequence element. *Plant J.* 37, 82–91. doi: 10.1046/j.1365-313x.2003.01941.x
- Slotte, T., Hazzouri, K. M., Agren, J. A., Koenig, D., Maumus, F., Guo, Y. L., et al. (2013). The *Capsella rubella* genome and the genomic consequences of rapid mating system evolution. *Nat. Genet.* 45, 831–835. doi: 10.1038/ng.2669
- Smith, C. W., Chu, T. T., and Nadal-Ginard, B. (1993). Scanning and competition between AGs are involved in 3' splice site selection in mammalian introns. *Mol. Cell Biol.* 13, 4939–4952. doi: 10.1128/mcb.13.8.4939
- Smith, C. W., Porro, E. B., Patton, J. G., and Nadal-Ginard, B. (1989). Scanning from an independently specified branch point defines the 3' splice site of mammalian introns. *Nature* 342, 243–247. doi: 10.1038/342243a0
- Sorek, R., Ast, G., and Graur, D. (2002). Alu-containing exons are alternatively spliced. *Genome Res.* 12, 1060–1067. doi: 10.1101/gr.229302
- Staiger, D., and Brown, J. W. (2013). Alternative splicing at the intersection of biological timing, development, and stress responses. *Plant Cell* 25, 3640–3656. doi: 10.1105/tpc.113.113803
- Su, Z., Wang, J., Yu, J., Huang, X., and Gu, X. (2006). Evolution of alternative splicing after gene duplication. *Genome Res.* 16, 182–189. doi: 10.1101/gr.4197006

- Szcześniak, M. W., Kabza, M., Pokrzywa, R., Gudyś, A., and Makalowska, I. (2013). ERISdb: a database of plant splice sites and splicing signals. *Plant Cell Physiol.* 54:e10. doi: 10.1093/pcp/pct001
- Thomas, J., Palusa, S. G., Prasad, K. V., Ali, G. S., Surabhi, G. K., Ben-Hur, A., et al. (2012). Identification of an intronic splicing regulatory element involved in auto-regulation of alternative splicing of SCL33 pre-mRNA. *Plant J.* 72, 935–946. doi: 10.1111/tpj.12004
- Tomato Genome Consortium (2012). The tomato genome sequence provides insights into fleshy fruit evolution. *Nature* 485, 635–641. doi: 10.1038/nature11119
- Trapnell, C., Pachter, L., and Salzberg, S. L. (2009). TopHat: discovering splice junctions with RNA-Seq. *Bioinformatics* 25, 1105–1111. doi: 10.1093/bioinformatics/btp120
- Trapnell, C., Roberts, A., Goff, L., Pertea, G., Kim, D., Kelley, D. R., et al. (2012). Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and Cufflinks. *Nat. Protoc.* 7, 562–578. doi: 10.1038/nprot.2012.016
- Vogan, K. J., Underhill, D. A., and Gros, P. (1996). An alternative splicing event in the Pax-3 paired domain identifies the linker region as a key determinant of paired domain DNA-binding activity. *Mol. Cell Biol.* 16, 6677–6686. doi: 10.1128/mcb.16.12.6677
- Wagner, G. P., Kin, K., and Lynch, V. J. (2012). Measurement of mRNA abundance using RNA-seq data: RPKM measure is inconsistent among samples. *Theory Biosci.* 131, 281–285. doi: 10.1007/s12064-012-0162-3
- Wang, B., Tseng, E., Regulski, M., Clark, T. A., Hon, T., Jiao, Y., et al. (2016). Unveiling the complexity of the maize transcriptome by single-molecule long-read sequencing. *Nat. Commun.* 7:11708. doi: 10.1038/ncomms11708
- Wang, X. T., Hu, L. J., Wang, X. F., Li, N., Xu, C. M., Gong, L., et al. (2016). DNA methylation affects gene alternative splicing in plants: an example from rice. *Mol. Plant* 9, 305–307. doi: 10.1016/j.molp.2015.09.016
- 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. doi: 10.1186/gb-2004-5-12-r102
- Warf, M. B., and Berglund, J. A. (2010). Role of RNA structure in regulating pre-mRNA splicing. *Trends Biochem. Sci.* 35, 169–178. doi: 10.1016/j.tibs.2009.10.004
- Weischenfeldt, J., Waage, J., Tian, G., Zhao, J., Damgaard, I., Jakobsen, J. S., et al. (2012). Mammalian tissues defective in nonsense-mediated mRNA decay display highly aberrant splicing patterns. *Genome Biol.* 13:R35. doi: 10.1186/gb-2012-13-5-r35
- Xu, Q., Modrek, B., and Lee, C. (2002). Genome-wide detection of tissue-specific alternative splicing in the human transcriptome. *Nucleic Acids Res.* 30, 3754–3766. doi: 10.1093/nar/gkf492
- Xu, S., Brockmoller, T., Navarro-Quezada, A., Kuhl, H., Gase, K., Ling, Z., et al. (2017). Wild tobacco genomes reveal the evolution of nicotine biosynthesis. *Proc. Natl. Acad. Sci. U.S.A.* 114, 6133–6138. doi: 10.1073/pnas.1700073114
- Yang, R., and Wang, X. (2013). Organ evolution in angiosperms driven by correlated divergences of gene sequences and expression patterns. *Plant Cell* 25, 71–82. doi: 10.1105/tpc.112.106716
- Yoshimura, K., Yabuta, Y., Ishikawa, T., and Shigeoka, S. (2002). Identification of a cis element for tissue-specific alternative splicing of chloroplast ascorbate peroxidase pre-mRNA in higher plants. *J. Biol. Chem.* 277, 40623–40632. doi: 10.1074/jbc.M201531200
- Zhang, X. T., Rosen, B. D., Tang, H. B., Krishnakumar, V., and Town, C. D. (2015). Polyribosomal RNA-Seq reveals the decreased complexity and diversity of the *Arabidopsis* translome. *PLoS One* 10:e0117699. doi: 10.1371/journal.pone.0117699

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2019 Ling, Brockmoller, Baldwin and Xu. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



Alternative Splicing and Protein Diversity: Plants Versus Animals

Saurabh Chaudhary^{1†}, Waqas Khokhar^{1†}, Ibtissam Jabre^{1†}, Anireddy S. N. Reddy², Lee J. Byrne¹, Cornelia M. Wilson¹ and Naeem H. Syed^{1*}

¹ School of Human and Life Sciences, Canterbury Christ Church University, Canterbury, United Kingdom, ² Department of Biology and Program in Cell and Molecular Biology, Colorado State University, Fort Collins, CO, United States

OPEN ACCESS

Edited by:

Ezequiel Pettillo,
CONICET Instituto de Fisiología,
Biología Molecular y Neurociencias
(IFIBYNE), Argentina

Reviewed by:

Federico Damian Ariel,
Instituto de Agrobiotecnología del
Litoral (IAL), Argentina
John William Slessor Brown,
University of Dundee, United Kingdom

*Correspondence:

Naeem H. Syed
naeem.syed@canterbury.ac.uk

[†]These authors have contributed
equally to this work

Specialty section:

This article was submitted to
Plant Physiology,
a section of the journal
Frontiers in Plant Science

Received: 08 March 2019

Accepted: 13 May 2019

Published: 12 June 2019

Citation:

Chaudhary S, Khokhar W, Jabre I,
Reddy ASN, Byrne LJ, Wilson CM
and Syed NH (2019) Alternative
Splicing and Protein Diversity: Plants
Versus Animals.
Front. Plant Sci. 10:708.
doi: 10.3389/fpls.2019.00708

Plants, unlike animals, exhibit a very high degree of plasticity in their growth and development and employ diverse strategies to cope with the variations during diurnal cycles and stressful conditions. Plants and animals, despite their remarkable morphological and physiological differences, share many basic cellular processes and regulatory mechanisms. Alternative splicing (AS) is one such gene regulatory mechanism that modulates gene expression in multiple ways. It is now well established that AS is prevalent in all multicellular eukaryotes including plants and humans. Emerging evidence indicates that in plants, as in animals, transcription and splicing are coupled. Here, we reviewed recent evidence in support of co-transcriptional splicing in plants and highlighted similarities and differences between plants and humans. An unsettled question in the field of AS is the extent to which splice isoforms contribute to protein diversity. To take a critical look at this question, we presented a comprehensive summary of the current status of research in this area in both plants and humans, discussed limitations with the currently used approaches and suggested improvements to current methods and alternative approaches. We end with a discussion on the potential role of epigenetic modifications and chromatin state in splicing memory in plants primed with stresses.

Keywords: alternative splicing, co-transcriptional splicing, protein diversity, intron retention, NMD, splicing memory, epigenetic modifications

INTRODUCTION

Plants have evolved various developmental and physiological strategies to control daily activities that respond to variable and extreme environmental conditions (Gratani, 2014; Becklin et al., 2016). To maximize efficiency under diverse conditions, the crosstalk between multiple layers of gene regulation including co-transcriptional, post-transcriptional, and post-translational regulation is crucial for plants (Reddy et al., 2013; Guerra et al., 2015; Skelly et al., 2016). Alternative splicing (AS) is one such mechanism, which is widespread in plants and humans, generates two or more mRNAs from the same precursor-mRNA (pre-mRNA) and is thought to significantly contribute toward protein diversity (Nilsen and Graveley, 2010; Syed et al., 2012; Reddy et al., 2013). The basic mechanism of AS in higher eukaryotes is similar, however, some differences in gene architecture, splicing and transcription machinery between plants and animals suggest plant-specific regulation of AS (Kornblihtt et al., 2013; Irimia and Roy, 2014; Wang et al., 2014).

The advances in next-generation sequencing (NGS) technology and omics approaches in plants have revealed that up to 70% of multi-exon genes undergo AS (Filichkin et al., 2010; Lu et al., 2010; Marquez et al., 2012; Shen et al., 2014; Thatcher et al., 2014; Chamala et al., 2015; Zhang et al., 2017). Among all AS events, intron retention (IR) is the predominant mode of AS in plants (Filichkin et al., 2010; Kalyna et al., 2012; Drechsel et al., 2013), whereas exon-skipping (ES) is the major type in humans (**Figure 1**) (Sammeth et al., 2008; Wang et al., 2008). Interestingly, IR generates mostly non-sense mRNAs harboring premature terminal codons (PTC+) and are either degraded by the non-sense-mediated mRNA decay (NMD) pathway, or escape NMD to produce truncated proteins, thereby regulating the function and abundance of their full-length counterparts (Filichkin and Mockler, 2012; Kalyna et al., 2012; Drechsel et al., 2013; Filichkin S.A. et al., 2015). The NMD pathway is a post-transcriptional mRNA quality control mechanism which acts to degrade PTC+ mRNAs. Some studies suggest alternative roles for transcripts with IR, which are either sequestered in the nucleus and released on demand (Filichkin S.A. et al., 2015; Filichkin S. et al., 2015) or function as protein-coding introns known as exintrons (**Figure 1**), a new class of retained introns with some features of exons (Marquez et al., 2015; Staiger and Simpson, 2015).

Plants modulate their gene expression patterns via AS coupled to NMD during different developmental stages, abiotic and/or biotic stresses and the circadian clock function (James et al., 2012; Kalyna et al., 2012; Drechsel et al., 2013; Kwon et al., 2014; Filichkin S.A. et al., 2015; Sureshkumar et al., 2016). Stressful conditions control not only the ratios but the timing of both sense and non-sense AS transcripts (Filichkin S.A. et al., 2015; Filichkin et al., 2018). However, it is unclear how environmental signals modulate splicing ratios and timing to help plants acclimate to such stresses in the short and long term. Furthermore, it is largely unknown to what extent AS transcripts are recruited for translation to be functionally significant at the proteomic level in plants.

Alternative splicing regulates essential functions in humans such as autophagy, apoptosis, protein localization, enzymatic activities and interaction with ligands, transcription factors activity and mRNA abundance, etc. (Kelemen et al., 2013; Paronetto et al., 2016; Gallego-Paez et al., 2017). Hence, it is not surprising that any aberrant or dysregulation in AS can cause several human diseases including cancer, neurological disorders, heart, and skeletal muscle abnormalities, and multiple genetic disorders (Matlin et al., 2005; Poulos et al., 2011; Kelemen et al., 2013; Sveen et al., 2016). Recent transcriptome (RNA-Seq), translome (ribosomal foot-printing), and proteome data have shown a significant contribution of AS toward protein diversity in humans (Weatheritt et al., 2016; Liu et al., 2017). On the other hand, some proteomic studies suggest that AS may not significantly contribute to protein diversity and only single dominant isoforms are represented at the protein level for most of the protein-coding genes (Ezkurdia et al., 2015; Tress et al., 2017a). Apparently, these contradictions stem from the lower depth and limitations of mass spectrometry (MS) techniques to detect changes in protein domains as a result of AS

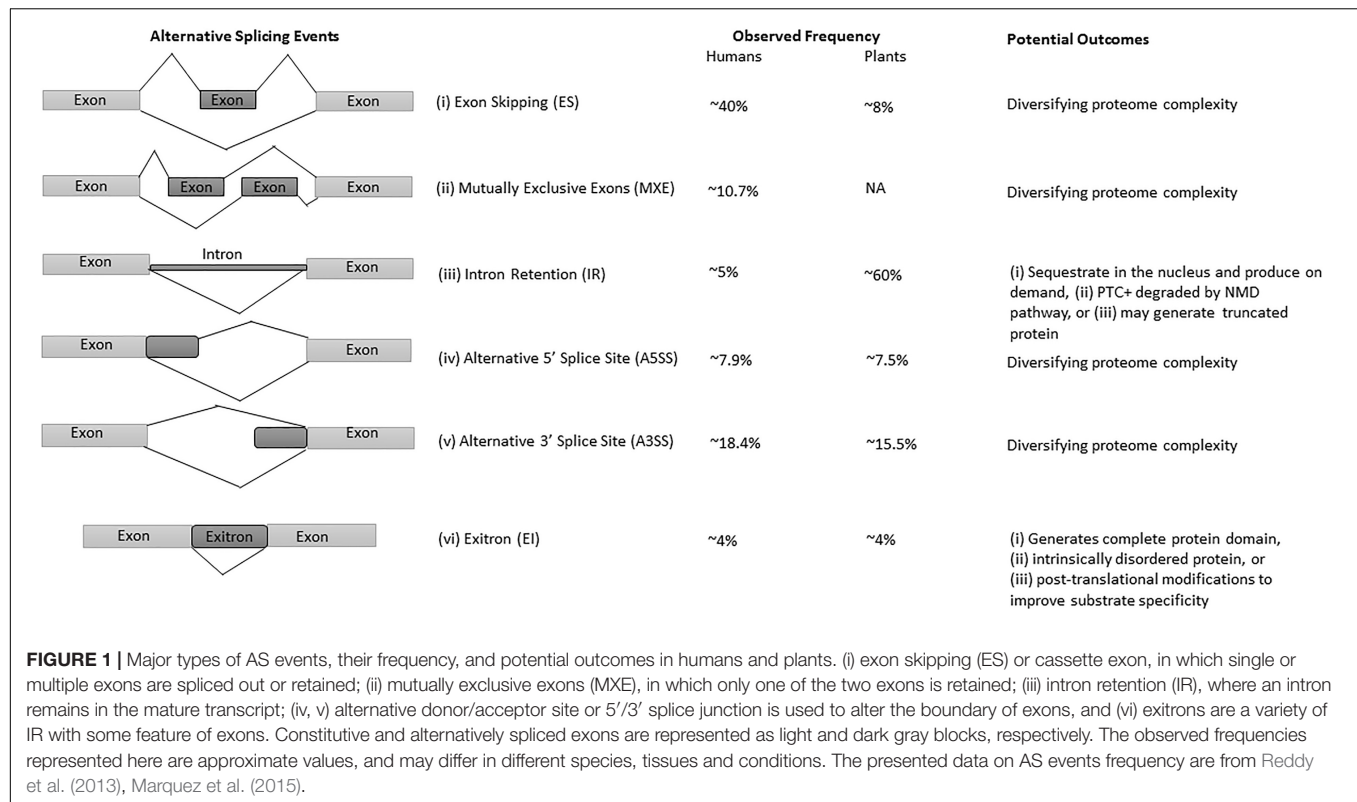
(Wang et al., 2018; Chaudhary et al., 2019). In this review, basic differences in the mechanism of AS and its contribution toward protein diversity in plants and humans are discussed. We also discuss some emerging aspects of IR, NMD pathway, chromatin structure, and splicing memory in plants.

COUPLING OF TRANSCRIPTION AND SPLICING IN PLANTS AND HUMANS

Plant spliceosome machinery is not well characterized due to the unavailability of *in vitro* systems. However, in a recent study, an attempt has been made to develop an *in vitro* pre-mRNA splicing assay using plant nuclear extracts, and it may help to delineate and characterize components of the plant spliceosome machinery (Albaqami and Reddy, 2018). Sequence similarity based analyses suggest conserved regulation of AS in higher eukaryotes. Briefly, splicing is carried out by the spliceosome, which consists of five small nuclear ribonucleoprotein particles (snRNPs) designated as U1, U2, U4, U5, and U6 and additional spliceosome-associated non-snRNP proteins (Will and Lührmann, 2011; Matera and Wang, 2014; Wang et al., 2014). The *cis*-acting elements present on pre-mRNA include 5' splice sites (5'SS), 3' splice sites (3'SS), polypyrimidine tracts (PPT) and branch point sequences, which are recognized by the *trans*-acting factors such as splicing factors (SFs) mainly SR proteins and hnRNPs. The *trans*-acting SFs and *cis*-regulatory elements guide and modulate the spliceosome to recognize differential splice sites present on pre-mRNA (Koncz et al., 2012; Reddy et al., 2013; Chen and Moore, 2015). The details on the assembly of the spliceosome and regulation of AS has been reviewed extensively and readers are referred to excellent articles on this topic (Will and Lührmann, 2011; Reddy et al., 2013; Chen and Moore, 2015).

Recent evidence from metazoans indicates that the process of splicing is largely co-transcriptional (Shukla and Oberdoerffer, 2012; Brugiolo et al., 2013; Merkhofer et al., 2014). Extensive studies in animals and emerging data in plants show that the splicing process for the majority of genes is predominantly co-transcriptional in nature (**Figure 2**) (Nojima et al., 2015, 2018; Wang et al., 2016; Wong et al., 2017; Zhu et al., 2018; Jabre et al., 2019). The co-transcriptional behavior of splicing means that the chromatin environment such as methylation status, histone modifications, nucleosome occupancy and RNA Polymerase II (RNAPII) processivity has a strong influence on splicing outcomes (Listerman et al., 2006; Khodor et al., 2011; Pajoro et al., 2017; Jabre et al., 2019).

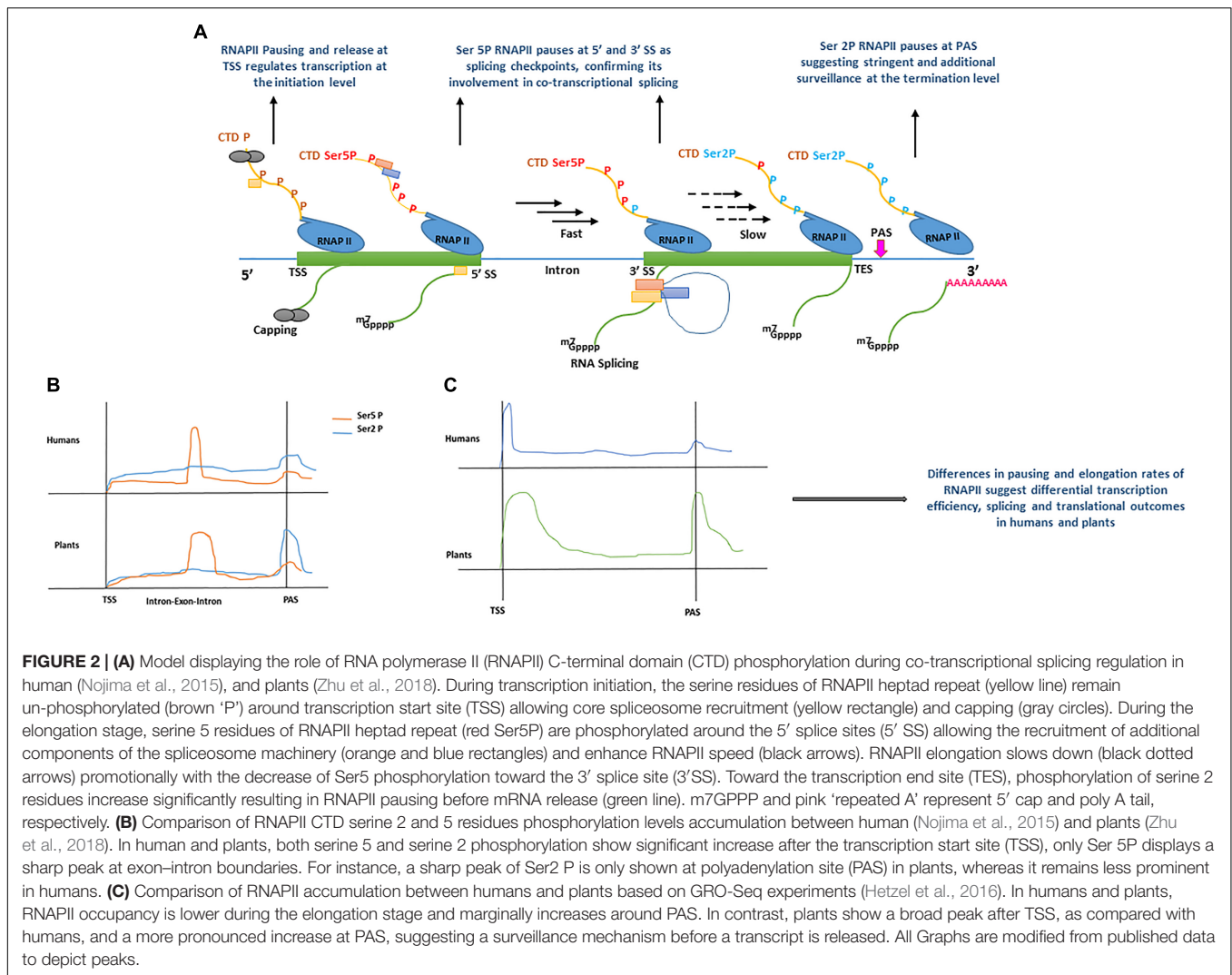
Interestingly, long non-coding RNAs (lncRNAs) can also influence the splicing dynamics of their target genes either directly and/or after processing into short interfering or micro RNAs (Romero-Barrios et al., 2018). Non-coding RNAs can affect AS via modulating chromatin structure (Luco et al., 2011; Romero-Barrios et al., 2018), splicing factor recruitment and altering the phosphorylation status of spliceosomal proteins (Misteli et al., 1998; Romero-Barrios et al., 2018). Circular RNAs which are generated by the so-called non-canonical "backsplicing" of pre-mRNAs are known to regulate AS in



animals and examples from plants are beginning to emerge as well. CircRNAs could make DNA:RNA hybrids with the genomic DNA to generate the so-called R-loop. Indeed, a circRNA derived from exon 6 of the *SEPALLATA3* (*SEP3*) gene forms an R-loop via direct interaction with the *SEP3* locus (Conn et al., 2017). The R-loop formation around exon 6 of the *SEP3* gene results in skipping of this exon and affects petal and stamen number in *Arabidopsis* (Conn et al., 2017).

Plant promoters are largely devoid of nucleosomes, as a result of lower GC content (high AT enrichment) as compared with humans (Narang et al., 2005; Yang et al., 2007; Hetzel et al., 2016). Therefore, the dynamics of transcription initiation are fundamentally different between humans and plants (Hetzel et al., 2016). Depending upon the chromatin context in animals and plants, RNAPII is recruited at a promoter to form the pre-initiation complex (PIC), however, its processivity is inherently dependent on the chromatin structure along gene bodies and influences RNA-processing during transcription (Guo and Price, 2013; Grasser and Grasser, 2018; Jabre et al., 2019). Techniques such as native elongation transcript sequencing (NET-Seq) (Churchman and Weissman, 2011) in mammals (mNET-Seq) (Nojima et al., 2015) and plants (pNET-Seq) (Zhu et al., 2018) and global run-on sequencing (GRO-seq) (Hetzel et al., 2016), have revealed some important aspects of RNAPII elongation and structural features during transcription and RNA-processing, in humans and plants, respectively. The carboxyl-terminal domain (CTD) of the largest subunit of RNAPII contains a heptad repeat "Tyr1-Ser2-Pro3-Thr4-Ser5-Pro6-Ser7." The

Ser2 and Ser5 of this heptad repeat undergoes phosphorylation and plays a key role in the coordination of transcription and other RNA processing activities (Harlen and Churchman, 2017). In mNET-Seq, phosphorylation-specific antibodies were used to study immunoprecipitated RNAPII transcripts in humans (Nojima et al., 2015, 2018). The comparative analysis of un-phosphorylated (unph) or low-phosphorylated and phosphorylated CTD of RNAPII revealed the accumulation of different forms at differential positions on protein-coding genes. For instance, the RNAPII unph-CTD shows a peak at the transcription start site (TSS), whereas RNAPII Ser5P CTD accumulates at the 5'SS of exon-intron boundaries and its density reduces as the RNAPII elongation proceeds downstream toward the 3' end of the intron (**Figure 2A**) (Nojima et al., 2015, 2018). Similarly, RNAPII Ser2P CTD spreads over gene bodies (GB) and shows accumulation at the transcription end site (TES) (**Figure 2A**) (Nojima et al., 2015, 2018). Moreover, genes that undergo co-transcriptional splicing, such as *TARS* in humans, show a major peak of RNAPII Ser5P CTD at 5'SS, suggesting pausing at the exon to allow time for the spliceosome to catalyze the first splicing reaction (Nojima et al., 2015). Similar to humans, the dynamics of RNAPII in plants is also established during transcription (Erhard et al., 2015; Hetzel et al., 2016; Zhu et al., 2018). As shown in the proposed model of co-transcriptional splicing in **Figure 2A**, plants RNAPII CTD is phosphorylated as transcription proceeds. However, in both humans and plants, unph RNAPII is recruited at the promoter region to form the PIC. After initiation, phosphorylation of RNAPII Ser5 CTD and Ser2 CTD begins as transcription



proceeds toward the 3' end. The RNAPII Ser5P CTD pauses at 5'SS, whereas RNAPII Ser2P CTD shows accumulation immediately after polyadenylation site (PAS), suggesting their role in splicing and transcription termination, respectively (Nojima et al., 2015; Zhu et al., 2018).

Despite similarities in the dynamics of RNAPII during transcription and co-transcriptional splicing among plants and humans, significant differences have also been reported, suggesting species-specific regulation of transcription and splicing (Hetzel et al., 2016). For instance, the engaged RNAPII profiles suggest, promoter-proximal pausing and divergent transcripts in *Arabidopsis* and maize are absent, whereas, these are prominent features of the human transcription (Core et al., 2008; Preker et al., 2008; Erhard et al., 2015; Zhu et al., 2015; Hetzel et al., 2016). In plants, the lack of promoter-proximal-pausing and a high correlation between transcription and steady-state RNA suggests initiation level regulation of transcription as compared to humans (Hetzel et al., 2016). In contrast to GRO-Seq analysis, the combination of GRO-Seq and pNET-Seq data in *Arabidopsis* show that

RNAPII pauses or slows down in some genes after initiation of transcription (Zhu et al., 2018). However, unlike humans, which show RNAPII pausing in narrow regions (20–25 nt), plant RNAPII pausing in the promoter-proximal-regions is much broader (Figure 2C) (Zhu et al., 2018). Additionally, a strong positive correlation has been observed between RNAPII pausing at PAS, CpG methylation and longer genes in plants than in humans, which further suggests plant-specific regulation of transcription and splicing regulation (Hetzel et al., 2016).

Many features of transcription are conserved between humans and plants, however, some important differences exist between them. For example, there is a higher RNAPII elongation rate and AS in the presence of light than dark, demonstrating coupling between AS transcription and growth conditions, which is an important mechanism for plants to respond to different environmental conditions (Petrillo et al., 2014; Godoy Herz et al., 2019). Thus, the role of RNAP II processivity and its impact on AS needs to be analyzed in a tissue- and condition-dependent manner in plants. In the last

decade, significant progress has been made to understand the co-transcriptional behavior of splicing/AS in animals, and yeast systems (Shukla and Oberdoerffer, 2012; Merkhofer et al., 2014; Saldi et al., 2016). However, this area is relatively new in plants and more studies are required to illuminate the co-transcriptional dynamics and its impact on RNA processing in tissue- and condition-specific manner.

ASPECTS OF IR AND NMD IN PLANTS AND HUMANS

Intron retention is the most prevalent AS event in plants with observed frequencies between 28% to as high as 64% (**Figure 1**) depending upon growth condition, tissue type and the coverage of transcriptome data (Filichkin et al., 2010; Kalyna et al., 2012; Marquez et al., 2012; Mandadi and Scholthof, 2015). In comparison with plants, only 5% of IR events were observed in humans (**Figure 1**) (Keren et al., 2010; Reddy et al., 2012), owing to the large size of animal introns, sequencing depth and bioinformatics challenges to detect them. As a consequence, IR had received limited interest in humans until recently (Wong et al., 2013; Braunschweig et al., 2014; Boutz et al., 2015), whereas in plants IR has been found to be an important regulator in growth, development, physiology, and stress responses (Kalyna et al., 2012; Syed et al., 2012; Drechsel et al., 2013; Filichkin S.A. et al., 2015). However, recent research is unveiling various menage regulatory functions of IR in humans. For example, in addition to physiologically regulated events, any mutation in the splice site or splicing regulatory sequences cause aberrant IR, which further results in perturbed splicing patterns and potentially cause diseases (Jung et al., 2015; Wong et al., 2015; Jacob and Smith, 2017).

In humans, possible causes of IR and its abundance in response to cell differentiation and stresses have been studied recently (Wong et al., 2013; Braunschweig et al., 2014; Boutz et al., 2015). For instance, to predict the prevalence of IR, and their regulation and biological significance, a deep quantitative survey using Poly(A⁺) RNA-Seq data from 40 human and mouse tissue samples was conducted (Braunschweig et al., 2014). This study involved the quantitative measurement and comparison of reads across unspliced (exon–intron) and spliced (exon–exon) junctions, as well as, reads within introns in terms of “percent intron retention” (PIR) (Braunschweig et al., 2014). These findings suggest a large number of multiexonic genes are affected by the variable frequency of IR events processed in different tissues, which is much higher in comparison with previously estimated values (Pan et al., 2008; Wang et al., 2008). Comparative analysis across various species revealed tissue-specific IR events in neurons and immune cells. Furthermore, IR in neurons is highly conserved as compared with other AS events (Barbosa-Morais et al., 2012; Merkin et al., 2012; Braunschweig et al., 2014). In contrast with previous studies, IR was prevalent and mainly enriched in untranslated regions (UTRs), non-coding RNAs, depleted protein coding regions, and/or at the 3′ end of RNAs among

different tissues in humans (Bicknell et al., 2012; Jacob and Smith, 2017). Moreover, the frequency of IR in the nucleus was observed to be higher than the cytoplasm, suggesting nuclear sequestration or coupling with the NMD pathway (Wong et al., 2013; Braunschweig et al., 2014; Boutz et al., 2015; Edwards et al., 2016).

In comparison with humans, the prevalence and significance of IR in plants and its role in development, stress and tissue-specific physiology are well documented. The observed frequency of IR in plants is as high as 64%, and potentially fine-tunes the transcriptome functionality (Filichkin et al., 2010, 2018; Kalyna et al., 2012; Drechsel et al., 2013; Filichkin S.A. et al., 2015). However, the mechanisms behind the high occurrence of IR in plants are still not very clear, yet many studies emphasize its significance in plants under normal, stress and various development and growth conditions. For example, the expression of *INTERMINATE DOMAIN 14 (IDD 14)* isoforms controlled via IR mediate starch accumulation and utilization under cold stress in *Arabidopsis* (Seo et al., 2011). Similarly, cold-dependent IR in clock genes such as *CIRCADIAN CLOCK ASSOCIATED 1 (CCA1)*, *LATE ELONGATED HYPOCOTYL (LHY)* and *PSEUDO-RESPONSE REGULATOR7 (PRR7)*, modulate their transcript and protein abundance for *CCA1* (Seo et al., 2011; James et al., 2012). In wheat, the *PECTIN METHYL ESTERASE INHIBITOR (PMEI)*, which secretes pectin for the cell wall, is also regulated by IR. Although, *PMEI* IR isoforms are found in almost all tissues but only anthers contained mature transcripts without IR, suggesting possible tissue-specific functionality of these transcripts (Rocchi et al., 2012). Similarly, studies in *Marsilea vestita* (Boothby et al., 2013) and *Arabidopsis* (Filichkin and Mockler, 2012; Filichkin S.A. et al., 2015) provide a useful model to explain unproductive AS via IR. It has been demonstrated in *M. vestita* that some NMD insensitive IR transcripts remain in the nucleus as un-spliced mRNAs. Subsequently, these IR transcripts could be spliced and their translation results in a specific function, such as gamete development (Boothby et al., 2013).

Interestingly, many of IR PTC⁺ transcripts are not subjected to NMD in plants (Kalyna et al., 2012), suggesting regulatory functions. Components of the NMD machinery are highly conserved between plants and humans and its efficiency is strongly influenced by the pioneer round of translation (activity of ribosomes) (Shaul, 2015). However, it is intriguing that NMD responses are much less pronounced under stressful conditions in humans and plants, affecting the expression and translation of stress-responsive genes and splice variants (Trcek et al., 2013; Shaul, 2015). For example, inhibition of NMD mediates plant defense response during pathogen attack in *Arabidopsis* NMD mutants as they constitutively make more salicylic acid (SA) and show a heightened response after infection with *Pseudomonas syringae* (Rayson et al., 2012). However, mechanistic details of AS and its role via protein diversity in subverting a pathogen attack is not clear. Since the NMD pathway is translation dependent, slow engagement of different non-canonical transcripts with the ribosomal machinery may be the cause of their degradation. Intriguingly, in several model

species including *Arabidopsis*, PTCs in the first and last intron appear earlier in their sequence than expected by chance alone, to keep the metabolic cost of producing truncated proteins and their subsequent degradation (Behringer and Hall, 2016). This data supports the notion that the appearance of earlier PTCs in introns seems to be favored by selection. Presence of PTCs in the first and last introns also points toward multiple features favoring degradation of non-sense transcripts (Behringer and Hall, 2016).

Interestingly, introns in plants UTRs also play a crucial role by affecting translation efficiency via a process called intron-mediated enhancement (IME). IME was proposed as a conserved phenomenon enhancing the translation efficiency of IR transcripts (Parra et al., 2011; Gallegos and Rose, 2015). For example, analysis of 5' UTR introns identified an intron element in transcripts of the Mg^{2+}/H^{+} ion exchange (MHX) gene in *Arabidopsis*, which further show an increase in translation efficiency (Akua and Shaul, 2013). In summary, differences in the frequencies of IR events suggest a varied mode of downstream processing and fates of IR transcripts in plants and humans. However, further work is needed to illuminate the mechanistic details of the IME mechanism.

AS AND PROTEIN DIVERSITY IN HUMANS: SUPPORTING EVIDENCE

Higher eukaryotes are diverse with varying degrees of biological complexity, nonetheless, the number of protein-coding genes is comparable between different species (Chen et al., 2014). Comparative sequencing and evolutionary studies between different eukaryotic species (including complex avian and mammals to species with fewer cell types) suggest a strong correlation between AS and organism complexity (Chen et al., 2014). AS plays a crucial role to enrich the expression of many genes and mediates various biological functions, pathways, and processes (Merkin et al., 2012; Weatheritt et al., 2012; Irimia et al., 2014). In humans, despite significant advancements in the field of transcriptome and proteome analysis techniques, the extent to which AS transcripts contribute to protein diversity remains unclear. However, renewed interest in humans has led to concerted efforts to illuminate this phenomenon in the recent past (Table 1). For example, isolation and sequencing of ribosome-bound transcripts have enabled researchers to delineate how the variety and abundance of mRNAs correlate with ribosomal recruitment (potentially translating mRNA). In a recent study, the ribosomal-engaged landscape of AS transcripts was surveyed using ribosomal-profiling in humans (Weatheritt et al., 2016). The ribosomal profiling data suggest transcripts with exon skipping events are present in medium to high abundance and thus likely to be translated. On the contrary, transcripts present in low abundance at the transcriptome level were not engaged with the ribosomes. This might be due to either the presence of introns in the low abundance transcripts, which remain in the nucleus (Braunschweig et al., 2014; Boutz et al., 2015) or incomplete RNA processing, preventing ribosomal

engagement (Weatheritt et al., 2016). Similarly, other studies using Frac-Seq (subcellular fractionation and RNA-sequencing) (Sterne-Weiler et al., 2013) and TrIP-Seq (transcript isoforms in polysomes sequencing) (Floor and Doudna, 2016), also detected a large proportion of splice variants in the polyribosome fractions suggesting spliced isoforms play a significant role in controlling protein output in human cells. However, the degree to which ribosomal bound AS transcripts are translated and represented at the protein level is unclear. For example, pre-mRNA processing in the nucleus influences an isoform's association with polyribosomes (Sterne-Weiler et al., 2013). Approximately 30% of mRNA processing events are differentially partitioned between cytoplasmic and polyribosome fractions (Sterne-Weiler et al., 2013). Moreover, differences in the polyribosome association are the result of a change in the *cis*-regulatory landscapes such as inclusion or exclusion of uORFs and Alu-elements in the 5'UTR, and microRNA target sites in the 3'UTR by AS (Sterne-Weiler et al., 2013). Similarly, TrIP-Seq analysis revealed that each transcript isoform harbors special regulatory features controlling ribosome occupancy and translation (Floor and Doudna, 2016). Floor and Doudna (2016) found robust translational control by 5' UTRs between cell lines, whereas 3' UTRs impact cell type-specific expression. This work also suggested that transcript isoform diversity must be considered when associating RNA and protein levels.

Some proteomic studies contradict ribosome profiling data and argue that only a small fraction of splice variants are represented at the protein level (Abascal et al., 2015; Ezkurdia et al., 2015; Tress et al., 2017a). Furthermore, the shotgun MS techniques used in many proteomic studies have their own limitations of coverage and sensitivity to detect low abundance splice variants at the protein level (Bensimon et al., 2012; Rost et al., 2015). To improve isoforms detection efficiency, alternative approaches need to be developed to overcome the limitations of the techniques used at present. Toward this goal, full-length ORFs of AS isoforms from a large number of human genes were cloned and protein-protein interaction (PPI) profiling was performed to demonstrate the functionality of hundreds of protein isoforms (Yang et al., 2016). This study demonstrated vastly different interaction profiles among isoforms as a result of AS. Strikingly, the isoforms encoded by the same genes exhibit widespread functional differences in the PPI network analysis. Since differences between protein isoforms are as high as observed between different genes, isoforms-specific partners could have different expression and functional characteristics. Yang et al. (2016) proposed that a vast diversity of "functional alloforms" are generated that contribute to different physiological and developmental processes (Yang et al., 2016).

In humans, a number of studies have been conducted to identify protein isoforms that result from AS by comparing transcriptome and proteome data (Brosch et al., 2011; Ezkurdia et al., 2012; Lopez-Casado et al., 2012; Sheynkman et al., 2013). However, most of these studies were carried out in a steady state manner and do not explain the consequences of perturbation in splicing to protein diversity. To overcome these limitations, an integrated approach was developed to illuminate

TABLE 1 | Major studies deciphering the role of AS in protein diversity in humans and plants using different technique.

Study	Organism	Major technique used	Conclusion	References
Stochastic noise in splicing machinery	Humans	Computational analysis	Most AS is a consequence of stochastic noise in the splicing machinery, and has no functional significance	Melamud and Moul, 2009
Assessing the contribution of alternative splicing to proteome diversity	Plants	Computational analysis	AS contributes to transcriptome diversity but its contribution to protein diversity is limited	Severing et al., 2009
Isoform-specific recruitment to polyribosomes	Humans	Frac-Seq	Addition to translation AS plays role in sequestration and mRNA-decay	Sterne-Weiler et al., 2013
Assess the role of AS in proteome diversity	Humans	Computational analysis	Most genes have a single dominant isoform at the protein level, whereas homologous exons have important cellular roles	Abascal et al., 2015
Expression of protein coding gene isoforms at protein level	Humans	Computational analysis	Most highly expressed gene have single dominant isoform represented at the protein level	Ezkurdia et al., 2015
Ribosomal-engaged landscape of AS transcripts	Humans and mouse	Ribo-Seq	Majority of splice variants are translated into proteins	Weatheritt et al., 2016
Tunable protein synthesis by transcript isoforms	Humans	TriP-Seq	Alternatively spliced isoform levels effects translation output	Floor and Doudna, 2016
AS mediated expansion of protein interaction capabilities	Humans	ORF-Seq and PPI	Large number of alternative isoforms in the human proteome are “functional alloforms”	Yang et al., 2016
Transcriptome survey and contribution of AS toward proteome diversity	Plants	RNA-Seq and Ribo-Seq	AS increases protein complexity, however, its contribution is lower in plants as compared to humans	Yu et al., 2016
Impact of AS on human proteome	Humans	RNA-Seq and SWATH-MS	IR reduces the protein diversity but fine-tunes the human proteome functionality	Liu et al., 2017
Relationship between AS and protein complexity	Humans	Computational analysis	Majority of alternatively spliced transcripts may not be translated into proteins	Tress et al., 2017a
AS contribution in transcriptome and proteome diversity in <i>Physcomitrella patens</i>	Plants	RNA-Seq and qRT-PCR	AS has a small effect on proteome diversity but shapes the transcriptome	Fesenko et al., 2017

how variation in mRNA splicing patterns could subsequently change the proteome composition in a systematic manner (Liu et al., 2017). Selectively depleted spliceosome U5 component PRPF8 (Wickramasinghe et al., 2015) orchestrated changes at the transcriptome and proteome level that were determined using RNA-Seq and Sequential Window Acquisition of all Theoretical Spectra-Mass Spectrometry (SWATH-MS), respectively. After PRPF8 depletion, quantification of splice variants and a large fraction of proteome identified 1,542 proteins that displayed at least one peptide with altered expression. Functional annotation revealed that transcripts with altered splicing patterns possess similar cellular functions and processes (such as RNA splicing, the mitotic cell cycle and ubiquitination) as those found in proteins with altered levels. Thus, splicing variants at the transcriptomic level were found to be functionally represented at the protein level (Liu et al., 2017). Furthermore, to identify the differentially spliced event at the transcriptome level, the authors used a transcript-centric approach, in which a transcript is considered as a whole unit (Liu et al., 2017). Firstly, transcript expression is estimated, followed by identification of differentially used transcripts and expressed genes. The correlation analysis

between fold changes in the expression level after PRPF8 depletion suggests protein expression levels are exclusively associated with the alternatively spliced transcripts involving differential transcripts usage (DTU). Interestingly, IR events, which are considered as one of the major regulatory events for gene expression, had reduced representation at the protein level (Liu et al., 2017). Although, around 75% of multi-exon genes are affected by IR and help in regulating transcript levels (Braunschweig et al., 2014), its impact on protein expression is inverse because an increase in the level of IR transcripts, throughout the genome, is associated with PRPF8 depletion (Wickramasinghe et al., 2015). The peptide expression of 270 genes with retained introns showed downregulation of protein expression coded by genes with IR. Moreover, the relative abundance of transcripts also plays a significant role in protein expression as the low abundance transcripts with IR do not affect the protein expression until they are present in high abundance. These observations suggest IR reduces the protein diversity but fine-tunes the human proteome functionality. However, this finding may not be strictly applicable to plants as IR is the predominant mode of AS and may fine-tune the

proteome function via modulating its abundance, especially in stressful conditions.

Collectively, various studies in the recent past such as ribosomal profiling (Weatheritt et al., 2016), PPI interaction analysis of spliced isoforms (Yang et al., 2016), and integrative analysis using perturbed systems (Liu et al., 2017) suggest a strong correlation between AS and protein diversity in humans. Moreover, these studies provide an alternative to MS techniques, which have limitations of coverage and sensitivity to detect low level splice isoforms at the protein level and could be useful to study plant systems in the future.

AS AND PROTEIN DIVERSITY IN HUMANS: OPPOSING EVIDENCE

The contribution of AS toward protein diversity in humans is well documented (Weatheritt et al., 2016; Yang et al., 2016; Blencowe, 2017; Liu et al., 2017). However, recent data from some proteomic studies in humans supports the opposing view and suggest that AS may not be the key contributor to protein diversity (Tress et al., 2017a,b). Substantial amount of AS data has been generated in various RNA-Seq experiments in humans, however, most of the alternative isoforms in proteomic experiments are undetectable even in large-scale MS-based analyses (Ezkurdia et al., 2015; Tress et al., 2017a,b). Moreover, some studies suggest that AS is the result of noise in the splicing machinery and does not contribute to protein diversity as expected. For example, Melamud and Moulton (2009) proposed a stochastic noise model of splicing machinery, which explained that AS events arise as a result of noise in the splicing machinery (Melamud and Moulton, 2009). The idea of noise in the splicing machinery has also been supported by other studies as well, suggesting a large proportion of alternative isoforms are non-functional (Modrek et al., 2001; Kan et al., 2002; Neverov et al., 2005). Further, it was recently demonstrated that the majority of expressed genes have a single major isoform represented at the protein level (Abascal et al., 2015; Ezkurdia et al., 2015). This was supported by monitoring peptide evidence from eight large-scale MS experiments and observing that only one main protein isoform was dominant at the protein level from almost all coding genes (Tress et al., 2017a). On the other hand, several reports have supported the presence of a small number of alternative protein isoforms in humans (Tanner et al., 2007), *Drosophila* (Tress et al., 2008), and mouse (Brosch et al., 2011) in large-scale proteomic studies. However, AS events such as ES detected in RNA-Seq studies have revealed subtle effects on the structure and function of proteins. Tress et al. argue that it is the gene expression that is conserved across species, have strong tissue dependence, and are translated to detectable proteins but not the alternatively spliced isoforms (Tress et al., 2017a,b). Clearly, more work and evidence is needed to illuminate the relationship between AS and protein diversity in tissue- and condition-dependent manner.

The efficiency of the MS also needs to be enhanced because current MS techniques cannot reliably detect changes in protein domains as a result of AS (Wang et al., 2018; Chaudhary et al., 2019). For example, lysine and arginine coding triplets are the

most abundant amino acids at the end of exons or exon-exon junctions (Wang et al., 2018), and are the preferential sites for trypsin, which is the most common enzyme used in MS analyses (Olsen et al., 2004). Since trypsin digests exon-exon junctions, it hinders with the detection of novel AS derived peptides in MS-based proteome analysis (Ning and Nesvizhskii, 2010; Sheynkman et al., 2013; Wang et al., 2013). To improve efficiency, enzymes such as chymotrypsin can be used as an alternative to improve the detection of AS-derived peptides in proteome studies (Wang et al., 2018; Chaudhary et al., 2019).

THE CONTRIBUTION OF AS TOWARD PROTEIN DIVERSITY IN PLANTS

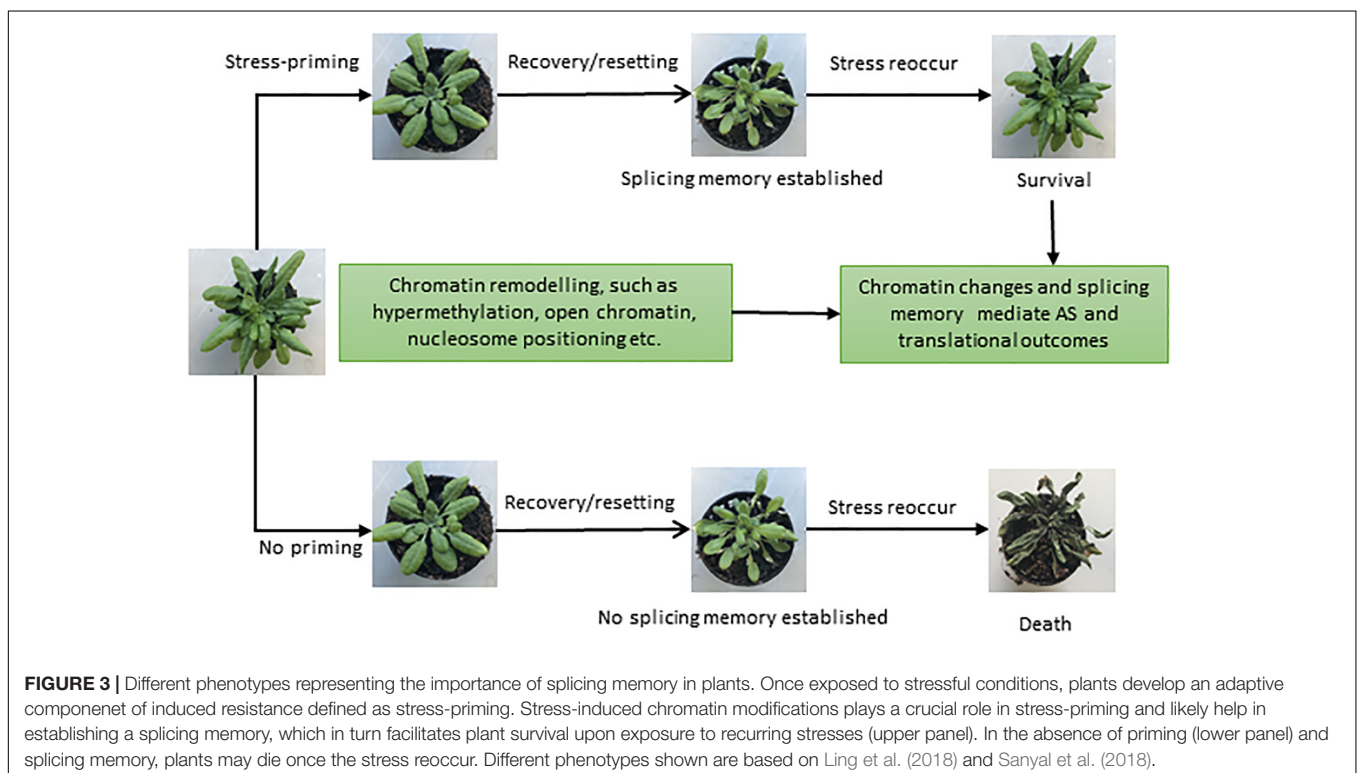
The role of AS in the expansion of functional protein diversity is less clear in plants as compared to humans (Kim et al., 2007). However, in the absence of in-depth proteomic studies to elucidate the role of AS toward protein diversity is tenuous. Recently, some studies have evaluated the influence of AS on protein diversity in plants. For example, hypoxia in *Arabidopsis* mediates an increase in the number of IR events in many mRNA isoforms, and show ribosomal engagement and potentially influence protein variety and abundance (Juntawong et al., 2014). Interestingly, transcriptome and translatome profiling among shoot apical meristem (SAM) and leaf domains, suggest 751 genes isoforms show domain-specific enrichment in the translatome data (Tian et al., 2019). Another study in *Arabidopsis* has shown that 35% of AS events are represented among the polysome-bound mRNAs and expected to undergo translation (Yu et al., 2016). Among all transcripts, IR is the least representative among translated transcripts, compared with untranslated transcripts, suggesting a variable role of IR in regulating transcript level via NMD machinery or sequestration in the nucleus and further processing on demand (Filichkin S.A. et al., 2015; Filichkin S. et al., 2015). In contrast, other splicing events such as ES, 5'AD, and 3'AA have higher proportions among transcripts that may be translated (Yu et al., 2016). Sequence analysis of translated transcripts suggests that any alteration in the CDS by AS could lead to a change in protein sequences (Yu et al., 2016). Interestingly, a large proportion of a new class of exon-like introns called exintrons (Marquez et al., 2015) (**Figure 1**) was found at the transcriptome as well as translatome level, suggesting these unique events of AS may contribute to protein diversity (Yu et al., 2016). A recent report in *Physcomitrella patens* suggests that AS shapes the transcriptome rather than the proteome (Fesenko et al., 2017), because only 85 isoform-specific peptides, representing only 25 differentially AS genes, were found in moss cells. Among all, only five genes unambiguously showed two or more protein isoforms from the same locus. The number of AS genes identified in this study was substantially large (approximately 66 times) as compared to proteomic datasets, nonetheless, only support a small contribution of AS on protein diversity. Collectively, these data support the view that AS increases protein complexity, however, its contribution is found to be lower as compared with humans (Yu et al., 2016). Further, supporting as well as

the opposing evidence presented above for the notion, “AS contributes toward protein diversity,” suggests that the exact number of splice isoforms represented at the proteome level in humans as well as in plants is still elusive. On the other hand, IR events are the predominant AS type in plants and may not be translated due to nuclear sequestration or degradation by the NMD pathway and thus remain poorly represented in MS experiments (Gohring et al., 2014; Hartmann et al., 2018). Since limited information is available at the proteome level, we envisage that strategies like cloning of spliced isoforms and PPI profiling (like in humans Yang et al., 2016), could be beneficial and may uncover different aspects of AS contribution toward protein diversity in plants.

SPLICING MEMORY AND PLANT STRESS TOLERANCE

Successful attempts have been made in plant systems to understand the impact of stress, its tolerance and the development of genetically engineered stress tolerant crops (Vinocur and Altman, 2005; Pereira, 2016). However, the majority of studies are restricted to acute and single stress only (Zhu, 2016). Since stresses are usually multiple, recurring and chronic, plants have evolved sophisticated defense mechanisms to deal with a variety of stresses. Plants have the ability to acquire tolerance to chronic stress through establishing “molecular stress memory” to confer tolerance through a phenomenon referred to as priming or acclimation, in response to previous exposure to a mild stress (Sani et al., 2013; Conrath et al., 2015;

Hilker et al., 2016). Priming establishes a new cellular state in plants, which is different from the naïve or unexposed plants (Sani et al., 2013; Conrath et al., 2015; Hilker et al., 2016). In recent years, it has become increasingly apparent that various epigenetic features, such as chromatin modifications, nucleosome positioning, and DNA methylation, are important components of adaptation and play a role in stress memory (Boyko et al., 2010; Ding et al., 2012; Lämke and Bäurle, 2017; Friedrich et al., 2018). Since the splicing process is largely co-transcriptional in nature, the chromatin structure has a strong influence on the transcriptional as well as the splicing processes (Listerman et al., 2006; Khodor et al., 2011; Jabre et al., 2019). Recent DNase I-Seq data suggest enrichment of IR in DNase I hypersensitivity sites (DHSS) in both Arabidopsis and rice (Ullah et al., 2018). Since RNAPII elongation speed is high in regions with open chromatin, the spliceosome machinery has less time to recognize introns, resulting in more IR during co-transcriptional splicing (Braunschweig et al., 2014; Naftelberg et al., 2015). Furthermore, condition-dependent variation in the chromatin environment under different stresses and environmental cues plays an additional regulatory and fine-tuning role (Struhl and Segal, 2013; Zentner and Henikoff, 2013). Moreover, along with the positioning and spacing of the nucleosome, posttranslational modifications and DNA methylation also affect the transcriptional and splicing dynamics (Naftelberg et al., 2015; Friedrich et al., 2018; Zhang et al., 2018). Hence, various epigenetic modifications may provide a basic regulatory mechanism to orchestrate stress and splicing memory (**Figure 3**) in the same or future generations to respond to recurring stress more efficiently.



Not surprisingly, a recent study uncovered splicing memory response to heat stress priming in *Arabidopsis* as revealed by genome-wide differentially expressed genes (DEGs) and AS patterns (Ling et al., 2018; Sanyal et al., 2018). DEGs in response to heat stress were identified for different stages of priming, and genes responsible for potentially controlling heat stress memory were selected. With the help of gene networking analysis, heat and abiotic responsive genes were found to be involved in stress memory (Ling et al., 2018). Importantly, IR was found to be the most prevalent event under heat stress and contributed significantly toward establishing the splicing memory in response to heat. The primed plants produced comparable splicing patterns and efficiency compared with control plants, which were not exposed to heat stress before. In contrast, non-primed plants showed a significant increase in IR and produced splicing variants in heat conditions. Therefore, the primed plants, after relief from the second exposure to heat stress, maintain the splicing memory and perform in a similar manner to the control plants under non-stressful conditions (Ling et al., 2018). Ling et al. (2018) suggested that heat stress priming might be established at the post-transcriptional level and maintains splicing memory, which is crucial for plant survival and adaptation under stress. It is tempting to speculate that exposure to multiple stresses and coordination of gene expression and splicing patterns mediated by the chromatin environment may influence predictable responses and adaptive solutions in the long term. However, further research is needed to explore splicing memory and the underlying molecular mechanisms in response to different stresses in plants. We envisage that in addition to its contribution to protein diversity, AS may also play regulatory roles, and after repeated episodes of stress, splicing memory may also fine-tune stress-specific protein diversity to enhance plants networking capability to cope with given stress.

CONCLUSION

Emerging evidence indicates that the splicing process is also predominantly co-transcriptional in plants as in humans (Zhu et al., 2018). In plants, environmental fluctuations modulate chromatin structure, which in turn, could influence the co-transcriptional splicing process. Intriguingly, recent work indicates that plants can establish splicing memory in response

to higher temperature conditions and thus may “remember” a particular stress, likely through specific epigenetic signatures. This strategy may allow plants to engender an appropriate and reproducible response to a given stress. Further, IR transcripts are prevalent in plants and a majority of these are “trapped” in the nucleus. In addition, IR and many other AS transcripts are NMD sensitive and potentially degraded by the NMD pathway. It is clear that AS modulates transcriptome composition and splicing ratios, however, its role in diversifying proteome complexity is far from being understood.

It was a surprising discovery to find that the human genome codes for only ~20,000 to 21,000 protein-coding genes (Willyard, 2018), which is comparable with a weed (*Arabidopsis*, which has over 27,000 protein-coding genes) with a much smaller genome (Swarbreck et al., 2008). Since 95% of human genes and over 70% of genes in some plants are alternatively spliced, they can potentially make multiple proteins from each gene and considerably increase their proteome complexity (Kim et al., 2007; Pan et al., 2008). Although it is clear that AS does increase proteome complexity, the extent to which it enhances proteome diversity is far from clear. Multiple proteomic studies do not support a linear relationship between splicing and proteome complexity in humans (Tress et al., 2017a,b). Therefore, in-depth proteome analyses in multiple tissues and conditions, in conjunction with the variable expression of corresponding genes, need to be performed to illuminate the relationship between AS and proteome complexity in plants.

AUTHOR CONTRIBUTIONS

All authors listed have made a substantial, direct and intellectual contribution to the work, and approved it for publication.

ACKNOWLEDGMENTS

We thank the funding agencies for research support. Leverhulme Trust (RPG-2016-014); DOE Office of Science, Office of Biological and Environmental Research (BER) (DE-SC0010733); National Science Foundation; and the U.S. Department of Agriculture (ASNR).

REFERENCES

- Abascal, F., Ezkurdia, I., Rodriguez-Rivas, J., Rodriguez, J. M., del Pozo, A., Vázquez, J., et al. (2015). Alternatively spliced homologous exons have ancient origins and are highly expressed at the protein level. *PLoS Comput. Biol.* 11:e1004325. doi: 10.1371/journal.pcbi.1004325
- Akua, T., and Shaul, O. (2013). The *Arabidopsis thaliana* MHX gene includes an intronic element that boosts translation when localized in a 5' UTR intron. *J. Exp. Bot.* 64, 4255–4270. doi: 10.1093/jxb/ert235
- Albaqami, M., and Reddy, A. S. N. (2018). Development of an in vitro pre-mRNA splicing assay using plant nuclear extract. *Plant Methods* 14:1. doi: 10.1186/s13007-017-0271-6
- Barbosa-Morais, N. L., Irimia, M., Pan, Q., Xiong, H. Y., Gueroussou, S., Lee, L. J., et al. (2012). The evolutionary landscape of alternative splicing in vertebrate species. *Science* 338, 1587–1593. doi: 10.1126/science.1230612
- Becklin, K. M., Anderson, J. T., Gerhart, L. M., Wadgymar, S. M., Wessinger, C. A., and Ward, J. K. (2016). Examining plant physiological responses to climate change through an evolutionary lens. *Plant Physiol.* 172, 635–649. doi: 10.1104/pp.16.00793
- Behringer, M. G., and Hall, D. W. (2016). Selection on position of nonsense codons in Introns. *Genetics* 204, 1239–1248. doi: 10.1534/genetics.116.189894
- Bensimon, A., Heck, A. J. R., and Aebersold, R. (2012). Mass spectrometry-based proteomics and network biology. *Annu. Rev. Biochem.* 81, 379–405. doi: 10.1146/annurev-biochem-072909-100424
- Bicknell, A. A., Cenik, C., Chua, H. N., Roth, F. P., and Moore, M. J. (2012). Introns in UTRs: why we should stop ignoring them. *BioEssays* 34, 1025–1034. doi: 10.1002/bies.201200073
- Blencowe, B. J. (2017). The relationship between alternative splicing and proteomic complexity. *Trends Biochem. Sci.* 42, 407–408. doi: 10.1016/j.tibs.2017.04.001

- Boothby, T. C., Zipper, R. S., Van der Wee, C. M., and Wolniak, S. M. (2013). Removal of retained introns regulates translation in the rapidly developing gametophyte of *Marsilea vestita*. *Dev. Cell* 24, 517–529. doi: 10.1016/j.devcel.2013.01.015
- Boutz, P. L., Bhutkar, A., and Sharp, P. A. (2015). Detained introns are a novel, widespread class of post-transcriptionally spliced introns. *Genes Dev.* 29, 63–80. doi: 10.1101/gad.247361.114
- Boyko, A., Blevins, T., Yao, Y., Golubov, A., Bilichak, A., Ilnytsky, Y., et al. (2010). Transgenerational adaptation of Arabidopsis to stress requires DNA methylation and the function of dicer-like proteins. *PLoS One* 5:e0009514. doi: 10.1371/journal.pone.0009514
- Braunschweig, U., Barbosa-Morais, N. L., Pan, Q., Nachman, E. N., Alipanahi, B., Gonatopoulos-Pournatzis, T., et al. (2014). Widespread intron retention in mammals functionally tunes transcriptomes. *Genome Res.* 24, 1774–1786. doi: 10.1101/gr.177790.114
- Brosch, M., Saunders, G. I., Frankish, A., Collins, M. O., Yu, L., Wright, J., et al. (2011). Shotgun proteomics aids discovery of novel protein-coding genes, alternative splicing, and “resurrected” pseudogenes in the mouse genome. *Genome Res.* 21, 756–767. doi: 10.1101/gr.114272.110
- Brugiolio, M., Herzel, L., and Neugebauer, K. M. (2013). Counting on co-transcriptional splicing. *F1000Prime Rep.* 5:9. doi: 10.12703/P5-9
- Chamala, S., Feng, G., Chavarro, C., and Barbazuk, W. B. (2015). Genome-wide identification of evolutionarily conserved alternative splicing events in flowering plants. *Front. Bioeng. Biotechnol.* 3:33. doi: 10.3389/fbioe.2015.00033
- Chaudhary, S., Jabre, I., Reddy, A. S. N., Staiger, D., and Syed, N. H. (2019). Perspective on alternative splicing and proteome complexity in plants. *Trends Plant Sci.* 24, 496–506. doi: 10.1016/j.tplants.2019.02.006
- Chen, L., Bush, S. J., Tovar-Corona, J. M., Castillo-Morales, A., and Urrutia, A. O. (2014). Correcting for differential transcript coverage reveals a strong relationship between alternative splicing and organism complexity. *Mol. Biol. Evol.* 31, 1402–1413. doi: 10.1093/molbev/msu083
- Chen, W., and Moore, M. J. (2015). Spliceosomes. *Curr. Biol.* 25, R181–R183. doi: 10.1016/j.cub.2014.11.059
- Churchman, L. S., and Weissman, J. S. (2011). Nascent transcript sequencing visualizes transcription at nucleotide resolution. *Nature* 469, 368–373. doi: 10.1038/nature09652
- Conn, V. M., Hugouvieux, V., Nayak, A., Conos, S. A., Capovilla, G., Cildir, G., et al. (2017). A circRNA from SEPALLATA3 regulates splicing of its cognate mRNA through R-loop formation. *Nat. Plants* 3:17053. doi: 10.1038/nplants.2017.53
- Conrath, U., Beckers, G. J. M., Langenbach, C. J. G., and Jaskiewicz, M. R. (2015). Priming for enhanced defense. *Annu. Rev. Phytopathol.* 53, 97–119. doi: 10.1146/annurev-phyto-080614-120132
- Core, L. J., Waterfall, J. J., and Lis, J. T. (2008). Nascent RNA sequencing reveals widespread pausing and divergent initiation at human promoters. *Science* 322, 1845–1848. doi: 10.1126/science.1162228
- Ding, Y., Fromm, M., and Avramova, Z. (2012). Multiple exposures to drought “train” transcriptional responses in Arabidopsis. *Nat. Commun.* 3:740. doi: 10.1038/ncomms1732
- Drechsel, G., Kahles, A., Kesarwani, A. K., Stauffer, E., Behr, J., Drewe, P., et al. (2013). Nonsense-mediated decay of alternative precursor mRNA splicing variants is a major determinant of the Arabidopsis steady state transcriptome. *Plant Cell* 25, 3726–3742. doi: 10.1105/tpc.113.115485
- Edwards, C. R., Ritchie, W., Wong, J. J.-L., Schmitz, U., Middleton, R., An, X., et al. (2016). A dynamic intron retention program in the mammalian megakaryocyte and erythrocyte lineages. *Blood* 127, e24–e34. doi: 10.1182/blood-2016-01-692764
- Erhard, K. F., Talbot, J. E. R. B., Deans, N. C., McClish, A. E., and Hollick, J. B. (2015). Nascent transcription affected by RNA polymerase IV in *Zea mays*. *Genetics* 199, 1107–1125. doi: 10.1534/genetics.115.174714
- Ezkurdia, I., Del Pozo, A., Frankish, A., Rodriguez, J. M., Harrow, J., Ashman, K., et al. (2012). Comparative proteomics reveals a significant bias toward alternative protein isoforms with conserved structure and function. *Mol. Biol. Evol.* 29, 2265–2283. doi: 10.1093/molbev/mss100
- Ezkurdia, I., Rodriguez, J. M., Carrillo-De Santa Pau, E., Vázquez, J., Valencia, A., and Tress, M. L. (2015). Most highly expressed protein-coding genes have a single dominant isoform. *J. Proteome Res.* 14, 1880–1887. doi: 10.1021/pr501286b
- Fesenko, I., Khazigaleeva, R., Kirov, I., Kniazev, A., Glushenko, O., Babalyan, K., et al. (2017). Alternative splicing shapes transcriptome but not proteome diversity in *Physcomitrella patens*. *Sci. Rep.* 7, 2698. doi: 10.1038/s41598-017-02970-z
- Filichkin, S. A., Priest, H. D., Givan, S. A., Shen, R., Bryant, D. W., Fox, S. E., et al. (2010). Genome-wide mapping of alternative splicing in *Arabidopsis thaliana*. *Genome Res.* 20, 45–58. doi: 10.1101/gr.093302.109
- Filichkin, S. A., and Mockler, T. C. (2012). Unproductive alternative splicing and nonsense mRNAs: a widespread phenomenon among plant circadian clock genes. *Biol. Direct* 7:20. doi: 10.1186/1745-6150-7-20
- Filichkin, S. A., Cumbie, J. S., Dharmawardhana, P., Jaiswal, P., Chang, J. H., Palusa, S. G., et al. (2015). Environmental stresses modulate abundance and timing of alternatively spliced circadian transcripts in Arabidopsis. *Mol. Plant* 8, 207–227. doi: 10.1016/j.molp.2014.10.011
- Filichkin, S., Priest, H. D., Megraw, M., and Mockler, T. C. (2015). Alternative splicing in plants: directing traffic at the crossroads of adaptation and environmental stress. *Curr. Opin. Plant Biol.* 24, 125–135. doi: 10.1016/j.pbi.2015.02.008
- Filichkin, S. A., Hamilton, M., Dharmawardhana, P. D., Singh, S. K., Sullivan, C., Ben-Hur, A., et al. (2018). Abiotic stresses modulate landscape of poplar transcriptome via alternative splicing, differential intron retention, and isoform ratio switching. *Front. Plant Sci.* 9:5. doi: 10.3389/fpls.2018.00005
- Floor, S. N., and Doudna, J. A. (2016). Tunable protein synthesis by transcript isoforms in human cells. *eLife* 5, e10921. doi: 10.7554/eLife.10921.001
- Friedrich, T., Faivre, L., Bäurle, I., and Schubert, D. (2018). Chromatin-based mechanisms of temperature memory in plants. *Plant. Cell Environ.* 42, 762–770. doi: 10.1111/pce.13373
- Gallego-Paez, L. M., Bordone, M. C., Leote, A. C., Saraiva-Agostinho, N., Ascensão-Ferreira, M., and Barbosa-Morais, N. L. (2017). Alternative splicing: the pledge, the turn, and the prestige: the key role of alternative splicing in human biological systems. *Hum. Genet.* 136, 1015–1042. doi: 10.1007/s00439-017-1790-y
- Gallegos, J. E., and Rose, A. B. (2015). The enduring mystery of intron-mediated enhancement. *Plant Sci.* 237, 8–15. doi: 10.1016/j.plantsci.2015.04.017
- Godoy Herz, M. A., Kubaczka, M. G., Brzyżek, G., Servi, L., Krzyszton, M., Simpson, C., et al. (2019). Light regulates plant alternative splicing through the control of transcriptional elongation. *Mol. Cell* 73:1066–1074.e3. doi: 10.1016/j.molcel.2018.12.005
- Gohring, J., Jacak, J., and Barta, A. (2014). Imaging of endogenous messenger RNA splice variants in living cells reveals nuclear retention of transcripts inaccessible to nonsense-mediated decay in Arabidopsis. *Plant Cell* 26, 754–764. doi: 10.1105/tpc.113.118075
- Grasser, M., and Grasser, K. D. (2018). The plant RNA polymerase II elongation complex: a hub coordinating transcript elongation and mRNA processing. *Transcription* 9, 117–122. doi: 10.1080/21541264.2017.1356902
- Gratani, L. (2014). Plant phenotypic plasticity in response to environmental factors. *Adv. Bot.* 2014, 1–17. doi: 10.1155/2014/208747
- 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 spiderâ€™s web of mechanisms. *Front. Plant Sci.* 6:57. doi: 10.3389/fpls.2015.00057
- Guo, J., and Price, D. H. (2013). RNA polymerase II transcription elongation control. *Chem. Rev.* 113, 8583–8603. doi: 10.1021/cr400105n
- Harlen, K. M., and Churchman, L. S. (2017). The code and beyond: transcription regulation by the RNA polymerase II carboxy-terminal domain. *Nat. Rev. Mol. Cell Biol.* 18, 263–273. doi: 10.1038/nrm.2017.10
- Hartmann, L., Wiefner, T., and Wachter, A. (2018). Subcellular compartmentation of alternatively-spliced transcripts defines SERINE/ARGININE-RICH PROTEIN 30 Expression. *Plant Physiol.* 176, 2886–2903. doi: 10.1104/pp.17.01260
- Hetzl, J., Dutke, S. H., Benner, C., and Chory, J. (2016). Nascent RNA sequencing reveals distinct features in plant transcription. *Proc. Natl. Acad. Sci. U.S.A.* 113, 12316–12321. doi: 10.1073/pnas.1603217113
- Hilker, M., Schwachtje, J., Baier, M., Balazadeh, S., Bäurle, I., Geiselhardt, S., et al. (2016). Priming and memory of stress responses in organisms lacking a nervous system. *Biol. Rev.* 91, 1118–1133. doi: 10.1111/brv.12215

- Irimia, M., and Roy, S. W. (2014). Origin of spliceosomal introns and alternative splicing. *Cold Spring Harb. Perspect. Biol.* 6, a016071. doi: 10.1101/cshperspect.a016071
- Irimia, M., Weatheritt, R. J., Ellis, J. D., Parikshak, N. N., Gonatopoulos-Pournatzis, T., Babor, M., et al. (2014). A highly conserved program of neuronal microexons is misregulated in autistic brains. *Cell* 159, 1511–1523. doi: 10.1016/j.cell.2014.11.035
- Jabre, I., Reddy, A. S. N., Kalyna, M., Chaudhary, S., Khokhar, W., Byrne, L. J., et al. (2019). Does co-transcriptional regulation of alternative splicing mediate plant stress responses? *Nucleic Acids Res.* 47, 2716–2726. doi: 10.1093/nar/gkz121
- Jacob, A. G., and Smith, C. W. J. (2017). Intron retention as a component of regulated gene expression programs. *Hum. Genet.* 136, 1043–1057. doi: 10.1007/s00439-017-1791-x
- James, A. B., Syed, N. H., Bordage, S., Marshall, J., Nimmo, G. A., Jenkins, G. I., et al. (2012). Alternative splicing mediates responses of the Arabidopsis circadian clock to temperature changes. *Plant Cell* 24, 961–981. doi: 10.1105/tpc.111.093948
- Jung, H., Lee, D., Lee, J., Park, D., Kim, Y. J., Park, W. Y., et al. (2015). Intron retention is a widespread mechanism of tumor-suppressor inactivation. *Nat. Genet.* 47, 1242–1248. doi: 10.1038/ng.3414
- Juntawong, P., Girke, T., Bazin, J., and Bailey-Serres, J. (2014). Translational dynamics revealed by genome-wide profiling of ribosome footprints in Arabidopsis. *Proc. Natl. Acad. Sci. U.S.A.* 111, E203–E212. doi: 10.1073/pnas.1317811111
- Kalyna, M., Simpson, C. G., Syed, N. H., Lewandowska, D., Marquez, Y., Kusenda, B., et al. (2012). Alternative splicing and nonsense-mediated decay modulate expression of important regulatory genes in Arabidopsis. *Nucleic Acids Res.* 40, 2454–2469. doi: 10.1093/nar/gkr932
- Kan, Z., States, D., and Gish, W. (2002). Selecting for functional alternative splices in ESTs. *Genome Res.* 12, 1837–1845. doi: 10.1101/gr.764102
- Kelemen, O., Convertini, P., Zhang, Z., Wen, Y., Shen, M., Falaleeva, M., et al. (2013). Function of alternative splicing. *Gene* 514, 1–30. doi: 10.1016/j.gene.2012.07.083
- Keren, H., Lev-Maor, G., and Ast, G. (2010). Alternative splicing and evolution: diversification, exon definition and function. *Nat. Rev. Genet.* 11, 345–355. doi: 10.1038/nrg2776
- Khodor, Y. L., Rodriguez, J., Abruzzi, K. C., Tang, C. H. A., Marr, M. T., and Rosbash, M. (2011). Nascent-seq indicates widespread cotranscriptional pre-mRNA splicing in Drosophila. *Genes Dev.* 25, 2502–2512. doi: 10.1101/gad.178962.111
- Kim, E., Magen, A., and Ast, G. (2007). Different levels of alternative splicing among eukaryotes. *Nucleic Acids Res.* 35, 125–131. doi: 10.1093/nar/gkl924
- 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. doi: 10.3389/fpls.2012.00009
- Kornblihtt, A. R., Schor, I. E., Alló, M., Dujardin, G., Petrillo, E., and Muñoz, M. J. (2013). Alternative splicing: a pivotal step between eukaryotic transcription and translation. *Nat. Rev. Mol. Cell Biol.* 14, 153–165. doi: 10.1038/nrm3525
- Kwon, Y.-J., Park, M.-J., Kim, S.-G., Baldwin, I. T., and Park, C.-M. (2014). Alternative splicing and nonsense-mediated decay of circadian clock genes under environmental stress conditions in Arabidopsis. *BMC Plant Biol.* 14:136. doi: 10.1186/1471-2229-14-136
- Lämke, J., and Bäurle, I. (2017). Epigenetic and chromatin-based mechanisms in environmental stress adaptation and stress memory in plants. *Genome Biol.* 18:124. doi: 10.1186/s13059-017-1263-6
- Ling, Y., Serrano, N., Gao, G., Atia, M., Mokhtar, M., Woo, Y. H., et al. (2018). Thermoprimer triggers splicing memory in Arabidopsis. *J. Exp. Bot.* 69, 2659–2675. doi: 10.1093/jxb/ery062
- Listerman, I., Sapra, A. K., and Neugebauer, K. M. (2006). Cotranscriptional coupling of splicing factor recruitment and precursor messenger RNA splicing in mammalian cells. *Nat. Struct. Mol. Biol.* 13, 815–822. doi: 10.1038/nsmb1135
- Liu, Y., González-Porta, M., Santos, S., Brazma, A., Marioni, J. C., Aebersold, R., et al. (2017). Impact of alternative splicing on the human proteome. *Cell Rep.* 20, 1229–1241. doi: 10.1016/j.celrep.2017.07.025
- Lopez-Casado, G., Covey, P. A., Bedinger, P. A., Mueller, L. A., Thannhauser, T. W., Zhang, S., et al. (2012). Enabling proteomic studies with RNA-Seq: the proteome of tomato pollen as a test case. *Proteomics* 12, 761–774. doi: 10.1002/pmic.201100164
- Lu, T., Lu, G., Fan, D., Zhu, C., Li, W., Zhao, Q., et al. (2010). Function annotation of the rice transcriptome at single-nucleotide resolution by RNA-seq. *Genome Res.* 20, 1238–1249. doi: 10.1101/gr.106120.110
- Luco, R. F., Allo, M., Schor, I. E., Kornblihtt, A. R., and Misteli, T. (2011). Epigenetics in alternative pre-mRNA splicing. *Cell* 144, 16–26. doi: 10.1016/j.cell.2010.11.056
- Mandadi, K. K., and Scholthof, K.-B. G. (2015). Genome-wide analysis of alternative splicing landscapes modulated during plant-virus interactions in *Brachypodium distachyon*. *Plant Cell* 27, 71–85. doi: 10.1105/tpc.114.133991
- Marquez, Y., Brown, J. W. S., 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. doi: 10.1101/gr.134106.111
- Marquez, Y., Höpfler, M., Ayatollahi, Z., Barta, A., and Kalyna, M. (2015). Unmasking alternative splicing inside protein-coding exons defines exons and their role in proteome plasticity. *Genome Res.* 25, 995–1007. doi: 10.1101/gr.186585.114
- Matera, A. G., and Wang, Z. (2014). A day in the life of the spliceosome. *Nat. Rev. Mol. Cell Biol.* 15, 108–121. doi: 10.1038/nrm3742
- Matlin, A. J., Clark, F., and Smith, C. W. J. (2005). Understanding alternative splicing: towards a cellular code. *Nat. Rev. Mol. Cell Biol.* 6, 386–398. doi: 10.1038/nrm1645
- Melamud, E., and Moul, J. (2009). Stochastic noise in splicing machinery. *Nucleic Acids Res.* 37, 4873–4886. doi: 10.1093/nar/gkp471
- Merkhofer, E. C., Hu, P., and Johnson, T. L. (2014). Introduction to cotranscriptional RNA splicing. *Methods Mol. Biol.* 1126, 83–96. doi: 10.1007/978-1-62703-980-2_6
- Merkin, J., Russell, C., Chen, P., and Burge, C. B. (2012). Evolutionary dynamics of gene and isoform regulation in mammalian tissues. *Science* 338, 1593–1599. doi: 10.1126/science.1228186
- Misteli, T., Cáceres, J. F., Clement, J. Q., Krainer, A. R., Wilkinson, M. F., and Spector, D. L. (1998). Serine phosphorylation of SR proteins is required for their recruitment to sites of transcription in vivo. *J. Cell Biol.* 143, 297–307. doi: 10.1083/jcb.143.2.297
- Modrek, B., Resch, A., Grasso, C., and Lee, C. (2001). Genome-wide detection of alternative splicing in expressed sequences of human genes. *Nucleic Acids Res.* 29, 2850–2859. doi: 10.1093/nar/29.13.2850
- Naftelberg, S., Schor, I. E., Ast, G., and Kornblihtt, A. R. (2015). Regulation of alternative splicing through coupling with transcription and chromatin structure. *Annu. Rev. Biochem.* 84, 165–198. doi: 10.1146/annurev-biochem-060614-034242
- Narang, V., Sung, W. K., and Mittal, A. (2005). Computational modeling of oligonucleotide positional densities for human promoter prediction. *Artif. Intell. Med.* 35, 107–119. doi: 10.1016/j.artmed.2005.02.005
- Neverov, A. D., Artamonova, I. I., Nurtudinov, R. N., Frishman, D., Gelfand, M. S., and Mironov, A. (2005). Alternative splicing and protein function. *BMC Bioinformatics* 6:266. doi: 10.1186/1471-2105-6-266
- Nilsen, T. W., and Graveley, B. R. (2010). Expansion of the eukaryotic proteome by alternative splicing. *Nature* 463, 457–463. doi: 10.1038/nature08909
- Ning, K., and Nesvizhskii, A. I. (2010). The utility of mass spectrometry-based proteomic data for validation of novel alternative splice forms reconstructed from RNA-Seq data: a preliminary assessment. *BMC Bioinformatics* 11(Suppl. 1):S14. doi: 10.1186/1471-2105-11-S11-S14
- Nojima, T., Gomes, T., Grosso, A. R. F., Kimura, H., Dye, M. J., Dhir, S., et al. (2015). Mammalian NET-seq reveals genome-wide nascent transcription coupled to RNA processing. *Cell* 161, 526–540. doi: 10.1016/j.cell.2015.03.027
- Nojima, T., Rebelo, K., Gomes, T., Grosso, A. R., Proudfoot, N. J., and Carmo-Fonseca, M. (2018). RNA polymerase II phosphorylated on CTD Serine 5 interacts with the spliceosome during co-transcriptional splicing. *Mol. Cell* 72:369–379.e4. doi: 10.1016/j.molcel.2018.09.004
- Olsen, J. V., Ong, S.-E., and Mann, M. (2004). Trypsin cleaves exclusively C-terminal to arginine and lysine residues. *Mol. Cell. Proteomics* 3, 608–614. doi: 10.1074/mcp.T400003-MCP200

- Pajoro, A., Severing, E., Angenent, G. C., and Immink, R. G. H. (2017). Histone H3 lysine 36 methylation affects temperature-induced alternative splicing and flowering in plants. *Genome Biol.* 18:102. doi: 10.1186/s13059-017-1235-x
- 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. doi: 10.1038/ng.259
- Paronetto, M. P., Passacantilli, I., and Sette, C. (2016). Alternative splicing and cell survival: from tissue homeostasis to disease. *Cell Death Differ.* 23, 1919–1929. doi: 10.1038/cdd.2016.91
- Parra, G., Bradnam, K., Rose, A. B., and Korf, I. (2011). Comparative and functional analysis of intron-mediated enhancement signals reveals conserved features among plants. *Nucleic Acids Res.* 39, 5328–5337. doi: 10.1093/nar/gk-r043
- Pereira, A. (2016). Plant abiotic stress challenges from the changing environment. *Front. Plant Sci.* 7:1123. doi: 10.3389/fpls.2016.01123
- Petrillo, E., Godoy Herz, M. A., Fuchs, A., Reifer, D., Fuller, J., Yanovsky, M. J., et al. (2014). A chloroplast retrograde signal regulates nuclear alternative splicing. *Science* 344, 427–430. doi: 10.1126/science.1250322
- Poulos, M. G., Batra, R., Charizanis, K., and Swanson, M. S. (2011). Developments in RNA splicing and disease. *Cold Spring Harb. Perspect. Biol.* 3, a000778. doi: 10.1101/cshperspect.a000778
- Preker, P., Nielsen, J., Kammler, S., Lykke-Andersen, S., Christensen, M. S., Mapendano, C. K., et al. (2008). RNA exosome depletion reveals transcription upstream of active human promoters. *Science* 322, 1851–1854. doi: 10.1126/science.1164096
- Rayson, S., Arciga-Reyes, L., Wootton, L., de Torres Zabala, M., Truman, W., Graham, N., et al. (2012). A role for nonsense-mediated mRNA decay in plants: pathogen responses are induced in *Arabidopsis thaliana* nmd mutants. *PLoS One* 7:e0031917. doi: 10.1371/journal.pone.0031917
- Reddy, A. S. N., Marquez, Y., Kalyna, M., and Barta, A. (2013). Complexity of the alternative splicing landscape in plants. *Plant Cell* 25, 3657–3683. doi: 10.1105/tpc.113.117523
- Reddy, A. S. N., Rogers, M. F., Richardson, D. N., Hamilton, M., and Ben-Hur, A. (2012). Deciphering the plant splicing code: experimental and computational approaches for predicting alternative splicing and splicing regulatory elements. *Front. Plant Sci.* 3:18. doi: 10.3389/fpls.2012.00018
- Rocchi, V., Janni, M., Bellincampi, D., Giardina, T., and D'Ovidio, R. (2012). Intron retention regulates the expression of pectin methyl esterase inhibitor (Pmei) genes during wheat growth and development. *Plant Biol.* 14, 365–373. doi: 10.1111/j.1438-8677.2011.00508.x
- Romero-Barrios, N., Legascue, M. F., Benhamed, M., Ariel, F., and Crespi, M. (2018). Splicing regulation by long noncoding RNAs. *Nucleic Acids Res.* 46, 2169–2184. doi: 10.1093/nar/gky095
- Rost, H. L., Malmstrom, L., and Aebersold, R. (2015). Reproducible quantitative proteotype data matrices for systems biology. *Mol. Biol. Cell* 26, 3926–3931. doi: 10.1091/mbc.E15-07-0507
- Saldi, T., Cortazar, M. A., Sheridan, R. M., and Bentley, D. L. (2016). Coupling of RNA polymerase II transcription elongation with Pre-mRNA splicing. *J. Mol. Biol.* 428, 2623–2635. doi: 10.1016/j.jmb.2016.04.017
- Sammeth, M., Foissac, S., and Guigó, R. (2008). A general definition and nomenclature for alternative splicing events. *PLoS Comput. Biol.* 4:e1000147. doi: 10.1371/journal.pcbi.1000147
- Sani, E., Herzyk, P., Perrella, G., Colot, V., and Amtmann, A. (2013). Hyperosmotic priming of *Arabidopsis* seedlings establishes a long-term somatic memory accompanied by specific changes of the epigenome. *Genome Biol.* 14, R59. doi: 10.1186/gb-2013-14-6-r59
- Sanyal, R. P., Misra, H. S., and Saini, A. (2018). Heat-stress priming and alternative splicing-linked memory. *J. Exp. Bot.* 69, 2431–2434. doi: 10.1093/jxb/er-y111
- Seo, P. J., Ryu, J., Kang, S. K., and Park, C. M. (2011). Modulation of sugar metabolism by an indeterminate domain transcription factor contributes to photoperiodic flowering in *Arabidopsis*. *Plant J.* 65, 418–429. doi: 10.1111/j.1365-3113.2010.04432.x
- Severing, E. I., van Dijk, A. D. J., Stiekema, W. J., and van Ham, R. C. H. J. (2009). Comparative analysis indicates that alternative splicing in plants has a limited role in functional expansion of the proteome. *BMC Genomics* 10:154. doi: 10.1186/1471-2164-10-154
- Shaul, O. (2015). Unique aspects of plant nonsense-mediated mRNA decay. *Trends Plant Sci.* 20, 767–779. doi: 10.1016/j.tplants.2015.08.011
- Shen, Y., Zhou, Z., Wang, Z., Li, W., Fang, C., Wu, M., et al. (2014). Global dissection of alternative splicing in paleopolyploid soybean. *Plant Cell* 26, 996–1008. doi: 10.1105/tpc.114.122739
- Sheynkman, G. M., Shortreed, M. R., Frey, B. L., and Smith, L. M. (2013). Discovery and mass spectrometric analysis of novel splice-junction peptides using RNA-Seq. *Mol. Cell. Proteomics* 12, 2341–2353. doi: 10.1074/mcp.O113.028142
- Shukla, S., and Oberdoerffer, S. (2012). Co-transcriptional regulation of alternative pre-mRNA splicing. *Biochim. Biophys. Acta* 1819, 673–683. doi: 10.1016/j.bbagr.2012.01.014
- Skelly, M. J., Frungillo, L., and Spoel, S. H. (2016). Transcriptional regulation by complex interplay between post-translational modifications. *Curr. Opin. Plant Biol.* 33, 126–132. doi: 10.1016/j.pbi.2016.07.004
- Staiger, D., and Simpson, G. G. (2015). Enter exons. *Genome Biol.* 16:136. doi: 10.1186/s13059-015-0704-3
- Sterne-Weiler, T., Martinez-Nunez, R. T., Howard, J. M., Cvitovik, I., Katzman, S., Tariq, M. A., et al. (2013). Frac-seq reveals isoform-specific recruitment to polyribosomes. *Genome Res.* 23, 1615–1623. doi: 10.1101/gr.148585.112
- Struhl, K., and Segal, E. (2013). Determinants of nucleosome positioning. *Nat. Struct. Mol. Biol.* 20, 267–273. doi: 10.1038/nsmb.2506
- Sureshkumar, S., Dent, C., Seleznev, A., Tasset, C., and Balasubramanian, S. (2016). Nonsense-mediated mRNA decay modulates FLM-dependent thermosensory flowering response in *Arabidopsis*. *Nat. Plants* 2:16055. doi: 10.1038/NPLANTS.2016.55
- Sveen, A., Kilpinen, S., Ruusulehto, A., Lothe, R. A., and Skotheim, R. I. (2016). Aberrant RNA splicing in cancer; expression changes and driver mutations of splicing factor genes. *Oncogene* 35, 2413–2427. doi: 10.1038/onc.2015.318
- Swarbreck, D., Wilks, C., Lamesch, P., Berardini, T. Z., Garcia-Hernandez, M., Foerster, H., et al. (2008). The *Arabidopsis* information resource (TAIR): gene structure and function annotation. *Nucleic Acids Res.* 36, 1009–1014. doi: 10.1093/nar/gkm965
- Syed, N. H., Kalyna, M., Marquez, Y., Barta, A., and Brown, J. W. S. (2012). Alternative splicing in plants - coming of age. *Trends Plant Sci.* 17, 616–623. doi: 10.1016/j.tplants.2012.06.001
- Tanner, S., Shen, Z., Ng, J., Florea, L., Guigó, R., Briggs, S. P., et al. (2007). Improving gene annotation using peptide mass spectrometry. *Genome Res.* 17, 231–239. doi: 10.1101/gr.5646507
- Thatcher, S. R., Zhou, W., Leonard, A., Wang, B.-B., Beatty, M., Zastrow-Hayes, G., et al. (2014). Genome-wide analysis of alternative splicing in *Zea mays*: landscape and genetic regulation. *Plant Cell Online* 26, 3472–3487. doi: 10.1105/tpc.114.130773
- Tian, C., Wang, Y., Yu, H., He, J., Wang, J., Shi, B., et al. (2019). A gene expression map of shoot domains reveals regulatory mechanisms. *Nat. Commun.* 10:141. doi: 10.1038/s41467-018-08083-z
- Trcek, T., Sato, H., Singer, R. H., and Maquat, L. E. (2013). Temporal and spatial characterization of nonsense-mediated mRNA decay. *Genes Dev.* 27, 541–551. doi: 10.1101/gad.209635.112
- Tress, M. L., Bodenmiller, B., Aebersold, R., and Valencia, A. (2008). Proteomics studies confirm the presence of alternative protein isoforms on a large scale. *Genome Biol.* 9, R162. doi: 10.1186/gb-2008-9-11-r162
- Tress, M. L., Abascal, F., and Valencia, A. (2017a). Alternative Splicing may not be the key to proteome complexity. *Trends Biochem. Sci.* 42, 98–110. doi: 10.1016/j.tibs.2016.08.008
- Tress, M. L., Abascal, F., and Valencia, A. (2017b). Most alternative isoforms are not functionally important. *Trends Biochem. Sci.* 42, 408–410. doi: 10.1016/j.tibs.2017.04.002
- Ullah, F., Hamilton, M., Reddy, A. S. N., and Ben-Hur, A. (2018). Exploring the relationship between intron retention and chromatin accessibility in plants. *BMC Genomics* 19:21. doi: 10.1186/s12864-017-4393-z
- Vinocur, B., and Altman, A. (2005). Recent advances in engineering plant tolerance to abiotic stress: achievements and limitations. *Curr. Opin. Biotechnol.* 16, 123–132. doi: 10.1016/j.copbio.2005.02.001
- Wang, E. T., Sandberg, R., Luo, S., Khrebukova, I., Zhang, L., Mayr, C., et al. (2008). Alternative isoform regulation in human tissue transcriptomes. *Nature* 456, 470–476. doi: 10.1038/nature07509
- Wang, X., Codreanu, S. G., Wen, B., Li, K., Chambers, M. C., Liebler, D. C., et al. (2018). Detection of proteome diversity resulted from alternative splicing

- is limited by trypsin cleavage specificity. *Mol. Cell. Proteomics* 17, 422–430. doi: 10.1074/mcp.RA117.000155
- Wang, X., Hu, L., Wang, X., Li, N., Xu, C., Gong, L., et al. (2016). DNA methylation affects gene alternative splicing in plants: an example from rice. *Mol. Plant* 9, 305–307. doi: 10.1016/j.molp.2015.09.016
- Wang, X., Zhang, B., and Wren, J. (2013). *customProDB*: an R package to generate customized protein databases from RNA-Seq data for proteomics search. *Bioinformatics* 29, 3235–3237. doi: 10.1093/bioinformatics/btt543
- Wang, Y., Liu, J., Huang, B., Xu, Y., Li, J., Huang, L., et al. (2014). Mechanism of alternative splicing and its regulation (Review). *Biomed. Reports* 3, 152–158. doi: 10.3892/br.2014.407
- Weatheritt, R. J., Davey, N. E., and Gibson, T. J. (2012). Linear motifs confer functional diversity onto splice variants. *Nucleic Acids Res.* 40, 7123–7131. doi: 10.1093/nar/gks442
- Weatheritt, R. J., Sterne-Weiler, T., and Blencowe, B. J. (2016). The ribosome-engaged landscape of alternative splicing. *Nat. Struct. Mol. Biol.* 23, 1117–1123. doi: 10.1038/nsmb.3317
- Wickramasinghe, V. O., González-Porta, M., Perera, D., Bartolozzi, A. R., Sibley, C. R., Hallegger, M., et al. (2015). Regulation of constitutive and alternative mRNA splicing across the human transcriptome by PRPF8 is determined by 5' splice site strength. *Genome Biol.* 16:201. doi: 10.1186/s13059-015-0749-3
- Will, C. L., and Lührmann, R. (2011). Spliceosome structure and function. *Cold Spring Harb. Perspect. Biol.* 3, 1–2. doi: 10.1101/cshperspect.a003707
- Willyard, C. (2018). New human gene tally reignites debate. *Nature* 558, 354–355. doi: 10.1038/d41586-018-05462-w
- Wong, J. J.-L., Au, A. Y. M., Ritchie, W., and Rasko, J. E. J. (2015). Intron retention in mRNA: no longer nonsense. *BioEssays* 38, 41–49. doi: 10.1002/bies.201500117
- Wong, J. J. L., Ritchie, W., Ebner, O. A., Selbach, M., Wong, J. W. H., Huang, Y., et al. (2013). Orchestrated intron retention regulates normal granulocyte differentiation. *Cell* 154, 583–595. doi: 10.1016/j.cell.2013.06.052
- Wong, M. M., Chong, G. L., and Verslues, P. E. (2017). Epigenetics and RNA processing: connections to drought, salt, and ABA? *Methods Mol. Biol.* 1631, 3–21. doi: 10.1007/978-1-4939-7136-7_1
- Yang, C., Bolotin, E., Jiang, T., Sladek, F. M., and Martinez, E. (2007). Prevalence of the initiator over the TATA box in human and yeast genes and identification of DNA motifs enriched in human TATA-less core promoters. *Gene* 389, 52–65. doi: 10.1016/j.gene.2006.09.029
- Yang, X., Coulombe-Huntington, J., Kang, S., Sheynkman, G. M., Hao, T., Richardson, A., et al. (2016). Widespread expansion of protein interaction capabilities by alternative splicing. *Cell* 164, 805–817. doi: 10.1016/j.cell.2016.01.029
- Yu, H., Tian, C., Yu, Y., and Jiao, Y. (2016). Transcriptome survey of the contribution of alternative splicing to proteome diversity in *Arabidopsis thaliana*. *Mol. Plant* 9, 749–752. doi: 10.1016/j.molp.2015.12.018
- Zentner, G. E., and Henikoff, S. (2013). Regulation of nucleosome dynamics by histone modifications. *Nat. Struct. Mol. Biol.* 20, 259–266. doi: 10.1038/nsmb.2470
- Zhang, H., Lang, Z., and Zhu, J. K. (2018). Dynamics and function of DNA methylation in plants. *Nat. Rev. Mol. Cell Biol.* 19, 489–506. doi: 10.1038/s41580-018-0016-z
- Zhang, R., Calixto, C. P. G., Marquez, Y., Venhuizen, P., Tzioutziou, N. A., Guo, W., et al. (2017). A high quality Arabidopsis transcriptome for accurate transcript-level analysis of alternative splicing. *Nucleic Acids Res.* 45, 5061–5073. doi: 10.1093/nar/gkx267
- Zhu, B., Zhang, W., Zhang, T., Liu, B., and Jiang, J. (2015). Genome-wide prediction and validation of intergenic enhancers in Arabidopsis using open chromatin signatures. *Plant Cell* 27, 2415–2426. doi: 10.1105/tpc.15.00537
- Zhu, J.-K. (2016). Abiotic stress signaling and responses in plants. *Cell* 167, 313–324. doi: 10.1016/j.cell.2016.08.029
- Zhu, J., Liu, M., Liu, X., and Dong, Z. (2018). RNA polymerase II activity revealed by GRO-seq and pNET-seq in Arabidopsis. *Nat. Plants* 4, 1112–1123. doi: 10.1038/s41477-018-0280-0

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2019 Chaudhary, Khokhar, Jabre, Reddy, Byrne, Wilson and Syed. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



A Role for Pre-mRNA-PROCESSING PROTEIN 40C in the Control of Growth, Development, and Stress Tolerance in *Arabidopsis thaliana*

OPEN ACCESS

Edited by:

Maria Kalyna,
University of Natural Resources and
Life Sciences Vienna, Austria

Reviewed by:

Marjori Matzke,
Academia Sinica, Taiwan
Federico Damian Ariel,
Instituto de Agrobiotecnología del
Litoral (IAL), Argentina

*Correspondence:

Marcelo Javier Yanovsky
mjyanovsky@gmail.com

[†]These authors contributed equally
to this work.

[‡]These authors have contributed
equally to this work as second authors.

Specialty section:

This article was submitted to
Plant Physiology,
a section of the journal
Frontiers in Plant Science

Received: 13 May 2019

Accepted: 22 July 2019

Published: 13 August 2019

Citation:

Hernando CE, García Hourquet M,
de Leone MJ, Careno D, Iserte J,
Mora García S and Yanovsky MJ
(2019) A Role for Pre-mRNA-
PROCESSING PROTEIN 40C in the
Control of Growth, Development,
and Stress Tolerance in
Arabidopsis thaliana.
Front. Plant Sci. 10:1019.
doi: 10.3389/fpls.2019.01019

Carlos Esteban Hernando[†], Mariano García Hourquet[†], María José de Leone[‡],
Daniel Careno[‡], Javier Iserte, Santiago Mora García and Marcelo Javier Yanovsky*

Comparative Genomics of Plant Development Laboratory, Instituto de Investigaciones Bioquímicas de Buenos Aires–Consejo
Nacional de Investigaciones Científicas y Técnicas de Argentina, Fundación Instituto Leloir, Buenos Aires, Argentina

Because of their sessile nature, plants have adopted varied strategies for growing and reproducing in an ever-changing environment. Control of mRNA levels and pre-mRNA alternative splicing are key regulatory layers that contribute to adjust and synchronize plant growth and development with environmental changes. Transcription and alternative splicing are thought to be tightly linked and coordinated, at least in part, through a network of transcriptional and splicing regulatory factors that interact with the carboxyl-terminal domain (CTD) of the largest subunit of RNA polymerase II. One of the proteins that has been shown to play such a role in yeast and mammals is pre-mRNA-PROCESSING PROTEIN 40 (PRP40, also known as CA150, or TCERG1). In plants, members of the PRP40 family have been identified and shown to interact with the CTD of RNA Pol II, but their biological functions remain unknown. Here, we studied the role of AtPRP40C, in *Arabidopsis thaliana* growth, development and stress tolerance, as well as its impact on the global regulation of gene expression programs. We found that the *prp40c* knockout mutants display a late-flowering phenotype under long day conditions, associated with minor alterations in red light signaling. An RNA-seq based transcriptome analysis revealed differentially expressed genes related to biotic stress responses and also differentially expressed as well as differentially spliced genes associated with abiotic stress responses. Indeed, the characterization of stress responses in *prp40c* mutants revealed an increased sensitivity to salt stress and an enhanced tolerance to *Pseudomonas syringae* pv. *maculicola* (Psm) infections. This constitutes the most thorough analysis of the transcriptome of a *prp40* mutant in any organism, as well as the first characterization of the molecular and physiological roles of a member of the PRP40 protein family in plants. Our results suggest that PRP40C is an important factor linking the regulation of gene expression programs to the modulation of plant growth, development, and stress responses.

Keywords: pre-mRNA processing protein 40, splicing, transcription, *Arabidopsis*, biotic stress, abiotic stress

INTRODUCTION

Plants are permanently subjected to environmental changes that adversely affect their growth and development either in natural or agricultural settings. Accurate control of gene expression and precursor mRNAs (pre-mRNAs) processing is essential for the adaptation to rapid changes in the environment. Synthesis of mRNA by the RNA polymerase II (RNA Pol II) is coordinated with subsequent RNA processing events such as 5' capping, splicing, cleavage, and polyadenylation (Darnell, 2013). The largest subunit of the RNA Pol II couples transcription and pre-mRNA processing *via* its C-terminal domain (CTD). The CTD comprises tandem Tyr1–Ser2–Pro3–Thr4–Ser5–Pro6–Ser7 heptapeptide repeats that are highly conserved in eukaryotes (Allison et al., 1988; Nawrath et al., 1990) and interacts with capping, splicing, and polyadenylation factors, acting as a scaffolding platform for mRNA processing factors (Greenleaf, 1993; McCracken et al., 1997; Misteli and Spector, 1999; Buratowski, 2003; Phatnani and Greenleaf, 2006; Muñoz et al., 2010; Braunschweig et al., 2013).

Pre-mRNA splicing occurs in two sequential transesterification steps catalyzed by the spliceosome, a large dynamic ribonucleoprotein complex present in the nucleus. The major spliceosome consists of five small nuclear ribonucleoproteins (snRNPs) named U1, U2, U4, U5, and U6, along with ~200 accessory proteins (Wahl et al., 2009). The splicing reaction begins with the initial recognition of the 5' and 3' splice sites (5'SS and 3'SS, respectively) at the exon–intron boundaries, the branch point sequence (BPS), and the polypyrimidine tract (Py) (Wahl et al., 2009; Meyer et al., 2015); the completion of the splicing process involves the removal of an intron and the ligation of the resulting exons.

Splicing can be classified as constitutive splicing (CS) or alternative splicing (AS), depending on splice site choice and usage. In constitutive splicing, canonical splice sites are always used for a given transcript. In contrast, alternative splicing involves variable splice site choice and/or usage, giving place to multiple mRNAs variants from a single gene. Splice sites can be strong or weak depending on the degree in which their sequences diverge from consensus sequences, and this determines their affinities for the spliceosomal machinery. In general, strong splice sites lead to constitutive splicing through full usage of the site. On the other side, weak splice sites are usually associated with alternative splicing, and the frequency of usage of the alternative splice sites varies depending on cellular context and environmental conditions (Kornblihtt et al., 2013). Alternative splicing events comprise intron retention (IR), exon skipping (ES), and alternative 5' donor and 3' acceptor splice site selection (Alt 5'SS and Alt 3'SS, respectively).

In plants of the model species *Arabidopsis thaliana*, current estimates indicate that ~61% of the intron-containing genes undergo alternative splicing (Marquez et al., 2012), which proved to be essential for proper growth and development, as well as for optimal responses to environmental changes (Reddy et al., 2013; Staiger and Brown, 2013; Ding et al., 2014; Kwon et al., 2014; Shikata et al., 2014; Filichkin et al., 2015; Deng and Cao, 2017). One of the earliest events in spliceosome assembly is the formation of an RNA duplex between the 5' SS and the 5'

end of the U1 snRNP, an event involving the activity of multiple proteins that stabilize the complex. PRE-mRNA-PROCESSING PROTEIN 40 (PRP40) was first discovered in yeast and proved to be essential in the early steps of spliceosome complex formation (Kao and Siliciano, 1996). PRP40 harbors two WW domains in the amino terminus and four FF domain repeats in the carboxyl terminus, which have been well characterized as mediators of protein–protein interactions (Bedford and Leder, 1999; Macias et al., 2002). In yeast, as well as in mammals, PRP40 helps to define the bridging interaction that links both ends of the intron, by interacting simultaneously with the BRANCHPOINT BINDING PROTEIN (BBP) and the U1 snRNP (Abovich and Rosbash, 1997). *Saccharomyces cerevisiae* PRP40 (ScPRP40) and its three human homologues, HsPRPF40A, HsPRPF40B, and HsTCERG1/CA150, have been extensively studied in the last two decades. While HsPRPF40A and HsPRPF40B were characterized as essential components of the early spliceosome assembly process (Lin et al., 2004; Wahl et al., 2009; Becerra et al., 2015), HsTCERG1/CA150 was initially discovered as a transcriptional modulator and later linked to pre-mRNA splicing (Sune et al., 1997; Pearson et al., 2008; Munoz-Cobo et al., 2017). In fact, HsTCERG1/CA150 has been found associated with elongation factors and is present in a complex with the RNA polymerase II *via* the FF domains (Sune et al., 1997; Carty et al., 2000; Sanchez-Alvarez et al., 2006). HsTCERG1/CA150 has also been linked to RNA splicing through its WW domain 2 (WW2), which interacts with the splicing factors SF1, U2AF, and components of the SF3 complex (Goldstrohm et al., 2001; Lin et al., 2004). Furthermore, HsTCERG1/CA150 has been identified in highly purified spliceosomes in multiple studies (Makarov et al., 2002; Rappsilber et al., 2002; Deckert et al., 2006). As previously mentioned, the processes of transcription and pre-mRNA processing are coordinated by the CTD of the RNA pol II. The modular structure of HsTCERG1/CA150, containing the splicing-factor associating WW domains present in the N-terminus and the CTD-associating FF repeats in the C terminus, confers the ideal structure for a protein that couples transcription and splicing. In accordance with this model, both halves of HsTCERG1/CA150 have been shown to be essential for the assembly of higher-order transcription-splicing complexes (Sanchez-Alvarez et al., 2006).

Plant homologues of ScPRP40 were first identified in 2009 by Kang et al. by means of a bioinformatic screening of proteins that interact with the RNA Pol II CTD in *Arabidopsis*. Three proteins were thus identified: AtPRP40A, AtPRP40B, and AtPRP40C (Kang et al., 2009), which interacted with the RNA Pol II CTD (both phosphorylated and nonphosphorylated); at least for AtPRP40B, the authors showed that the WW domains at the amino terminus mediated this interaction (Kang et al., 2009). So far, a characterization of the biological and molecular roles of plant PRP40s is missing. Here, we decided to focus on the characterization of AtPRP40C using a reverse genetics approach and performed a physiological and molecular analysis of the function of this protein in the control of growth, development and stress tolerance, as well as in the regulation of gene expression. We found that AtPRP40C is involved in the regulation of flowering time and photomorphogenic responses. In addition, an RNA-seq

analysis revealed that AtPRP40C is associated with the proper control of expression and splicing of abiotic and biotic stress-related transcripts, and, indeed, physiological assays showed that *prp40c* mutants display altered tolerance to salt stress and *Pseudomonas syringae* pv. *maculicola* (Psm) infections. Our work is the first physiological and molecular characterization of a member of the PRP40 protein family in plants, which reveals an important role for PRP40C linking the regulation of gene expression and pre-mRNA splicing, and modulating plant growth, development, and stress responses.

MATERIALS AND METHODS

Plant Material

All of the *Arabidopsis* lines used in this study were on the Columbia (Col-0) ecotype background, and the *prp40c-1* (SALK_148319) and *prp40c-2* (SALK-205357) mutants were obtained from the Arabidopsis Biological Research Center (ABRC). Genotypic verification of the mutants was performed via PCR analysis using primers as detailed in **Supplementary Table S1**. PRP40C expression levels in the mutant alleles, as well as in wild-type plants, were assessed by reverse transcription followed by a semiquantitative PCR (RT-PCR) using Actin 2 as expression control; primers used are described in the **Supplementary Table S1**.

Growth Conditions

Plants were grown on soil at 22°C under long days (LD; 16-h light/8-h dark cycles; 80 $\mu\text{mol m}^{-2} \text{s}^{-1}$ of white light), 12:12 days (LD 12:12; 12-h light/12-h dark cycles; 80 $\mu\text{mol m}^{-2} \text{s}^{-1}$ of white light), short days (SD; 8-h light/16-h dark cycles; 140 $\mu\text{mol m}^{-2} \text{s}^{-1}$ of white light), or continuous light (LL; 50 $\mu\text{mol m}^{-2} \text{s}^{-1}$ of white light), depending on the experiment.

Phylogenetic Analysis

Evolutionary analysis was conducted using the maximum likelihood method implemented in MEGA X (Kumar et al., 2018). Sequences were retrieved using TBLASTN with different query sequences from Phytozome (phytozome.jgi.doe.gov), Onekp project at China National GeneBank (db.cngb.org/bblast/bblast/tblastn/), www.fernbase.org, and the *Klebsormidium nitens* genome portal (<http://www.plantmorphogenesis.bio.titech.ac.jp/>). All sequences were manually inspected to search for annotation errors. Retrieved sequences with long stretches of indeterminate bases were excluded from the analysis. In the case of sequences derived from transcriptomic surveys (namely, Onekp), care was taken so that only species with the longest contigs for all the genes searched were included in the analysis.

Flowering Time Analysis

For flowering time experiments, the plants were grown on soil at 22°C under standard long days (LD; 16-h light/8-h dark cycles; 80 $\mu\text{mol m}^{-2} \text{s}^{-1}$ of white light), 12:12 days (12-h light/12-h dark cycles; 80 $\mu\text{mol m}^{-2} \text{s}^{-1}$ of white light), continuous light (LL;

50 $\mu\text{mol m}^{-2} \text{s}^{-1}$ of white light) or short days (SD; 8-h light/16-h dark cycles; 140 $\mu\text{mol m}^{-2} \text{s}^{-1}$ of white light) depending on the experiment. Flowering time was estimated by counting the number of rosette leaves at the time of bolting. These experiments were performed in triplicate with $n = 16$ for each genotype. The statistical analysis was done using a two-tailed Student's *t*-test.

Hypocotyl Length Characterization

For hypocotyl length measurements, seedlings were grown on 0.8% agar under complete darkness, continuous red light (1 $\mu\text{mol m}^{-2} \text{s}^{-1}$), continuous blue light (1 $\mu\text{mol m}^{-2} \text{s}^{-1}$), continuous white light (LL), or under cycles of white light in SD or LD (all white light treatments 1 $\mu\text{mol m}^{-2} \text{s}^{-1}$). The final length of the hypocotyls was measured 4 days after germination. Light effects on hypocotyl elongation under continuous red and blue light were calculated normalizing hypocotyl length under each light regime to the hypocotyl length of the same genotype under constant dark conditions. For the white light experiments, absolute values of hypocotyl elongation are shown. These experiments were performed in triplicate with $n = 20$ for each genotype. The statistical analysis was done using a two-tailed Student's *t*-test.

Circadian Leaf Movement Analysis

For leaf movement analysis, plants were grown under 16-h light/8-h dark cycles until the appearance of the first pair of leaves. This period is referred to as the entrainment period. In order to measure circadian rhythms in leaf movement, plants were transferred to continuous white light (20 $\mu\text{mol m}^{-2} \text{s}^{-1}$) at 22°C. The position of the first pair of leaves was recorded every 2 h for 5–6 days using digital cameras, and the leaf angle was determined using ImageJ software (Schneider et al., 2012). Period estimates were calculated with Brass 3.0 software (Biological Rhythms Analysis Software System, available from <http://www.amillar.org>) and analyzed with fast Fourier transform nonlinear least squares (FFT-NLLS) using Brass 3.0 software. These experiments were performed in triplicate, with $n = 8$ for each genotype. The statistical analysis was done using a two-tailed Student's *t*-test.

Growth Conditions and Protocol Used for cDNA Library Preparation and High-Throughput Sequencing

Three biological replicates with seeds of wild-type (Col-0) and *prp40c-1* mutant allele were sown onto Murashige and Skoog medium containing 0.8% agarose, stratified for 4 days in the dark at 4°C and then grown at 22°C in continuous light. Whole plants were harvested after 12 days, and total RNA was extracted with RNeasy Plant Mini Kit (QIAGEN) following the manufacturer's protocols. To estimate the concentration and quality of samples, NanoDrop 2000c (Thermo Scientific) and the Agilent 2100 Bioanalyzer (Agilent Technologies) with the Agilent RNA 6000 Nano Kit were used, respectively. Libraries were prepared following the TruSeq RNA Sample Preparation Guide (Illumina). Briefly, 3 μg of total RNA was polyA-purified and fragmented, first-strand cDNA synthesized by

reverse transcriptase (SuperScript II, Invitrogen) using random hexamers. This was followed by RNA degradation and second-strand cDNA synthesis. End repair process and addition of a single A nucleotide to the 3' ends allowed ligation of multiple indexing adapters. Then, an enrichment step of 12 cycles of PCR was performed. Library validation included size and purity assessment with the Agilent 2100 Bioanalyzer and the Agilent DNA1000 kit (Agilent Technologies). Samples were pooled to create six multiplexed DNA libraries, which were pair-end sequenced with an Illumina HiSeq 1500 at INDEAR Argentina, providing 100-bp pair-end reads. Three replicates for each genotype were sequenced. Sequencing data have been uploaded to the Gene Expression Omnibus database and are available under accession number (GSE129932).

Processing of RNA Sequencing Reads

Sequence reads were mapped to *A. thaliana* TAIR10 (Lamesch et al., 2012) genome using TopHat v2.1.1 (Trapnell et al., 2009) with default parameters, except of maximum intron length set at 5,000. Count tables for different feature levels were obtained from bam files using custom R scripts and considering TAIR10 transcriptome.

Differential Gene Expression Analysis

Before differential expression analysis, we decided to discard genes with fewer than 10 reads on average per condition. Differential gene expression was estimated using the edgeR package version 3.4.2 (Robinson et al., 2010), and resulting *p* values were adjusted using a false discovery rate (FDR) criterion (Benjamini and Hochberg, 1995). Genes with FDR values <0.05 and absolute log two-fold change > 0.58 were deemed differentially expressed. Overlapping analysis were performed using Venny (Oliveros, 2007).

Differential Alternative Splicing

For the analysis of alternative splicing, the transcriptome was partitioned into subgenic joint features called “bins,” as proposed on DEXseq (Anders et al., 2012). Because of our special interest in new intron retention events, not only exons but also introns were considered in our analysis. The transcriptome was partitioned into 281,321 bins; 152,631 corresponding exclusively to exonic regions, 120,717 to intronic regions, and 7,973 to DNA regions directly associated with alternatively spliced events. We labeled these three kinds of bins as exon-bin, intron-bin, or AS-bins, respectively. In addition, AS-bins were further classified as defined by Mancini et al. (2019). For our analysis, we discarded bins from monoexonic genes and with mean count values lower than five reads per condition. We used edgeR exact test for the identification of differential use of bins corresponding to AS events or introns and FDR-corrected *p* values. We also computed read densities to have a relationship between the bin and its corresponding gene. Only genes with read densities > 0.05 in all genotypes were used for the analysis. AS events as well as all introns with an absolute log2 fold change (bin read density in the mutant/bin

read density in wild type) value >0.58, with FDR values <0.15 were deemed differentially spliced. Overlapping analysis was performed using Venny (Oliveros, 2007).

Functional Category Enrichment Analysis

Functional categories associated with specific groups of genes were identified using the BioMaps tool from the virtual plant software (<http://virtualplant.bio.nyu.edu/cgi-bin/vpweb>). This tool allowed us to determine which functional categories were statistically overrepresented in particular lists of genes compared to the entire genome (Katari et al., 2010). We analyzed 14 functional categories of our interest, and for each one, we determined the genes in common with our data sets, finally calculating a representation factor and the probability of finding an overlap simply by chance. The representation factor is the number of overlapping genes divided by the expected number of overlapping genes drawn from two independent groups. A representation factor > 1 indicates more overlap than expected by chance for two independent groups of genes or events, a representation factor <1 indicates less overlap than expected. The probability of each overlapping was determined using the hypergeometric probability formula.

Analysis of Splice-Site Sequences

To evaluate possible changes in the splice-site sequences of the most significantly affected splicing events in the *prp40c-1* mutants, we obtained the donor and acceptor splice site sequences of all the intron retention events that were deemed differentially spliced (absolute log2 fold value >0.58, with FDR values <0.15) in *prp40c-1* mutants compared to wild-type plants and compared them to the consensus splice-site sequences of the total 30,142 introns sequenced in our RNA-seq experiment. The frequency of each nucleotide for each position was obtained using custom R scripts and was represented using the R package Seqlogo (Bembom, 2014). The over- or underrepresentation of a particular nucleotide relative to its genome-wide frequency was determined, and a *p* value for the analysis was obtained using the hypergeometric test. The custom R scripts used here are available upon request.

Salt and Osmotic Stress Assays

For germination under salt stress assays, 40 seeds of wild-type (Col-0), *prp40c-1*, and *prp40c-2* plants, and three replicates of each, were sown on 3-mm filter paper (Whatman) and were imbibed in different concentrations of sodium chloride (NaCl) solution (30, 50, 70, and 90 mM) or distilled water. After sowing, the plates were irradiated with far-red light and stratified overnight. On the next day, the plates were irradiated with white light for 1 h and then transferred to dark conditions for 3 days at 22°C. A seed was considered as germinated when the embryonic root protruded through the seed envelope. These experiments were performed in triplicate. The statistical analysis was done using a two-way ANOVA. For growth and survival assays, seeds were germinated on 1/2 MS agar medium. In the case of the root growth assay, 4-day-old seedlings were transferred to 1/2 MS agar containing 100 mM of NaCl or

200 mM of D-mannitol (isosmotic concentrations) and grown vertically for 5 days. The plants were photographed after this period of growth, and their root length was assessed. These experiments were performed in triplicate with $n = 30$ for each genotype. For the NaCl tolerance assay, 4-day-old seedlings were transferred from the germination medium to 1/2 MS agar containing 0 or 160 mM of NaCl, and the survival rate was determined 30 days after the seedlings were transferred. These experiments were performed in triplicate with $n = 50$ for each genotype. The statistical analysis was done using a two-tailed Student's *t*-test.

Evaluation of Cold Tolerance

Seeds were sown onto 1/2 MS agar medium, stratified for 3 days at 4°C, and grown at 22°C under continuous light. Four-day-old seedlings were transferred to 1/2 MS agar plates and grown vertically for 5 days at 22°C (control) or for 10 days at 10°C under continuous light. The plants were photographed after this period of growth, and their root length was assessed. These experiments were performed in triplicate with $n = 30$ for each genotype.

Pathogen Infection Assay

Pseudomonas syringae pv. *maculicola* ES4326 strains were grown at 28°C with King's B medium (20 g protease peptone, 1.5 g K₂HPO₄, 6.09 ml 1 M MgSO₄, and 10 g glycerol per liter) supplemented with rifampicin (100mg/L) and kanamycin (50mg/L) for selection. Freshly cultured bacteria were collected and resuspended to a final concentration of OD₆₀₀ = 0.0002 in 10 mM MgSO₄. The bacterial solution was then pressured infiltrated with a 1-ml needleless syringe into the abaxial side of the 8th to 10th leaves of 5- to 6-week-old plants grown under SD conditions. Bacterial growth assays were performed at 48 h postinfection (hpi). Leaves were surface washed in sterile water before carrying out bacterial counts. A disc was punched from each leaf, and then, the three discs corresponding to each plant were placed in 750 µl of 10 mM MgCl₂ and crushed to release the bacteria. The resulting solution was serially diluted and spot plated on KB plates containing the corresponding antibiotics. The plates were incubated at 28°C for 48 h before counting of the colonies. Log-transformed bacterial growth was statistically analyzed. All bacterial infection experiments were repeated at least three times. The statistical analysis was done using a two-tailed Student's *t*-test.

PCR Alternative Splicing and qPCR Differential Expression Assessment

For alternative splicing and differential expression assessment, 12-day-old plants were grown in Murashige and Skoog 0.8% agar medium under continuous white light at 22°C. Total RNA was extracted using TRIzol reagent (Invitrogen, Carlsbad, CA, USA). One microgram of RNA was treated with RQ1 RNase-Free DNase (Promega, Madison, WI, USA) and subjected to retro-transcription with Super Script II Reverse Transcriptase (SSII RT) (Thermo Fisher Scientific, Waltham, MA, USA) and oligo-dT according to manufacturer's instructions. For alternative splicing, PCR amplification was performed using 1.5 U of Taq polymerase (Invitrogen). Primers used for amplification are detailed in

Supplementary Table S1. RT-PCR products were electrophoresed and detected by SYBR Green 2%. For qPCR differential expression assessment, cDNAs were then amplified with FastStart Universal SYBR Green Master (Roche, Basel, Switzerland) using the Mx3000P Real Time PCR System (Agilent Technologies, Santa Clara, CA, USA) cyclers. The IPP2 (AT3G02780) transcript was used as a housekeeping gene. Quantitative RT-PCR (qRT-PCR) analysis was conducted using the 2- $\Delta\Delta$ CT method (Schmittgen and Livak, 2008). Primer sequences and conditions are available on **Supplementary Table S1.** Three biological samples were measured for each genotype. The statistical analysis was done using a two-tailed Student's *t*-test.

RESULTS

Phylogenetic Analysis of the Pre-mRNA Processing Protein 40 Family

We initially performed a phylogenetic analysis of the pre-mRNA processing protein 40 family (PRP40) (**Figure 1A**). We found that PRP40s from both human and *Arabidopsis* genomes cluster in two separate clades, one that includes the so-called PRP40A and B types and another that includes HsTCERG1/CA150 and AtPRP40C. In order to assess whether this clustering reflects an ancestral origin, or may be an effect of sequence affinity unrelated to descent, we retrieved and analyzed plant PRP40 sequences sampled from phylogenetically relevant species (**Figure 1B**). We found that members of the PRP40A/B clade are found in all Viridiplantae, both uni- and multicellular. However, the occurrence of A and B forms appear to be restricted to eudicots, since neither monocots nor basal angiosperms show this type of duplication. PRP40C-type sequences are also widespread among plants, usually found as single copy genes. Still, we could not find *bona fide* homologues in Chlorophytes; consequently, PRP40C genes appear to be restricted to streptophytes. Taken together, our analysis shows that 1) type A PRP40s appear to be an ancestral feature of eukaryotic organisms, 2) the occurrence of duplicated forms in this clade (types A and B) is lineage specific, and 3) genes of the HsTCERG1/PRP40C type may also have independent origins; in Viridiplantae, they appear to be an ancient innovation of streptophytes.

Thus, both AtPRP40C and HsTCERG1/CA150 appear to have diverged from the others members of the family in the respective organism. This observation, together with the fact that HsTCERG1/CA150 is a well-characterized protein that links transcription and pre-mRNA processing, suggested that this gene could have evolved to modulate gene expression networks in complex eukaryotic organisms and prompted us to analyze its biological roles and molecular functions in *A. thaliana*.

PRP40C Mutants Display a Late Flowering Phenotype and Mild Alterations in Other Light Controlled Process

In order to characterize the role of AtPRP40C (hereinafter referred to as PRP40C) using a reverse genetics approach, we identified plants with T-DNA insertions in the *PRP40C* loci, *prp40c-1* (SALK_148319), and *prp40c-2* (SALK_205357). Both mutants

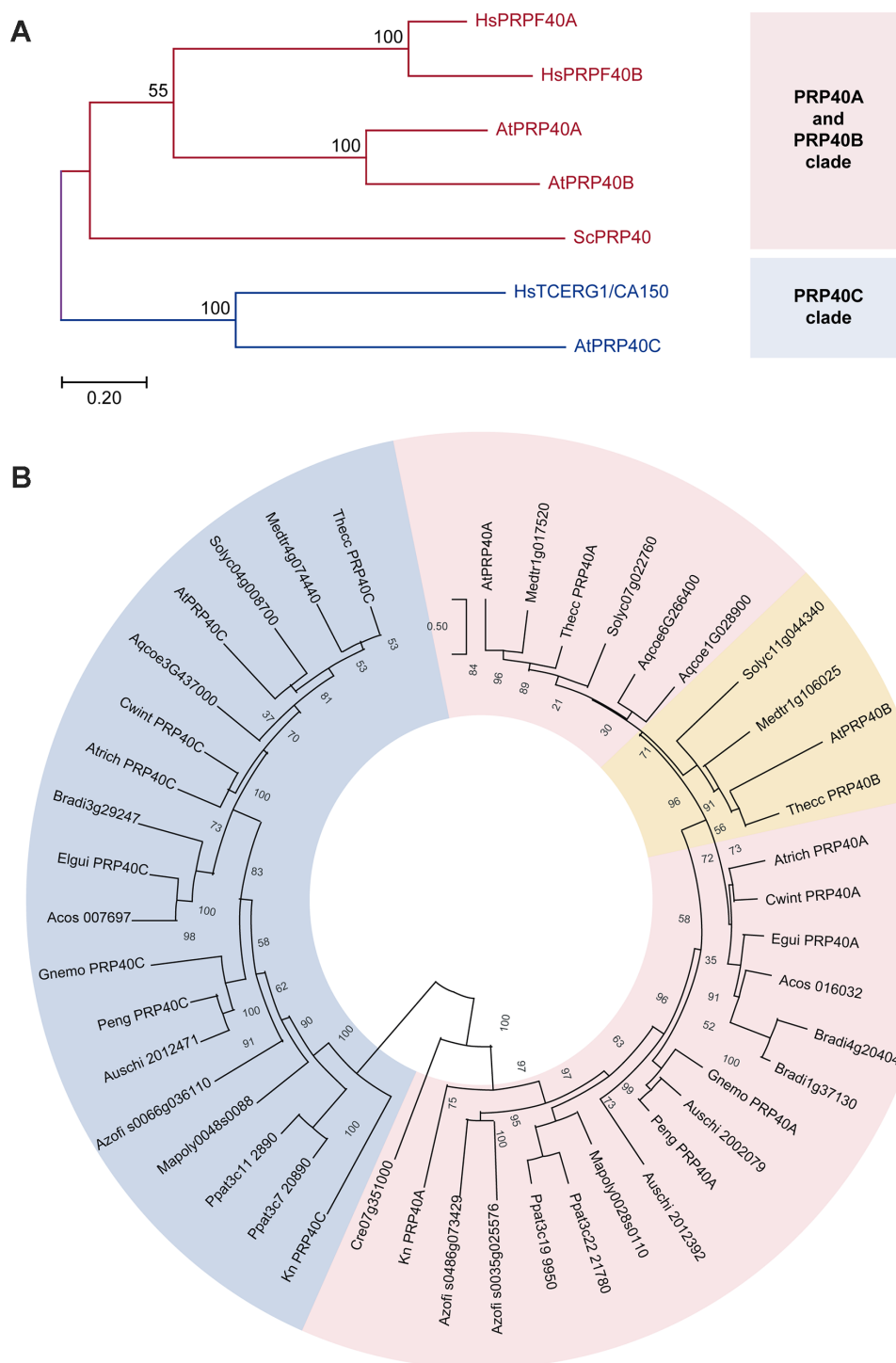


FIGURE 1 | Phylogenetic analysis of the PRP40 family. Pink and blue shades indicate the PRP40A and PRP40C groups, respectively. Numbers indicate the percentage support for each node with 1,000 bootstrap iterations. **(A)** Comparison between the corresponding sequences of yeast (Sc, *Saccharomyces cerevisiae*), humans (Hs, *Homo sapiens*), and Arabidopsis (At, *Arabidopsis thaliana*). **(B)** Phylogenetic relationships of the PRP40 family in the Viridiplantae. The eudicot-specific duplication between A- and B-type sequences is indicated in orange; note that the two sequences from *A. coerulea*, a basal eudicot, do not clearly cluster with either group. Species identification: Medtr, *Medicago truncatula*; Thecc, *Theobroma cacao*; Solyc, *Solanum lycopersicum*; Aqcoe, *Aquilegia coerulea*; Atrich, *Amborella trichopoda*; Cwint, *Canella winterana*; Egui, *Elaeis guineensis*; Acos, *Ananas cosmosus*; Bradi, *Brachipodium distachyon*; Gnemo, *Gnetum montanum*; Auschi, *Austrocedrus chilensis*; Peng, *Picea engelmannii*; Azofi, *Azolla filiculoides*; Ppat, *Physcomitrella patens*; Mapoly, *Marchantia polymorpha*; Kn, *Klebsormidium nitens*; Cr, *Chlamydomonas reinhardtii*.

were genotyped to verify the homozygosity of the T-DNA insertion, and null expression of the wild-type mRNA was evaluated by RT-PCR in both mutant alleles. We then conducted an analysis of the effect of PRP40C in the control of several developmental and physiological processes. Initially, we did not observe severe phenotypic alterations in growth and morphology during the vegetative stage (**Supplementary Figure S1**). Then, we studied flowering time, a key developmental trait that marks the transition from the vegetative to the reproductive stage. We found that the *prp40c* mutants flower later than wild-type plants in both 12-h light/12-h dark and 16-h light/8-h dark long day photoperiods (**Figure 2A, B**). No delay in flowering time was observed in a short-day photoperiod or under continuous light (**Supplementary Figure S2**). Light perception and the circadian clock are two key factors that ensure proper flowering time regulation. To evaluate if any of these could be related to the late flowering phenotype, we analyzed photomorphogenic responses and clock function in the *prp40c* mutants. Seedling photomorphogenesis was assessed measuring the inhibition of hypocotyl elongation under different light treatments. Interestingly, *prp40c* mutants displayed shorter hypocotyls than wild-type plants under continuous red light, indicating hypersensitivity to this wavelength (**Figure 2C**). No differences were detected under continuous blue light (**Figure 2D**). We then assessed the effect of different photoperiods on hypocotyl elongation. No significant differences were detected between wild-type plants and *prp40c* mutants in any photoperiodic condition under whitelight (**Supplementary Figure S3**). Finally, we monitored circadian rhythms in leaf movements in wild-type plants as well as in *prp40c-1* and *prp40c-2* mutants. The rhythms observed for both mutant alleles were similar to those of wild-type plants, exhibiting a 24.5-h period (**Figure 2E, F**). These data suggest that the circadian function is unaffected, at least under standard growth conditions, in these plants. Collectively, our results indicate that PRP40C has a role in flowering time control, which could be associated, at least in part, with alterations in the red light signaling pathway.

Impact of PRP40C on Genome-Wide Gene Expression and Pre-mRNA Splicing

In order to study the regulatory impact of PRP40C on gene expression, we analyzed the transcriptome of wild-type and *prp40c-1* mutant plants grown under standard nonstressful conditions (continuous white light at 22°C) using RNA-seq. We found 869 differentially expressed genes (DEG) in *prp40c-1* mutants relative to wild-type plants, 642 overexpressed (73.9%) and 227 underexpressed (26.1%). To characterize the role of PRP40C in pre-mRNA splicing, we evaluated both constitutive (CS) as well as alternative splicing (AS) events, which included annotated or novel AS events. We found a total of 680 differentially spliced transcripts (DST) in *prp40c-1* mutants when compared to wild-type plants. These events occurred on 553 transcripts, indicating that only a few transcripts had more than one differentially spliced event. Interestingly, only 39 transcripts were differentially expressed and differentially spliced at the same time (**Figure 3A**). We next studied the abundance of the different annotated AS events among the differentially spliced transcripts. In wild-type plants, the most

abundant AS events were those associated with the use of alternative 3' splice sites (Alt 3'SS, 36.4%), followed by intron retention (IR, 32%), alternative 5' splice sites (Alt 5'SS, 20.3%), and exon skipping (ES, 11.2%). Among the annotated AS events affected in *prp40c-1*, we found an increase in the proportion of both IR events (IR, 47.4%) and alternative 5' splice sites (Alt 5'SS, 31.6%) and a decrease in alternative 3' splice sites (Alt 3'SS, 18.4%) and exon skipping events (ES, 2.6%), relative to their frequency in wild-type plants (**Figure 3B**). We also evaluated splicing of all introns present in expressed genes. Among the 105,555 introns analyzed, we detected alterations in splicing of 397 introns from which 18 were already annotated as alternatively spliced, and 379 had no previous evidence of being alternatively spliced (i.e., they were considered constitutively spliced introns). Interestingly, the proportion of increased intron retention events detected in the *prp40c-1* mutants among alternatively spliced introns was larger than the proportion observed among constitutively spliced introns (**Figure 3C**). This observation indicates that PRP40C acts as a splicing modulator rather than as an essential splicing factor. We also evaluated whether there was any change in the splice-site sequences of the intron retention events affected in *prp40c-1* mutants compared to the consensus sequence of all the introns belonging to genes expressed in our RNA-seq experiment. Interestingly, no significant deviation from the consensus sequence was observed for the donor splice site or for the acceptor splice site for the introns whose splicing was affected in *prp40c-1* mutants (**Supplementary Figure S4**). In order to get a broader landscape of the changes in the transcriptome induced by PRP40C, both the DEG and the DST were categorized into functional groups based on Gene Ontology (GO). Fourteen functional categories of our interest were examined in detail determining the representation factor for each category (data available in **Supplementary Table S2**). The representation factor (RF) is the number of overlapping genes observed in the selected group divided by the expected number of overlapping genes drawn randomly from two independent groups. Among the upregulated genes, we found a significant enrichment (representation factor > 1; $p < 0.05$) for categories corresponding to RNA metabolic process, response to light stimulus, response to hormone stimulus, response to stress, response to salt stress, immune response, signal transduction, phosphorylation, sequence-specific DNA binding transcription factor activity, and signal transducer activity. For downregulated genes, no enrichment was found among the categories studied. Among the differentially spliced genes, significant enrichment was found for categories such as primary metabolic process, response to stress, response to salt stress, phosphorylation, metal ion transport, ion transmembrane transporter activity, and protein kinase activity (**Figure 3D**). None of the categories studied here displayed a statistically significant under-representation (representation factor < 1; $p < 0.05$). These data taken together indicate that, instead of being a global transcription and pre-mRNA splicing regulator, PRP40C regulates transcription and splicing of a defined subset of transcripts particularly related to biotic and abiotic stress responses.

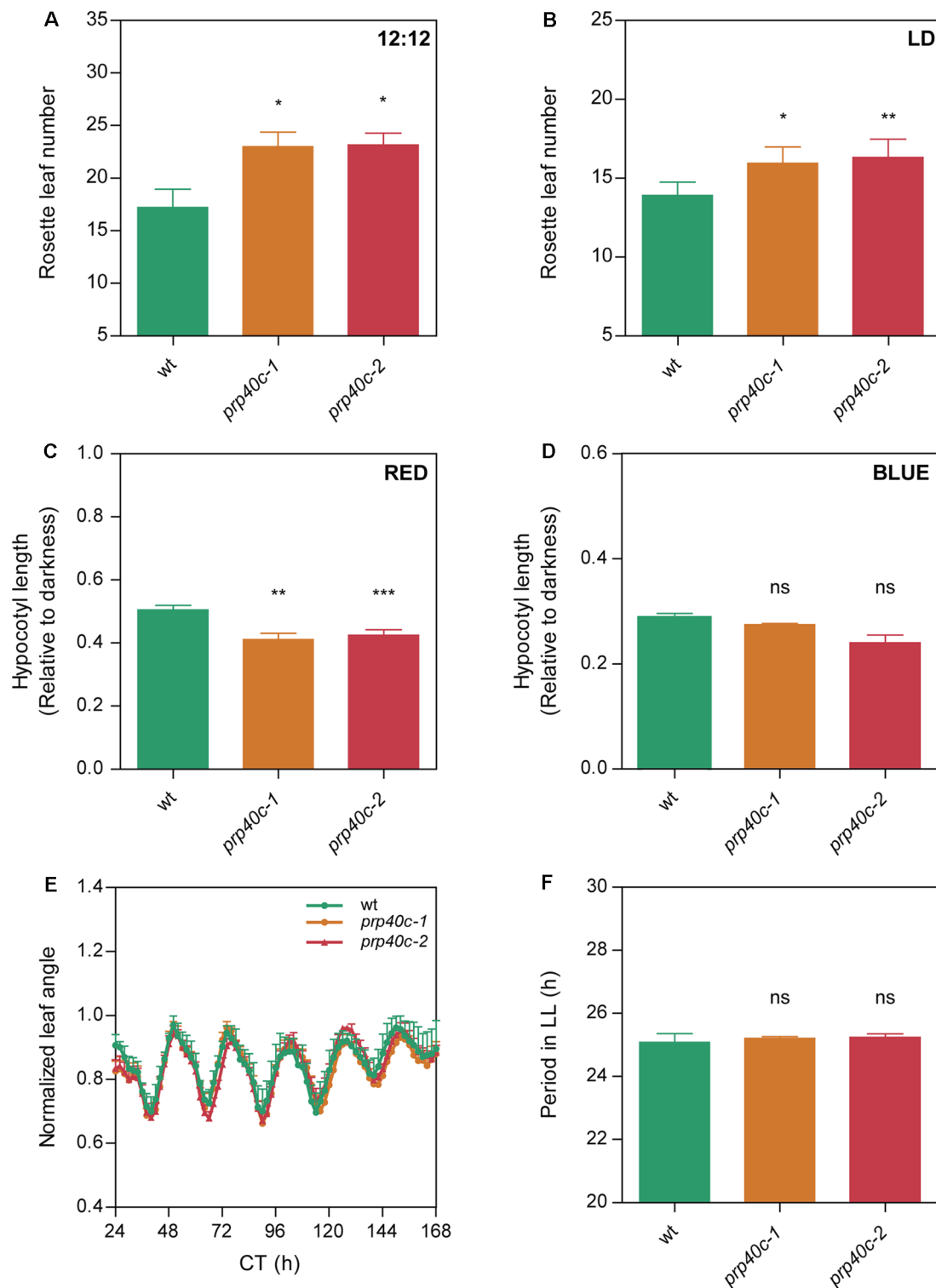


FIGURE 2 | PRP40C plays a role in the photoperiodic control of flowering and photomorphogenesis. Flowering time measured as the number of rosette leaves at bolting in (A) 12-h light/12-h dark cycles (LD; 12:12), (B) long days (LD; 16:8). Hypocotyls of WT and PRP40C mutants grown under different wavelengths, measurements are expressed relative to the dark control. (C) Continuous red light. (D) Continuous blue light. (E) Circadian rhythms of leaf movement in continuous light (LL), after entrainment under long-day conditions. (F) Periods of circadian rhythms in leaf movement were estimated with BRASS 3.0 software. Error bars indicate SEM. Student's *t*-test was performed between mutants and wild type (*significantly different, $p \leq 0.05$; **significantly different, $p \leq 0.01$; ***significantly different, $p \leq 0.001$; ns, not significant).

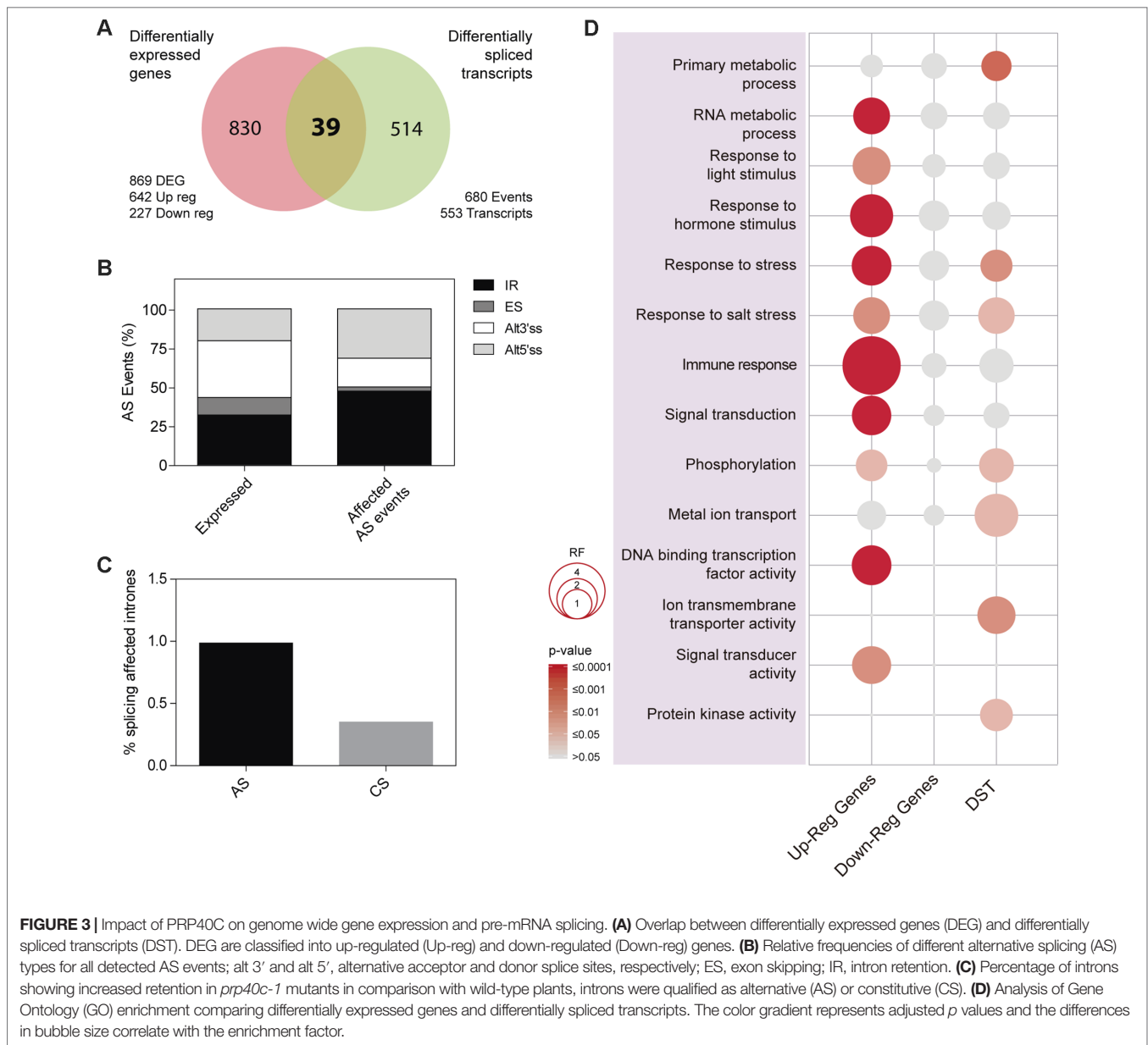


FIGURE 3 | Impact of PRP40C on genome wide gene expression and pre-mRNA splicing. **(A)** Overlap between differentially expressed genes (DEG) and differentially spliced transcripts (DST). DEG are classified into up-regulated (Up-reg) and down-regulated (Down-reg) genes. **(B)** Relative frequencies of different alternative splicing (AS) types for all detected AS events; alt 3' and alt 5', alternative acceptor and donor splice sites, respectively; ES, exon skipping; IR, intron retention. **(C)** Percentage of introns showing increased retention in *prp40c-1* mutants in comparison with wild-type plants, introns were qualified as alternative (AS) or constitutive (CS). **(D)** Analysis of Gene Ontology (GO) enrichment comparing differentially expressed genes and differentially spliced transcripts. The color gradient represents adjusted *p* values and the differences in bubble size correlate with the enrichment factor.

PRP40C Plays a Role in Both Salt and Biotic Stress Tolerance

The global transcriptome analysis revealed a significant enrichment for genes related to salt and biotic stress responses among the differentially expressed and differentially spliced transcripts in *prp40c-1* mutants. We therefore performed physiological analyses of the mutant plants under both stress conditions. We first compared germination rates between wild-type plants and *prp40c* mutants under salt stress. There was no difference in germination rates between wild-type, *prp40c-1*, and *prp40c-2* seeds germinated on filter paper embedded with water, whereas the germination rate significantly decreased in *prp40c* mutants when the seeds were placed on a sodium chloride (NaCl) solution (Figure 4A). We next examined the root growth of wild-type plants and *prp40c* mutants under salt (NaCl), mannitol, and cold stress conditions, as well

as survival rate under salt stress. No significant differences were found between genotypes in these experiments (Supplementary Figure S5). These results suggest that PRP40C is essential for proper germination under salt stress but is probably not involved in NaCl, mannitol, and cold tolerance in the early vegetative stage. In order to assess if *prp40c* mutants display alterations in response to biotic stress, we performed an infection assay with *Pseudomonas syringae* pv. *maculicola* ES4326. Leaves from 5-week-old wild-type and *prp40c* plants were pressure infiltrated with the virulent bacterium. Bacterial growth assays at 2 days postinfection (2 dpi) revealed a significant degree of enhanced resistance in both *prp40c-1* and *prp40c-2* mutant alleles when compared to wild-type plants (Figure 4B). Finally, we confirmed by RT-PCR some of the alterations in splicing detected in transcripts related to abiotic and biotic stress in the RNA-seq experiment (Figure 5).

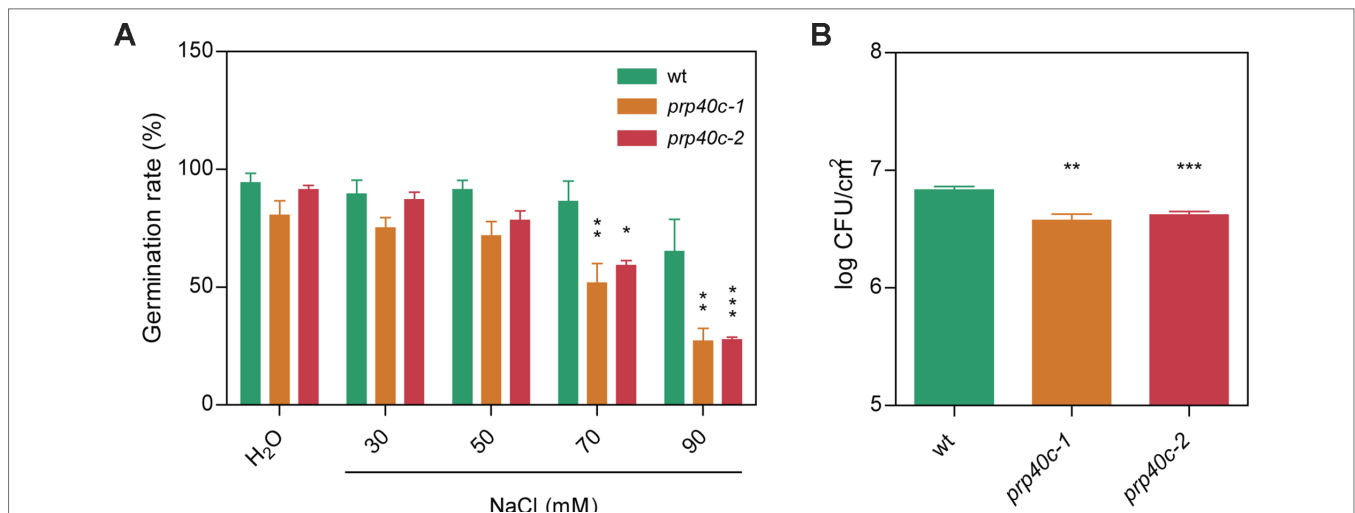


FIGURE 4 | *prp40c* mutants show altered responses to both salt and biotic stress. **(A)** Seed germination rates in pure water and different NaCl concentrations were quantified. **(B)** Five- to six-week-old plants grown in short-day conditions (SD; 8-h light/16-h darkness) were infected by infiltration with *Pseudomonas syringae* pv. *maculicola* ES4326. Bacterial growth was assessed at 2 dpi. CFU, colony-forming units. Data represent the average of log-transformed bacterial growth. Error bars indicate SEM. An ANOVA followed by a Bonferroni comparison test was performed between mutants and wild type for the salt stress experiment; a Student's *t*-test was performed between mutants and wild type for the infection assays (*significantly different, $p \leq 0.05$; **significantly different, $p \leq 0.01$; ***significantly different, $p \leq 0.001$; ns, not significant).

DISCUSSION

Proper control of gene expression and pre-mRNAs processing is crucial for eukaryotic organisms, and, indeed, many pre-mRNA processing events occur cotranscriptionally. PRP40 proteins appear to couple transcript elongation to the control of pre-mRNA splicing, but the global extent of their contribution to the regulation of gene expression networks, as well as their physiological roles, have only recently started to be characterized in higher eukaryotes (Munoz-Cobo et al., 2017). The founding member of this family (ScPRP40) was identified in yeast (Kao and Siliciano, 1996) and shown to be an essential splicing factor. Interestingly, PRP40 constitutes a small gene family in most eukaryotes, suggesting that duplication of the ancestral PRP40 gene followed by sequence divergence might have contributed to the evolution of novel molecular and/or physiological roles for some members of this protein family (Becerra et al., 2016). A comparison between PRP40s from *Saccharomyces cerevisiae*, *Homo sapiens*, and *A. thaliana* showed that A-type PRP40s are probably an ancestral feature of eukaryotic cells. Interestingly, AtPRP40C clusters with human TCERG1/CA150, but this clustering most likely reflects the divergent sequences in this group compared to other family members, rather than a common evolutionary origin. Still, C-type PRP40s proved to be conserved across all plant species, including streptophyte algae. This prompted us to explore a possible role for AtPRP40C as a modulator of gene expression programs, which could act linking transcription and splicing to optimize plant growth and development.

Loss-of-function *prp40c* mutants were viable and showed no severe phenotypical alterations compared to wild-type plants, suggesting that indeed PRP40C is a modulator of gene expression networks rather than an essential splicing factor. Whether

other PRP40 genes are essential in *Arabidopsis* remains to be determined. It is possible that different members of this gene family play both redundant as well as specific functions in the control of gene expression and that the essentiality of the PRP40 gene family could only be observed in a triple mutant lacking all PRP40 homologues.

In support of a modulatory function for PRP40C, *prp40c* mutants displayed a late flowering phenotype only under long-day photoperiods, revealing that the mutant does not have a global developmental alteration, but rather a defect in the environmental control of specific developmental traits. Furthermore, the finding that *prp40c* mutants only flower late under long-day conditions reveals that the photoperiodic flowering pathway may be specifically affected by this mutation; this well-known signaling pathway is controlled by light perception and the circadian clock through their effects on gene expression. *prp40c* mutants in fact exhibit hypersensitivity to red light but not alterations in circadian clock function, suggesting that a defect in light signaling rather than in time measurement could be associated with its altered flowering time phenotype.

To analyze the molecular cause of the late flowering phenotype, we looked for genes related to the control of the photoperiodic flowering pathway among the DEG in the mutant. We found four candidates: AT1G01060 (*LHY*), AT2G46830 (*CCA1*), AT1G25560 (*TEM1*), and AT1G68840 (*TEM2*). Given that the circadian function was not affected in the mutant, the flowering phenotype was most likely due, at least in part, to the overexpression of *TEM1* and *TEM2*, rather than to changes in *LHY* and *CCA1* clock genes (**Supplementary Figure S6**). In fact, the transcription factors *TEM1* and *TEM2* are floral repressors of the photoperiodic and the gibberellin

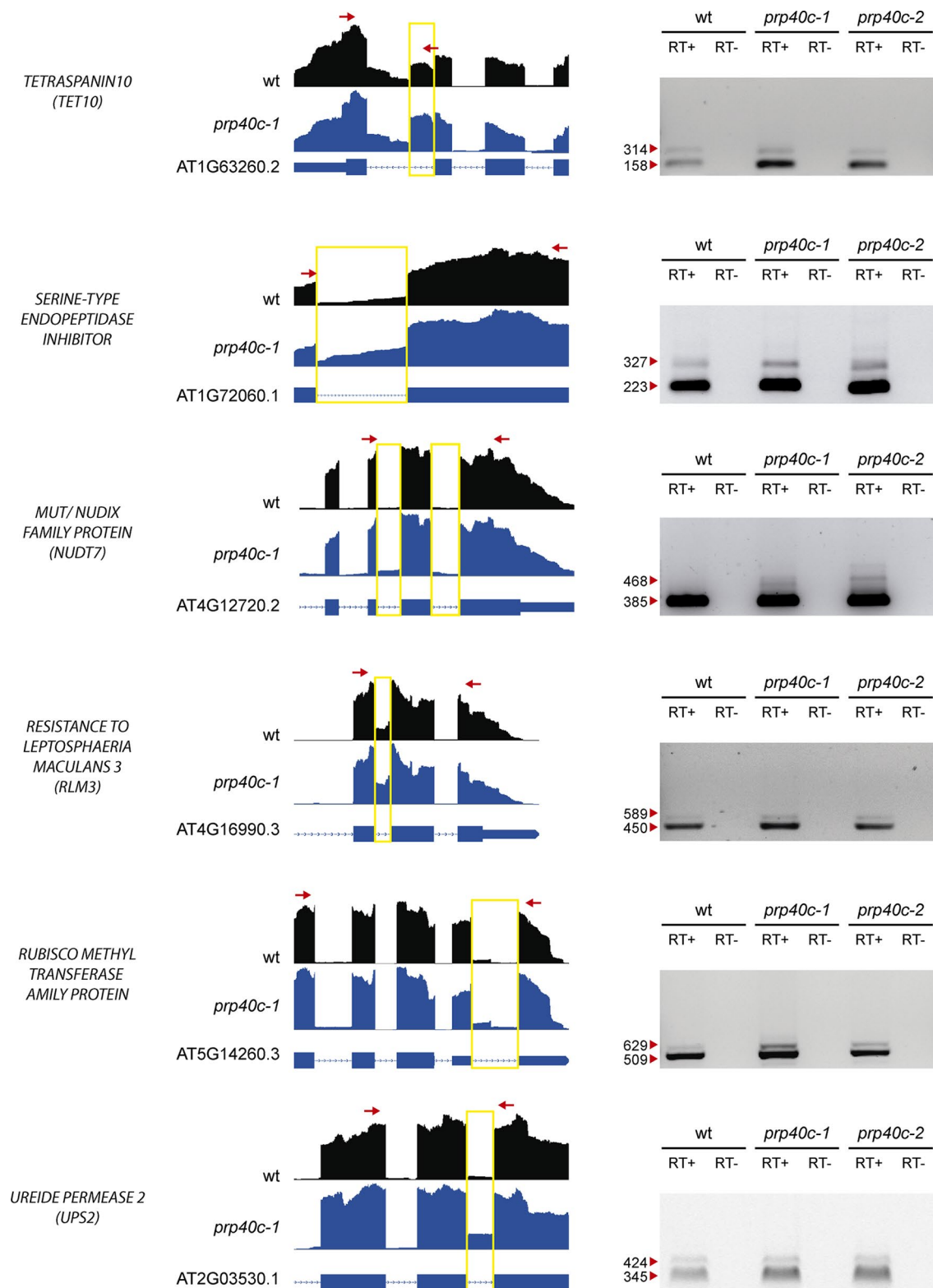


FIGURE 5 | Analysis of alternative splicing events. RT-PCR validation of six events identified as alternatively spliced through RNA-seq. The read density map for each event evaluated is displayed (WT: black, *prp40c-1*: blue). A scheme describing each gene is displayed below the read density maps, with exons and introns displayed as boxes and lines, respectively. A yellow square encloses the measured event, while red arrows display the position of the oligos used for the RT-PCR measurement. An image of the agarose gel with the RT-PCR amplicons is displayed next to the read density maps. Red arrowheads indicate amplicon sizes. +, retrotranscriptase added; –, retrotranscriptase not added.

flowering induction pathways (Osnato et al., 2012; Matias-Hernandez et al., 2014).

Almost three quarters of the differentially expressed genes in *prp40c-1* mutants were upregulated when compared to the wild type, suggesting that PRP40C may be acting as a negative transcriptional regulator. It has been proposed that HsTCERG1/CA150 is a promoter-specific negative regulator of RNA Pol II transcription elongation, and it is possible that plant PRP40C proteins have a similar repressive role (see **Supplementary Table S2**) (Sune et al., 1997; Sune and Garcia-Blanco, 1999).

Regarding alterations in pre-mRNA splicing in *prp40c-1* mutants, we detected 680 events as differentially spliced between mutant and wild-type plants. Although this is a significant number of differentially spliced events, it is relatively small compared with the numbers of altered events that we have previously observed in other *Arabidopsis* mutants affected in core spliceosomal components, such as *lsm5* (6,049 events) and *lsm4* (6,377 events), or in mutants affected in spliceosome assembly, such as *gemin2* (6680 events) (Perez-Santangelo et al., 2014; Schlaen et al., 2015). Among the different types of annotated AS events, IR was clearly overrepresented in the group of events that were differentially spliced between *prp40c* mutants and wild-type plants. In addition, *prp40c* had a stronger impact on the splicing of alternative compared to constitutive introns. These results, taken together, suggest that AtPRP40C is a splicing modulator, which affects a specific subset of splicing events, rather than an essential splicing factor.

In contrast to what was reported for HsPRPF40B, we found no significant deviations in the sequences of the global consensus donor and acceptor splice sites compared to the sequences of donor or acceptors sites present in the differentially spliced events (i.e., the affected events were not associated with weak splice site signals). This observation suggests that AtPRP40C may contribute to the regulation of pre-mRNA splicing events through a mechanism that is at least partially different from its previously proposed role in the stabilization of weak RNA–RNA interactions (Becerra et al., 2015).

When DEG and the DST were categorized into functional groups several interesting results came to light. Among the DEG, upregulated genes were enriched in several GO categories (**Supplementary Table S2**). Some of the overrepresented categories may help to explain, at least in part, some of the phenotypes observed in the mutants, such as red light hypersensitivity (Response to light stimulus) and the splicing alterations (RNA metabolic process). This suggests that some of the splicing alterations observed in the *prp40c* mutants could take place through the differential expression of other splicing regulators rather than through a direct action of PRP40C on pre-mRNA processing.

Other GO terms significantly overrepresented were immune response (RF = 4.45), response to hormone stimulus (RF = 2.4), and response to stress (RF = 1.9). Particularly, we noticed that among the 34 upregulated genes related to

the immune response ontology, 21 (60%) were also related to signal transduction (RF = 1.9). Among the DST signal-transduction related genes that were affected, categories such as phosphorylation (RF = 1.45) and protein kinase activity (RF = 1.38) were significantly enriched. Abiotic stress-related GO categories such as salt stress (RF = 1.73), metal ion transport (RF = 2.46), and ion transmembrane transporter activity (RF = 1.80) were also significantly enriched among DST. The ion-transport related categories were enriched among the DST but not among DEG, suggesting a level of specificity for AtPRP40C in splicing regulation.

Both biotic and abiotic stresses were differentially affected in *prp40c* mutants when compared to wild-type plants. Interestingly, seed germination was affected in the mutants under salt stress, but growth was not affected by this stress at the vegetative stage. In addition, cold stress during the early vegetative stage did not affect *prp40c* root growth. According to the data available in the Bio-Analytic Resource for Plant Biology of the University of Toronto (Toronto BAR—www.bar.utoronto.ca/), AtPRP40C expression is highest at the dry seed stage (**Supplementary Table S3**), which may explain why PRP40C plays a role in salt tolerance during germination. Nonetheless, a more thorough evaluation of stress tolerance phenotypes throughout a variety of life stages is needed to unveil in more detail the regulatory role of PRP40C in abiotic stress responses. On the other hand, *prp40c* mutants proved to be more resistant to *P. syringae* infection. We found that 15 out of the 34 DEG (44%) grouped in the immune response GO were TIR-NBS-LRR (Toll interleukin1 receptor–nucleotide binding site–leu-rich repeat) class of R (resistance) proteins, and some of these TIR-NBS-LRR proteins are known to enhance disease resistance when overexpressed (Parker et al., 1996; Tang et al., 1999; Stokes et al., 2002). These data suggest that AtPRP40C may regulate this specific group of genes negatively in order to orchestrate an appropriate defense response. Interestingly, according to the GENEVESTIGATOR database, the expression of the *PRP40C* gene appears to be modulated by both abiotic (salt and cold) as well as biotic (*P. syringae*) stresses, suggesting that PRP40C acts as a node within the stress response gene regulatory network in *Arabidopsis* (**Supplementary Figure S7**) (Hruz et al., 2008).

In summary, our work constitutes the first physiological and molecular characterization of a member of the PRP40 protein family in plants and the most thorough characterization of the global effects of PRP40 on gene expression and pre-mRNA splicing in eukaryotic organisms. We provide evidence that PRP40C is an important factor linking the regulation of gene expression programs to the modulation of plant growth, development, and biotic and abiotic stress responses. We propose that *PRP40C* evolved from an ancestral *PRP40* gene that had an essential role in the control of pre-mRNA splicing, into a global modulator of gene expression and splicing that targets specific genes and processes to improve physiological adjustments to environmental challenges. Although it is tempting to speculate that PRP40C modulates gene expression by coupling the regulation of transcription rates to the control of

pre-mRNA splicing, its precise mechanism of action remains to be determined.

DATA AVAILABILITY

The datasets generated for this study can be found in the Gene Expression Omnibus (<https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE129932>).

AUTHOR CONTRIBUTIONS

Most of the investigation and experiments were performed by CEH and MGH. Pathogen infection assay was performed by MJdL and PCR alternative splicing and qPCR differential expression assessment was performed by DC. RNA-Seq data analysis was performed by JI and sequence retrieval and phylogenetic analysis was performed by SMG. Writing, review, and editing were done by CEH, SMG, and MJY. All authors read and approved the final manuscript.

REFERENCES

- Abovich, N., and Rosbash, M. (1997). Cross-intron bridging interactions in the yeast commitment complex are conserved in mammals. *Cell* 89 (3), 403–412. doi: 10.1016/S0092-8674(00)80221-4
- Allison, L. A., Wong, J. K., Fitzpatrick, V. D., Moyle, M., and Ingles, C. J. (1988). The C-terminal domain of the largest subunit of RNA polymerase II of *Saccharomyces cerevisiae*, *Drosophila melanogaster*, and mammals: a conserved structure with an essential function. *Mol. Cellular Biol.* 8 (1), 321–329. doi: 10.1128/MCB.8.1.321
- Anders, S., Reyes, A., and Huber, W. (2012). Detecting differential usage of exons from RNA-seq data. *Genome Res.* 22 (10), 2008–2017. doi: 10.1101/gr.133744.111
- Becerra, S., Andres-Leon, E., Prieto-Sanchez, S., Hernandez-Munain, C., and Sune, C. (2016). Prp40 and early events in splice site definition. *Wiley Interdiscip. Rev. RNA* 7 (1), 17–32. doi: 10.1002/wrna.1312
- Becerra, S., Montes, M., Hernandez-Munain, C., and Sune, C. (2015). Prp40 pre-mRNA processing factor 40 homolog B (PRPF40B) associates with SF1 and U2AF65 and modulates alternative pre-mRNA splicing *in vivo*. *RNA* 21 (3), 438–457. doi: 10.1261/rna.047258.114
- Bedford, M. T., and Leder, P. (1999). The FF domain: a novel motif that often accompanies WW domains. *Trends Biochem. Sci.* 24 (7), 264–265. doi: 10.1016/S0968-0004(99)01417-6
- Bembom, O. (2014). “seqLogo: Sequence Logos for DNA Sequence Alignments. R Package Version 1.32.1.” in *R Foundation for Statistical Computing*.
- Benjamini, Y., and Hochberg, Y. (1995). Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J. R. Stat. Soc. Series B Stat. Methodol.* 57 (1), 289–300. doi: 10.1111/j.2517-6161.1995.tb02031.x
- Braunschweig, U., Gueroussou, S., Plocik, A. M., Graveley, B. R., and Blencowe, B. J. (2013). Dynamic integration of splicing within gene regulatory pathways. *Cell* 152 (6), 1252–1269. doi: 10.1016/j.cell.2013.02.034
- Buratowski, S. (2003). The CTD code. *Nat. Struct. Biol.* 10, 679. doi: 10.1038/nsb0903-679
- Carty, S. M., Goldstrohm, A. C., Sune, C., Garcia-Blanco, M. A., and Greenleaf, A. L. (2000). Protein-interaction modules that organize nuclear function: FF domains of CA150 bind the phosphoCTD of RNA polymerase II. *Proc. Natl. Acad. Sci. U. S. A.* 97 (16), 9015–9020. doi: 10.1073/pnas.160266597
- Darnell, J. E., Jr. (2013). Reflections on the history of pre-mRNA processing and highlights of current knowledge: a unified picture. *RNA* 19 (4), 443–460. doi: 10.1261/rna.038596.113
- Deckert, J., Hartmuth, K., Boehringer, D., Behzadnia, N., Will, C. L., Kastner, B., et al. (2006). Protein composition and electron microscopy structure of affinity-purified human spliceosomal B complexes isolated under physiological conditions. *Mol. Cell. Biol.* 26 (14), 5528–5543. doi: 10.1128/MCB.00582-06
- Deng, X., and Cao, X. (2017). Roles of pre-mRNA splicing and polyadenylation in plant development. *Curr. Opin. Plant Biol.* 35, 45–53. doi: 10.1016/j.pbi.2016.11.003
- Ding, F., Cui, P., Wang, Z., Zhang, S., Ali, S., and Xiong, L. (2014). Genome-wide analysis of alternative splicing of pre-mRNA under salt stress in *Arabidopsis*. *BMC Genom.* 15, 431. doi: 10.1186/1471-2164-15-431
- Filichkin, S., Priest, H. D., Megraw, M., and Mockler, T. C. (2015). Alternative splicing in plants: directing traffic at the crossroads of adaptation and environmental stress. *Curr. Opin. Plant Biol.* 24, 125–135. doi: 10.1016/j.pbi.2015.02.008
- Goldstrohm, A. C., Albrecht, T. R., Sune, C., Bedford, M. T., and Garcia-Blanco, M. A. (2001). The transcription elongation factor CA150 interacts with RNA polymerase II and the pre-mRNA splicing factor SF1. *Mol. Cell. Biol.* 21 (22), 7617–7628. doi: 10.1128/MCB.21.22.7617-7628.2001
- Greenleaf, A. L. (1993). Positive patches and negative noodles: linking RNA processing to transcription? *Trends Biochem. Sci.* 18 (4), 117–119. doi: 10.1016/0968-0004(93)90016-G
- Hruz, T., Laule, O., Szabo, G., Wessendorp, F., Bleuler, S., Oertle, L., et al. (2008). Genevestigator v3: a reference expression database for the meta-analysis of transcriptomes. *Adv. Bioinformatics* 2008, 420747. doi: 10.1155/2008/420747
- Kang, C. H., Feng, Y., Vikram, M., Jeong, I. S., Lee, J. R., Bahk, J. D., et al. (2009). *Arabidopsis thaliana* PRP40s are RNA polymerase II C-terminal domain-associated proteins. *Arch. Biochem. Biophys.* 484 (1), 30–38. doi: 10.1016/j.abb.2009.01.004
- Kao, H. Y., and Siliciano, P. G. (1996). Identification of Prp40, a novel essential yeast splicing factor associated with the U1 small nuclear ribonucleoprotein particle. *Mol. Cell. Biol.* 16 (3), 960–967. doi: 10.1128/MCB.16.3.960
- Katari, M. S., Nowicki, S. D., Aceituno, F. F., Nero, D., Kelfer, J., Thompson, L. P., et al. (2010). VirtualPlant: a software platform to support systems biology research. *Plant Physiol.* 152 (2), 500–515. doi: 10.1104/pp.109.147025
- Kornblihtt, A. R., Schor, I. E., Allo, M., Dujardin, G., Petrillo, E., and Munoz, M. J. (2013). Alternative splicing: a pivotal step between eukaryotic transcription and translation. *Nat. Rev. Mol. Cell. Biol.* 14 (3), 153–165. doi: 10.1038/nrm3525
- Kumar, S., Stecher, G., Li, M., Knyaz, C., and Tamura, K. (2018). MEGA X: Molecular Evolutionary Genetics Analysis across computing platforms. *Mol. Biol. Evol.* 35 (6), 1547–1549. doi: 10.1093/molbev/msy096
- Kwon, Y. J., Park, M. J., Kim, S. G., Baldwin, I. T., and Park, C. M. (2014). Alternative splicing and nonsense-mediated decay of circadian clock genes under environmental stress conditions in *Arabidopsis*. *BMC Plant Biol.* 14, 136. doi: 10.1186/1471-2229-14-136

FUNDING

This research was funded by grants from Agencia Nacional de Promoción Científica y Tecnológica to MJY.

ACKNOWLEDGMENTS

We thank the Consejo Nacional de Investigaciones Científicas y Tecnológicas (CONICET) for the scholarships of MJL, CEH, DC, and JI and Agencia Nacional de Promoción Científica y Tecnológica for the scholarship of MGH.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fpls.2019.01019/full#supplementary-material>

- Lamesch, P., Berardini, T. Z., Li, D., Swarbreck, D., Wilks, C., Sasidharan, R., et al. (2012). The Arabidopsis Information Resource (TAIR): improved gene annotation and new tools. *Nucleic Acids Res.* 40, D1202–D1210. doi: 10.1093/nar/gkr1090
- Lin, K. T., Lu, R. M., and Tarn, W. Y. (2004). The WW domain-containing proteins interact with the early spliceosome and participate in pre-mRNA splicing *in vivo*. *Mol. Cell. Biol.* 24 (20), 9176–9185. doi: 10.1128/MCB.24.20.9176-9185.2004
- Macias, M. J., Wiesner, S., and Sudol, M. (2002). WW and SH3 domains, two different scaffolds to recognize proline-rich ligands. *FEBS Lett.* 513 (1), 30–37. doi: 10.1016/S0014-5793(01)03290-2
- Makarov, E. M., Makarova, O. V., Urlaub, H., Gentzel, M., Will, C. L., Wilm, M., et al. (2002). Small nuclear ribonucleoprotein remodeling during catalytic activation of the spliceosome. *Science* 298 (5601), 2205–2208. doi: 10.1126/science.1077783
- Mancini, E., Iserte, J., Yanovsky, M., and Chernomoretz, A. (2019). *ASpli: Analysis of alternative splicing using RNA-Seq*. 1.8.1. ed. Bioconductor: Bioconductor.
- Marquez, Y., Brown, J. W. S., Simpson, C., Barta, A., and Kalyna, M. (2012). Transcriptome survey reveals increased complexity of the alternative splicing landscape in *Arabidopsis*. *Genome Res.* 22 (6), 1184–1195. doi: 10.1101/gr.134106.111
- Matias-Hernandez, L., Aguilar-Jaramillo, A. E., Marin-Gonzalez, E., Suarez-Lopez, P., and Pelaz, S. (2014). RAV genes: regulation of floral induction and beyond. *Ann. Bot.* 114 (7), 1459–1470. doi: 10.1093/aob/mcu069
- McCracken, S., Fong, N., Yankulov, K., Ballantyne, S., Pan, G., Greenblatt, J. et al. (1997). The C-terminal domain of RNA polymerase II couples mRNA processing to transcription. *Nature* 385 (6614), 357–361. doi: 10.1038/385357a0
- 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 (3), 1717–1740. doi: 10.3390/biom5031717
- Misteli, T., and Spector, D. L. (1999). RNA polymerase II targets pre-mRNA splicing factors to transcription sites *in vivo*. *Mol. Cell.* 3 (6), 697–705. doi: 10.1016/S1097-2765(01)80002-2
- Munoz-Cobo, J. P., Sanchez-Hernandez, N., Gutierrez, S., El Yousfi, Y., Montes, M., Gallego, C., et al. (2017). Transcriptional elongation regulator 1 affects transcription and splicing of genes associated with cellular morphology and cytoskeleton dynamics and is required for neurite outgrowth in neuroblastoma cells and primary neuronal cultures. *Mol. Neurobiol.* 54 (10), 7808–7823. doi: 10.1007/s12035-016-0284-6
- Muñoz, M. J., de la Mata, M., and Kornblihtt, A. R. (2010). The carboxy terminal domain of RNA polymerase II and alternative splicing. *Trends Biochem. Sci.* 35 (9), 497–504. doi: 10.1016/j.tibs.2010.03.010
- Nawrath, C., Schell, J., and Koncz, C. (1990). Homologous domains of the largest subunit of eucaryotic RNA polymerase II are conserved in plants. *Mol. Gen. Genet. MGG* 223 (1), 65–75. doi: 10.1007/BF00315798
- Oliveros, J. C. (2007). VENN. An interactive tool for comparing lists with Venn Diagrams. <http://bioinfogp.cnb.csic.es/tools/venny/index.html>.
- Osnato, M., Castillejo, C., Matias-Hernandez, L., and Pelaz, S. (2012). TEMPRANILLO genes link photoperiod and gibberellin pathways to control flowering in *Arabidopsis*. *Nat. Commun.* 3, 808. doi: 10.1038/ncomms1810
- Parker, J. E., Holub, E. B., Frost, L. N., Falk, A., Gunn, N. D., and Daniels, M. J. (1996). Characterization of eds1, a mutation in *Arabidopsis* suppressing resistance to *Peronospora parasitica* specified by several different RPP genes. *Plant Cell.* 8 (11), 2033–2046. doi: 10.1105/tpc.8.11.2033
- Pearson, J. L., Robinson, T. J., Munoz, M. J., Kornblihtt, A. R., and Garcia-Blanco, M. A. (2008). Identification of the cellular targets of the transcription factor TCERG1 reveals a prevalent role in mRNA processing. *J. Biol. Chem.* 283 (12), 7949–7961. doi: 10.1074/jbc.M709402200
- Perez-Santangelo, S., Mancini, E., Francey, L. J., Schlaen, R. G., Chernomoretz, A., Hogenesch, J. B., et al. (2014). Role for LSM genes in the regulation of circadian rhythms. *Proc. Natl. Acad. Sci. U. S. A.* 111 (42), 15166–15171. doi: 10.1073/pnas.1409791111
- Phatnani, H. P., and Greenleaf, A. L. (2006). Phosphorylation and functions of the RNA polymerase II CTD. *Genes Dev.* 20 (21), 2922–2936. doi: 10.1101/gad.1477006
- Rappsilber, J., Ryder, U., Lamond, A. I., and Mann, M. (2002). Large-scale proteomic analysis of the human spliceosome. *Genome Res.* 12 (8), 1231–1245. doi: 10.1101/gr.473902
- Reddy, A. S., Marquez, Y., Kalyna, M., and Barta, A. (2013). Complexity of the alternative splicing landscape in plants. *Plant Cell.* 25 (10), 3657–3683. doi: 10.1105/tpc.113.117523
- Robinson, M. D., McCarthy, D. J., and Smyth, G. K. (2010). edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* 26 (1), 139–140. doi: 10.1093/bioinformatics/btp616
- Sanchez-Alvarez, M., Goldstrohm, A. C., Garcia-Blanco, M. A., and Sune, C. (2006). Human transcription elongation factor CA150 localizes to splicing factor-rich nuclear speckles and assembles transcription and splicing components into complexes through its amino and carboxyl regions. *Mol. Cell. Biol.* 26 (13), 4998–5014. doi: 10.1128/MCB.01991-05
- Schlaen, R. G., Mancini, E., Sanchez, S. E., Perez-Santangelo, S., Rugnone, M. L., Simpson, C. G., et al. (2015). The spliceosome assembly factor GEMIN2 attenuates the effects of temperature on alternative splicing and circadian rhythms. *Proc. Natl. Acad. Sci. U. S. A.* 112 (30), 9382–9387. doi: 10.1073/pnas.1504541112
- Schmittgen, T. D., and Livak, K. J. (2008). Analyzing real-time PCR data by the comparative C(T) method. *Nat. Protoc.* 3 (6), 1101–1108. doi: 10.1038/nprot.2008.73
- Schneider, C. A., Rasband, W. S., and Eliceiri, K. W. (2012). NIH Image to ImageJ: 25 years of image analysis. *Nat. Methods* 9 (7), 671–675. doi: 10.1038/nmeth.2089
- Shikata, H., Hanada, K., Ushijima, T., Nakashima, M., Suzuki, Y., and Matsushita, T. (2014). Phytochrome controls alternative splicing to mediate light responses in *Arabidopsis*. *Proc. Natl. Acad. Sci. U. S. A.* 111 (52), 18781–18786. doi: 10.1073/pnas.1407147112
- Staiger, D., and Brown, J. W. (2013). Alternative splicing at the intersection of biological timing, development, and stress responses. *Plant Cell.* 25 (10), 3640–3656. doi: 10.1105/tpc.113.113803
- Stokes, T. L., Kunkel, B. N., and Richards, E. J. (2002). Epigenetic variation in *Arabidopsis* disease resistance. *Genes Dev.* 16 (2), 171–182. doi: 10.1101/gad.952102
- Sune, C., and Garcia-Blanco, M. A. (1999). Transcriptional cofactor CA150 regulates RNA polymerase II elongation in a TATA-box-dependent manner. *Mol. Cell. Biol.* 19 (7), 4719–4728. doi: 10.1128/MCB.19.7.4719
- Sune, C., Hayashi, T., Liu, Y., Lane, W. S., Young, R. A., and Garcia-Blanco, M. A. (1997). CA150, a nuclear protein associated with the RNA polymerase II holoenzyme, is involved in Tat-activated human immunodeficiency virus type 1 transcription. *Mol. Cell. Biol.* 17 (10), 6029–6039. doi: 10.1128/MCB.17.10.6029
- Tang, X., Xie, M., Kim, Y. J., Zhou, J., Klessig, D. F., and Martin, G. B. (1999). Overexpression of Pto activates defense responses and confers broad resistance. *Plant Cell.* 11 (1), 15–29. doi: 10.1105/tpc.11.1.15
- Trapnell, C., Pachter, L., and Salzberg, S. L. (2009). TopHat: discovering splice junctions with RNA-Seq. *Bioinformatics* 25 (9), 1105–1111. doi: 10.1093/bioinformatics/btp120
- Wahl, M. C., Will, C. L., and Luhrmann, R. (2009). The spliceosome: design principles of a dynamic RNP machine. *Cell* 136 (4), 701–718. doi: 10.1016/j.cell.2009.02.009

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2019 Hernando, García Hourquet, de Leone, Careno, Iserte, Mora Garcia and Yanovsky. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



Alternative Splicing Regulation During Light-Induced Germination of *Arabidopsis thaliana* Seeds

Rocío Soledad Tognacca^{1,2}, Lucas Servi², Carlos Esteban Hernando³, Maite Saura-Sanchez¹, Marcelo Javier Yanovsky³, Ezequiel Petrillo^{2*} and Javier Francisco Botto^{1*}

¹ Consejo Nacional de Investigaciones Científicas y Técnicas, Instituto de Investigaciones Fisiológicas y Ecológicas Vinculadas a la Agricultura (IFEVA), Facultad de Agronomía, Universidad de Buenos Aires, Buenos Aires, Argentina, ² Facultad de Ciencias Exactas y Naturales, Departamento de Fisiología, Biología Molecular y Celular and CONICET-UBA, Instituto de Fisiología, Biología Molecular y Neurociencias (IFIBYNE), Universidad de Buenos Aires (UBA), Buenos Aires, Argentina, ³ Fundación Instituto Leloir, IIBBA-CONICET, Buenos Aires, Argentina

OPEN ACCESS

Edited by:

Anna N. Stepanova,
North Carolina State University,
United States

Reviewed by:

Wilco Ligterink,
Wageningen University & Research,
Netherlands
Dora Szakonyi,
Gulbenkian Institute of Science,
Portugal

*Correspondence:

Ezequiel Petrillo
petry@fbmc.fcen.uba.ar
Javier Francisco Botto
botto@agro.uba.ar

Specialty section:

This article was submitted to
Plant Physiology,
a section of the journal
Frontiers in Plant Science

Received: 26 April 2019

Accepted: 07 August 2019

Published: 10 September 2019

Citation:

Tognacca RS, Servi L, Hernando CE, Saura-Sanchez M, Yanovsky MJ, Petrillo E and Botto JF (2019) Alternative Splicing Regulation During Light-Induced Germination of *Arabidopsis thaliana* Seeds. *Front. Plant Sci.* 10:1076. doi: 10.3389/fpls.2019.01076

Seed dormancy and germination are relevant processes for a successful seedling establishment in the field. Light is one of the most important environmental factors involved in the relief of dormancy to promote seed germination. In *Arabidopsis thaliana* seeds, phytochrome photoreceptors tightly regulate gene expression at different levels. The contribution of alternative splicing (AS) regulation in the photocontrol of seed germination is still unknown. The aim of this work is to study gene expression modulated by light during germination of *A. thaliana* seeds, with focus on AS changes. Hence, we evaluated transcriptome-wide changes in stratified seeds irradiated with a pulse of red (Rp) or far-red (FRp) by RNA sequencing (RNA-seq). Our results show that the Rp changes the expression of ~20% of the transcriptome and modifies the AS pattern of 226 genes associated with mRNA processing, RNA splicing, and mRNA metabolic processes. We further confirmed these effects for some of the affected AS events. Interestingly, the reverse transcriptase–polymerase chain reaction (RT–PCR) analyses show that the Rp modulates the AS of splicing-related factors (*At-SR30*, *At-RS31a*, *At-RS31*, and *At-U2AF65A*), a light-signaling component (*At-PIF6*), and a dormancy-related gene (*At-DRM1*). Furthermore, while the phytochrome B (phyB) is responsible for the AS pattern changes of *At-U2AF65A* and *At-PIF6*, the regulation of the other AS events is independent of this photoreceptor. We conclude that (i) Rp triggers AS changes in some splicing factors, light-signaling components, and dormancy/germination regulators; (ii) phyB modulates only some of these AS events; and (iii) AS events are regulated by R and FR light, but this regulation is not directly associated with the intensity of germination response. These data will help in boosting research in the splicing field and our understanding about the role of this mechanism during the photocontrol of seed germination.

Keywords: dormancy, germination, light, phytochrome B (phyB), alternative splicing (AS), *Arabidopsis*

INTRODUCTION

Seed dormancy is a developmental checkpoint that allows plants to regulate when and where they grow. Temperature, light, and nitrates are the most relevant environmental factors regulating the relief of seed dormancy to promote seed germination (Benecch-Arnold et al., 2000). These cues can trigger molecular responses including hormone signaling, mainly those of abscisic acid (ABA) and gibberellin (GA). The balance between the contents and sensitivity of these hormones is key for the regulation of the dormancy status of the seeds. ABA promotes primary dormancy induction and later maintenance, whereas GA promotes seed germination. Environmental signals regulate this balance by modifying the expression of metabolic enzymes as well as those of positive and negative regulators of both hormones, many of which are feedback regulated (Finkelstein et al., 2008).

Light has been one of the most characterized factors regulating the relief of dormancy. Phytochromes are the best-known photoreceptors perceiving red (R) and far-red (FR) light. They are synthesized in its inactive form, Pr (with maximum absorption in R), and are photo-converted into their active form, Pfr (with maximum absorption in FR). The *Arabidopsis* genome encodes five phytochromes, named phyA to phyE. Among them, phytochrome B (phyB) has a prominent role as the main photoreceptor regulating the R/FR reversible response (Shinomura et al., 1994; Botto et al., 1995), and phyD and phyE can contribute to this regulation in *phyA/phyB* double mutant seeds (Hennig et al., 2002; Arana et al., 2014), suggesting some redundancy in phytochrome functions in the R-light-mediated seed germination. In the soils, weed seeds change their light sensitivity according with after-ripening and burial conditions (Scopel et al., 1991; Botto et al., 1998a; Botto et al., 1998b; Botto et al., 2000), being phyA responsible for the detection of very brief light stimulus-promoting seed germination (Botto et al., 1996; Shinomura et al., 1996).

Light tightly regulates the expression (transcript levels) of thousands of genes (Casal and Yanovsky, 2005; Galvao and Fankhauser, 2015) and also many other layers of gene expression such as mRNA splicing, translation, and stability (Mano et al., 1999; Yakir et al., 2007; Simpson et al., 2008; Jung et al., 2009; Juntawong and Bailey-Serres, 2012; Liu et al., 2012; Paik et al., 2012; Liu et al., 2013; Petrillo et al., 2014; Shikata et al., 2014; Tsai et al., 2014; Wang et al., 2014; Wu et al., 2014; Mancini et al., 2016). RNA splicing is a co-transcriptional molecular event that is carried out by a macromolecular complex called spliceosome. Alternative splicing (AS) is the process that generates multiple transcripts from a single gene by using different combinations of available splice sites. AS leads to different outcomes and produces transcripts encoding for proteins that may have altered or lost function. Several investigations have demonstrated the importance of AS in processes like photosynthesis, defense responses, circadian clock, hormone signaling, flowering time, and metabolism (Lightfoot et al., 2008; Matsumura et al., 2009; Sanchez et al., 2010; Martin-Trillo et al., 2011; James et al., 2012; Jones et al., 2012; Rosloski et al., 2013).

AS is also relevant during the early (Fouquet et al., 2011) and late stages of embryo development (Sugliani et al., 2010).

At the seed level, the work by Srinivasan et al. (2016) analyzed the AS during *Arabidopsis* seed development at a global level, both before and after seed desiccation. They identified 4,003 genes that are alternatively spliced, and 1,408 of those genes showing a differential pattern of splicing between both stages. Remarkably, most of these alternatively spliced transcripts had not been found in other tissues. More recently, another report highlights the relevance of AS in *Arabidopsis* seeds (Narsai et al., 2017). These authors analyzed the dynamics of gene expression over 10 developmental time points during seed germination and identified 620 genes undergoing AS. The regulation of these AS events during seed germination is time specific and/or tissue specific (Narsai et al., 2017). Interestingly, they also found complex variations in the relative abundance of *PIF6* (*PHYTOCHROME INTERACTING FACTOR 6*, AT3G62090) isoforms, as previously demonstrated (Penfield et al., 2010). *PIF6* transcription factor is expressed during seed development, and its expression is dramatically reduced during imbibition. It has four known AS isoforms (Narsai et al., 2017), one of them originating from an exon skipping (ES) event that creates a premature stop codon and encodes for a protein that lacks the DNA-binding domain (Penfield et al., 2010). As expected, the overexpression of this *PIF6* isoform reduces seed dormancy (Penfield et al., 2010). *DOG1* (*DELAY OF GERMINATION1*, AT5G45830), another main regulator of seed germination, is also affected at the AS level. *DOG1* protein accumulates during seed maturation, and its abundance in freshly harvested seeds determines the dormancy status of the seed (Graeber et al., 2014). *DOG1* transcripts are extensively alternatively spliced, giving place to five AS isoforms that are all functional but unstable if not expressed in combination (Bentsink et al., 2006; Nakabayashi et al., 2012; Nakabayashi et al., 2015). These findings clearly show the importance of AS control in the regulation of seed germination.

Photosensory-protein pathways are not the only way to sense light and regulate gene expression accordingly. Petrillo et al. (2014) have shown that *At-RS31* (*ARGININE/SERINE-RICH SPLICING FACTOR 31*, AT3G61860), *At-SR30* (*ARGININE/SERINE-RICH SPLICING FACTOR 30*, AT1G09140), and *At-U2AF65A* (*U2 SNRNP AUXILIARY FACTOR*, AT4G36690) AS patterns are modulated by different light conditions through retrograde signals arising from chloroplasts. They demonstrated that the chloroplast, which is able to sense and to communicate light signals to the nucleus, is the main actor triggering the AS changes in response to light (Petrillo et al., 2014). All these experimental evidences suggest that AS is a relevant mechanism even though there is still no information available about how light-promoting seed germination affects this process in *Arabidopsis thaliana* seeds. Here, we showed that (i) red pulse (Rp) triggers AS changes in some splicing factors, light-signaling components, and dormancy/germination regulators; and (ii) phyB is involved in the regulation of only some of these AS events. Furthermore, (iii) AS events are regulated by R and FR light, but this regulation is not directly associated with the intensity of germination response. We conclude that AS is a source of gene expression diversity, and a proper regulation of this process might be of key relevance for seed germination modulation under different light conditions.

MATERIALS AND METHODS

Plant Material and Growth Conditions

A. thaliana plants were grown under long day conditions [16-h L/8-h D, photosynthetically active radiation (PAR) = $100 \mu\text{mol}\cdot\text{m}^{-2}\cdot\text{s}^{-1}$] with an average temperature of $21 \pm 2^\circ\text{C}$. Plants were grown together, and their mature seeds were harvested at the same time to avoid differences in post-maturation, which can affect seed germination. Seeds of each genotype were harvested as a single bulk that consisted of at least five plants. Seeds were stored in tubes with small holes inside a closed box and maintained in darkness with silica gel at 4°C until the experiments were performed. *A. thaliana* Columbia-0 (Col-0) and Landsberg *erecta* (Ler) were used as wild type (WT). Seeds of *phyB-9* (Col-0 background) and *phyB-5* (Ler background) (218790) were obtained from the ABRC (www.arabidopsis.org/abrc/).

Germination Conditions and Light Treatments

Samples of 20 seeds per genotype were sown in clear plastic boxes, each containing 10 ml of 0.8% (w/v) agar in demineralized water. To establish a minimum and equal photo-equilibrium, seeds were imbibed for 2 h in darkness and then irradiated for 20 min with a saturated far-red pulse (FRp; calculated Pfr/P = 0.03, $42 \mu\text{mol}\cdot\text{m}^{-2}\cdot\text{s}^{-1}$) in order to minimize the quantities of Pfr that formed during their development in the mother plant. Seeds were then stratified at 5°C in darkness for 3 days, prior to the 20 min with a saturated Rp (calculated Pfr/P = 0.87, $0.05 \mu\text{mol}\cdot\text{m}^{-2}\cdot\text{s}^{-1}$) or FRp. After light treatments, the boxes containing the seeds were wrapped again with black plastic bags and incubated at 25°C for 3 days before germination was determined. The criterion for germination was the emergence of the radicle.

For experiments with hormones, seeds were sown in clear plastic boxes, each containing filter papers imbibed with 750 μl of fluridone 100 μM (Sigma-Aldrich, Steinheim, Germany) or ABA 1 μM supplemented with fluridone 100 μM (Ibarra et al., 2016) until the end of the experiment. We have previously performed calibration curves to determine the optimal ABA concentration to counteract the promotion of germination triggered by fluridone (Supplementary Figure 1).

cDNA Library Preparation and High-Throughput Sequencing

Seed samples were sown in clear plastic boxes, each containing 10 ml of 0.8% (w/v) agar in de-mineralized water. Three biological replicates of each condition were collected 12 h after the corresponding R and FR light pulses. After sampling, seeds were immediately frozen in liquid nitrogen and stored at -80°C . RNA was extracted using the Spectrum Plant Total RNA Kit (Sigma-Aldrich, Steinheim, Germany) according to manufacturer's protocol. To estimate the concentration and quality of the samples, spectrophotometry and agarose gel were used, respectively. RNA samples were processed at the Instituto de Agrobiotecnología de Rosario (INDEAR, Rosario, Argentina). Samples were pooled to create six multiplexed DNA libraries, which were pair-end sequenced with an Illumina HiSeq 1500.

Processing of RNA Sequencing Reads

Sequence reads were aligned with the *A. thaliana* genome TAIR10 (Langmead et al., 2009) with TopHat v2.1.1 (Trapnell et al., 2009) with default parameters, except in the case of the maximum intron length parameter, which was set at 5,000. Count tables for the different feature levels were obtained from bam files using custom R scripts and considering the TAIR10 transcriptome.

Differential Gene Expression Analysis

Differential gene expression analysis was conducted for genes whose expression was above a minimum threshold level [>10 reads and a read density (RD) >0.05] in at least one experimental condition. RD was computed as the number of reads in each gene divided by its effective width. The term effective width corresponds to the sum of the length of all the exons of a given gene. Differential gene expression was estimated using the edgeR package version 3.4.2 (Robinson et al., 2010), and resulting *p*-values were adjusted using a false discovery rate (FDR) criterion (Benjamini and Hochberg, 1995). Genes with FDR values lower than 0.05 and an absolute fold change >1.5 were considered to be differentially expressed (DE). This dataset was labeled as DE genes (Supplementary Table 1).

Differential AS Analysis

For the analysis of differential AS, multiexonic genes were partitioned into features defined as “bins,” corresponding to exonic, intronic, and alternatively spliced regions. We labeled these three kinds of bins as exon bins, intron bins, or AS bins, respectively. In addition, AS bins were further classified as ES, alternative 5' splice site (Alt5'SS), alternative 3' splice site (Alt3'SS), and intron retention (IR). Bins with three or more different AS events in the same subgenic region were labeled as multiple. Read summarization was performed at those three levels: exon, intron, and AS bins. These datasets were then filtered according to several criteria applied at the gene and bin levels. First, defined subgenic regions (i.e., bins) were considered for differential AS analysis only if the genes with which they are associated with were expressed above a minimum threshold level (more than 10 reads per gene and RD >0.05) in all experimental conditions. Next, bins were considered for differential AS analysis only if they had more than five reads and an RD bin/RD gene ratio >0.05 , in at least one experimental condition. After these filters were applied, reads summarized at the bin level were normalized to the read counts of their corresponding gene. This was done to avoid the influence of changes in gene expression on the differential AS analysis at the bin level. Then, similarly to the approach used for the differential expression analysis, differential AS analysis was conducted at the bin levels using the edgeR package version 3.4.2. Bins with FDR values lower than 0.15 were considered to undergo differential AS. Finally, we restricted the selection of AS bins to those bins for which differential AS analysis was supported by expected changes in the numbers of splice junctions. In order to do this, we obtained information on the number of reads associated with each splice junction, both annotated and novel. Junction coordinates were

extracted from gap containing aligned reads. Junctions with fewer than five reads were discarded. We then computed the metrics PSI (percent spliced-in) and PIR (percent IR), which were used as a final filtering criteria for the AS analysis. PSI was defined as the percentage of the number of junction reads supporting bin inclusion relative to the combined number of reads supporting inclusion and exclusion (Pervouchine et al., 2013). PSI values were computed for ES, Alt5'SS, and Alt3'SS. PIR values, calculated as previously described (Braunschweig et al., 2014), were used for the IR analysis. Briefly, PIR is defined for each experimental condition as the percentage of the number of reads supporting IR (E1I + IE2) relative to the combined number of reads supporting IR and exclusion (E1I + IE2 + 2 exclusion junction [JE1E2]), where E is the exonic bin, I the intronic bin, and J the junction (Mancini et al., 2016). AS, exon, and intron bins were considered to be differentially spliced (DS) if, in addition to fulfilling the filtering criteria described above, the difference in PSI or PIR between experimental conditions was >0.5%. Bins corresponding to alternatively spliced regions identified through novel splice junctions were considered to be differentially alternatively spliced if there was a difference in the PSI value larger than 10% between experimental conditions. This dataset was labeled as DS genes (**Supplementary Table 1**).

Gene Ontology Analysis

Gene ontology (GO) terms assignment for the DE genes and DS genes datasets were obtained using the BioMaps tool from the virtual plant software (<http://virtualplant.bio.nyu.edu/cgi-bin/vpweb/>). An enrichment test was performed for the following categories: BP (biological process), MF (molecular function), and CC (cellular component). *p*-values were obtained using the Fisher exact test and corrected for multiple testing using FDR. The enrichment factor (EF) was estimated as the ratio between the proportion of genes associated with a particular GO category present in the dataset under analysis, relative to the number of genes in this category in the whole genome (**Supplementary Table 1**). Bubble plots were generated, using a custom script written in R language, for all those categories for which the adjusted *p*-value was lower than 0.01 in at least one dataset.

Semi-Quantitative Reverse Transcriptase–Polymerase Chain Reaction

Seed samples were sown in clear plastic boxes, each containing 10 ml of 0.8% (w/v) agar in de-mineralized water. Three biological replicates of each condition were collected 12 h after the corresponding R and FR light pulses. After sampling, seeds were immediately frozen in liquid nitrogen and stored at -80°C . RNA was extracted using the Spectrum Plant Total RNA Kit (Sigma-Aldrich, Steinheim, Germany) according to manufacturer's protocol. cDNA derived from the extracted RNA was synthesized using M-MLV reverse transcriptase (Promega, Madison, WI, USA) and oligo-dT primers. polymerase chain reaction (PCR) analyses were conducted using a mix with Taq DNA Pol (Inbio Highway, Tandil, Buenos Aires, Argentina), 10 \times buffer, 10 \times polyvinylpyrrolidone

(PVP), 25 mM of MgCl_2 , 10 mM of dNTPs, and gene-specific primers according to the manufacturer's instructions. The PCR program was as follows: 95°C for 3 min and the respective number of cycles (28–31) at 95°C for 30 s, 60°C for 30 s, and 72°C for 1.5 min. The amplified products were resolved by 1% or 2% agarose gel electrophoreses. The SI, defined as the abundance of the longest splicing isoform relative to the levels of all possible isoforms, was calculated from the relative levels of the corresponding reverse transcriptase (RT)–PCR products quantified using densitometry by the ImageJ software (<https://imagej.net/Welcome>). Gene models and gel images are shown in **Supplementary File 1**. The specific primers used are described in **Supplementary Table 2**.

Statistical Analysis

To test for significant differences in the response of the seeds, we conducted two-way analyses of variance (ANOVAs) for each WT and mutant group, using the angular transformation of the percentage of germination and the InfoStat Software version 2017 (Grupo InfoStat, FCA, Universidad Nacional de Córdoba, Argentina). Fisher post-test was used to test differences between genotypes, when significant treatment-by-genotype interactions were observed.

To test for significant differences in the splicing index (SI), we also conducted two-way ANOVAs using the same software, and Fisher post-test was used to test for differences, when significant treatment-by-genotype interactions were observed. When indicated, we used Student's *t*-test.

RESULTS

Seed Germination Induced by Red Light Modulates the AS Pattern of 226 genes

WT Col-0 seeds were imbibed in darkness at 5°C for 3 days prior to irradiation with an Rp or an FRp. Germination was counted after 3 days at 25°C in darkness (**Figure 1A**). The Rp significantly induces seed germination $\sim 85\%$ in Col-0 seeds, while the FRp does not promote seed germination (**Figure 1B**), indicating that the phytochrome system is involved in this response. We used this seed population to evaluate changes at the transcriptome level triggered by light, with special focus in AS modulation. Seed samples for RNA isolation were collected 12 h after the light treatments (**Figure 1A**), and R light effects on mRNA levels and AS were analyzed (**Supplementary Table 1**). We identified a total of 5,785 genes whose mRNA levels are significantly affected, either increased or decreased, more than 1.5-fold ($\text{FDR} < 0.05$) in response to the Rp (**Figure 1C** and **Supplementary Table 1**). We defined this group as DE genes. Interestingly, DE genes are enriched in genes associated with response to temperature, light, and ABA stimuli (**Figure 1D** and **Supplementary Table 1**). We then evaluated the effects of R light on AS and identified a total of 226 genes with AS events that are regulated by the light pulse (**Figure 1C** and **Supplementary Table 1**). We defined this group as DS genes. Less than half of the genes whose AS patterns are affected by R light show alterations at the total mRNA levels too (102 genes,

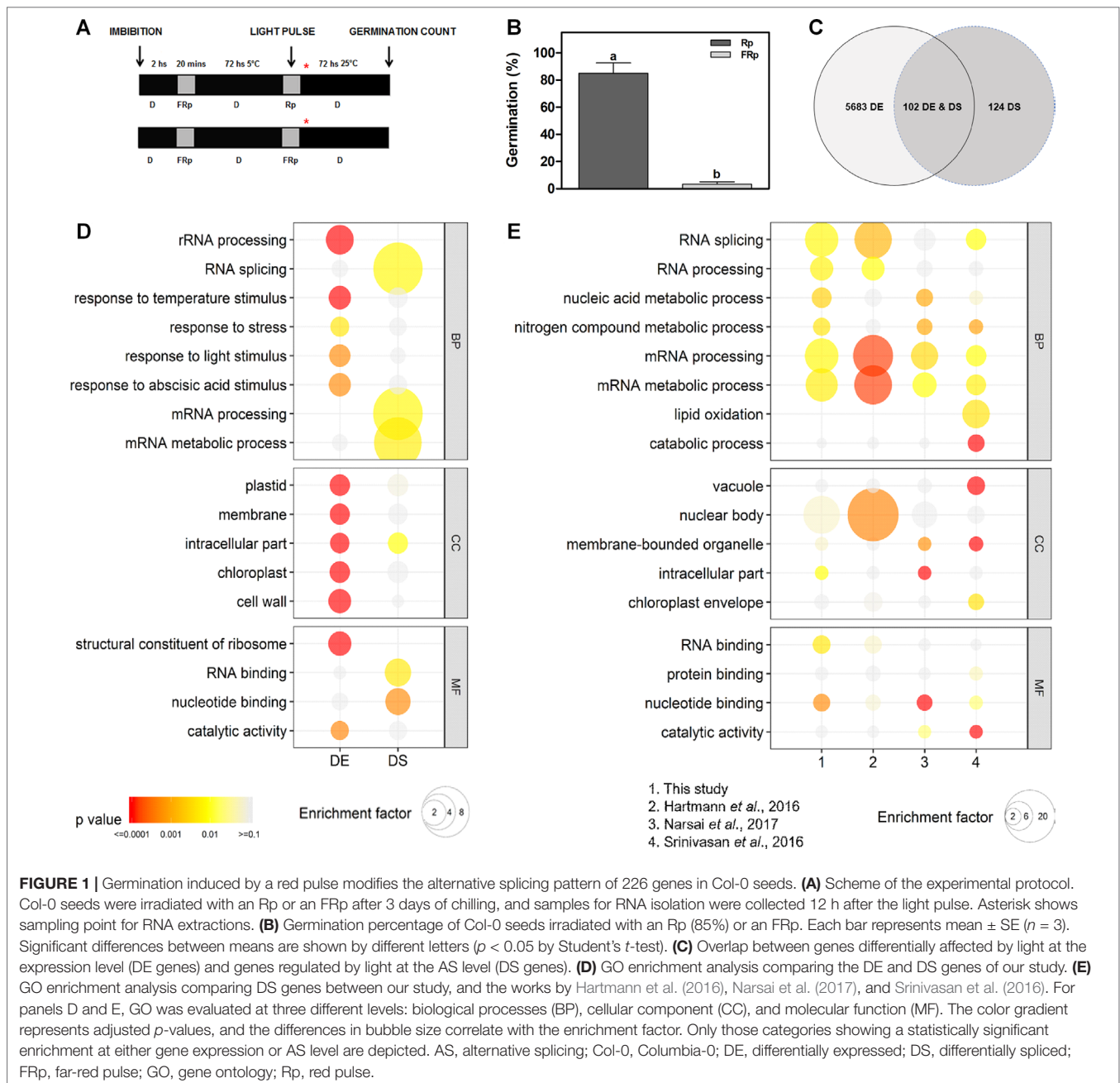


Figure 1C). We also observed a strong enrichment in GO categories associated with mRNA processing, RNA splicing, and mRNA metabolic processes among these DS genes (**Figure 1D** and **Supplementary Table 1**). These categories are not enriched among the genes whose mRNA levels, rather than AS patterns, are affected by R light, supporting the idea that light regulates AS patterns mostly through its effect on the AS of splicing factors themselves (**Figure 1C** and **Supplementary Table 1**).

Additionally, we performed a comparative transcriptomic analysis to evaluate if differentially alternatively spliced genes identified here, as responsive to an Rp that induced germination, were also regulated in other developmental processes or conditions,

like seed maturation (Srinivasan et al., 2016), germination induced by a white light/dark photoperiod (Narsai et al., 2017), and/or seedling de-etiolation (Hartmann et al., 2016). We found two common genes affected at the AS level in the four transcriptome studies (*At-RS41* and *At-HYP1*), but a higher number of common genes are found when comparing different combinations of specific transcriptomes with those in our study (**Supplementary Table 1**). We also observed a strong enrichment in GO categories associated with RNA splicing, RNA processing, mRNA processing, and mRNA metabolic processes among these different physiological processes (**Figure 1E** and **Supplementary Table 1**). These data are in agreement with previous studies showing that genes related to

RNA metabolism are among the most affected by AS (Syed et al., 2012; Reddy et al., 2013).

Red Pulse Reduces the SI of *At-SR30*, *At-RS31a*, *At-RS31*, and *At-U2AF65A*

To validate the RNA sequencing (RNA-seq) data, we selected four genes of our interest and studied their responses by RT-PCR: (a) *At-SR30* and *At-RS31*, which are among the down-regulated and AS affected genes (DE and DS) according to the RNA-seq (Supplementary Table 1), and (b) *At-U2AF65A* and *At-RS31a*, which belong to the down-regulated category (DE) but are known to be light-responsive events in seedlings (Petrillo et al., 2014). All these candidate genes are related to the splicing process itself: three are SR genes (*At-SR30*, *At-RS31a* and *At-RS31*) and one is an auxiliary splicing factor (*At-U2AF65A*). We used the SI, defined as the abundance of the longest splicing isoform relative to the levels of all possible isoforms, as the parameter to evaluate AS changes induced by the Rp. Figure 2 shows that the Rp reduces the SI of *At-SR30*, *At-RS31a*, *At-RS31*, and *At-U2AF65A* between 2- and 5-folds than does the FRp. These results confirm that the AS patterns of these events are modulated by red light when seed germination is induced. Moreover, these results suggest that the total number of affected AS events by red light in germinating seeds is probably higher than those considered in our list (Supplementary Table 1). Further research in this direction is needed to know the actual impact of AS in gene expression regulation by red light during seed germination.

Phytochrome B Regulates the AS Pattern of *At-PIF6* and *At-U2AF65A*

R-light-induced germination can be mediated by different phytochromes (Botto et al., 1996; Hennig et al., 2002; Arana et al., 2014), and both phyA and phyB are the main photoreceptors. The phyB contribution to AS regulation is still unknown. Thus, we analyzed germination of Col-0 and *Ler* WT seeds and *phyB-9* (Col-0 background) and *phyB-5* (*Ler* background) mutant seeds under our experimental conditions. The FRp does not promote seed germination in any of the analyzed genotypes (Figures 3A, B). On the contrary, the Rp induces 98% of germination in Col-0 seeds (Figure 3A) and 78% in *Ler* seeds (Figure 3B), while *phyB-9* and *phyB-5* mutant seeds germinate at 10% and 32%, respectively, suggesting that phyB is the main phytochrome controlling seed germination under these conditions (Figures 3A, B).

We then asked whether the light-induced AS changes could be subjected to phyB regulation. The Rp reduces the SI of *At-PIF6*, a phytochrome interacting factor, and increases the SI of *At-DRM1* (DORMANCY-ASSOCIATED PROTEIN 1, AT1G28330), a dormancy-related gene in Col-0 and *Ler* seeds (Figures 3C, D). The phyB is responsible for the AS changes of *At-PIF6* since light differences associated with the SI are lost in both *phyB-5* and *phyB-9* mutant seeds (Figure 3C). In opposition, the AS pattern of *At-DRM1* is not regulated by the phyB (Figure 3D). With respect to the splicing related factors, the SI of *At-SR30*, *At-RS31a*, and *At-RS31* are significantly

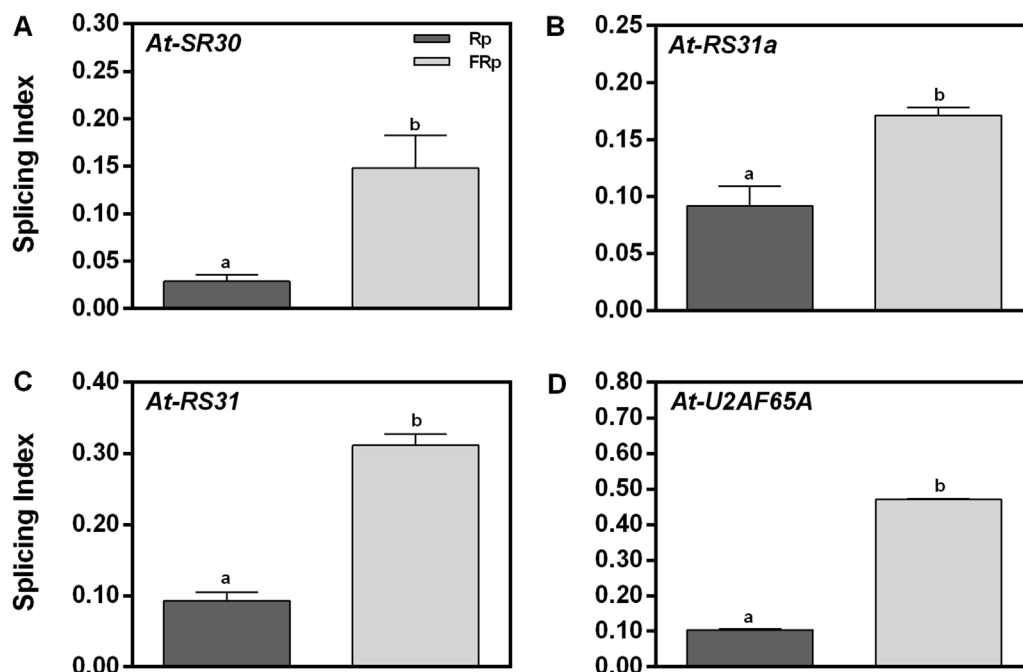


FIGURE 2 | Germination and alternative splicing regulation by red/far-red light. (A–D) Germination induced by an Rp reduces the splicing index (SI) of *At-U2AF65A*, *At-SR30*, *At-RS31*, and *At-RS31a* in Col-0 seeds. We selected these genes from the RNA-seq data based on two simultaneous criteria: (1) genes involved in RNA splicing and (2) genes previously documented to have AS (Petrillo et al., 2014; Shikata et al., 2014). RNA Col-0 samples were harvested 12 h after the Rp or FRp. AS alterations as SI in response to the Rp or FRp are shown. Gene expression levels were analyzed by RT-PCR, and each PCR product was quantified by densitometry using ImageJ. Each bar represents mean \pm SE ($n = 3$). Significant differences between means are shown by different letters ($p < 0.05$ by Student's t -test). AS, alternative splicing; Col-0, Columbia-0; FRp, far-red pulse; Rp, red pulse; RT-PCR, reverse transcriptase–polymerase chain reaction.

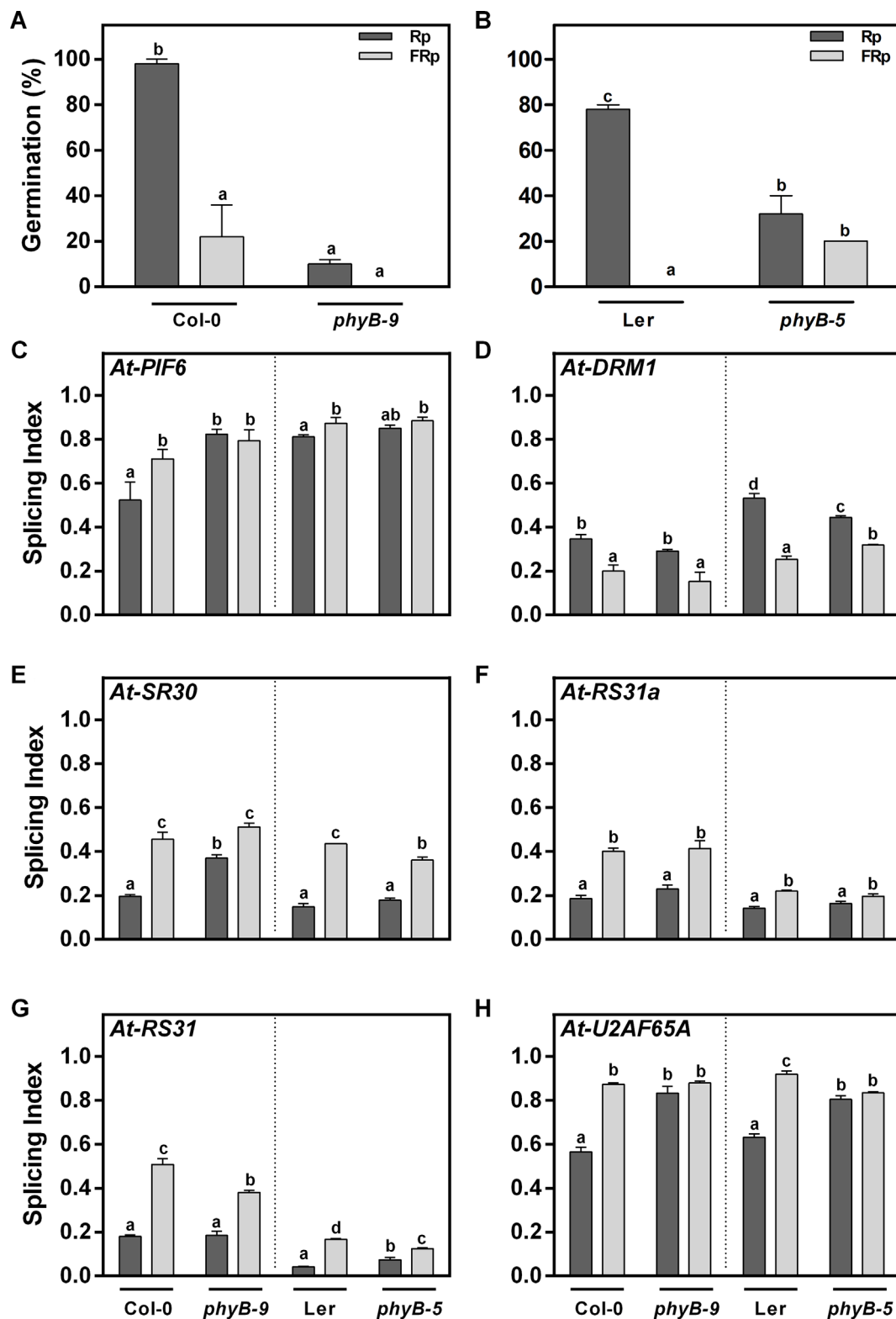


FIGURE 3 | *phyB* differentially regulates the AS pattern of *At-U2AF65A* and *At-PIF6*. **(A–B)** Germination percentages of Col-0, *phyB-9*, Ler, and *phyB-5* irradiated with an Rp or an FRp. **(C–H)** AS alterations quantified as splicing index (SI) in response to the Rp or FRp are shown. Seed samples were harvested 12 h after the Rp or FRp. Splice variants of the indicated genes were analyzed by RT-PCR, and each PCR product was quantified by densitometry using ImageJ. Each bar represents mean \pm SE ($n = 3$). Means with the same letters are not significantly different ($p > 0.05$ by ANOVA followed by Fisher post-test). ANOVA, analysis of variance; AS, alternative splicing; Col-0, Columbia-0; FRp, far-red pulse; Rp, red pulse; RT-PCR, reverse transcriptase-polymerase chain reaction.

reduced in response to the Rp in both Col-0 and *Ler* seeds. Even though *phyB* mutants show some changes in the SI of these AS events, similar differences between Rp and FRp are still observed in both *phyB-9* and *phyB-5* mutant seeds, suggesting the AS control of these splicing factors by light does not involve *phyB* regulation (Figures 3E–G). On the other hand, even though the Rp significantly changes the SI of *At-U2AF65A* in Col-0 and *Ler* seeds (~30%), this difference is completely abolished in the *phyB* mutant seeds in both genetic backgrounds (Figure 3H). We conclude that *phyB* regulates the AS pattern of *At-U2AF65A* and *At-PIF6* in *Arabidopsis* seeds germinating with an Rp.

Red Light Perception and Alternative Splicing Are Directly Linked

We have clearly shown that red light promotes seed germination and affects AS (Figures 1–3). One remaining question is whether these AS changes are directly triggered by the red-light signal or if they are a consequence of the germination process *per se*. We hypothesized that if light is directly controlling the AS response, changes in splicing patterns should not be affected by different levels of germination under the same light condition (Rp or FRp). Since ABA is a known hormone regulating seed germination, we sowed WT Col-0 seeds in water (control); fluridone (F), an inhibitor of ABA synthesis; and fluridone supplemented with ABA (F + ABA). The Rp induces 100% germination in water and F but only 40% in F + ABA. The FRp reduces the germination to ~40%, ~80%, and 0%, respectively (Figure 4A). As expected, the F and F + ABA treatments modulate the germination response induced by red light. Hence, we used these seeds to analyze the AS patterns of *At-PIF6*, *At-DRM1*, *At-SR30*, *At-RS31a*, *At-RS31*, and *At-U2AF65A*. We found that the SI of the six genes changes significantly between seeds exposed to Rp and FRp in water, and these changes in the SI are still present in F and F + ABA-treated seeds (Figures 4B–G). These results clearly show that changes in AS patterns induced by light are independent of the level of germination. This conclusion is also supported by our previous results showing that the *phyB* mutant, which presents reduced germination under Rp, displayed similar AS changes than the WT seeds for the vast majority of the analyzed events (Figure 3). Taken altogether, our data suggest that the Rp directly controls AS processes and that these effects are independent of the level of germination induced by light.

DISCUSSION

Seed dormancy and germination are processes of extreme relevance for a successful seedling establishment in the field. Light is one of the most important and best characterized environmental factors involved in the relief of seed dormancy to promote seed germination (Benech-Arnold et al., 2000). Plants possess a wide variety of photoreceptors capable of sensing this environmental cue (van Gelderen et al., 2018). The *Arabidopsis* genome encodes for five different phytochromes, and *phyB* is the one with the most prominent role in controlling the R/FR reversible response inducing seed germination (Shinomura et al., 1994; Botto et al., 1995). Moreover, light shapes plants' transcriptomes by affecting

every possible level of gene expression regulation (Petrillo et al., 2015; Merchante et al., 2017; Kaiserli et al., 2018; Godoy Herz et al., 2019). AS, a powerful mechanism that allows rapid changes in transcriptome and proteome complexity during development and in response to changes in the environment (Reddy et al., 2013), is also dramatically affected by this environmental cue at different developmental stages. Nowadays, it is becoming evident that AS substantially increases transcriptome complexity and plays an important role in modulating gene expression in response to internal and external cues (Lightfoot et al., 2008; Matsumura et al., 2009; Sanchez et al., 2010; Martin-Trillo et al., 2011; James et al., 2012; Jones et al., 2012; Rosloski et al., 2013). While the information concerning the molecular basis of light-induced germination is abundant at the level of gene expression/transcription (Penfield et al., 2005; Oh et al., 2006; Oh et al., 2007; Oh et al., 2009; Gabriele et al., 2010; Park et al., 2012; Ibarra et al., 2013), that concerning global studies describing light effects on AS has only recently started to appear. Here, we provide evidence showing that (i) Rp triggers AS changes in some splicing factors, light-signaling components, and dormancy/germination regulators; (ii) Rp exerts these effects on AS through the action of *phyB* in some few events; and (iii) AS events are regulated by R/FR light, but this regulation is not directly associated with the intensity of germination response.

We identified 226 genes with AS events regulated by light (Figure 1C and Supplementary Table 1) and that the Rp reduces the SI of *At-SR30*, *At-RS31a*, *At-RS31*, and *At-U2AF65A* (Figure 2). Since *At-SR30*, *At-RS31*, and *At-RS31a* are members of the RS subfamily (like *At-RS41*), these results may have implications for all the SR genes in *Arabidopsis*. Previous studies have evaluated light effects on AS at a global level in *Physcomitrella patens* and in etiolated *A. thaliana* seedlings using RNA-seq (Shikata et al., 2014; Wu et al., 2014). These studies found several hundreds of light-regulated AS events, many of which were associated with genes encoding splicing factors and light-signaling components. Interestingly, in both reports, the effects of brief light treatments on AS were modulated to a great extent, although not exclusively, by the phytochromes (Shikata et al., 2014; Wu et al., 2014). Furthermore, it was previously shown that *RRC1* (REGULATOR OF CHROMOSOME CONDENSATION 1), an SR-like protein, is required for normal seedling development under R light (Shikata et al., 2014) and, more recently, that *RRC1* interacts with SFPS (SPLICING FACTOR FOR PHYTOCHROME SIGNALING) and *phyB* to coordinately regulate the splicing of genes involved in light signaling and circadian clock pathways to promote photomorphogenesis in *A. thaliana* (Xin et al., 2019). The fact that we found common genes whose AS is regulated in different developmental processes (Shikata et al., 2014; Hartmann et al., 2016; Srinivasan et al., 2016; Narsai et al., 2017; this study) strongly points towards the existence and relevance of an AS regulatory network that is active throughout the whole life cycle of a plant.

The R/FR reversible response of seed germination is mainly mediated by *phyB* (Shinomura et al., 1994; Botto et al., 1995). Interestingly, we found that changes on some AS events (*At-PIF6* and *At-U2AF65A* out of six genes analyzed) are mediated by this photoreceptor (Figures 3C, H). However, the observation that the SI values of *At-SR30*, *At-RS31a*, *At-RS31*, and *At-DRM1* were not altered by the absence of *phyB* suggests that other stable

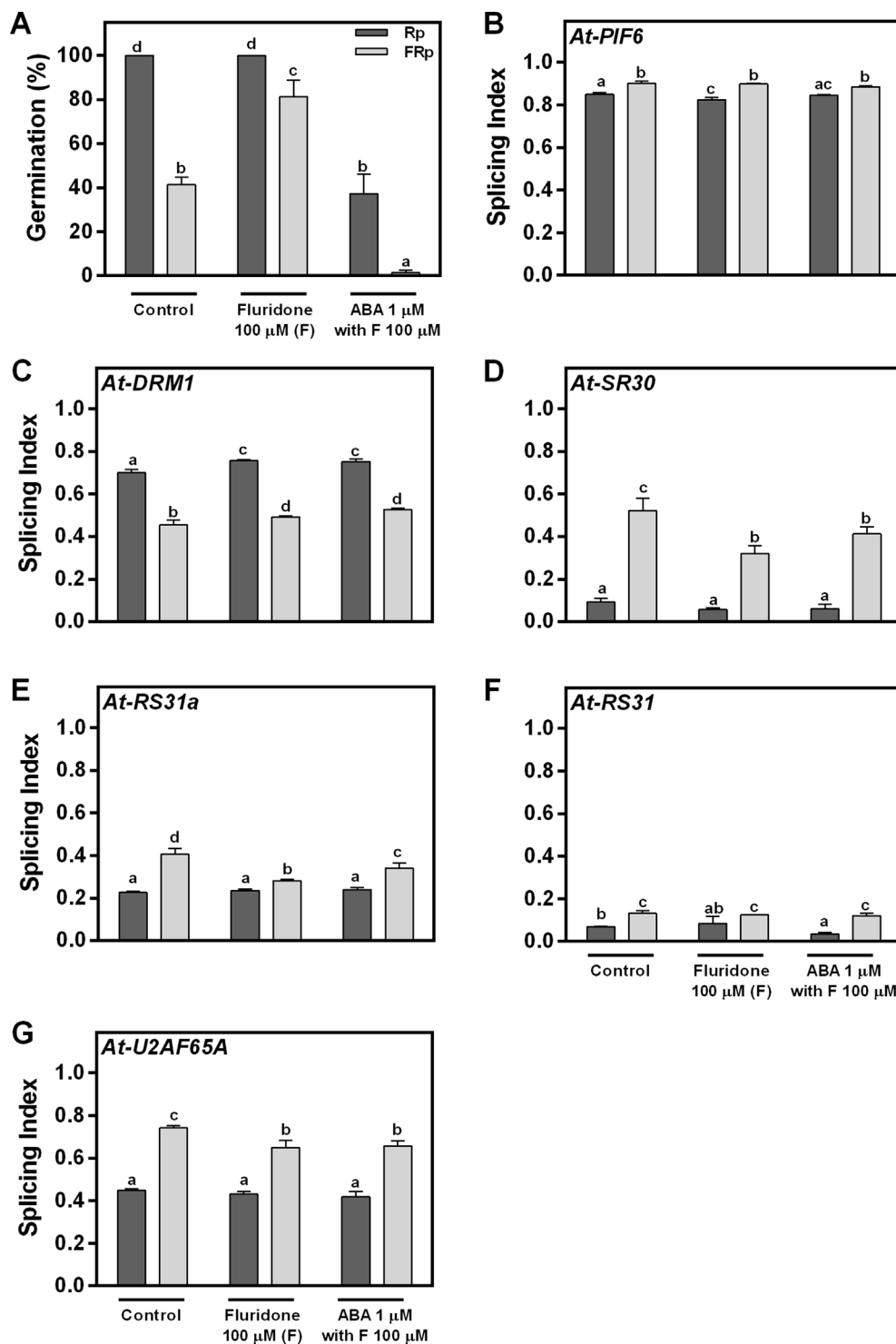


FIGURE 4 | Red light perception and alternative splicing are directly linked. **(A)** Germination percentages of Col-0 seeds imbibed in water (control), fluridone, or ABA supplemented with fluridone and irradiated with an Rp or an FRp. **(B–G)** AS alterations for each hormone condition quantified as splicing index (SI) in response to the Rp or FRp are shown. Seed samples were harvested 12 h after the Rp or FRp. Splice variants of the indicated genes were analyzed by RT-PCR, and each PCR product was quantified by densitometry using ImageJ. Each bar represents mean \pm SE ($n = 3$). Means with same letters are not significantly different ($p > 0.05$ by ANOVA followed by Fisher post-test). ABA, abscisic acid; ANOVA, analysis of variance; AS, alternative splicing; Col-0, Columbia-0; FRp, far-red pulse; Rp, red pulse; RT-PCR, reverse transcriptase–polymerase chain reaction.

phytochromes, mediating the R/FR response, might be responsible for the regulation of the AS pattern of these genes. Interestingly, *At-U2AF65A* and *At-RS31* AS is regulated by a common light pathway in seedlings, involving chloroplast's retrograde signals (Petrillo et al., 2014). Hence, these results suggest seeds can have completely different gene regulatory networks. Further research is needed to know whether other photoreceptors, or other signaling pathways (i.e., retrograde signals from organelles), could be involved in the control of AS patterns in light-germinating seeds. Moreover, since red light promotes seed germination (Figure 1B), we could argue that AS changes are a consequence of the germination process. However, we provide compelling evidence showing that Rp and FRp can directly affect AS regulation since the analyzed AS changes are not correlated with the intensity of germination response (Figure 4). As previously shown, retrograde signals arising from the organelles are important for triggering AS changes during the transition from darkness to light in *Arabidopsis* seedlings (Petrillo et al., 2014; Riegler et al., 2018). Taking this into account, it is possible to think that retrograde signals arising from different organelles could be associated with signals from photoreceptors to fine-tune the AS process in some genes during light induction of seed germination.

Finally, we conclude that AS is a source of gene expression diversity that potentially leads to different proteins involved in the promotion of seed germination by light, and a proper regulation of this process might be of key relevance for the adjustment of seed germination in the correct place and time. The next step would be to determine which signaling pathway(s) is (are) controlling the splicing of these events and identify the targets of the splicing factors regulated by light. Moreover, the question whether these changes in the AS of different splicing regulators are physiologically relevant is still open. Unraveling these AS regulatory networks will help us understand the mechanisms through which these genes may be responsible in the promotion of germination in crop seeds.

DATA AVAILABILITY

All datasets for this study are included in the manuscript and the Supplementary files. The RNA-seq data is available at the Gene Expression Omnibus (GEO) with this accession GSE134019

REFERENCES

- Arana, M. V., Sánchez-Lamas, M., Strasser, B., Ibarra, S. E., Cerdán, P. D., and Sánchez, R. A. (2014). Functional diversity of phytochrome family in the control of light and gibberellin-mediated germination in *Arabidopsis*. *Plant Cell Environ.* 37, 2014–2023. doi: 10.1111/pce.12286
- Benech-Arnold, R. L., Sánchez, R. A., Forcella, F., Kruk, B. C., and Ghera, C. M. (2000). Environmental control of dormancy in weed seed banks in soil. *Field Crops Res.* 67, 105–122. doi: 10.1016/S0378-4290(00)00087-3
- Benjamini, Y., and Hochberg, Y. (1995). Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J. R. Stat. Soc. Ser. B.* 57, 289–300. doi: 10.1111/j.2517-6161.1995.tb02031.x
- Bentsink, L., Jowett, J., Hanhart, C. J., and Koornneef, M. (2006). Cloning of DOG1, a quantitative trait locus controlling seed dormancy in

AUTHOR CONTRIBUTIONS

EP and JB conceived the project. RT, LS, and EP performed the experiments. CH and MY processed the transcriptome data. MS-S performed the GO analyses. RT, EP, and JB analyzed and interpreted the data and wrote the manuscript. All authors read and approved the final version of the manuscript.

FUNDING

The research was supported by Agencia Nacional de Promoción Científica y Tecnológica of Argentina and the University of Buenos Aires (UBA) grants to JB (UBACYT 20020170100265BA and PICT2014-1074) and to EP (PICT2016-4366). RT, CH, and MS-S are fellows from Consejo Nacional de Investigaciones Científicas y Técnicas (CONICET). JB and EP are career investigators from CONICET. LS's work was supported by the University of Buenos Aires (UBA undergraduate fellowship) and is currently a PhD fellow from CONICET.

ACKNOWLEDGMENTS

We deeply appreciate the help, discussions, and work atmospheres at IFEVA, FIL, and IFIBYNE. We also thank the assistance and suggestions from Barta and Kalyna labs at Vienna and the seeds and support from the P. Cerdán, J. Casal, and M. Yanovsky labs.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fpls.2019.01076/full#supplementary-material>

SUPPLEMENTARY FIGURE 1 | Calibration curves for the hormone experiments.

SUPPLEMENTARY TABLE 1 | Differential gene expression and alternative splicing analyses.

SUPPLEMENTARY TABLE 2 | Primers used in the study.

SUPPLEMENTARY FILE 1 | Gene models and gel images.

- Arabidopsis*. *Proc. Natl. Acad. Sci. U. S. A.* 103, 17042–17047. doi: 10.1073/pnas.0607877103
- Botto, J. F., Sánchez, R. A., and Casal, J. J. (1995). Role of phytochrome B in the induction of seed germination by light in *Arabidopsis thaliana*. *J. Plant Physiol.* 146, 307–312. doi: 10.1016/S0176-1617(11)82059-6
- Botto, J. F., Scopel, A. L., Ballaré, C. L., and Sánchez, R. A. (1998a). The effect of light during and after soil cultivation with different tillage implements on weed seedling emergence. *Weed Sci.* 46, 351–357. doi: 10.1017/S0043174500089530
- Botto, J. F., Sánchez, R. A., and Casal, J. J. (1998b). Burial conditions affect light responses of *Datura ferox* seeds. *Seed Sci. Res.* 8, 423–429. doi: 10.1017/S0960258500004384
- Botto, J. F., Sánchez, R. A., Whitelam, G. C., and Casal, J. J. (1996). Phytochrome A mediates the promotion of seed germination by very low fluences of light and canopy shade-light in *Arabidopsis*. *Plant Physiol.* 110, 439–444. doi: 10.1104/pp.110.2.439

- Botto, J. F., Scopel, A. L., and Sánchez, R. A. (2000). Water constrains on the photoinduction on weed seed germination during tillage. *Funct. Plant Biol.* 27, 463–471. doi: 10.1071/PP99199
- Braunschweig, U., Barbosa-Morais, N. L., Pan, Q., Nachman, E. N., Alipanahi, B., Gonatopoulos-Pournatzis, T., et al. (2014). Widespread intron retention in mammals functionally tunes transcriptomes. *Genome Res.* 24, 1774–1786. doi: 10.1101/gr.177790.114
- Casal, J. J., and Yanovsky, M. J. (2005). Regulation of gene expression by light. *Int. J. Dev. Biol.* 49, 501–511. doi: 10.1387/ijdb.051973jc
- Finkelstein, R., Reeves, W., Ariizumi, T., and Steber, C. (2008). Molecular aspects of seed dormancy. *Ann. Rev. Plant Biol.* 59, 387–415. doi: 10.1146/annurev.arplant.59.032607.092740
- Fouquet, R., Martin, F., Fajardo, D. S., Gault, C. M., Gomez, E., Tseung, C. W., et al. (2011). Maize *rough endosperm3* encodes an RNA splicing factor required for endosperm cell differentiation and has a nonautonomous effect on embryo development. *Plant Cell* 23, 4280–4297. doi: 10.1105/tpc.111.092163
- Gabriele, S., Rizza, A., Martone, J., Circelli, P., Costantino, P., and Vittorioso, P. (2010). The Dof protein DAG1 mediates PIL5 activity on seed germination by negatively regulating GA biosynthetic gene AtGA3ox1. *Plant J.* 61, 312–323. doi: 10.1111/j.1365-313X.2009.04055.x
- Galvao, V. C., and Fankhauser, C. (2015). Sensing the light environment in plants: photoreceptors and early signaling steps. *Curr. Opin. Neurobiol.* 34C, 46–53. doi: 10.1016/j.conb.2015.01.013
- Godoy Herz, M. A., Kubacka M. G., Brzyzek G., Servi, L., Krzyszton M., Simpson, C., et al. (2019). Light regulates plant alternative splicing through the control of transcriptional elongation. *Mol. Cell* 73, 1066–1074
- Graeber, K., Linkies, A., Steinbrecher, T., Mummenhoff, K., Tarkowska, D., Turečková, V., et al. (2014). *DELAY OF GERMINATION 1* mediates a conserved coat-dormancy mechanism for the temperature- and gibberellin-dependent control of seed germination. *Proc. Natl. Acad. Sci. U. S. A.* 111, 3571–3580. doi: 10.1073/pnas.1403851111
- Hartmann, L., Drewe-Boß, P., Wiefner, T., Wagner, G., Geue, S., Lee, H. C., et al. (2016). Alternative splicing substantially diversifies the transcriptome during early photomorphogenesis and correlates with the energy availability in *Arabidopsis*. *Plant Cell* 28, 2715–2734. doi: 10.1105/tpc.16.00508
- Hennig, L., Stoddart, W. M., Dieterle, M., Whitelam, G. C., and Schäfer, E. (2002). Phytochrome E controls light-induced germination of *Arabidopsis*. *Plant Physiol.* 128, 194–200. doi: 10.1104/pp.010559
- Ibarra, S. E., Auge, G., Sánchez, R. A., and Botto, J. F. (2013). Transcriptional programs related to phytochrome A function in *Arabidopsis* seed germination. *Mol. Plant* 6, 1261–1273. doi: 10.1093/mp/sst001
- Ibarra, S. E., Tognacca, R. S., Dave, A., Graham, I. A., Sanchez, R. A., and Botto, J. F. (2016). Molecular mechanisms underlying the entrance in secondary dormancy of *Arabidopsis* seeds. *Plant Cell Environ.* 39, 213–221. doi: 10.1111/pce.12607
- James, A. B., Syed, N. H., Bordage, S., Marshall, J., Nimmo, G. A., Jenkins, G. I., et al. (2012). Alternative splicing mediates responses of the *Arabidopsis* circadian clock to temperature changes. *Plant Cell* 24, 961–981. doi: 10.1105/tpc.111.093948
- Jones, M. A., Williams, B. A., McNicol, J., Simpson, C. G., Brown, J. W. S., and Harmer, S. L. (2012). Mutation of *Arabidopsis* *SPLICEOSOMAL TIMEKEEPER LOCUS1* causes circadian clock defects. *Plant Cell* 24, 4066–4082. doi: 10.1105/tpc.112.104828
- Jung, K. H., Bartley, L. E., Cao, P., Canlas, P. E., and Ronald, C. P. (2009). Analysis of alternatively spliced rice transcripts using microarray data. *Rice* 2, 44–55. doi: 10.1007/s12284-008-9020-9
- Juntawong, P., and Bailey-Serres, J. (2012). Dynamic light regulation of translation status in *Arabidopsis thaliana*. *Front. Plant Sci.* 3, 66. doi: 10.3389/fpls.2012.00066
- Kaiserli, E., Perrella, G., and Davidson, M. L. (2018). Light and temperature shape nuclear architecture and gene expression. *Curr. Opin. Plant Biol.* 45 (Pt A), 103–111. doi: 10.1016/j.pbi.2018.05.018
- Langmead, B., Trapnell, C., Pop, M., and Salzberg, S. L. (2009). Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol.* 10, R25.
- Lightfoot, D. J., Malone, K. M., Timmis, J. N., and Orford, S. J. (2008). Evidence for alternative splicing of MADS-box transcripts in developing cotton fibre cells. *Mol. Genet. Genomics* 279, 75–85. doi: 10.1007/s00438-007-0297-y
- Liu, M. J., Wu, S. H., Wu, J. F., Lin, W. D., Wu, Y. C., Tsai, T. Y., et al. (2013). Translational landscape of photomorphogenic *Arabidopsis*. *Plant Cell* 25, 3699–3710. doi: 10.1105/tpc.113.114769
- Liu, M. J., Wu, S. H., and Chen, H. M. (2012). Widespread translational control contributes to the regulation of *Arabidopsis* photomorphogenesis. *Mol. Syst. Biol.* 8, 566. doi: 10.1038/msb.2011.97
- Mancini, E., Sanchez, S. E., Romanowski, A., Schlaen, R. G., Sanchez-Lamas, M., Cerdán, P. D., et al. (2016). Acute effects of light on alternative splicing in light-grown plants. *Photochem. Photobiol.* 92, 126–133. doi: 10.1111/php.12550
- Mano, S., Hayashi, M., and Nishimura, M. (1999). Light regulates alternative splicing of hydroxypyruvate reductase in pumpkin. *Plant J.* 17, 309–320. doi: 10.1046/j.1365-313X.1999.00378.x
- Martin-Trillo, M., Grandio, E. G., Serra, F., Marcel, F., Rodriguez-Buey, M. L., Schmitz, G., et al. (2011). Role of tomato *BRANCHED1-like* genes in the control of shoot branching. *Plant J.* 67, 701–714. doi: 10.1111/j.1365-313X.2011.04629.x
- Matsumura, H., Kitajima, H., Akada, S., Abe, J., Minaka, N., and Takahashi, R. (2009). Molecular cloning and linkage mapping of cryptochrome multigene family in soybean. *Plant Genome* 2, 271–281. doi: 10.3835/plantgenome.2009.06.0018
- Merchante, C., Stepanova, A. N., and Alonso, J. M. (2017). Translation regulation in plants: an interesting past, an exciting present and a promising future. *Plant J.* 90, 628–653. doi: 10.1111/tpl.13520
- Nakabayashi, K., Bartsch, M., Ding, J., and Soppe, W. J. J. (2015). Seed dormancy in *Arabidopsis* requires self-binding ability of DOG1 protein and the presence of multiple isoforms generated by alternative splicing. *PLOS Genet.* 11, e1005737. doi: 10.1371/journal.pgen.1005737
- Nakabayashi, K., Bartsch, M., Xiang, Y., Miatton, E., Pellengahr, S., Yano, R., et al. (2012). The time required for dormancy release in *Arabidopsis* is determined by *DELAY OF GERMINATION1* protein levels in freshly harvested seeds. *Plant Cell* 24, 2826–2838. doi: 10.1105/tpc.112.100214
- Narsai, R., Gouil, Q., Secco, D., Srivastava, A., Karpievitch, Y. V., Liew, L. C., et al. (2017). Extensive transcriptomic and epigenomic remodelling occurs during *Arabidopsis thaliana* germination. *Genome Biol.* 18, 172. doi: 10.1186/s13059-017-1302-3
- Oh, E., Kang, H., Yamaguchi, S., Park, J., Lee, D., Kamiya, Y., et al. (2009). Genome-wide analysis of genes targeted by PHYTOCHROME INTERACTING FACTOR 3-LIKE5 during seed germination in *Arabidopsis*. *Plant Cell* 21, 403–419. doi: 10.1105/tpc.108.064691
- Oh, E., Yamaguchi, S., Hu, J., Usuke, J., Jung, B., Paik, I., et al. (2007). PIL5, a phytochrome-interacting bHLH protein, regulates gibberellin responsiveness by binding directly to the GAI and RGA promoters in *Arabidopsis* seeds. *Plant Cell* 19, 1192–1208. doi: 10.1105/tpc.107.050153
- Oh, E., Yamaguchi, S., Kamiya, Y., Bae, G., Chung, W. I., and Choi, G. (2006). Light activates the degradation of PIL5 protein to promote seed germination through gibberellin in *Arabidopsis*. *Plant J.* 47, 124–139. doi: 10.1111/j.1365-313X.2006.02773.x
- Paik, I., Yang, S., and Choi, G. (2012). Phytochrome regulates translation of mRNA in the cytosol. *Proc. Natl. Acad. Sci. U. S. A.* 109, 1335–1340. doi: 10.1073/pnas.1109683109
- Park, E., Park, J., Kim, J., Nagatani, A., Lagarias, J. C., and Choi, G. (2012). Phytochrome B inhibits binding of phytochrome-interacting factors to their target promoters. *Plant J.* 72, 537–546. doi: 10.1111/j.1365-313X.2012.05114.x
- Penfield, S., Josse, E. M., Kannangara, R., Gilday, A. D., Halliday, K. J., and Graham, I. A. (2005). Cold and light control seed germination through the bHLH transcription factor SPATULA. *Curr. Biol.* 15, 1998–2006.
- Penfield, S., Josse, E. M., and Halliday, K. J. (2010). A role for an alternative splice variant of *PIF6* in the control of *Arabidopsis* primary seed dormancy. *Plant Mol. Biol.* 73, 89–95. doi: 10.1007/s11103-009-9571-1
- Pervouchine, D. D., Knowles, D. G., and Guigo, R. (2013). Intron-centric estimation of alternative splicing from RNA-seq data. *Bioinformatics* 29, 273–274. doi: 10.1093/bioinformatics/bts678
- Petrillo, E., Godoy Herz, M. A., Barta, A., Kalyna, M., and Kornblihtt, A. R. (2015). Let there be light: regulation of gene expression in plants. *RNA Biol.* 11, 1215–1220. doi: 10.4161/15476286.2014.972852
- Petrillo, E., Godoy Herz, M. A., Fuchs, A., Reifer, D., Fuller, J., Yanovsky, M. J., et al. (2014). A chloroplast retrograde signal regulates nuclear alternative splicing. *Science* 344, 427–430. doi: 10.1126/science.1250322
- Reddy, A. S. N., Marquez, Y., Kalyna, M., and Barta, A. (2013). Complexity of the alternative splicing landscape in plants. *Plant Cell* 25, 3657–3683. doi: 10.1105/tpc.113.117523

- Riegler, S., Servi, L., Fuchs, A., Godoy Herz, M. A., Kubaczka, M. G., Venhuizen, P., et al. (2018). Light remote control of alternative splicing in roots through TOR kinase. Pre-print at bioRxiv. Available at: <https://www.biorxiv.org/content/10.1101/472126v2>.
- Robinson, M. D., McCarthy, D. J., and Smyth, G. K. (2010). edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics*. 26, 139–140. doi: 10.1093/bioinformatics/btp616
- Rosloski, S. M., Singh, A., Jali, S. S., Balasubramanian, S., Weigel, D., and Grbic, V. (2013). Functional analysis of splice variant expression of MADS AFFECTING FLOWERING 2 of *Arabidopsis thaliana*. *Plant Mol. Biol.* 81, 57–69. doi: 10.1007/s11030-012-9982-2
- Sanchez, S. E., Petrillo, E., Beckwith, E. J., Zhang, X., Rugnone, M. L., Hernando, C. E., et al. (2010). A methyl transferase links the circadian clock to the regulation of alternative splicing. *Nature*. 468, 112–116. doi: 10.1038/nature09470
- Scopel, A. L., Ballaré, C. L., and Sánchez, R. A. (1991). Induction of extreme light sensitivity in buried weed seeds and its role in the perception of soil cultivations. *Plant Cell Environ.* 14, 501–508. doi: 10.1111/j.1365-3040.1991.tb01520.x
- Shikata, H., Hanada, K., Ushijima, T., Nakashima, M., Suzuki, Y., and Matsushita, T. (2014). Phytochrome controls alternative splicing to mediate light responses in *Arabidopsis*. *Proc. Natl. Acad. Sci. U. S. A.* 111, 18781–18786. doi: 10.1073/pnas.1407147112
- Shikata, H., Shibata, M., Ushijima, T., Nakashima, M., Kong, S. G., Matsuoka, K., et al. (2012). The RS domain of *Arabidopsis* splicing factor RRC1 is required for phytochrome B signal transduction. *Plant J.* 70, 727–738. doi: 10.1111/j.1365-3113.2012.04937.x
- Shinomura, T., Nagatani, A., Chory, J., and Furuya, M. (1994). The induction of seed germination in *Arabidopsis thaliana* is regulated principally by phytochrome B and secondarily by phytochrome A. *Plant Physiol.* 104, 363–371. doi: 10.1104/pp.104.2.363
- Shinomura, T., Nagatani, A., Hanzawa, H., Kubota, M., Watanabe, M., and Furuya, M. (1996). Action spectra for phytochrome A- and B-specific photoinduction of seed germination in *Arabidopsis thaliana*. *Proc. Natl. Acad. Sci. U. S. A.* 93, 8129–8133. doi: 10.1073/pnas.93.15.8129
- Simpson, C. G., Fuller, J., Maronova, M., Kalyna, M., Davidson, D., McNicol, J., et al. (2008). Monitoring changes in alternative precursor messenger RNA splicing in multiple gene transcripts. *Plant J.* 53, 1035–1048. doi: 10.1111/j.1365-3113.2007.03392.x
- Srinivasan, A., Jiménez-Gómez, J. M., Fornara, F., Soppe, W. J. J., and Brambilla, V. (2016). Alternative splicing enhances transcriptome complexity in desiccating seeds. *JIPB*. 58, 947–958. doi: 10.1111/jipb.12482
- Sugliani, M., Brambilla, V., Clerckx, E. J. M., Koornneef, M., and Soppe, W. J. J. (2010). The conserved splicing factor SUA controls alternative splicing of the developmental regulator ABI3 in *Arabidopsis*. *Plant Cell*. 22, 1936–1946. doi: 10.1105/tpc.110.074674
- Syed, N. H., Kalyna, M., Marquez, Y., Barta, A., and Brown, J. W. S. (2012). Alternative splicing in plants—coming of age. *Trends Plant Sci.* 17, 616–623. doi: 10.1016/j.tplants.2012.06.001
- Trapnell, C., Pachter, L., and Salzberg, S. L. (2009) TopHat: discovering splice junctions with RNA-Seq. *Bioinformatics* 25, 1105–1111.
- Tsai, H. L., Li, Y. H., Hsieh, W. P., Lin, M. C., Ahn, J. H., and Wu, S. H. (2014). HUA ENHANCER1 is involved in posttranscriptional regulation of positive and negative regulators in *Arabidopsis* photomorphogenesis. *Plant Cell*. 26, 2858–2872. doi: 10.1105/tpc.114.126722
- van Gelderen, K., Kang, C., and Pierik, R. (2018). Light signaling, root development, and plasticity. *Plant Physiol.* 176, 1049–1060. doi: 10.1104/pp.17.01079
- Wang, Y., Fan, X., Lin, F., He, G., Terzaghi, W., Zhu, D., et al. (2014). *Arabidopsis* noncoding RNA mediates control of photomorphogenesis by red light. *Proc. Natl. Acad. Sci. U. S. A.* 111, 10359–10364. doi: 10.1073/pnas.1409457111
- Wu, H. P., Su, Y. S., Chen, H. C., Chen, Y. R., Wu, C. C., Lin, W. D., et al. (2014). Genome-wide analysis of light-regulated alternative splicing mediated by photoreceptors in *Physcomitrella patens*. *Genome Biol.* 15, R10. doi: 10.1186/gb-2014-15-1-r10
- Xin, R., Kathare, P. K., and Huq, E. (2019). SFPS forms a complex with the SR-like protein RRC1. *Plant Cell*. doi: 10.1105/tpc.18.00786
- Yakir, E., Hilman, D., Hassidim, M., and Green, R. M. (2007). CIRCADIAN CLOCK ASSOCIATED1 transcript stability and the entrainment of the circadian clock in *Arabidopsis*. *Plant Physiol.* 145, 925–932. doi: 10.1104/pp.107.103812.

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2019 Tognacca, Servi, Hernando, Saura-Sanchez, Yanovsky, Petrillo and Botto. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



Quantitative Proteomics Reveals a Role for SERINE/ARGININE-Rich 45 in Regulating RNA Metabolism and Modulating Transcriptional Suppression *via* the ASAP Complex in *Arabidopsis thaliana*

Samuel L. Chen¹, Timothy J. Rooney², Anna R. Hu², Hunter S. Beard³, Wesley M. Garrett⁴, Leann M. Mangalath⁵, Jordan J. Powers², Bret Cooper³ and Xiao-Ning Zhang^{2,5*}

OPEN ACCESS

Edited by:

Craig G. Simpson,
The James Hutton Institute,
United Kingdom

Reviewed by:

R. Glen Uhrig,
University of Alberta, Canada
Xia Wu,
University of Washington,
United States

*Correspondence:

Xiao-Ning Zhang
xzhang@sbsu.edu

Specialty section:

This article was submitted to
Plant Proteomics,
a section of the journal
Frontiers in Plant Science

Received: 29 March 2019

Accepted: 14 August 2019

Published: 19 September 2019

Citation:

Chen SL, Rooney TJ, Hu AR,
Beard HS, Garrett WM,
Mangalath LM, Powers JJ,
Cooper B and Zhang X-N (2019)
Quantitative Proteomics Reveals a
Role for SERINE/ARGININE-Rich
45 in Regulating RNA Metabolism
and Modulating Transcriptional
Suppression *via* the ASAP Complex
in *Arabidopsis thaliana*.
Front. Plant Sci. 10:1116.
doi: 10.3389/fpls.2019.01116

¹ Bioinformatics Program, St. Bonaventure University, St. Bonaventure, NY, United States, ² Biochemistry Program, St. Bonaventure University, St. Bonaventure, NY, United States, ³ Soybean Genomics and Improvement Laboratory, USDA-ARS, Beltsville, MD, United States, ⁴ Animal Biosciences & Biotechnology Laboratory, USDA-ARS, Beltsville, MD, United States, ⁵ Department of Biology, St. Bonaventure University, St. Bonaventure, NY, United States

Pre-mRNA alternative splicing is a conserved mechanism for eukaryotic cells to leverage existing genetic resources to create a diverse pool of protein products. It is regulated in coordination with other events in RNA metabolism such as transcription, polyadenylation, RNA transport, and nonsense-mediated decay *via* protein networks. SERINE/ARGININE-RICH 45 (SR45) is thought to be a neutral splicing regulator. It is orthologous to a component of the apoptosis and splicing-associated protein (ASAP) complex functioning to regulate RNA metabolism at multiple levels. Within this context, we try to understand why the *sr45-1* mutant *Arabidopsis* has malformed flowers, delayed flowering time, and increased disease resistance. Prior studies revealed increased expression for some disease resistance genes and the flowering suppressor *Flowering Locus C* (*FLC*) in *sr45-1* mutants and a physical association between SR45 and reproductive process-related RNAs. Here, we used Tandem Mass Tag-based quantitative mass spectrometry to compare the protein abundance from inflorescence between *Arabidopsis* wild-type (Col-0) and *sr45-1* mutant plants. A total of 7,206 proteins were quantified, of which 227 proteins exhibited significantly different accumulation. Only a small percentage of these proteins overlapped with the dataset of RNAs with altered expression. The proteomics results revealed that the *sr45-1* mutant had increased amounts of enzymes for glucosinolate biosynthesis which are important for disease resistance. Furthermore, the mutant inflorescence had a drastically reduced amount of the Sin3-associated protein 18 (SAP18), a second ASAP complex component, despite no significant reduction in *SAP18* RNA. The third ASAP component protein, ACINUS, also had lower abundance without significant RNA changes in the *sr45-1* mutant. To test the effect of SR45 on SAP18, a SAP18-GFP fusion protein was overproduced in transgenic *Arabidopsis* Col-0 and *sr45-1* plants. SAP18-GFP has less accumulation in the nucleus, the site of activity for the ASAP complex, without SR45.

Furthermore, transgenic *sr45-1* mutants overproducing SAP18-GFP expressed even more *FLC* and had a more severe flowering delay than non-transgenic *sr45-1* mutants. These results suggest that SR45 is required to maintain the wild-type level of SAP18 protein accumulation in the nucleus and that *FLC*-regulated flowering time is regulated by the correct expression and localization of the ASAP complex.

Keywords: ACINUS, apoptosis and splicing-associated protein complex, *Arabidopsis thaliana*, inflorescence, quantitative proteomics, RNA metabolism, Sin3-associated protein 18, SERINE/ARGININE-rich 45

INTRODUCTION

In eukaryotic cells, pre-mRNA alternative splicing is a conserved mechanism to increase the diversity of mature transcripts and their protein products. The spliced mature mRNAs with proper 5'-7-methylguanosine capping and 3'-polyadenylation are resistant to immediate degradation by cellular machinery and are viable templates for translation. A successful splicing event consists of several sequential steps: splicing factors recruiting spliceosome components; spliceosome components aggregating in sequence to recognize splice sites; catalysis involving the 5' splice site, branch site, and 3' splice site; the release of the excised intron and spliced mRNA; and a conclusion with spliceosome disassembling. This process is energy-dependent and is regulated in coordination with other events in RNA metabolism such as transcription, polyadenylation, nuclear export, and nonsense-mediated decay (NMD). These related events are coordinated through a network of protein players, including the SR proteins, a family of known splicing regulators (Fu and Ares, 2014).

Evidence suggests that the SERINE/ARGININE-rich 45 (SR45) protein in *Arabidopsis thaliana* acts as a neutral splicing regulator that could trigger nearby splicing activation or suppression events (Zhang et al., 2017). SR45, an RNA-binding protein, is orthologous to RNPS1 in humans and other animals (Zhang and Mount, 2009). RNPS1 is a component of the apoptosis and splicing-associated protein (ASAP) complex (Schwerk et al., 2003), which functions to regulate RNA metabolism at multiple levels (Deka and Singh, 2017). The two other core proteins in the ASAP complex are SAP18 and ACINUS (Schwerk et al., 2003). A general understanding of the function of the ASAP complex has been mostly focused on transcriptional repression because SAP18 can bind to the mSIN3 transcriptional repressor to recruit histone deacetylases (HDACs) to induce transcriptional silencing in mammalian cells (Zhang et al., 1997).

In addition to functioning in the ASAP complex, RNPS1 is a peripheral component of the conserved RNA quality control machinery exon junction complex (EJC) and is involved in splicing regulation and communication with NMD (Lykke-Andersen et al., 2001), a surveillance process that removes mRNA transcripts harboring premature stop codons (PTCs). The multifaceted involvement of RNPS1 in transcriptional regulation, splicing, and RNA quality control denotes its significance in the regulatory network for RNA metabolism. Due to SR45's orthology to RNPS1, it is likely that SR45 has these RNPS1 functions.

A. thaliana with the *sr45-1* null mutation exhibits both vegetative and reproductive defects such as smaller stature with narrower leaves and flower petals, delayed root growth, late flowering, and a mild sterility during seed formation (Ali et al., 2007; Zhang et al., 2017). The pre-mRNA of the *SR45* gene is alternatively spliced into two functional isoforms, *SR45.1* and *SR45.2*. The protein product of these alternative transcripts has distinct functions: *SR45.1* is mostly involved in flower development and *SR45.2* plays a bigger role in proper root growth (Zhang and Mount, 2009). The two isoforms differ by 7 amino acids which are missing in *SR45.2*. Within this alternative fragment, a phosphorylation event on threonine 218 is instrumental in bringing about the separate functions of the two transcripts (Zhang et al., 2014). Independent efforts have been put forth to understand the molecular mechanisms that *SR45.1* employs during plant reproduction (Ali et al., 2007; Ausin et al., 2012; Zhang et al., 2014; Questa et al., 2016; Zhang et al., 2017). The current understanding suggests that in the inflorescence *SR45* is associated with RNAs functioning in a wide range of processes, from splicing to reproduction, and that *SR45*-dependent alternative splicing events are overrepresented in transcripts for RNA binding proteins and RNA splicing (Zhang et al., 2017). Although the exact mechanisms have not been proven, it is possible that the fate of a splicing event is determined *via* direct *SR45*-RNA interaction or indirectly by association with other *SR45*-associated proteins, such as splicing factors and spliceosome components (Golovkin and Reddy, 1999; Day et al., 2012; Baldwin et al., 2013; Zhang et al., 2014; Stankovic et al., 2016).

There are, however, other *A. thaliana* phenotypes of the *sr45-1* mutant that are not as easily explained by altered RNA splicing events. For example, there is an elevated RNA level of *FLOWER LOCUS C (FLC)* in the *sr45-1* mutant (Ali et al., 2007). A recent study discovered the presence of the ASAP complex at the *FLC* locus, which suggests that the ASAP complex recruits HDACs to the *FLC* locus for transcriptional repression of *FLC* (Questa et al., 2016). This finding expanded the narrow focus of *SR45* as a splicing regulator to a chromatin-level transcriptional control factor. In mammalian cells, it has been found that epigenetic changes could affect the rate of transcription and the subsequent outcome in RNA splicing (Warns et al., 2016). We have found that *SR45* is physically associated with the *FLC* RNA, but *FLC* RNA was not alternatively spliced in the *sr45-1* mutant (Zhang et al., 2017). Questions still remain as to whether there is any coordination between *SR45*-regulated splicing events and chromatin modification for the same gene. Interestingly, *FLC* is

not the only flowering-related gene that SR45 regulates. The *sr45-1* mutant also displays a lower frequency of DNA methylation at the *FLOWERING WAGENINGEN* (*FWA*) locus and a reduction of the RNA-dependent DNA methylation (RdDM) pathway, which is associated with the delayed flowering phenotype in the *sr45-1* mutant (Ausin et al., 2012). Thus, SR45 may regulate RdDM components that coordinate gene silencing for flowering.

In addition, the *sr45-1* mutant has enhanced resistance to *Pseudomonas syringae* pv. *maculicola* strain DG3 and *Hyaloperonospora parasitica* isolate Noco2 and has higher levels of callose deposition at the cell wall, reactive oxygen species, and salicylic acid (Zhang et al., 2017). SR45 also exhibits a strong preference in suppressing plant innate immunity genes such as *PR1*, *PR5*, *ACD6*, and *PAD4* (Zhang et al., 2017) which cannot be explained by alternative splicing alone. Consequently, SR45 is considered a suppressor of innate immunity in *A. thaliana*.

These observations urge an exploration of the possibility that SR45 is more than just a splicing regulator. To better understand the proteomic landscape in which SR45 acts to affect RNA metabolism in inflorescence tissue, we evaluated inflorescence proteins from wild-type *A. thaliana* (Col-0) and *sr45-1* mutants by quantitative tandem mass spectrometry. We identified 227 differentially accumulated proteins and predicted their roles in RNA metabolism and other biological processes. Our data shows that SR45 likely functions through the ASAP complex to suppress *FLC* and immunity genes.

MATERIALS AND METHODS

Plant Growth Condition

All *A. thaliana* plants used in this study are in the *Colombia* (Col-0) background. The *sr45-1* (SALK_004132) mutant plant was originally obtained from the *Arabidopsis Biological Resource Center* (ABRC). Primers used to confirm T-DNA insertion in *sr45-1* were described (Zhang and Mount, 2009) and are listed in **Supplemental Table S1**. All plants were grown in soil (Sunshine #8, Griffin Greenhouse & Nursery Supplies), under a long-day (LD) condition of a 16/8 h photoperiod with light intensity of 100 $\mu\text{mol m}^{-2} \text{s}^{-1}$ at 22 °C or otherwise specified.

Total Protein Extraction

Healthy inflorescence tissues were harvested from 6-week old plants, and ground into fine powder using liquid nitrogen. Total protein was extracted according to the protocol from the Plant Total Protein Extraction Kit (PE0230, Sigma).

Peptide Preparation

Protein concentration was determined by bicinchoninic assay (Pierce, Rockford, IL, USA). Proteins (~300 μg), dissolved in urea, were reduced in 5 mM Tris(2-carboxyethyl)phosphine for 20 min, carboxyamidomethylated with 20 mM iodoacetamide for 20 min, and digested overnight at 37°C with Poroszyme immobilized trypsin (Thermo Fisher Scientific, Waltham, MA). The digested peptides were purified by reverse phase chromatography using SPEC-PLUS PT C18 columns (Varian, Lake Forrest, CA, USA). Eighty

micrograms of peptides from each sample was labeled with TMT 6-plex reagents according to manufacturer instructions (Thermo Fisher Scientific). The samples were dried, resuspended in 0.1% trifluoroacetic acid, and desalted, and the peptide concentrations measured with the Pierce Quantitative Colorimetric Peptide Assay (Thermo Fisher Scientific). Small, equivalent volumes of samples were combined and a 500 ng aliquot was analyzed by mass spectrometry (below) to determine label incorporation percentage (>99%) and to estimate quantitative ratios between samples. The labeled samples were then mixed in equal amounts based on the quantitative ratios and separated by high pH reverse-phase HPLC through a Waters Xbridge 3.5 μm C18 column (4.6 \times 15 cm) with a Dionyx UltiMate 3000 pump controlling a 38 min linear gradient from 4% to 28% acetonitrile and 0.5% triethylamine pH 10.7 (Wang et al., 2011). Seventy-five fractions were pooled by concatenation, dried, and resuspended in 5% acetonitrile and 0.1% formic acid, and the peptide concentrations were measured for 13 pools (Wang et al., 2011).

Mass Spectrometry

Peptides (~500 ng to 1 μg per pool) were separated on a 75 μm (inner diameter) fused silica capillary emitter packed with 22 cm of 2.5 μm Synergi Hydro-RP C18 (Phenomenex, Torrance, CA) coupled directly to a Dionex UltiMate 3000 RSLCnano System (Thermo Fisher Scientific) controlling a 180 min linear gradient from 3.2% to 40% acetonitrile and 0.1% formic acid at a flow rate of 300 nl per min. Peptides were electrosprayed at 2.4 kV into an Orbitrap Fusion Lumos Tribrid mass spectrometer (ThermoFisher) operating in data-dependent mode with positive polarity and using *m/z* 445.12003 as an internal mass calibrant. Quadrupole isolation was enabled and survey scans were recorded in the Orbitrap at 120,000 resolution over a mass range of 400–1,600 *m/z*. The instrument was operated in Top Speed mode using the multinotch MS³ method with a cycle time of 3 s (McAlister et al., 2014; Isasa et al., 2015). The automatic gain control (AGC) target was set to 200,000 and the maximum injection time was set to 50 ms. The most abundant precursor ions (intensity threshold 5,000) were fragmented by collision-induced dissociation (35% energy) and fragment ions were detected in the linear ion trap (AGC 10,000, 50 ms maximum injection). Analyzed precursors were dynamically excluded for 45 s. Multiple MS² fragment ions were captured using isolation waveforms with multiple frequency notches and fragmented by high energy collision-induced dissociation (65% normalized collision energy). MS³ spectra were acquired in the Orbitrap (AGC 100,000; isolation window 2.0 *m/z*, maximum injection time 120 ms, 60,000 resolution scanning from 100–500 *m/z*).

Peptide Matching and Statistics

Mass spectrometry data files were processed with Proteome Discoverer 2.1 (Thermo Fisher Scientific) which extracted MS² spectra for peptide identification and MS³ spectra for peptide quantitation. MS² spectra were searched with Mascot 2.5.1. (Perkins et al., 1999) against the *A. thaliana* protein database (TAIR10 with 35,386 records including splice variants) appended

with a list of 172 sequences to detect common contaminants. Search parameters were for tryptic digests with two possible missed cleavages, fixed amino acid modification for chemically modified cysteine and labeled N-terminal and internal lysine (+57.021 Da, C; +229.163 Da, K), variable oxidized methionine (+15.995 Da, M), monoisotopic mass values, ± 10 ppm parent ion mass tolerance, and ± 0.6 Da fragment ion mass tolerance. Peptide spectrum matches (PSM) were processed by Percolator (Kall et al., 2007) using delta Cn (0.05), strict false discovery rate (FDR) (0.01), relaxed FDR (0.05) and PEP (0.05) settings. Additional filters limited Mascot Ions scores (greater than or equal to 13) and PSM and peptide PEPs (strict 0.01; relaxed 0.05). Peptides were assigned to logical protein groups using parsimony. Proteins were quantified on summed signal-to-noise (S/N) ratios for each TMT channel for qualified PSMs for unique peptides (isolation interference < 25%, average reporter S/N > 8). The most confident centroid within 0.003 Da of the expected mass of the reporter ions was used. TMT signals were also corrected for isotope impurities (lot specific data provided by the manufacturer). Missing values were replaced with a minimum value. Matches to contaminants and decoys were removed from the dataset as were proteins with quantitative signal sums across all 6 channels < 150 or proteins with less than 2 PSMs contributing to the qualified quantitative signal sum. Protein quantitative values for each channel were normalized and then scaled to 100 across the channels. Plotting the log₂ fold changes after normalization (**Supplemental Figure S1A**) revealed a normal distribution. A t-test was used to measure significant differences and the Benjamini and Hochberg correction was applied to limit the FDR to 0.05. All proteins with an *FDR* < 0.01 and an abundance ratio (*sr45-1*/Col-0) > 1.20 or < 0.80 were defined as SR45-dependent differently accumulated proteins.

RNA-Protein Expression Comparison

Comparisons were performed between the transcriptome data from our previous study (Zhang et al., 2017) and the proteome data from this study. All SR45-differentially regulated (SDR) RNAs identified by at least two independent pipelines (Tophat2, STAR and Lasergene v12) were pooled together to generate two RNA lists with either higher expression in Col-0 or with higher expression in *sr45-1* (**Supplemental Table S2**). Each of these two lists was compared with the respective protein list with either higher accumulation in Col-0 or with higher accumulation in *sr45-1*. Identities of SR45-associated RNAs (SARs), SR45-dependent alternative splicing (SAS) RNAs and SR45-dependent differentially accumulated (SDA) proteins found in inflorescences were also compared for overlap. All comparisons were performed using R.

Functional Enrichment Analysis

PANTHER v14.0 (Mi et al., 2017) was used for GO term enrichment analysis. A list of proteins with greater accumulation in Col-0 and a list of proteins with greater accumulation in the *sr45-1* mutant were each submitted to STRING version 11.0 (Szklarczyk et al., 2015) for functional term enrichment analysis and visualization. All available evidence through STRING was used to define

protein-protein associations. The evidence includes known interactions that were determined experimentally (from curated databases), predicted interactions (gene neighborhood, gene fusions, and gene co-occurrence), and other evidence (textmining, co-expression, and protein homology). A high confidence score of 0.700 was used as the minimum required score for filtering. Disconnected nodes in the network were not displayed due to the lack of evidence for their association with other proteins. Proteins belonging to enriched GO terms and/or KEGG Pathways that are highly relevant to RNA metabolism or known functions of SR45 were highlighted in different colors. All proteins with available KEGG IDs were mapped to their corresponding biological pathways using the KEGG mapping tool (www.genome.jp/kegg/).

Predicted Protein Sequence and Structure Alignment

The amino acid sequences for SR45 (AtRNPS1, At1g16610), AtSAP18 (At2g45640) and AtACINUS (At4g39680) were used for sequence alignment with the sequences for animal ASAP complex protein model (PDB 4A8X) using ClustalW 2.1 (<https://www.genome.jp/tools-bin/clustalw>). The conserved sequences were then submitted to I-Tasser (Yang et al., 2015) for protein structure prediction. The predicted protein models were used to perform a structural alignment between *A. thaliana* and animal ASAP complex proteins using the PyMOL Molecular Graphics System, version 1.3 (Schrödinger, LLC).

Total RNA Extraction and Real Time-qPCR

The RNeasy Plus Mini Kit (Qiagen) was used to extract RNA. About 5 μ g of RNA from each sample was treated by DNase (ThermoFisher) followed by reverse transcription with Superscript IV (ThermoFisher). Real time-qPCR was performed using Power SYBR Master Mix (ThermoFisher) on a CFX96 machine (Bio-Rad). Expression levels were normalized to the expression of *GAPDH*. Primer sequences are listed in **Supplemental Table S1**.

SAP18 Cloning

The CDS of *SAP18* and the genomic *SAP18* (*gSAP*) sequences were amplified from either Col-0 inflorescence cDNA or genomic DNA using primers *SAP18ATGXhoI* and *SAP18nonstopKpnI*. The PCR products were inserted into *XhoI/KpnI* sites in the same GFP overexpression vector (pGlobug) as used before (Zhang and Mount, 2009) to create a 35S::*SAP18CDS-GFP* or 35S::*gSAP18-GFP* fusion, respectively. The overexpression cassettes 35S::*SAP18CDS-GFP-NOS3'* and 35S::*gSAP18-GFP-NOS3'* were isolated by *NotI* and cloned into a binary vector pMLBart, separately. All primers used in the cloning process are listed in **Supplemental Table S1**.

Plant Transformation, Screening, and Verification of Transgenic Plants

DNA plasmids *pMLBart-35S::SAP18CDS-GFP-NOS3'* and *pMLBart-35S::gSAP18-GFP-NOS3'* were individually transformed into *Agrobacterium tumefaciens* GB3101 and used to transform Col-0 and *sr45-1* mutant plants by flower-dipping (Bent, 2000).

All T1 plants were screened for *Basta* resistance by Finale (1:1,000 dilution) spray and examined for the GFP signal using a Nikon D-Eclipse C1 confocal microscope. At least 20 independent transgenic lines were selected for genotype confirmation using primers for the *SAP18-GFP* fusion (*SAP18ATGXhoI* and *GFPR* in **Supplemental Table S1**) and validated by confocal imaging using an Eclipse Ti confocal microscope (Nikon).

Quantification of GFP Signal Intensity in Transgenic Root Cells

Eight days old 35S::g*SAP18-GFP* transgenic seedlings were used for quantifications of GFP signal intensity in root cells using an Eclipse Ti confocal microscope (Nikon). For each root, a Z-stack was generated with 0.35 μ m per section for a total of 12 sections. The maximum signal was obtained by NIS Element (Nikon). Then the GFP signal intensity in nucleus and cytoplasm in each cell was quantified by ImageJ. A total of 15 cells per seedling were used for measurement in 3 seedlings. The average ratio of nucleus-to-cytoplasm GFP intensity per seedling was used for statistical analysis.

Statistical Analysis

Normal distribution of samples was tested by Shapiro-Wilk normality test. For experiments that passed the normality test, one-way ANOVA followed by Tukey's HSD test was performed when comparing more than two groups to each other. Unpaired Student t-test was used for two groups comparison. For experiments that did not pass the normality test, Kruskal-Wallis followed by post-hoc Dunn test was performed when comparing more than two groups to each other, Benjamini-Hochberg FDR method was used to calculate adjusted *p*-values.

RESULTS

SR45 Modulates Proteins Functioning in RNA Metabolism

We previously compared RNA sequences from inflorescence from *A. thaliana* Col-0 and *sr45-1* mutants (Zhang et al., 2017). There were 358 differentially expressed RNAs (SDR). The SDR transcript gene ontologies (GO) did not explain flower development, but there was an elevated number of transcripts involved in immunity in the *sr45-1* mutants. The analysis also identified 542 SR45-dependent alternative splicing events (SAS) that, for the most part, did not overlap with gene expression changes. Again, GO analysis of these transcripts did not explain flower development, but rather revealed the breadth of transcripts that SR45 influences through alternative splicing. Immunoprecipitation, however, revealed a set of RNAs physically associated with SR45 (SAR) that included transcripts encoding nucleic acid binding proteins involved in the regulation of flowering, flower development, and embryo development in seeds. These results prompted us to investigate *sr45-1* mutant inflorescence at a proteomic level.

We extracted protein from the remaining bulked inflorescence tissue used for the RNA sequencing study. This included three biological replicates from Col-0 and the *sr45-1* null mutants.

Tandem Mass Tag-based quantitative mass spectrometry was employed to generate quantifiable results. After searching the MS2 spectra against the *A. thaliana* protein database of 35,386 records in TAIR10, including splicing variants, a total of 58,962 peptides were determined from 101,605 peptide-spectrum matches (PSMs) (**Supplemental Figure S1B**). From these peptides, 10,120 *A. thaliana* proteins were identified (**Supplemental Figure S1B**). The quantification methods yielded 7,206 proteins with high-quality quantification information (**Supplemental Figure S1C**). On the basis of the quantified relative abundances of proteins, the three biological replicates of Col-0 closely clustered together and were distinct from the cluster of three biological replicates of the *sr45-1* mutant (**Supplemental Figure S1C**). This provides high confidence within the rigors of the experimental procedure and subsequent data analysis.

A total of 227 proteins exhibited statistically different accumulation with at least 20% fold change (**Supplemental Table S3**). Among these proteins was SR45 which had a 93% decrease in accumulation in *sr45-1*. It has been confirmed that *sr45-1* is a T-DNA knock-out, and it does not produce a full-length transcript. However it does produce a truncated transcript at 8% the rate of the full-length transcript in wild type (Ali et al., 2007). It would be unlikely for this truncated transcript to produce a viable peptide because it lacks a stop codon and 3' UTR for polyadenylation. The reason why SR45 did not appear to be absent in the mutant is because TMT signal reporting is relative, not absolute, and because signal values, even small ones, are normalized and scaled for analysis (Rauniyar and Yates, 2014). Nonetheless, SR45 exhibited the greatest decrease of all proteins measured, which is consistent with our expectation.

The other 226 proteins, defined as SR45-dependent differentially accumulated (SDA) proteins, were mostly distinct from the SDR RNAs identified from the same tissue (Zhang et al., 2017). In order to increase the coverage of SDR RNAs, we combined all SDR RNAs identified by any two of the three independent pipelines (Tophat2, STAR and Lasergene v12) to create larger datasets of 444 SR45-upregulated RNAs and 776 SR45-downregulated RNAs (**Supplemental Table S2**), whereas in our previous methods, SDR RNAs were determined only when all three pipelines agreed (Zhang et al., 2017). Even with the larger SDR RNA datasets, only 19 (8% of 227) of SDA proteins overlapped (**Figure 1A, Table 1**). However, 57 (25% of 227) of SDA proteins were found to be SAR products, SAS products, or both (**Figure 1B**), although their RNAs were not differentially expressed. In the *sr45-1* mutant, the splicing pattern of 11 SARs was altered and there was a lower steady state level of their corresponding protein products (**Figure 1B, Table 2**). It is quite possible that after SR45 binding to RNA targets, the splicing pattern of these RNAs was influenced, which provided the template for the formation of protein products. Some of the RNAs for the SDA proteins mentioned above have been confirmed for differential expression and/or alternative splicing changes (Zhang et al., 2017). Most of these 11 genes code for enzymes, one of which is a jmjC domain-containing lysine-specific histone demethylase (IBM1, **Table 2**). IBM1 reduces histone H3K9 methylation and prevents CHG hypermethylation in active genes in *A. thaliana* and *Populus* (Miura et al., 2009; Fan et al., 2018).

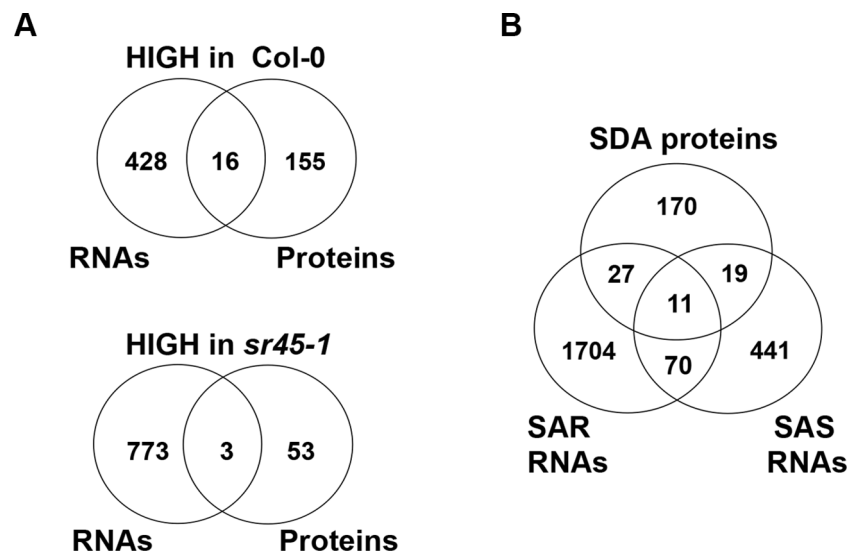


FIGURE 1 | An RNA-protein expression association study. **(A)** A comparison between SR45-differentially regulated RNAs and SR45-dependent differentially accumulated (SDA) proteins; **(B)** a comparison between SR45-dependent differentially accumulated (SDA) proteins, SR45-dependent alternatively spliced (SAS) RNAs and SR45-associated RNAs (SARs), published previously (Zhang et al., BMC Genomics, 2017).

TABLE 1 | A summary of genes that are differentially expressed in an SR45-dependent manner at both RNA and protein levels. Their identity as either SAS or SAR genes is also listed below.

AGI	Gene names	RNA FC	Protein FC	SAS/SAR
		(<i>sr45-1/Col-0</i>)	(<i>sr45-1/Col-0</i>)	
AT1G11930	At1g11930/F12F1_20	0.476	0.348	SAS
AT1G16610	SR45	0.156	0.072	SAS/SAR
AT1G20120	GDSL esterase/lipase	0.411	0.276	SAS
AT1G21650	Protein translocase subunit SECA2, chloroplastic	0.385	0.587	
AT2G21385	ATCGLD11, BFA3, BIOGENESIS FACTORS REQUIRED FOR ATP SYNTHASE 3, CGLD11	0.436	0.268	
AT2G21590	APL4	0.459	0.543	
AT2G22400	ATTRM4B, TRM4B	0.352	0.498	
AT2G27880	AGO5, ARGONAUTE 5, ATAGO5	0.344	0.437	SAR
AT3G14150	Peroxisomal (S)-2-hydroxy-acid oxidase GLO3	0.416	0.591	SAS
AT3G42850	ARA2, ARABINOKINASE 2	0.456	0.611	SAR
AT4G38350	Patched family protein	0.468	0.332	SAR
AT4G39460	S-adenosylmethionine carrier 1, chloroplastic/mitochondrial;SAMC1	0.470	0.461	
AT5G13380	Auxin-responsive GH3 family protein	0.464	0.782	
AT5G23960	Alpha-humulene/(-)-(E)-beta-caryophyllene synthase;TPS21	0.351	0.629	
AT5G45140	NRPC2, NUCLEAR RNA POLYMERASE C2	0.496	0.569	SAR
AT5G63810	ATBGAL10, BETA-GALACTOSIDASE 10, BGAL10	0.456	0.442	
AT2G43620	Chitinase family protein	3.358	3.472	
AT4G25000	ALPHA-AMYLASE-LIKE, AMY1, ATAMY1	2.353	1.852	
AT5G25450	Cytochrome bd ubiquinol oxidase, 14kDa subunit	3.581	1.385	

Here, IBM1 protein had reduced accumulation (72.5% of Col-0) in the *sr45-1* mutant, which suggests that at least some of its target genes could be hypermethylated and therefore expressed at a lower level in the *sr45-1* mutant. Indeed, genes encoding 57 proteins, one-third of 171 proteins that have decreased abundance in the *sr45-1* mutant, were found to be hypermethylated in the *ibm-1* loss-of-function mutant (**Supplemental Table S4**) (Miura et al., 2009). Five of them are SARs, SAS products and code for

SDA proteins (**Table 2**). These hypermethylated regions were mostly found in the gene body, not necessarily in the promoter regions (Miura et al., 2009). Nevertheless, the comparison between RNA and protein from the same tissues suggests that the majority of the SDA protein accumulation changes did not arise from transcriptional changes.

Bioinformatics programs PANTHER (Mi et al., 2017), KEGG (www.genome.jp/kegg/) and STRING (Szklarczyk et al., 2015)

TABLE 2 | A summary of genes of which their RNAs are identified as SR45-associated (SAR) and SR45-dependent alternatively spliced (SAS), and of which their proteins are differentially accumulated in an SR45-dependent manner. The RNA for these proteins is not differentially expressed in the *sr45-1* mutant (Zhang et al., 2017).

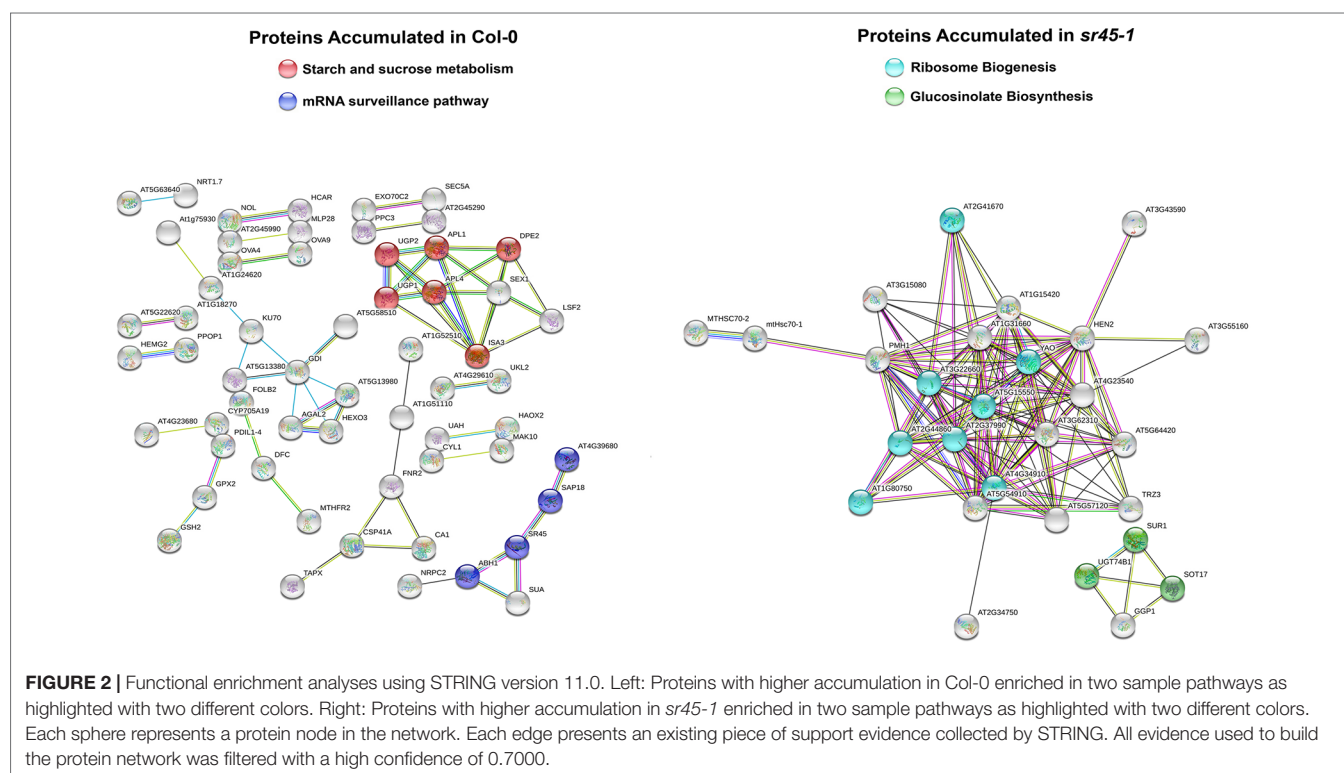
AGI	Gene names	Protein FC	Hypermethylated gene in <i>ibm1</i> mutant
		(<i>sr45-1</i> /Col-0)	
AT1G16610	Serine/arginine-rich splicing factor SR45;SR45;ortholog	0.072	
AT1G65540	LETM1-like protein;At1g65540;ortholog	0.244	Yes
AT5G13980	Alpha-mannosidase; At5g13980;ortholog	0.350	Yes
AT3G10160	Folylpolyglutamate synthase;FPGS2;ortholog	0.448	Yes
AT2G30170	Probable protein phosphatase 2C 26;At2g30170;ortholog	0.457	
AT4G30310	FGGY family of carbohydrate kinase;At4g30310;ortholog	0.505	
AT5G13690	Alpha-N-acetylglucosaminidase;CYL1;ortholog	0.515	Yes
AT4G10060	Non-lysosomal glucosylceramidase; At4g10060;ortholog	0.535	
AT4G10030	Hydrolase, alpha/beta fold family protein; T5L19.160;ortholog	0.633	
AT3G07610	Lysine-specific demethylase JMJ25;JMJ25;ortholog	0.725	Yes
AT1G01710	Acyl-CoA thioesterase family protein;At1g01710;ortholog	0.741	

were used to gain insights to the SDA protein functions. The 171 proteins with decreased accumulation in the *sr45-1* mutant were overrepresented in starch and sucrose metabolism and mRNA surveillance pathways, whereas the 56 proteins with increased accumulation appeared to be overrepresented in ribosome biogenesis and glucosinolate biosynthesis pathways (Figure 2, Supplemental Table S5).

mRNA Surveillance

The mRNA surveillance pathway includes RNA quality control through the EJC and the degradation of aberrant RNAs. It identifies PTC-containing mRNAs and prevents them from

being used for translation (Popp and Maquat, 2013). Four proteins in this pathway had significantly less accumulation in the *sr45-1* mutant compared to Col-0 (Supplemental Table S3). One was SR45, as expected. The next was AtSAP18, exhibiting the next greatest fold decrease of 0.128 (an 87% reduction, see Supplemental Figure S2A). This suggests that the mutational absence of SR45 led to a near absence of AtSAP18. The other two proteins, AtACINUS and ABA HYPERSENSITIVE 1 (ABH1), had fold reductions of 0.664 and 0.796 respectively (Supplemental Figure S2A and Supplemental Table S3). SR45, AtSAP18, and AtACINUS appear to be the counterparts of the ASAP core proteins in animals. Meanwhile, ABH1 functions as a cap-binding protein [cap-binding protein 80



(CBP80)] and stabilizes CBP20 in the nucleus when binding to the 7-methylguanosine cap at the 5' end of mature mRNAs (Kierzkowski et al., 2009). When the CBP80/20 complex binds to the 5'-end of capped target transcripts, it plays dual roles by directly influencing alternative splicing, mostly at the 5' splice site of the first intron, and pri-miRNA processing (Laubinger et al., 2008; Raczynska et al., 2010). Hence, the results suggest that the absence of SR45 led to the substantial reduction of other ASAP core proteins, likely through complex instability. This might have a substantial effect on associated processes covering different steps of RNA metabolism from transcription to alternative splicing. It is possible that these proteomic differences contributed to the transcriptome level differences between Col-0 and the *sr45-1* mutant reported previously (Zhang et al., 2017) and to the other protein differences that follow.

Ribosome Biogenesis

The biogenesis of ribosomes produces the machinery for translation. During ribosome biogenesis, both ribosome RNAs (rRNAs) and ribosome proteins need to be mature and assembled. This is an energy-consuming process and requires the coordination and regulation control among RNA polymerases I, II, III, and the splicing of introns (Nerurkar et al., 2015). A total of 8 proteins in the ribosome biogenesis pathway exhibited moderately elevated fold increase (1.214–1.450) in the *sr45-1* mutant. They are involved in rRNA processing, maturation and ribosome assembly (Supplemental Table S5). Although the change in each of them was relatively mild, the aggregated catalytic outcome in ribosome biogenesis could be more notable than each individual increase represents. Taken with RNA surveillance, SR45 seems to have a preference in modulating the abundance of protein factors functioning in more than one aspect of RNA metabolism. As of now, however, there are no clear explanations for how the loss of SR45 could cause the increase in ribosome biogenesis proteins in inflorescence.

Plant Defense

Proteins with greater accumulation in the *sr45-1* mutant included several that may be related to immunity to disease. Three proteins, UGT74B1 (At1g24100), SUR1 (At2g20610) and SOT17 (At1g18590), had fold increases ranging from 1.272 to 1.841 in the *sr45-1* mutant (Supplemental Table S5). These three proteins catalyze the last three consecutive steps of the glucosinolate biosynthetic pathway (Grubb and Abel, 2006). Glucosinolates are used in plant defense (Grubb and Abel, 2006; Bednarek et al., 2009). An aggregated increase of all three of the enzymes could result in more glucosinolate production in the *sr45-1* plants. Meanwhile, the protein with the greatest increase (a fold increase of 3.472) in the mutant was an endochitinase, while two others that also significantly increased in the mutant were catalase (increased to 1.760) and peroxidase (increased to 1.750) (Supplemental Table S3). These three enzymes often increase in accumulation during pathogen attack or during cell wall metabolism. These results agree with increased pathogen resistance, increased reactive oxygen species, and increased cell wall callose deposition in *sr45-1* mutants (Zhang et al., 2017).

The Structure of the *Arabidopsis* ASAP Complex Closely Resembles the Animal Core ASAP Complex

In animal models, the ASAP complex has three core components, RNPS1, SAP18, and ACINUS. A crystal structure of the core ASAP complex (4A8X) is available in the Protein Data Bank. It comprises the conserved RNA Recognition Motif (RRM) in human RNPS1 (HsRNPS1), a ubiquitin-like (UBL) domain in mouse SAP18 (MmSAP18), and a RNPS1-SAP18-binding (RSB) motif in *Drosophila* ACINUS (DmACIN) (Murachelli et al., 2012). To examine how closely these conserved domains in the three *A. thaliana* proteins resemble those in their animal counterparts, the corresponding amino acid segments were used for pair-wise alignment (Supplemental Figure S3). SR45 protein has two RS domains flanking an RRM, which is distinct from all other *A. thaliana* SR proteins (Barta et al., 2010); rather, it resembles the domain structure of RNPS1 (Zhang and Mount, 2009). The SR45 RRM sequence had 36.4% amino acid identity and 73.9% similarity to the HsRNPS1 RRM (Supplemental Figure S3A). The AtSAP18 fragment sequence for the UBL domain had 53.6% amino acid identity and 80.0% of similarity to the MmSAP18 UBL domain (Supplemental Figure S3B). The AtACINUS sequence RSB motif had 64.0% amino acid identity and 84.0% similarity to the DmACINUS RSB sequence (Supplemental Figure S3C). The predicted domain structure for the *A. thaliana* ASAP complex core proteins aligned with their animal counterparts in 4A8X. Specifically, SR45 and AtSAP18 aligned closely with HsRNPS1 and MmSAP18, respectively (Supplemental Figure S2B). The alignment of the RNPS1 RRM structure yielded an all-atom root-mean-square deviation of atomic positions (RMSD) of 2.350 Å (Supplemental Figure S2C). The predicted SR45 RRM protein model lacked 4 small beta sheets compared to HsRNPS1 RRM, which seemed to have little effect on the overall structure of the RRM itself. The alignment of the SAP18 UBL domain structure yielded an all-atom RMSD of 4.300 Å (Supplemental Figure S2C). In comparison to MmSAP18, AtSAP18 had a different small alpha helix. This caused minimal disturbance in the overall UBL domain structure. The alignment of ACINUS yielded an all-atom RMSD of 6.888 Å (Supplemental Figure S2C). However, the predicted AtACINUS RSB structure did not align as well with DmACINUS RSB. In this small region, AtACINUS was missing 2 small beta sheets compared to DmACINUS, which seemed to influence the orientation of the overall RSB structure. Nevertheless, both sequence and structure alignments supported the hypothesis that the *A. thaliana* ASAP complex is orthologous to the ASAP complex found in animals. In the rest of the text, we will refer to AtSAP18 and AtACINUS as SAP18 and ACINUS.

SR45 Maintains the Wild-Type Level of the ASAP Complex Core Component Proteins

To evaluate the function of SAP18 with respect to SR45, we produced transgenic plants overexpressing an in-frame fusion of gSAP18-GFP in Col-0 and in the *sr45-1* mutant (Figure 3). In the inflorescence tissue, bright and distinct nucleoplasmic GFP signal was detected in carpels and ovules in the Col-0, while only very dim nucleoplasmic GFP signal was detected

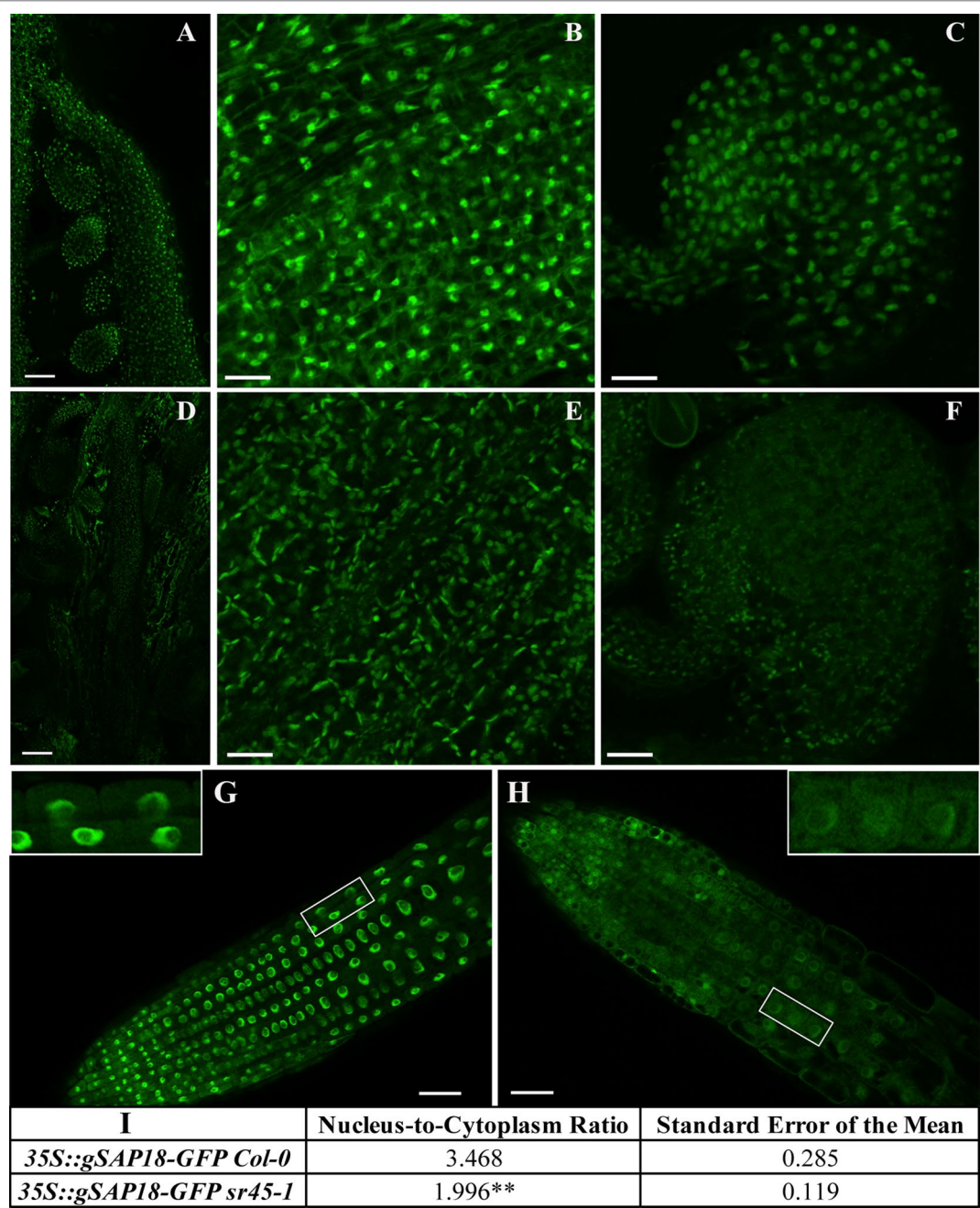


FIGURE 3 | A comparison of SAP18-GFP expression between Col-0 and *sr45-1*. **A–C** and **G** represent Col-0 transgenic; **D–F** and **H** represent *sr45-1* transgenic. **(A)** and **(D)**: carpel containing ovule inside with scale bars = 50 μ m. **(B)** and **(E)**: a close-up view of carpel cells with scale bars = 25 μ m. **(C)** and **(F)**: a close-up view of ovules with scale bars = 25 μ m. **(G)** and **(H)**: root tip with scale bars = 50 μ m. The inserts showed a close-up view of root cells in the boxed area. **(I)**: qualifications of root GFP signal intensity in nucleus vs. cytoplasm. A total of 15 cells were measured per seedling. Three seedling were used for student t-test **; $p < 0.01$.

in *sr45-1* in a few carpel cells overshadowed by chloroplasts. If it was not for the chloroplasts, the *sr45-1* ovule would have been barely noticeable (**Figures 3A–F**). To visualize the subcellular distribution of SAP18-GFP more clearly without the strong background from chloroplasts, root tips of seedlings were compared between the transgenic Col-0 and *sr45-1* lines.

In the transgenic Col-0 root tip, the SAP18-GFP signal was prominent and strong in the nucleoplasm, and it was much lower and diffused in the cytoplasm of every cell in the meristematic zone. In the transgenic *sr45-1* mutant, however, the overall SAP18-GFP signal was much weaker and more diffused. More specifically, there was much less nuclear SAP18-GFP in the

transgenic *sr45-1* mutant root compared to the transgenic Col-0, even though the intensities of their cytosolic SAP18-GFP signals were similar (Figures 3G, H). A statistically significant 1.7-fold reduction in nucleus-to-cytoplasm ratios of GFP signal intensity (3.468 in Col-0 background vs. 1.996 in *sr45-1* background) was observed when comparing 35S::gSAP18-GFP transgenic seedlings in the Col-0 background to transgenic seedlings in the *sr45-1* background (Figure 3I). RT-qPCR results revealed that there was no significant difference in the transcript for both the *SAP18-GFP* transgene and the endogenous *SAP18* between the transgenic *sr45-1* and transgenic Col-0. (Supplemental Figure S4). Thus, these data support the notion that SR45 protein is required to maintain nuclear SAP18 protein at the wild-type level, a hypothesis corroborated by the proteomics data. RT-qPCR also revealed that the expression of *ACINUS* was unchanged in

the *sr45-1* mutant (Figure 4A), indicating SR45's possible role in a protein control for *ACINUS* as well.

SAP18 Participates in the Suppression of a Subset of SR45 Differentially Regulated RNAs

In our prior study, we identified 358 SDR RNAs differentially expressed in *sr45-1* inflorescence (Zhang et al., 2017). To understand whether SAP18 is involved in the regulation of SDR RNAs and RNAs for ASAP core components, RT-qPCR was performed in Col-0, *sr45-1*, and the *SAP18-GFP* overexpression lines. Although there was no statistically significant difference in the RNA level of *SAP18* between Col-0 and the *sr45-1* mutant, there was a substantial 10–22 fold increase in the overall *SAP18* RNA, including the *SAP18-GFP* transgene and endogenous *SAP18*, due to *SAP18-GFP* transgenic

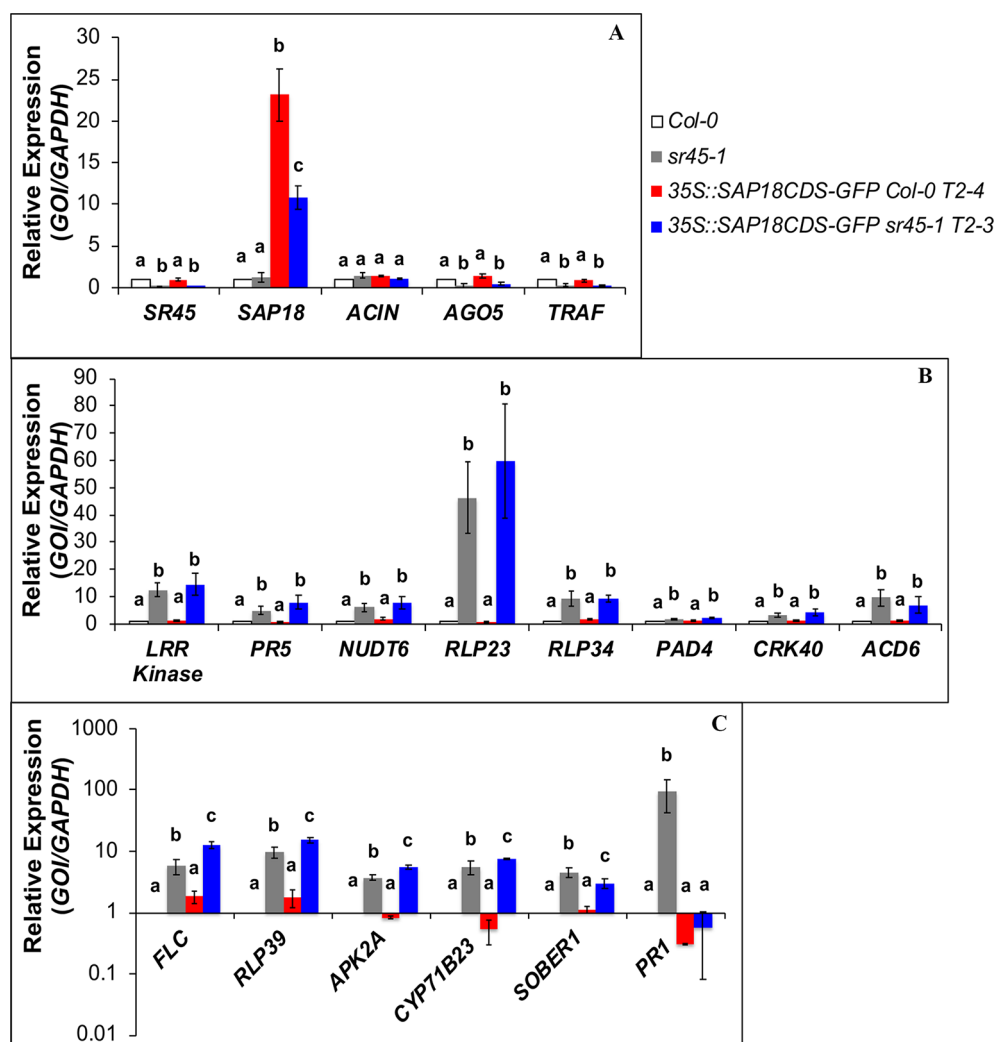


FIGURE 4 | RT-qPCR on SDR genes in Col-0, *sr45-1*, 35S::SAP18CDS-GFP Col-0 and 35S::SAP18CDS-GFP *sr45-1*. (A) SR45-upregulated genes; (B) SR45-downregulated genes that are not affected by SAP18-GFP overexpression; (C) SR45-downregulated genes that are affected by SAP18-GFP overexpression. One-way ANOVA followed by Tukey HSD was used for statistical analysis. Letters a–c represent different levels of statistical significance with $p < 0.05$. $n = 3$. Error bars represent standard deviation.

overexpression (Figure 4A). This dramatic change, however, did not seem to affect the expression of *SR45* and *ACINUS* (Figure 4A). Neither did it change the expression pattern for two previously confirmed *SR45*-upregulated SDR RNAs, *AGO5*, and *TRAF* (Figure 4A), nor eight *SR45*-downregulated SDR genes known to play roles in plant immunity (Figure 4B). However, the expressions of defense genes *SOBER1* and *PR1* decreased or returned to a level observed in non-transgenic Col-0 (Figure 4C), while the expressions for three other defense-response genes, *RLP39*, *APK2A*, and *CYP71B23*, increased (Figure 4C). These findings suggest that *SAP18* is involved in the control of the expression of a subset of SDR defense genes.

Overexpression of *SAP18* Further Delays Flowering in the *sr45-1* Mutant

The *sr45-1* mutant is delayed in flowering, and it has increased expression of *FLC*, a flowering suppressor (Ali et al., 2007). Interestingly, overexpression of *SAP18-GFP* in the *sr45-1* mutant resulted in a modest but statistically significant increase in the RNA levels of *FLC* compared to the non-transgenic *sr45-1* mutant (Figure 4C). To determine the physiological effect of this, we evaluated the flowering time of transgenic plants with the *SAP18CDS-GFP* or the *gSAP18-GFP* gene fusion versions. Transgenic Col-0 plants overexpressing either *SAP18-GFP* transgene versions had no detectable difference in flowering time compared to non-transgenic Col-0. Surprisingly, transgenic *sr45-1* plants overexpressing either of the two *SAP18-GFP* transgenes were further delayed in flowering by 4 days compared to the non-transgenic *sr45-1* parent which already has a late flowering phenotype (Figure 5). This phenotype

was consistent in independent transgenic lines. These observations suggest that *SR45* may be required to control *SAP18* in flowering time suppression. A recent study has shown that all ASAP core complex proteins were associated with the vernalization complex via *VAL1* to induce transcriptional silencing at the *FLC* locus (Questa et al., 2016). Therefore, it is possible that *SR45* and *SAP18* work together to achieve an optimal suppression of *FLC* and the initiation of flowering.

DISCUSSION

RNA metabolism is a complex and concerted process that requires both protein factors and RNA substrates. Our previous study described *SR45*-dependent RNA level changes in the inflorescence tissue (Zhang et al., 2017). In this study, using the same batch of inflorescence tissue as used in the transcriptome study, we found that proteins differentially accumulated in the *sr45-1* mutant likely participate in different steps of RNA metabolism, especially mRNA surveillance (Figure 2 and Supplemental Table S3). This provides a protein context in RNA metabolism for the transcriptome difference we described before (Zhang et al., 2017). The fact that there is little overlap between the differentially expressed RNAs and differentially accumulated proteins strongly suggests that most SDA proteins are not direct products of the differentially expressed RNAs. Alternative splicing does not sufficiently explain this discordance either. Rather, the proteins coexisted with the differentially expressed RNAs in the same inflorescence tissue and are likely a product of transcriptional and post-translational regulation that arises from the loss of the *SR45* protein.

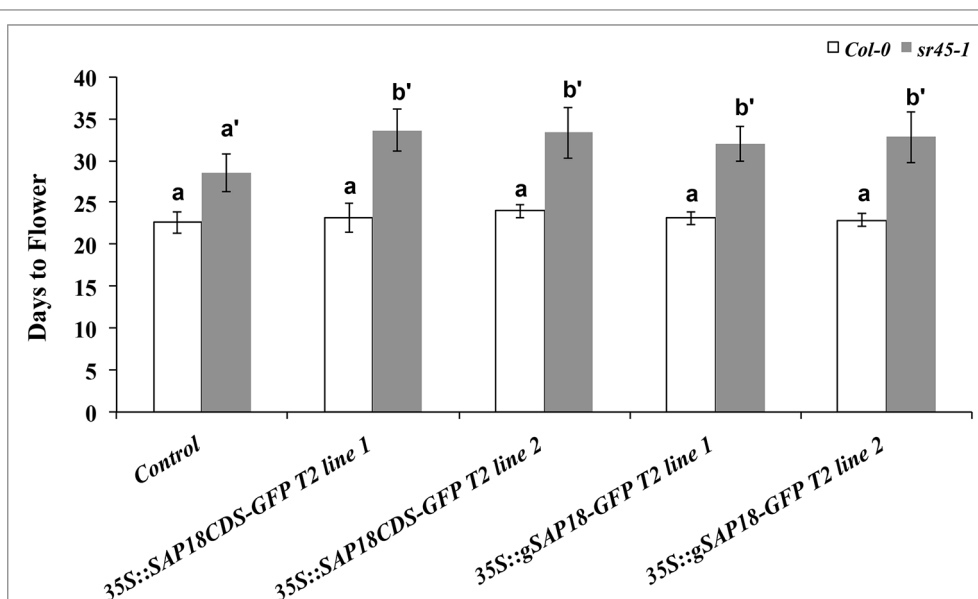


FIGURE 5 | A comparison of flowering time among different genotypes grown under long-day condition (L:D = 16:8). The number of days before bolting is used to indicate days to flower. Transgenic lines were compared with their corresponding controls (Col-0 or *sr45-1*), respectively. Kruskal-Wallis test followed by post-hoc Dunn test was used for statistical analysis. Benjamini-Hochberg FDR method was used to calculate adjusted *p*-values. Letter a represents statistical level of Col-0 and its transgenic lines; letters a' and b' represent different levels of statistical significance for *sr45-1* and its transgenic lines. *n* = 10. *FDR* < 0.05. Error bars represent standard deviation.

Despite the evidence that SR45 is a splicing regulator, there is also evidence that SR45, but not other splicing factors, may regulate RdDM-mediated DNA methylation *via* unknown mechanisms (Ausin et al., 2012). Our data indicates that hypermethylation in gene body of IBM targets may be one such mechanism (**Supplemental Table S4**). In our study, all core proteins of a conserved ASAP complex, SR45, SAP18, and ACINUS, exhibited a significantly lower level of accumulation in the *sr45-1* mutant. Their orthology to the ASAP complex proteins in animal models, shown here through amino acid sequence homology and predicted protein domain structures (**Supplemental Figures S2 and S3**), supports the hypothesis that SR45 not only functions as a splicing regulator to affect pre-mRNA splicing, but can influence transcriptional control *via* associations with other proteins that are not splicing factors (Song and Galbraith, 2006; Hill et al., 2008; Zhang and Mount, 2009; Day et al., 2012; Xing et al., 2015; Questa et al., 2016; Deka and Singh, 2017; Zhang et al., 2017). Such an example is its interaction with SAP18 in the ASAP complex.

Understanding the effect of SR45 on SAP18 is key to revealing the function of the ASAP complex on affected processes. Although there is limited knowledge on the role of SR45 in the *A. thaliana* ASAP complex, SR45 seems to maintain the wild-type level of SAP18 protein in the nucleus. Without SR45, the level of nuclear SAP18 protein is drastically decreased (**Figure 3**). This could have a broad impact on processes not directly related to splicing. Since SAP18 recruits HDACs to chromatin for gene silencing, less nuclear SAP18 protein could reduce the abundance of HDACs on the affected loci leading to a leaky gene expression. Specifically, a transcriptional repressor VAL1 can interact with the ASAP complex and promote HDA19 docking at the *FLC* locus to silence *FLC* (Questa et al., 2016). The RNA level of *FLC* was slightly higher in an HDAC mutant *hda19-1*, an *sr45* T-DNA insertion mutant and a *sap18* T-DNA insertion mutant. We overexpressed *SAP18-GFP* in the *sr45-1* mutant and observed a correlation between the increased expression level of *FLC* and increased time to flower compared to the non-transgenic *sr45-1* mutant (**Figures 4C and 5**). It is unexpected to see that the overexpression of *SAP18-GFP* protein did not lead to a stronger suppression of *FLC* expression. On the contrary, it exacerbated the effect of the *sr45-1* mutation on *FLC* and caused a further delay in flowering. Together, these results imply that SR45 is required for a wild-type level of SAP18 accumulation in the nucleus and that FLC-regulated flowering time is controlled by the nuclear presence of two components of the ASAP complex.

The requirement of SR45 for proper SAP18 functions can also be seen in the expression of a subset of RNAs involved in defense that were previously found to have increased expression in *sr45-1* mutants (**Figure 4C**). This subset consists of five plant defense genes, among which is *PRI*, a marker for activation of the salicylic acid signaling pathway (Gaffney et al., 1993; van Loon et al., 2006). The expression of *PRI* RNA is returned to even lower level than in Col-0 by overexpressing *SAP18-GFP* in the *sr45-1* mutant, which indicates a possible reversal of the elevated SA pathway. Although enzymes for glucosinolate biosynthesis had higher accumulation in the *sr45-1* mutant to support the notion of a stronger immunity

in the *sr45-1* mutant, it is unclear whether there is an actual change in glucosinolate production in the *sr45-1* mutant, and whether overexpressing *SAP18-GFP* would have any impact on the level of these enzymes and defense response. Future experiments in these areas will be necessary to answer the questions mentioned above.

In conclusion, this study provides evidence for the pleiotropic effects of SR45 on protein expression in *A. thaliana* inflorescence with an emphasis on different aspects of RNA metabolism. In addition to being a main component of the ASAP complex, SR45 also regulates the other two ASAP component proteins to achieve optimal transcriptional suppression of substrate target genes. Our findings also provided evidence for a new explanation beyond alternative splicing for the regulation of SR45-suppressed RNAs. Further studies on histone modification at SAP18-targeted loci will help elucidate whether and how a conserved ASAP complex functions to regulate these genes and proteins during flowering.

DATA AVAILABILITY

The RNA-seq datasets used for RNA-protein comparison for this study can be found as a BioProject (PRJNA382852) in the NCBI Sequence Read Archive (SRA) [<https://www.ncbi.nlm.nih.gov/bioproject/PRJNA382852/>]. The mass spectrometry data files can be retrieved from massive.ucsd.edu (MSV000083728). All the peptide-spectrum match for the mass spectrometry data is available in **Supplemental Table S6**.

AUTHOR CONTRIBUTIONS

X-NZ conceived, designed, conducted, and supervised the study. HB, WG, and BC performed the quantitative proteomics experiment and subsequent statistical analyses. SC, TR, and X-NZ performed bioinformatics analysis and functional enrichment analysis. SC, TR, AH, LM, and JP conducted the study. X-NZ, SC, TR, AH, and BC wrote the paper.

FUNDING

This work was supported by National Science Foundation (DBI-1146300 to XZ), USDA-ARS to BC, and research funds in Department of Biology at St. Bonaventure University.

ACKNOWLEDGMENTS

We thank the Spring 2018 Molecular Biology class at St. Bonaventure University for their contribution to preliminary studies.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fpls.2019.01116/full#supplementary-material>

REFERENCES

- 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 (5), e471. doi: 10.1371/journal.pone.0000471
- 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 (1), 29–33. doi: 10.4161/epi.7.1.18782
- Baldwin, K. L., Dinh, E. M., Hart, B. M., and Masson, P. H. (2013). CACTIN is an essential nuclear protein in *Arabidopsis* and may be associated with the eukaryotic spliceosome. *FEBS Letters* 587 (7), 873–879. doi: 10.1016/j.febslet.2013.02.041
- 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. *Plant Cell* 22 (9), 2926–2929. doi: 10.1105/tpc.110.078352
- Bednarek, P., Pislewska-Bednarek, M., Svatos, A., Schneider, B., Doubek, J., Mansurova, M., et al. (2009). A glucosinolate metabolism pathway in living plant cells mediates broad-spectrum antifungal defense. *Science* 323 (5910), 101–106. doi: 10.1126/science.1163732
- Bent, A. F. (2000). *Arabidopsis* in planta transformation. Uses, mechanisms, and prospects for transformation of other species. *Plant Physiol.* 124 (4), 1540–1547. doi: 10.1104/pp.124.4.1540
- Day, I. S., Golovkin, M., Palusa, S. G., Link, A., Ali, G. S., Thomas, J., et al. (2012). Interactions of SR45, an SR-like protein, with spliceosomal proteins and an intronic sequence: insights into regulated splicing. *Plant J.* 71 (6), 936–947. doi: 10.1111/j.1365-313X.2012.05042.x
- Deka, B., and Singh, K. K. (2017). Multifaceted regulation of gene expression by the apoptosis- and splicing-associated protein complex and its components. *Int. J. Biol. Sci.* 13 (5), 545–560. doi: 10.7150/ijbs.18649
- Fan, D., Wang, X., Tang, X., Ye, X., Ren, S., Wang, D., et al. (2018). Histone H3K9 demethylase JM25 epigenetically modulates anthocyanin biosynthesis in poplar. *Plant J.* 96 (6), 1121–1136. doi: 10.1111/tpj.14092
- Fu, X. D., and Ares, M. Jr. (2014). Context-dependent control of alternative splicing by RNA-binding proteins. *Nat. Rev. Genet.* 15 (10), 689–701. doi: 10.1038/nrg3778
- Gaffney, T., Friedrich, L., Vernooij, B., Negrotto, D., Nye, G., Uknes, S., et al. (1993). Requirement of salicylic acid for the induction of systemic acquired resistance. *Science* 261 (5122), 754–756. doi: 10.1126/science.261.5122.754
- 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 (51), 36428–36438. doi: 10.1074/jbc.274.51.36428
- Grubb, C. D., and Abel, S. (2006). Glucosinolate metabolism and its control. *Trends Plant Sci.* 11 (2), 89–100. doi: 10.1016/j.tplants.2005.12.006
- Hill, K., Wang, H., and Perry, S. E. (2008). A transcriptional repression motif in the MADS factor AGL15 is involved in recruitment of histone deacetylase complex components. *Plant J.* 53 (1), 172–185. doi: 10.1111/j.1365-313X.2007.03336.x
- Isasa, M., Rose, C. M., Elsasser, S., Navarrete-Perea, J., Paulo, J. A., Finley, D. J., et al. (2015). Multiplexed, proteome-wide protein expression profiling: yeast deubiquitylating enzyme knockout strains. *J. Proteome Res.* 14 (12), 5306–5317. doi: 10.1021/acs.jproteome.5b00802
- Kall, L., Canterbury, J. D., Weston, J., Noble, W. S., and MacCoss, M. J. (2007). Semi-supervised learning for peptide identification from shotgun proteomics datasets. *Nat. Methods* 4 (11), 923–925. doi: 10.1038/nmeth1113
- Kierzkowski, D., Kmiecik, M., Piontek, P., Wojtaszek, P., Szwejkowska-Kulinska, Z., and Jarmolowski, A. (2009). The *Arabidopsis* CBP20 targets the cap-binding complex to the nucleus, and is stabilized by CBP80. *Plant J.* 59 (5), 814–825. doi: 10.1111/j.1365-313X.2009.03915.x
- Laubinger, S., Sachsenberg, T., Zeller, G., Busch, W., Lohmann, J. U., Ratsch, G., et al. (2008). Dual roles of the nuclear cap-binding complex and SERRATE in pre-mRNA splicing and microRNA processing in *Arabidopsis thaliana*. *Proc. Natl. Acad. Sci. U S A* 105 (25), 8795–8800. doi: 10.1073/pnas.0802493105
- Lykke-Andersen, J., Shu, M. D., and Steitz, J. A. (2001). Communication of the position of exon-exon junctions to the mRNA surveillance machinery by the protein RNPS1. *Science* 293 (5536), 1836–1839. doi: 10.1126/science.1062786
- McAlister, G. C., Nusinow, D. P., Jedrychowski, M. P., Wuhr, M., Huttlin, E. L., Erickson, B. K., et al. (2014). MultiNotch MS3 enables accurate, sensitive, and multiplexed detection of differential expression across cancer cell line proteomes. *Anal. Chem.* 86 (14), 7150–7158. doi: 10.1021/ac502040v
- Mi, H., Huang, X., Muruganujan, A., Tang, H., Mills, C., Kang, D., et al. (2017). PANTHER version 11: expanded annotation data from Gene Ontology and Reactome pathways, and data analysis tool enhancements. *Nucleic Acids Res.* 45 (D1), D183–D189. doi: 10.1093/nar/gkw1138
- Miura, A., Nakamura, M., Inagaki, S., Kobayashi, A., Saze, H., and Kakutani, T. (2009). An *Arabidopsis* jmjC domain protein protects transcribed genes from DNA methylation at CHG sites. *EMBO J.* 28 (8), 1078–1086. doi: 10.1038/emboj.2009.59
- Murachelli, A. G., Ebert, J., Basquin, C., Le Hir, H., and Conti, E. (2012). The structure of the ASAP core complex reveals the existence of a Pinin-containing PSAP complex. *Nat. Struct. Mol. Biol.* 19 (4), 378–386. doi: 10.1038/nsmb.2242
- Nerurkar, P., Altwater, M., Gerhardt, S., Schutz, S., Fischer, U., Weirich, C., et al. (2015). Eukaryotic Ribosome Assembly and Nuclear Export. *Int. Rev. Cell Mol. Biol.* 319, 107–140. doi: 10.1016/bs.ircmb.2015.07.002
- Perkins, D. N., Pappin, D. J., Creasy, D. M., and Cottrell, J. S. (1999). Probability-based protein identification by searching sequence databases using mass spectrometry data. *Electrophoresis* 20 (18), 3551–3567. doi: 10.1002/(SICI)1522-2683(19991201)20:18<3551::AID-ELPS3551>3.0.CO;2-2
- Popp, M. W., and Maquat, L. E. (2013). Organizing principles of mammalian nonsense-mediated mRNA decay. *Annu. Rev. Genet.* 47, 139–165. doi: 10.1146/annurev-genet-111212-133424
- Questa, J. I., Song, J., Geraldo, N., An, H., and Dean, C. (2016). *Arabidopsis* transcriptional repressor VAL1 triggers Polycomb silencing at FLC during vernalization. *Science* 353 (6298), 485–488. doi: 10.1126/science.aaf7354
- Raczynska, K. D., Simpson, C. G., Ciesiolka, A., Szewc, L., Lewandowska, D., McNicol, J., et al. (2010). Involvement of the nuclear cap-binding protein complex in alternative splicing in *Arabidopsis thaliana*. *Nucleic Acids Res.* 38 (1), 265–278. doi: 10.1093/nar/gkp869
- Rauniyar, N., and Yates, J. R. 3rd (2014). Isobaric labeling-based relative quantification in shotgun proteomics. *J. Proteome Res.* 13 (12), 5293–5309. doi: 10.1021/pr500880b
- Schwerk, C., Prasad, J., Degenhardt, K., Erdjument-Bromage, H., White, E., Tempst, P., et al. (2003). ASAP, a novel protein complex involved in RNA processing and apoptosis. *Mol. Cell Biol.* 23 (8), 2981–2990. doi: 10.1128/MCB.23.8.2981-2990.2003
- Song, C. P., and Galbraith, D. W. (2006). AtSAP18, an orthologue of human SAP18, is involved in the regulation of salt stress and mediates transcriptional repression in *Arabidopsis*. *Plant Mol. Biol.* 60 (2), 241–257. doi: 10.1007/s11103-005-3880-9
- Stankovic, N., Schloesser, M., Joris, M., Sauvage, E., Hanikenne, M., and Motte, P. (2016). dynamic distribution and interaction of the *Arabidopsis* SRSF1 subfamily splicing factors. *Plant Physiol.* 170 (2), 1000–1013. doi: 10.1104/pp.15.01338
- Szklarczyk, D., Franceschini, A., Wyder, S., Forslund, K., Heller, D., Huerta-Cepas, J., et al. (2015). STRING v10: protein-protein interaction networks, integrated over the tree of life. *Nucleic Acids Res.* 43 (Database issue), D447–D452. doi: 10.1093/nar/gku1003
- van Loon, L. C., Rep, M., and Pieterse, C. M. (2006). Significance of inducible defense-related proteins in infected plants. *Annu. Rev. Phytopathol.* 44, 135–162. doi: 10.1146/annurev.phyto.44.070505.143425
- Wang, Y., Yang, F., Gritsenko, M. A., Wang, Y., Clauss, T., Liu, T., et al. (2011). Reversed-phase chromatography with multiple fraction concatenation strategy for proteome profiling of human MCF10A cells. *Proteomics* 11 (10), 2019–2026. doi: 10.1002/pmic.201000722
- Warns, J. A., Davie, J. R., and Dhasarathy, A. (2016). Connecting the dots: chromatin and alternative splicing in EMT. *Biochem. Cell. Biol.* 94 (1), 12–25. doi: 10.1139/bcb-2015-0053
- 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 (12), 3294–3308. doi: 10.1105/tpc.15.00641
- Yang, J., Yan, R., Roy, A., Xu, D., Poisson, J., and Zhang, Y. (2015). The I-TASSER Suite: protein structure and function prediction. *Nat. Methods* 12 (1), 7–8. doi: 10.1038/nmeth.3213
- 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. doi: 10.4161/psb.29134

- 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 (3), 1450–1458. doi: 10.1104/pp.109.138180
- Zhang, X. N., Shi, Y., Powers, J. J., Gowda, N. B., Zhang, C., Ibrahim, H. M. M., et al. (2017). Transcriptome analyses reveal SR45 to be a neutral splicing regulator and a suppressor of innate immunity in Arabidopsis thaliana. *BMC Genomics* 18 (1), 772. doi: 10.1186/s12864-017-4183-7
- Zhang, Y., Iratni, R., Erdjument-Bromage, H., Tempst, P., and Reinberg, D. (1997). Histone deacetylases and SAP18, a novel polypeptide, are components of a human Sin3 complex. *Cell* 89 (3), 357–364 doi: 10.1016/S0092-8674(00)80216-0

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2019 Chen, Rooney, Hu, Beard, Garrett, Mangalath, Powers, Cooper and Zhang. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



To Splice or to Transcribe: SKIP-Mediated Environmental Fitness and Development in Plants

Ying Cao and Ligeng Ma*

College of Life Sciences, RNA Center, Capital Normal University, Beijing, China

OPEN ACCESS

Edited by:

Craig G. Simpson,
The James Hutton Institute,
United Kingdom

Reviewed by:

Misato Ohtani,
Nara Institute of Science and
Technology (NAIST), Japan
Yingxiang Wang,
Fudan University, China

*Correspondence:

Ligeng Ma
ligeng.ma@cnu.edu.cn

Specialty section:

This article was submitted to
Plant Physiology,
a section of the journal
Frontiers in Plant Science

Received: 17 June 2019

Accepted: 04 September 2019

Published: 03 October 2019

Citation:

Cao Y and Ma L (2019) To Splice
or to Transcribe: SKIP-Mediated
Environmental Fitness and
Development in Plants.
Front. Plant Sci. 10:1222.
doi: 10.3389/fpls.2019.01222

Gene expression in eukaryotes is controlled at multiple levels, including transcriptional and post-transcriptional levels. The transcriptional regulation of gene expression is complex and includes the regulation of the initiation and elongation phases of transcription. Meanwhile, the post-transcriptional regulation of gene expression includes precursor messenger RNA (pre-mRNA) splicing, 5' capping, and 3' polyadenylation. Among these events, pre-mRNA splicing, conducted by the spliceosome, plays a key role in the regulation of gene expression, and the efficiency and precision of pre-mRNA splicing are critical for gene function. Ski-interacting protein (SKIP) is an evolutionarily conserved protein from yeast to humans. In plants, SKIP is a bifunctional regulator that works as a splicing factor as part of the spliceosome and as a transcriptional regulator *via* interactions with the transcriptional regulatory complex. Here, we review how the functions of SKIP as a splicing factor and a transcriptional regulator affect environmental fitness and development in plants.

Keywords: SKIP, alternative splicing, transcriptional regulator, splicing factor, environmental fitness, plant development

INTRODUCTION AND GENE EXPRESSION REGULATION

Due to their sessile nature, plants must respond to both the external environment and internal signals to regulate their environmental fitness and development. To respond to these signals in a precise manner, gene expression must be tightly controlled both temporally and spatially. Gene expression is regulated at multiple levels, but most regulation occurs at the transcriptional and post-transcriptional levels (reviewed in Licatalosi and Darnell, 2010). This allows a gene to be expressed at the correct time, in specific cells, and with the appropriate abundance to support its function.

Transcriptional regulation is crucial for controlling the temporal and spatial expression of a gene, as well as the abundance of precursor messenger RNA (pre-mRNA) molecules. In eukaryotes, messenger RNAs (mRNAs) are transcribed by RNA polymerase II (Pol II) in a complicated process that includes initiation, elongation, and termination steps. The regulation of gene expression at the transcriptional level occurs mainly at the initiation and elongation stages (reviewed in Kwak and Lis, 2013; Jonkers and Lis, 2015; Sainsbury et al., 2015). In the initiation stage, Pol II with an unphosphorylated C-terminal domain (CTD) forms a pre-initiation complex by associating with transcription factors and mediators (reviewed in Hsin and Manley, 2012; Sainsbury et al., 2015; Hantsche and Cramer, 2017). Transcription initiation also requires interactions with *cis*-elements in the genomic DNA sequence and changes in chromatin structure and nucleosome position *via* epigenetic modifications (reviewed in Kouzarides, 2007; Matzke et al., 2009; Bonasio and Shiekhattar, 2014; Voss and Hager, 2014; Sainsbury et al., 2015; Lawrence et al., 2016; Jeronimo and Robert, 2017). The elongation phase is regulated by multiple elongation factors, including Pol

II-associated factor 1 complex (Paf1c), and additional factors that influence the epigenetic modification and higher-order structure of chromatin, the phosphorylation status of the Pol II CTD, and the eventual pause and release of Pol II (reviewed in Li et al., 2007; Gilchrist et al., 2010; Hajheidari et al., 2013; Tessarz and Kouzarides, 2014; Van Lijsebettens and Grasser, 2014; Lawrence et al., 2016; Van Oss et al., 2017).

The products transcribed by Pol II from a DNA template require processing to form stable, mature mRNAs. The regulation of pre-mRNA processing at the post-transcriptional level affects the abundance of functional mature mRNAs; thus, it affects both gene expression and function. Post-transcriptional pre-mRNA processing involves 5' capping mediated by 5' capping enzymes at the 5' end of the pre-mRNA, splicing by the spliceosome to remove introns from the pre-mRNA, and 3' polyadenylation mediated by the 3' polyadenylation complex at the 3' end of the pre-mRNA. In addition to these events, pre-mRNA alternative splicing plays a key role in the post-transcriptional regulation of gene expression. By using different splice sites, one pre-mRNA can be processed to multiple transcripts, thus increasing the complexity of the transcriptome and proteome (reviewed in Reddy, 2007; Keren et al., 2010; Syed et al., 2012; Lee and Rio, 2015). Consequently, incorrect splicing of a pre-mRNA can decrease the amount of functional mature mRNA or lead to the production of toxic proteins that may perturb normal cellular processes (reviewed in Reddy, 2007; Braunschweig et al., 2013). The incorrectly spliced variants with a premature termination codon may activate mRNA degradation through the nonsense-mediated decay (NMD) pathway to prevent the formation of nonfunctional or aberrant proteins (reviewed in He and Jacobson, 2015). Therefore, efficient and precise pre-mRNA splicing is crucial to protect gene function (reviewed in Reddy, 2007; Moore and Proudfoot, 2009; Syed et al., 2012; Staiger and Brown, 2013). Accurate splicing of an intron depends on both short consensus sequence elements around the intron and correct assembly of the components of the spliceosome around the intron's splice sites (reviewed in Wahl et al., 2009; Lee and Rio, 2015).

The spliceosome, which is responsible for pre-mRNA splicing, is a large and highly dynamic protein complex. Specific splicing factors are sequentially recruited to and released from splice sites to mediate efficient splicing. Ski-interacting protein (SKIP), a component of the spliceosome-associated NineTeen complex (NTC), is required to catalyze the first and second transesterification reactions of pre-mRNA splicing in yeast and human cells (Albers et al., 2003; Figueroa and Hayman, 2004; Bessonov et al., 2008; Chen et al., 2011; Schneider et al., 2015; Zhang et al., 2017). In addition, SKIP is a transcriptional coregulator for the expression of some genes in human cells (Laduron et al., 2004; Brès et al., 2005; Brès et al., 2009; Chen et al., 2011). SKIP protein is conserved from yeast to humans including plants. It acts both as a splicing factor to regulate precise and efficient pre-mRNA splicing and as a transcriptional regulator of gene transcription in *Arabidopsis*. This review focuses on the regulatory functions of SKIP that control gene expression at the transcriptional and post-transcriptional levels to mediate the environmental fitness and development of plants.

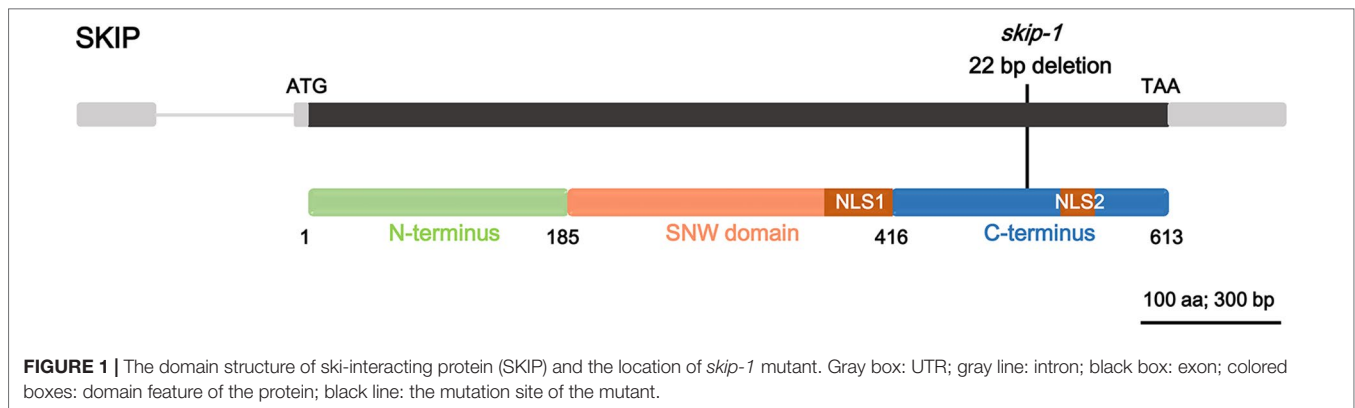
SKIP MEDIATES PLANT ENVIRONMENTAL FITNESS BY REGULATING ALTERNATIVE SPLICING

Flowering is an important developmental phase transition in higher plants. To regulate flowering time, plants integrate endogenous and environmental signals, which are important for survival and crop productivity. To find new components of the flowering time control pathway in *Arabidopsis*, a genetic screen was performed using a T-DNA insertion library for altered flowering time mutants, and a mutant, *eip1-1*, that exhibits an early flowering phenotype under long- and short-day conditions was isolated (Wang et al., 2012). Such a photoperiod-insensitive flowering time defect is characteristic of circadian clock-defective mutants. Consistent with this, *eip1-1* exhibits a lengthened circadian period in a temperature-sensitive manner. Compared to the ~24-h circadian period of wild-type plants, the circadian period of *eip1-1* is lengthened by ~2.4 h due to changes in the rhythmic expression of the core oscillator genes *CIRCADIAN CLOCK-ASSOCIATED 1*, *LATE ELONGATED HYPOCOTYL*, and *TIMING OF CAB EXPRESSION 1* (Wang et al., 2012). Map-based cloning revealed that a mutation in At1g77180, which encodes SKIP, is responsible for the flowering time and circadian period defects observed in *eip1-1* (which was therefore renamed *skip-1*). There is a 22-nucleotide deletion at the C-terminus of the *SKIP* locus, which disrupts the integrity of the SKIP protein and impairs SKIP function in *skip-1* plants (Wang et al., 2012).

SKIP is a single-copy gene in the *Arabidopsis* genome encoding a protein of 613 amino acids with three structural domains: the N-terminus (amino acids 1–185), SNW domain (amino acids 186–416), and C-terminus (amino acids 417–613) (Figure 1; Li et al., 2016). The plant SKIP protein sequence is highly similar to that of its ortholog (SKIP) in humans and pre-mRNA processing (Prp)45 in yeast (Wang et al., 2012). SKIP localizes to the nucleus using two nuclear localization signals (NLSs), which lie in the SNW domain and C-terminus, respectively (Figure 1; Lim et al., 2010; Li et al., 2016).

In *Arabidopsis*, SKIP co-localizes with the spliceosome components U1 SMALL NUCLEAR RIBONUCLEOPROTEIN-70K (U1-70K) (Golovkin and Reddy, 1996) and SERINE/ARGININE RICH 45 (SR45) (Day et al., 2012) in nuclear bodies (Wang et al., 2012). SKIP associates closely with SR45 and other NTC components, facilitating its integration into the spliceosome (Wang et al., 2012; Li et al., 2016). Mutations in SKIP decrease the splicing efficiency of the spliceosome and can cause genome-wide alternative splicing defects (Wang et al., 2012; Feng et al., 2015). SKIP is required for 5' and 3' splice site recognition or cleavage; novel splicing events have been reported in *skip* mutant plants with decreased usage of the dominant GU and AG, respectively, at 5' and 3' splice sites (Wang et al., 2012; Feng et al., 2015). Therefore, SKIP is a splicing factor that regulates the efficient and precise splicing of pre-mRNAs on a genome-wide scale in *Arabidopsis*.

In addition, SKIP binds directly to the pre-mRNAs of clock genes, including *PSEUDO-RESPONSE REGULATOR 7* (*PRR7*)



and *PRR9*, to regulate their accurate splicing and mRNA maturation (Wang et al., 2012). Compared to wild-type plants, *skip-1* plants show increased levels of aberrantly spliced variants of *PRR7* and *PRR9* and decreased levels of functional, fully spliced *PRR7* and *PRR9* mRNAs. The reduced levels of functional *PRR7* and *PRR9* mRNAs in *skip-1* contribute to its lengthened circadian period phenotype (Wang et al., 2012). Therefore, SKIP mediates the circadian clock by regulating the alternative splicing of clock genes. These findings demonstrate that post-transcriptional regulation plays vital roles in controlling the circadian clock (Sanchez et al., 2010; Jones et al., 2012; Wang et al., 2012; Li et al., 2019).

SKIP also regulates plant response to abiotic stress (Hou et al., 2009; Zhang Y, et al., 2013; Feng et al., 2015; Li et al., 2016; Li et al., 2019). Mutations in *SKIP* have been shown to cause hypersensitivity to salt or osmotic stress in *Arabidopsis*. Compared to wild-type plants, *skip-1* plants exhibit a significantly decreased germination rate, survival rate, and relative root growth under high-salt or drought conditions (Feng et al., 2015; Li et al., 2016). Meanwhile, ectopic expression of *SKIP* results in increased tolerance to salt or dehydration (Lim et al., 2010). In *Arabidopsis*, salt stress induces genome-wide alternative splicing events, most of which are regulated by SKIP (Feng et al., 2015). SKIP mediates the recognition or cleavage of 5' alternative donor sites and 3' alternative acceptor sites, and it is essential for alternative gene splicing under conditions of salt stress (Feng et al., 2015). Transcripts of several salt tolerance-related genes, including *NA⁺/H⁺ EXCHANGER 1 (NHX1)*, *CALCINEURIN B-LIKE PROTEIN 1 (CBL1)*, *DELTA1-PYRROLINE-5-CARBOXYLATE SYNTHASE 1 (P5CS1)*, *RARE-COLD-INDUCIBLE 2A (RCI2A)*, and *PROTEIN S-ACYL TRANSFERASE 10 (PAT10)*, are aberrantly spliced in *skip-1* under salt stress conditions, decreasing the abundance of fully spliced mRNAs. Premature termination during the translation of these aberrantly spliced variants in *skip-1* reduces the level of functional proteins, resulting in salt hypersensitivity (Feng et al., 2015; Li et al., 2016; Li et al., 2019). Therefore, SKIP is necessary for plants to respond to salt or drought stress, and alternative gene splicing is crucial for plants to respond to environmental cues (Hou et al., 2009; Zhang Y, et al., 2013; Feng et al., 2015; Li et al., 2016; Li et al., 2019; reviewed in Filichkin et al., 2015; Laloum et al., 2018).

In summary, SKIP is a splicing factor that is essential for the precise and efficient splicing of pre-mRNAs on a genome-wide scale, and it mediates the circadian clock and resistance to salt or drought stress by regulating the alternative splicing of clock and salt tolerance-related genes in plants.

SKIP MEDIATES THE FLORAL TRANSITION BY REGULATING TRANSCRIPTION

Initially, *skip-1* was characterized as a photoperiod-insensitive early flowering mutant. As defects in the circadian clock may cause changes in the temporal expression of *CONSTANS (CO)*, which regulates *FLOWERING LOCUS T (FT)* transcription and affects flowering time (reviewed in Yanovsky and Kay, 2003; Song et al., 2015; Shim et al., 2017), some speculate whether SKIP regulates flowering time by regulating the circadian expression of *CO*. However, no obvious change in the expression of *CO* was observed in *skip-1* (Cao et al., 2015), suggesting that the early flowering phenotype of *skip-1* is not caused by a circadian clock defect.

FLOWERING LOCUS C (FLC) is a MADS-box transcription factor that dose-dependently suppresses the floral transition (reviewed in He and Amasino, 2005; Michaels, 2009; Romera-Branchat et al., 2014; Whittaker and Dean, 2017). Both sense and antisense (*COOLAIR*) transcripts of *FLC* undergo alternative splicing (Marquardt et al., 2014; Mahrez et al., 2016); moreover, temperature-dependent alternative splicing of *FLOWERING LOCUS M (FLM)*, which changes the ratio of *FLM-β* to *FLM-δ*, plays a vital role in regulating the temperature-dependent floral transition (Lee et al., 2013; Posé et al., 2013). Given that SKIP is a splicing factor that regulates genome-wide pre-mRNA splicing in *Arabidopsis*, it was suggested that SKIP is essential for the alternative splicing of sense and antisense *FLC* transcripts, or *FLM* pre-mRNA, to regulate the level of functional, mature mRNAs in the control of flowering time. However, splicing defects in *FLC* sense and *COOLAIR* transcripts were not observed in *skip-1* (Cao et al., 2015). In addition, even though SKIP is required for the alternative splicing of *FLM*, the alternative splicing pattern of *FLM* pre-mRNA (i.e., the ratio of *FLM-β* to *FLM-δ*) in response

to temperature changes was not obviously affected in *skip-1* (Cao et al., 2015). Thus, SKIP is not required for the accurate splicing of *FLC* or *COOLAIR* pre-mRNA. Instead, *FLC* transcription (i.e., the level of unspliced *FLC* mRNA) is significantly repressed in *skip-1*; this is reflected in the obvious repression of mature *FLC* transcripts and the early flowering phenotype observed in *skip-1* under different photoperiod and temperature conditions (Cao et al., 2015; Li et al., 2016; Li et al., 2019). Thus, SKIP activates *FLC* expression at the transcriptional level in a photoperiod- and temperature-independent manner, and it represses flowering time by promoting *FLC* transcription.

To determine how SKIP activates *FLC* transcription, thereby repressing flowering, a yeast two-hybrid screen was performed to identify factors that interact with SKIP. ELF7, a Paf1c component that regulates transcription elongation, was found to interact with SKIP (Cao et al., 2015). Paf1c represses flowering by promoting the trimethylation of histone H3 at lysine 4 on *FLC* chromatin and activating *FLC* transcription (He et al., 2004; Oh et al., 2004; Yu and Michaels, 2010). It was verified that SKIP interacts physically and genetically with ELF7 to regulate flowering time in *Arabidopsis* (Cao et al., 2015; Li et al., 2016). SKIP and ELF7 bind directly to *FLC/MAFs* chromatin and promote histone H2B monoubiquitination, increasing the trimethylation of histone H3 at lysine 4 and *FLC/MAFs* transcriptional activation, thereby repressing the floral transition in wild-type plants (Cao et al.,

2015). During this process, SKIP functions as a co-transcriptional activator, mediating plant flowering *via* the regulation of *FLC/MAFs* transcription. Therefore, SKIP can promote the transcription of specific genes as a co-transcriptional activator in plants.

CONCLUSION AND PROSPECTS

Plants respond to internal and environmental signals affecting their development and environmental fitness by accurately regulating gene expression at the transcriptional and post-transcriptional levels. SKIP has a dual function in plants: it acts as a splicing factor to control efficient and precise pre-mRNA splicing on a genome-wide scale by interacting with other spliceosome components (e.g., SR45) and integrating into the spliceosome, and it is required for 5' and 3' splice site recognition or cleavage (Wang et al., 2012; Feng et al., 2015). SKIP affects the circadian clock and plant responses to salt stress by regulating the accurate splicing of certain clock- and salt tolerance-related genes (Wang et al., 2012; Feng et al., 2015; Li et al., 2016; Li et al., 2019). SKIP also functions as a transcriptional co-regulator by interacting with other transcriptional regulators (e.g., Paf1c), and it mediates the floral transition (Cao et al., 2015; Li et al., 2016; Li et al., 2019). Therefore, SKIP precisely regulates gene expression at the transcriptional and post-transcriptional levels to mediate plant development and environmental fitness (Figure 2).

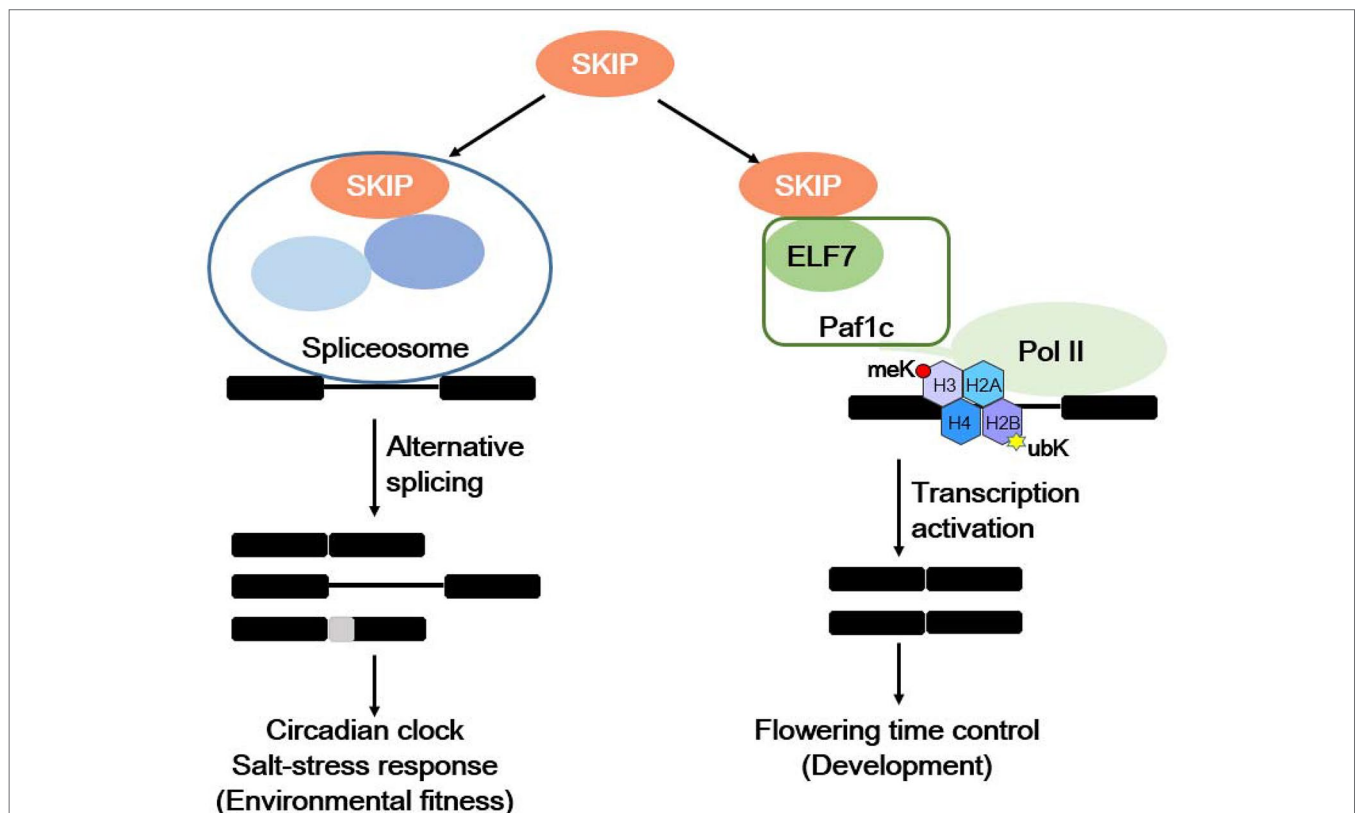


FIGURE 2 | The dual functions of SKIP in mediating environmental fitness and development in plants. SKIP is a splicing factor that regulates the alternative splicing of clock and stress tolerance-related genes in order to mediate the circadian clock and stress responses. It also functions as a transcriptional coactivator by interacting with RNA polymerase II-associated factor 1 complex (Paf1c) to activate *FLC/MAFs* transcription and mediate the floral transition.

In addition to SKIP, several other splicing factors are believed to have dual functions in splicing and transcription in *Arabidopsis*. For example, the RNA-binding protein SR45, first identified as an interacting partner of U1-70K and an essential splicing factor in plants (Golovkin and Reddy, 1999; Ali et al., 2007; Carvalho et al., 2016), was reported to be recruited to *FLC* chromatin by VIVIPAROUS 1/ABI3-LIKE factor 1 (VAL1), a transcriptional repressor that further recruits the transcriptional repression complex plant homeodomain–polycomb repressive complex 2 (PHD-PRC2), resulting in decreased histone acetylation of *FLC* chromatin and *FLC* transcriptional silencing during vernalization (Qüesta et al., 2016). This suggests that SR45 is a component of the transcriptional repression complex that suppresses *FLC* expression in *Arabidopsis*. In addition, SR45 participates in a small interfering RNA-directed DNA methylation (RdDM) pathway that mediates gene silencing in *Arabidopsis* (Ausin et al., 2012). ZINC-FINGER AND OCRE DOMAIN-CONTAINING PROTEIN 1 (ZOP1), a pre-mRNA splicing factor, associates with such typical spliceosome components as U1-70K, STABILIZED 1 (STA1, a PRP6-like splicing factor), and RNA-DIRECTED DNA METHYLATION 16 (RDM16, pre-mRNA-splicing factor 3) and is required for RdDM-mediated transcriptional gene silencing (Dou et al., 2013; Huang et al., 2013; Zhang CJ, et al., 2013). Furthermore, PRP31, a conserved pre-mRNA splicing factor, associates with STA1, ZOP1, and RDM16 to regulate transcriptional gene silencing in a manner independent of the RdDM pathway (Du et al., 2015). Therefore, it appears that SKIP is not the only factor required for the regulation of gene expression at both the transcriptional and post-transcriptional levels. Bifunctional splicing factors may provide an effective way for plants to coordinate their responses to environmental and internal signals in order to adjust their development and environmental adaptation *via*

accurate gene expression regulation at the transcriptional and post-transcriptional levels.

Resolved structures of the spliceosome from yeast and human cells indicate that SKIP is intrinsically highly disordered; it serves as a scaffold protein and interacts directly with other necessary components to facilitate splicing by promoting spliceosome assembly (Yan et al., 2015; Bai et al., 2017; Zhang et al., 2017; Zhang et al., 2018). It will be interesting to determine the functions of SKIP in transcriptional complex assembly (e.g., Paf1c-SKIP) and transcriptional regulation in plants. In addition, as transcription is usually coupled with splicing (reviewed in Naftelberg et al., 2015; Saldi et al., 2016; Herzel et al., 2017) and given that SKIP is integrated into both the spliceosome and transcriptional complexes (Wang et al., 2012; Cao et al., 2015; Li et al., 2016; Li et al., 2019), it should be investigated whether SKIP mediates the coupling of transcription and splicing in *Arabidopsis*.

AUTHOR CONTRIBUTIONS

YC and LM wrote the manuscript.

FUNDING

This work was supported by grants from the National Natural Science Foundation of China (31770247 and 31600248) and the Innovation Research Team of Beijing Municipal Government Science Foundation (IDHT20170513).

ACKNOWLEDGMENTS

We thank Dr. Jessica Habashi for critical reading of the manuscript.

REFERENCES

- Albers, M., Diment, A., Muraru, M., Russell, C. S., and Beggs, J. D. (2003). Identification and characterization of Prp45p and Prp46p, essential pre-mRNA splicing factors. *RNA* 9, 138–150. doi: 10.1261/rna.2119903
- 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. doi: 10.1371/journal.pone.0000471
- 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. doi: 10.4161/epi.7.1.18782
- Bai, R., Yan, C., Wan, R., Lei, J., and Shi, Y. (2017). Structure of the post-catalytic spliceosome from *Saccharomyces cerevisiae*. *Cell* 171, 1589–1598. doi: 10.1016/j.cell.2017.10.038
- Bessonov, S., Anokhina, M., Will, C. L., Urlaub, H., and Lührmann, R. (2008). Isolation of an active step I spliceosome and composition of its RNP core. *Nature* 452, 846–850. doi: 10.1038/nature06842
- Bonasio, R., and Shiekhattar, R. (2014). Regulation of transcription by long noncoding RNAs. *Annu. Rev. Genet.* 48, 433–455. doi: 10.1146/annurev-genet-120213-092323
- Braunschweig, U., Gueroussov, S., Plocik, A. M., Graveley, B. R., and Blencowe, B. J. (2013). Dynamic integration of splicing within gene regulatory pathways. *Cell* 152, 1252–1269. doi: 10.1016/j.cell.2013.02.034
- Brès, V., Gomes, N., Pickle, L., and Jones, K. A. (2005). A human splicing factor, SKIP, associates with P-TEFb and enhances transcription elongation by HIV-1 Tat. *Genes Dev.* 19, 1211–1226. doi: 10.1101/gad.1291705
- Brès, V., Yoshida, T., Pickle, L., and Jones, K. A. (2009). SKIP interacts with c-Myc and Menin to promote HIV-1 Tat transactivation. *Mol. Cell.* 36, 75–87. doi: 10.1016/j.molcel.2009.08.015
- Cao, Y., Wen, L. G., Wang, Z., and Ma, L. G. (2015). SKIP interacts with the Paf1 complex to regulate flowering *via* the activation of *FLC* transcription in *Arabidopsis*. *Mol. Plant* 8, 1816–1819. doi: 10.1016/j.molp.2015.09.004
- Carvalho, R. F., Szakonyi, D., Simpson, C. G., Barbosa, I. C., Brown, J. W., Baena-González, E., et al. (2016). The *Arabidopsis* SR45 splicing factor, a negative regulator of sugar signaling, modulates SNF1-related protein kinase 1 stability. *Plant Cell* 28, 1910–1925. doi: 10.1105/tpc.16.00301
- Chen, Y., Zhang, L., and Jones, K. A. (2011). SKIP counteracts p53-mediated apoptosis *via* selective regulation of p21^{Cip1} mRNA splicing. *Genes Dev.* 25, 701–716. doi: 10.1101/gad.2002611
- Day, I. S., Golovkin, M., Palusa, S. G., Link, A., Ali, G. S., Thomas, J., et al. (2012). Interactions of SR45, an SR-like protein, with spliceosomal proteins and an intronic sequence: insights into regulated splicing. *Plant J.* 71, 936–947. doi: 10.1111/j.1365-313X.2012.05042.x
- Dou, K., Huang, C. F., Ma, Z. Y., Zhang, C. J., Zhou, J. X., Huang, H. W., et al. (2013). The PRP6-like splicing factor STA1 is involved in RNA-directed DNA methylation by facilitating the production of Pol V-dependent scaffold RNAs. *Nucleic Acids Res.* 41, 8489–8502. doi: 10.1093/nar/gkt639
- Du, J. L., Zhang, S. W., Huang, H. W., Cai, T., Li, L., Chen, S., et al. (2015). The splicing factor PRP31 is involved in transcriptional gene silencing and stress response in *Arabidopsis*. *Mol. Plant* 8, 1053–1068. doi: 10.1016/j.molp.2015.02.003

- Feng, J., Li, J., Gao, Z., Lu, Y., Yu, J., Zheng, Q., et al. (2015). SKIP confers osmotic tolerance during salt stress by controlling alternative gene splicing in *Arabidopsis*. *Mol. Plant*. 8, 1038–1052. doi: 10.1016/j.molp.2015.01.011
- Figuerola, J. D., and Hayman, M. J. (2004). The human ski-interacting protein functionally substitutes for the yeast *PRP45* gene. *Biochem. Biophys. Res. Commun.* 319, 1105–1109. doi: 10.1016/j.bbrc.2004.05.096
- Filichkin, S., Priest, H. D., Megraw, M., and Mockler, T. C. (2015). Alternative splicing in plants: directing traffic at the crossroads of adaptation and environmental stress. *Curr. Opin. Plant Biol.* 24, 125–135. doi: 10.1016/j.pbi.2015.02.008
- Gilchrist, D. A., Dos Santos, G., Fargo, D. C., Xie, B., Gao, Y., Li, L., et al. (2010). Pausing of RNA polymerase II disrupts DNA-specified nucleosome organization to enable precise gene regulation. *Cell* 143, 540–551. doi: 10.1016/j.cell.2010.10.004
- Golovkin, M., and Reddy, A. S. (1996). Structure and expression of a plant U1 snRNP 70K gene: alternative splicing of U1 snRNP 70K pre-mRNAs produces two different transcripts. *Plant Cell*. 8, 1421–1435. doi: 10.1105/tpc.8.8.1421
- 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. doi: 10.1074/jbc.274.51.36428
- Hajheidari, M., Koncz, C., and Eick, D. (2013). Emerging roles for RNA polymerase II CTD in *Arabidopsis*. *Trends Plant Sci.* 18, 633–643. doi: 10.1016/j.tplants.2013.07.001
- Hantsche, M., and Cramer, P. (2017). Conserved RNA polymerase II initiation complex structure. *Curr. Opin. Struct. Biol.* 47, 17–22. doi: 10.1016/j.sbi.2017.03.013
- He, F., and Jacobson, A. (2015). Nonsense-mediated mRNA decay: degradation of defective transcripts is only part of the story. *Annu. Rev. Genet.* 49, 339–366. doi: 10.1146/annurev-genet-112414-054639
- He, Y., and Amasino, R. M. (2005). Role of chromatin modification in flowering-time control. *Trends Plant Sci.* 10, 30–35. doi: 10.1016/j.tplants.2004.11.003
- He, Y., Doyle, M. R., and Amasino, R. M. (2004). PAF1-complex-mediated histone methylation of *FLOWERING LOCUS C* chromatin is required for the vernalization-responsive, winter-annual habit in *Arabidopsis*. *Genes Dev.* 18, 2774–2784. doi: 10.1101/gad.1244504
- Herzel, L., Ottoz, D. S. M., Alpert, T., and Neugebauer, K. M. (2017). Splicing and transcription touch base: co-transcriptional spliceosome assembly and function. *Nat. Rev. Mol. Cell. Biol.* 18, 637–650. doi: 10.1038/nrm.2017.63
- Hou, X., Xie, K., Yao, J., Qi, Z., and Xiong, L. (2009). A homolog of human ski-interacting protein in rice positively regulates cell viability and stress tolerance. *Proc. Natl. Acad. Sci. U.S.A.* 106, 6410–6415. doi: 10.1073/pnas.0901940106
- Hsin, J. P., and Manley, J. L. (2012). The RNA polymerase II CTD coordinates transcription and RNA processing. *Genes Dev.* 26, 2119–2137. doi: 10.1101/gad.200303.112
- Huang, C. F., Miki, D., Tang, K., Zhou, H. R., Zheng, Z., Chen, W., et al. (2013). A pre-mRNA splicing factor is required for RNA-directed DNA methylation in *Arabidopsis*. *PLoS Genet.* 9, e1003779. doi: 10.1371/journal.pgen.1003779
- Jeronimo, C., and Robert, F. (2017). The mediator complex: at the nexus of RNA polymerase II transcription. *Trends Cell Biol.* 27, 765–783. doi: 10.1016/j.tcb.2017.07.001
- Jones, M. A., Williams, B. A., McNicol, J., Simpson, C. G., Brown, J. W., and Harmer, S. L. (2012). Mutation of *Arabidopsis* spliceosomal timekeeper locus1 causes circadian clock defects. *Plant Cell*. 24, 4066–4082. doi: 10.1105/tpc.112.104828
- Jonkers, I., and Lis, J. T. (2015). Getting up to speed with transcription elongation by RNA polymerase II. *Nat. Rev. Mol. Cell. Biol.* 16, 167–177. doi: 10.1038/nrm3393
- Keren, H., Lev-Maor, G., and Ast, G. (2010). Alternative splicing and evolution: diversification, exon definition and function. *Nat. Rev. Genet.* 11, 345–355. doi: 10.1038/nrg2776
- Kouzarides, T. (2007). Chromatin modifications and their function. *Cell* 128, 695–705. doi: 10.1016/j.cell.2007.02.005
- Kwak, H., and Lis, J. T. (2013). Control of transcriptional elongation. *Annu. Rev. Genet.* 47, 483–508. doi: 10.1146/annurev-genet-110711-155440
- Laduron, S., Deplus, R., Zhou, S., Kholmanskikh, O., Godelaine, D., De Smet, C., et al. (2004). MAGE-A1 interacts with adaptor SKIP and the deacetylase HDAC1 to repress transcription. *Nucleic Acids Res.* 32, 4340–4350. doi: 10.1093/nar/gkh735
- Laloum, T., Martin, G., and Duque, P. (2018). Alternative splicing control of abiotic stress responses. *Trends Plant Sci.* 23, 140–150. doi: 10.1016/j.tplants.2017.09.019
- Lawrence, M., Daujat, S., and Schneider, R. (2016). Lateral thinking: how histone modifications regulate gene expression. *Trends Genet.* 32, 42–56. doi: 10.1016/j.tig.2015.10.007
- Lee, J. H., Ryu, H. S., Chung, K. S., Posé, D., Kim, S., Schmid, M., et al. (2013). Regulation of temperature-responsive flowering by MADS-box transcription factor repressors. *Science* 342, 628–632. doi: 10.1126/science.1241097
- Lee, Y., and Rio, D. C. (2015). Mechanisms and regulation of alternative pre-mRNA splicing. *Annu. Rev. Biochem.* 84, 291–323. doi: 10.1146/annurev-biochem-060614-034316
- Li, B., Carey, M., and Workman, J. L. (2007). The role of chromatin during transcription. *Cell* 128, 707–719. doi: 10.1016/j.cell.2007.01.015
- Li, Y., Xia, C., Feng, J., Yang, D., Wu, F., Cao, Y., et al. (2016). The SNW domain of SKIP is required for its integration into the spliceosome and its interaction with the Paf1 complex in *Arabidopsis*. *Mol. Plant*. 9, 1040–1050. doi: 10.1016/j.molp.2016.04.011
- Li, Y., Yang, J., Shang, X. D., Lv, W. Z., Xia, C. C., Wang, C., et al. (2019). SKIP regulates environmental fitness and floral transition by forming two distinct complexes in *Arabidopsis*. *New Phytol.* 224, 321–335. doi: 10.1111/nph.15990
- Licatalosi, D. D., and Darnell, R. B. (2010). RNA processing and its regulation: global insights into biological networks. *Nat. Rev. Genet.* 11, 75–87. doi: 10.1038/nrg2673
- Lim, G. H., Zhang, X., Chung, M. S., Lee, D. J., Woo, Y. M., Cheong, H. S., et al. (2010). A putative novel transcription factor, AtSKIP, is involved in abscisic acid signalling and confers salt and osmotic tolerance in *Arabidopsis*. *New Phytol.* 185, 103–113. doi: 10.1111/j.1469-8137.2009.03032.x
- Mahrez, W., Shin, J., Muñoz-Viana, R., Figueiredo, D. D., Trejo-Arellano, M. S., Exner, V., et al. (2016). BRR2a affects flowering time via *FLC* splicing. *PLoS Genet.* 12, e1005924. doi: 10.1371/journal.pgen.1005924
- Marquardt, S., Raitskin, O., Wu, Z., Liu, F., Sun, Q., and Dean, C. (2014). Functional consequences of splicing of the antisense transcript *COOLAIR* on *FLC* transcription. *Mol. Cell*. 54, 156–165. doi: 10.1016/j.molcel.2014.03.026
- Matzke, M., Kanno, T., Daxinger, L., Huettel, B., and Matzke, A. J. (2009). RNA-mediated chromatin-based silencing in plants. *Curr. Opin. Cell. Biol.* 21, 367–376. doi: 10.1016/j.ccb.2009.01.025
- Michaels, S. D. (2009). Flowering time regulation produces much fruit. *Curr. Opin. Plant Biol.* 12, 75–80. doi: 10.1016/j.pbi.2008.09.005
- Moore, M. J., and Proudfoot, N. J. (2009). Pre-mRNA processing reaches back to transcription and ahead to translation. *Cell* 136, 688–700. doi: 10.1016/j.cell.2009.02.001
- Naftelberg, S., Schor, I. E., Ast, G., and Kornblihtt, A. R. (2015). Regulation of alternative splicing through coupling with transcription and chromatin structure. *Annu. Rev. Biochem.* 84, 165–198. doi: 10.1146/annurev-biochem-060614-034242
- Oh, S., Zhang, H., Ludwig, P., and van Nocker, S. (2004). A mechanism related to the yeast transcriptional regulator Paf1c is required for expression of the *Arabidopsis* *FLC*/MAF MADS box gene family. *Plant Cell*. 16, 2940–2953. doi: 10.1105/tpc.104.026062
- Posé, D., Verhage, L., Ott, F., Yant, L., Mathieu, J., Angenent, G. C., et al. (2013). Temperature-dependent regulation of flowering by antagonistic *FLM* variants. *Nature* 503, 414–417. doi: 10.1038/nature12633
- Qüesta, J. I., Song, J., Geraldo, N., An, H., and Dean, C. (2016). *Arabidopsis* transcriptional repressor VAL1 triggers polycomb silencing at *FLC* during vernalization. *Science* 353, 485–488. doi: 10.1126/science.aaf7354
- Reddy, A. S. (2007). Alternative splicing of pre-messenger RNAs in plants in the genomic era. *Annu. Rev. Plant Biol.* 58, 267–294. doi: 10.1146/annurev.arplant.58.032806.103754
- Romera-Branchat, M., Andrés, F., and Coupland, G. (2014). Flowering responses to seasonal cues: what's new? *Curr. Opin. Plant Biol.* 21, 120–127. doi: 10.1016/j.pbi.2014.07.006
- Sainsbury, S., Bernecky, C., and Cramer, P. (2015). Structural basis of transcription initiation by RNA polymerase II. *Nat. Rev. Mol. Cell. Biol.* 16, 129–143. doi: 10.1038/nrm3952
- Saldi, T., Cortazar, M. A., Sheridan, R. M., and Bentley, D. L. (2016). Coupling of RNA polymerase II transcription elongation with pre-mRNA splicing. *J. Mol. Biol.* 428, 2623–2635. doi: 10.1016/j.jmb.2016.04.017

- Sanchez, S. E., Petrillo, E., Beckwith, E. J., Zhang, X., Rugnone, M. L., Hernando, C. E., et al. (2010). A methyl transferase links the circadian clock to the regulation of alternative splicing. *Nature* 468, 112–116. doi: 10.1038/nature09470
- Schneider, C., Agafonov, E., Schmitzova, J., Hartmuth, K., Fabrizio, P., and Luhrmann, R. (2015). Dynamic contacts of U2, RES, Cwc25, Prp8 and Prp45 proteins with the pre-mRNA branch-site and 3' splice site during catalytic activation and step 1 catalysis in yeast spliceosomes. *PLoS Genet.* 11, e1005539. doi: 10.1371/journal.pgen.1005539
- Shim, J. S., Kubota, A., and Imaizumi, T. (2017). Circadian clock and photoperiodic flowering in *Arabidopsis*: CONSTANS is a hub for signal integration. *Plant Physiol.* 173, 5–15. doi: 10.1104/pp.16.01327
- Song, Y. H., Shim, J. S., Kinmonth-Schultz, H. A., and Imaizumi, T. (2015). Photoperiodic flowering: time measurement mechanisms in leaves. *Annu. Rev. Plant Biol.* 66, 441–464. doi: 10.1146/annurev-arplant-043014-115555
- Staiger, D., and Brown, J. W. (2013). Alternative splicing at the intersection of biological timing, development, and stress responses. *Plant Cell.* 25, 3640–3656. doi: 10.1105/tpc.113.113803
- 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. doi: 10.1016/j.tplants.2012.06.001
- Tessarz, P., and Kouzarides, T. (2014). Histone core modifications regulating nucleosome structure and dynamics. *Nat. Rev. Mol. Cell. Biol.* 15, 703–708. doi: 10.1038/nrm3890
- Van Lijsebettens, M., and Grasser, K. D. (2014). Transcript elongation factors: shaping transcriptomes after transcript initiation. *Trends Plant Sci.* 19, 717–726. doi: 10.1016/j.tplants.2014.07.002
- Van Oss, S. B., Cucinotta, C. E., and Arndt, K. M. (2017). Emerging insights into the roles of the Paf1 complex in gene regulation. *Trends Biochem. Sci.* 42, 788–798. doi: 10.1016/j.tibs.2017.08.003
- Voss, T. C., and Hager, G. L. (2014). Dynamic regulation of transcriptional states by chromatin and transcription factors. *Nat. Rev. Genet.* 15, 69–81. doi: 10.1038/nrg3623
- Wahl, M. C., Will, C. L., and Luhrmann, R. (2009). The spliceosome: design principles of a dynamic RNP machine. *Cell* 136, 701–718. doi: 10.1016/j.cell.2009.02.009
- Wang, X., Wu, F., Xie, Q., Wang, H., Wang, Y., Yue, Y., et al. (2012). SKIP is a component of the spliceosome linking alternative splicing and the circadian clock in *Arabidopsis*. *Plant Cell.* 24, 3278–3295. doi: 10.1105/tpc.112.100081
- Whittaker, C., and Dean, C. (2017). The *FLC* locus: a platform for discoveries in epigenetics and adaptation. *Annu. Rev. Cell Dev. Biol.* 33, 555–575. doi: 10.1146/annurev-cellbio-100616-060546
- 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. doi: 10.1126/science.aac7629
- Yanovsky, M. J., and Kay, S. A. (2003). Living by the calendar: how plants know when to flower. *Nat. Rev. Mol. Cell. Biol.* 4, 265–275. doi: 10.1038/nrm1077
- Yu, X., and Michaels, S. D. (2010). The *Arabidopsis* Paf1c complex component CDC73 participates in the modification of *FLOWERING LOCUS C* chromatin. *Plant Physiol.* 153, 1074–1084. doi: 10.1104/pp.110.158386
- Zhang, C. J., Zhou, J. X., Liu, J., Ma, Z. Y., Zhang, S. W., Dou, K., et al. (2013). The splicing machinery promotes RNA-directed DNA methylation and transcriptional silencing in *Arabidopsis*. *EMBO J.* 32, 1128–1140. doi: 10.1038/emboj.2013.49
- Zhang, X., Yan, C., Hang, J., Finci, L. I., Lei, J., and Shi, Y. (2017). An atomic structure of the human spliceosome. *Cell* 169, 918–929. doi: 10.1016/j.cell.2017.04.033
- Zhang, X., Yan, C., Zhan, X., Li, L., Lei, J., and Shi, Y. (2018). Structure of the human activated spliceosome in three conformational states. *Cell Res.* 28, 307–322. doi: 10.1038/cr.2018.14
- Zhang, Y., Zhao, L., Li, H., Gao, Y., Li, Y., Wu, X., et al. (2013). GmGBP1, a homolog of human ski interacting protein in soybean, regulates flowering and stress tolerance in *Arabidopsis*. *BMC Plant Biol.* 13, 21. doi: 10.1186/1471-2229-13-21

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2019 Cao and Ma. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



Genome-Wide Identification of Splicing Quantitative Trait Loci (sQTLs) in Diverse Ecotypes of *Arabidopsis thaliana*

Waqas Khokhar¹, Musa A. Hassan^{2,3}, Anireddy S. N. Reddy⁴, Saurabh Chaudhary¹, Ibtissam Jabre¹, Lee J. Byrne¹ and Naeem H. Syed^{1*}

OPEN ACCESS

Edited by:

Kranthi Kiran Mandadi,
Texas A&M University,
United States

Reviewed by:

William Brad Barbazuk,
University of Florida,
United States
Stephen M. Mount,
University of Maryland,
College Park, United States
Renesh Bedre,
Texas A&M University,
United States

*Correspondence:

Naeem Hasan Syed
naeem.syed@canterbury.ac.uk

Specialty section:

This article was submitted to
Plant Physiology,
a section of the journal
Frontiers in Plant Science

Received: 15 March 2019

Accepted: 26 August 2019

Published: 03 October 2019

Citation:

Khokhar W, Hassan MA, Reddy ASN,
Chaudhary S, Jabre I, Byrne LJ
and Syed NH (2019) Genome-Wide
Identification of Splicing Quantitative
Trait Loci (sQTLs) in Diverse Ecotypes
of *Arabidopsis thaliana*.
Front. Plant Sci. 10:1160.
doi: 10.3389/fpls.2019.01160

¹ School of Human and Life Sciences, Canterbury Christ Church University, Canterbury, United Kingdom, ² Division of Infection and Immunity, The Roslin Institute, University of Edinburgh, Edinburgh, United Kingdom, ³ Centre for Tropical Livestock Genetics and Health, University of Edinburgh, Edinburgh, United Kingdom, ⁴ Department of Biology and Program in Cell and Molecular Biology, Colorado State University, Fort Collins, CO, United States

Alternative splicing (AS) of pre-mRNAs contributes to transcriptome diversity and enables plants to generate different protein isoforms from a single gene and/or fine-tune gene expression during different development stages and environmental changes. Although AS is pervasive, the genetic basis for differential isoform usage in plants is still emerging. In this study, we performed genome-wide analysis in 666 geographically distributed diverse ecotypes of *Arabidopsis thaliana* to identify genomic regions [splicing quantitative trait loci (sQTLs)] that may regulate differential AS. These ecotypes belong to different microclimatic conditions and are part of the relict and non-relict populations. Although sQTLs were spread across the genome, we observed enrichment for *trans*-sQTL (*trans*-sQTLs hotspots) on chromosome one. Furthermore, we identified several sQTL (911) that co-localized with trait-linked single nucleotide polymorphisms (SNP) identified in the *Arabidopsis* genome-wide association studies (AraGWAS). Many sQTLs were enriched among circadian clock, flowering, and stress-responsive genes, suggesting a role for differential isoform usage in regulating these important processes in diverse ecotypes of *Arabidopsis*. In conclusion, the current study provides a deep insight into SNPs affecting isoform ratios/genes and facilitates a better mechanistic understanding of trait-associated SNPs in GWAS studies. To the best of our knowledge, this is the first report of sQTL analysis in a large set of *Arabidopsis* ecotypes and can be used as a reference to perform sQTL analysis in the Brassicaceae family. Since whole genome and transcriptome datasets are available for these diverse ecotypes, it could serve as a powerful resource for the biological interpretation of trait-associated loci, splice isoform ratios, and their phenotypic consequences to help produce more resilient and high yield crop varieties.

Keywords: splicing quantitative trait loci (sQTL), *Arabidopsis thaliana*, alternative splicing, isoform usage, GWAS, adaptation

INTRODUCTION

Plants have evolved diverse and efficient genetic and physiological strategies to cope with environmental fluctuations. For an appropriate response, plants employ different regulatory mechanisms that can modulate genomic architecture and transcriptome composition to generate phenotypic diversity, allowing them to engender appropriate responses and occupy diverse niches (Frachon et al., 2018). The majority of protein-coding genes (~90%) in plants contain introns, which must be removed by a process called pre-mRNA splicing to produce mature messenger RNAs (mRNAs) for translation. Due to differential exon and splice site usage, ~70% of plant genes can be alternatively spliced (AS) to generate few to thousands of structurally and functionally different mRNA isoforms with different fates (Filichkin et al., 2010; Marquez et al., 2012; Shen et al., 2014; Klepikova et al., 2016; Zhang et al., 2017). Interestingly, the majority of splicing regulator genes in plants are subject to extensive AS and change the profile of their splicing patterns in response to various environmental stresses (Palusa et al., 2007; Zhang et al., 2013; Calixto et al., 2018). AS events in some plant-specific SR genes are highly conserved from single-cell green alga *Chlamydomonas reinhardtii* or moss *Physcomitrella patens* to Arabidopsis, suggesting the importance of their regulation in plants (Iida and Go, 2006; Kalyna et al., 2006). Plants employ AS to fine-tune their physiology and metabolism to maintain a balance between carbon fixation and resource allocation under normal as well as biotic and/or abiotic stress conditions such as pathogen infection, temperature, salt, drought, wounding, and light (Feng et al., 2015; Huang et al., 2017; Calixto et al., 2018; Filichkin et al., 2018; Hartmann et al., 2018; Liu et al., 2018; Seaton et al., 2018). For example, short read (Illumina) and long read (Iso-seq) RNA-Seq data from poplar leaf, root, and stem xylem tissues under drought, salt, and temperature fluctuations revealed changes in AS profiles that modulate plant transcriptome under abiotic stresses (Filichkin et al., 2018). Moreover, intron retention (IR) was found to be the predominant and variable type of AS across all treatments and tissue types (Filichkin et al., 2018). In wheat, global profiling of the AS landscape responsive to drought, heat, and their combination showed significant AS variation (Liu et al., 2018). Recent studies have shown that salt stress, high temperature, disease, and cold-stress alter AS patterns in Arabidopsis, grape, rice, and soybean (Liu et al., 2013; Feng et al., 2015; Filichkin et al., 2015; Lee et al., 2016; Jiang et al., 2017; Calixto et al., 2018). Similarly, AS plays a key role in biotic stress responses; for example, data from soybean show that PsAvr3c, a *Phytophthora sojae* pathogen effector, can manipulate host spliceosomal machinery to shift splicing profiles and overcome the host immune system (Huang et al., 2017). However, it is not clear to what extent sequence variation influences splicing variation and whether chromatin environment also contributes towards it, in a condition- and stress-dependent manner (Syed et al., 2012; Pajoro et al., 2017; Jabre et al., 2019).

Genome and transcriptome sequencing in multiple accessions of Arabidopsis revealed that genetic variations influence the expression and splicing of several genes, including stress-responsive genes (Gan et al., 2011). For instance, a strong

association between genetic variations and spliced isoform accumulation was observed in sunflower (Smith et al., 2018). Interestingly, a significant proportion of splicing variation was associated with variants that harbor *trans*-QTLs, of which the majority were associated with spliceosomal proteins (Smith et al., 2018). Further examination of splicing variation in the wild and cultivated sunflowers revealed that higher frequency of AS was triggered during the domestication process. Some of the genetic variations can also influence important life-history traits like flowering and may have an influence on the geographical distribution of different accessions. For example, insertion polymorphisms in the first intron of the *FLOWERING LOCUS M (FLM)* gene influence AS and accelerate flowering in a temperature-dependent manner in many accessions of Arabidopsis (Lutz et al., 2015). Taken together, structural variation in the *FLM* gene can change ratios of different splice variants and influence a highly adaptive trait-like flowering in Arabidopsis (Lutz et al., 2015).

Insight into trait-associated genetic variants and their distribution pattern can delineate the mechanisms of genome regulation (Alonso-Blanco et al., 2016). Since AS can increase transcriptome/proteome complexity, the genetic underpinnings of natural sequence variations and AS are strongly associated with each other. Genetic variants, such as single nucleotide polymorphisms (SNPs), can substantially regulate the expression of transcript isoforms by modulating splice sites, which can impact phenotypic diversity and susceptibility to diseases in humans (Singh and Cooper, 2012; Takata et al., 2017). Recent studies showed that 22% of SNPs that are associated with different human diseases affect splicing (Qu et al., 2017; Park et al., 2018). The advances in RNA-Seq and genotyping have augmented the opportunities to monitor genetic variants and quantify transcriptomic features that allow us to understand the genetic landscapes of AS (Smith et al., 2018). Recent studies in animals and plants have elucidated the association of genetic variants with trait-associated loci at a population level (Monlong et al., 2014; Chen et al., 2018). In Arabidopsis, many studies have identified expression quantitative trait loci (eQTL) to explain trait-associated loci (Zhang et al., 2011). However, there have been very few studies on the genome-wide investigation of genetic variants affecting splicing patterns termed as splicing quantitative trait loci (sQTLs) in Arabidopsis (Yoo et al., 2016). To illuminate the role of genetic variations on AS in a large collection of highly diverse Arabidopsis lines, we sought to map sQTLs influencing AS. sQTLs spread across the genome can either act in *cis*- to disrupt the splicing of a proximal pre-mRNA by modulating splicing factors binding affinity to the pre-mRNA or in *trans* by regulating the splicing of distal pre-mRNA through altered expression of splicing regulators (Yoo et al., 2016; Qu et al., 2017).

We have used 666 diverse natural inbred Arabidopsis accessions including 'relicts' that occupied postglacial Eurasia first and were later invaded by 'non-relicts', which demographically spread along the east-west axis of Eurasia owing to its higher latitudinal regional diversity, human disturbance and climatic pressure (François et al., 2008; Alonso-Blanco et al., 2016). These accessions are of immense significance as they hold a huge amount of diversity and their

expansion leaves traces of admixture in the north and south of the species range that facilitated colonization to new habitats (Lee et al., 2017). In order to illuminate the relationship between splicing variants, phenotypic diversity and geographical distribution of these lines, we have performed sQTL analysis to reveal the functional impact of genetic variations on AS and adaptive consequences (Han et al., 2017). This analysis will provide a solid platform in the form of a useful and well-enriched dataset for sQTL in Arabidopsis to develop more resilient plant species in the face of climatic challenges to crop production.

MATERIALS AND METHODS

Genotype Datasets and Quality Control

High-quality genetic variant (SNPs) data for 1,135 globally distributed natural inbred lines of Arabidopsis representing relicts (accessions that hold ancestral habitats) and non-relicts (accessions range from native Eurasia to recently colonized North America) were downloaded in variant calling format (vcf) from the 1001 Genomes Data Centre (<https://1001genomes.org/data/GMI-MPI/releases/v3.1/>) (Alonso-Blanco et al., 2016). The genetic variants were pre-processed using stringent filtering criteria as follows: (i) having at least two genotypes across accessions, and (ii) each genotype has at least five occurrences across all accessions (Yang et al., 2017). Genotypes that occurred in less than five samples were converted to NA values to avoid their consideration in linkage analysis. The pre-processed high-quality genetic variant data was then used as input in mapping sQTLs (Supplementary Figure 1).

RNA-Sequencing Analysis

Single-ended 100 bp long RNA sequencing (RNA-Seq) reads, generated by Illumina HiSeq 2500 platform, for 727 ecotypes of Arabidopsis (without biological replicates) were downloaded from GEO dataset under accession number GSE80744 and SRA study SRP074107 (Kawakatsu et al., 2016). Read quality assessment was performed using FastQC and reads with Phred score < 20 were removed *via* Trim Galore version 0.5.0 (Andrews, 2010; Krueger, 2015). Transcript abundance level in terms of transcripts per million (TPM), was estimated for 82,910 isoforms in 34,212 genes using Arabidopsis reference transcript dataset (AtRTD2) and Salmon (Zhang et al., 2015; Patro et al., 2017) (Supplementary Dataset 1). The filtered RNA-seq reads were also mapped to the TAIR10 genome assembly using the STAR aligner version 2.7.0e with modified parameters (Supplementary Dataset 1) (Dobin et al., 2013). Transcripts were assembled and assemblies were merged using StringTie (Pertea et al., 2015) (Supplementary Dataset 4). The merged assembly was then used as a reference gene model to perform the second assembly to generate expression dataset for 124,422 expressed transcripts using StringTie (Pertea et al., 2015). The TPM values for both (transcriptome- and genome-based) expression datasets were then used to compute splicing ratios of each isoform for all genes. Genes with less than two isoforms or splicing variability <0.01 were filtered out using the core functionality of ulfasQTL method (Yang et al., 2017).

Identification of sQTLs

The genotype data from 1,135 and RNA-seq data from 727 Arabidopsis accessions were initially processed to filter 666 accessions that have both genotype and transcriptomic data (Supplementary Table 1; Supplementary Figure 1). The processed genetic variants and expression datasets were used to perform a genome-wide scan for sQTLs using the ulfasQTL package (v 0.1) – a composite sQTL analysis package that takes expression and genotype dataset to test splicing QTLs at genome-wide scale (Yang et al., 2017). It uses the core functionality of sQTLseeker approach, which is a multivariate model and calculates splicing ratios variability of a gene across samples using a distance-based approach. It estimates intra- and inter-genotype splicing variability using a non-parametric analog to the ANOVA (Monlong et al., 2014). ulfasQTL identifies and outputs a list of significant sQTLs (p -value ≤ 0.05) and their cognate genes across the genome. We used sQTL cognate genes to derive modes of AS events present in these genes using SUPPA version 2.2.1 (Alamancos et al., 2014; Alamancos et al., 2015), which provides an estimate of the inclusion level of AS events across all samples.

Colocalization of sQTL With GWAS Hits

The list of trait-associated loci was downloaded from the Arabidopsis genome-wide association studies (AraGWAS) catalog, which is a manually curated and standardized database that holds GWAS results for 167 publicly available phenotypes of Arabidopsis (Togninalli et al., 2018). It contains around 222,000 SNP-trait associations (GWAS hits), of which 3,887 are highly significant (p -value < 10^{-4}). The list of unique sQTLs was then matched with trait-associated loci (identical SNPs) to identify the significant association with important phenotypic traits associated sQTLs.

Gene Enrichment Analysis

Functional enrichment analysis was performed on the parent genes with significant sQTL using Database for Annotation, Visualization and Integrated Discovery (DAVID version 6.8) with default parameters that work on the principle of Fisher exact test for statistical analysis (Huang et al., 2009). The gene ontology (GO) terms [biological process (BP), molecular function (MF), and cellular components (CC)] were identified to provide biological insights into the significant sQTLs using false discovery rate (FDR) ≤ 0.05 .

Functional Annotation of sQTL and Non-sQTL SNPs

Publicly available functionally annotated SNPs dataset for 666 Arabidopsis accessions was downloaded and overlapped with sQTL-SNPs to obtain functional annotation of sQTLs and termed the other unmatched SNPs as non-sQTL SNPs. The SNPs were further classified into exonic-SNPs (nonsense, start-loss, frameshift, splice site, missense, synonymous, splice region, 5-UTR, 3-UTR, and non-coding exon variants) and non-exonic SNPs (intron and intergenic variants).

RESULTS

Majority of Splicing Events in *A. thaliana* Are Regulated as *Trans*

We performed genome-wide sQTL analysis using transcriptomic and genomic datasets and identified 6,406 and 6484 unique sQTLs that are associated with 6,129 and 7653 non-redundant genes, respectively (Table 1; Supplementary Tables 2 and 3). The comparison of sQTL analysis based on two expression datasets (AtRTD2 and genome assemblies) showed significant overlapping of sQTLs (6181) between two strategies (Supplementary Figure 1). The number of cognate genes increased for the genomic dataset, possibly due to the presence of novel genes/transcripts coming from known/novel genes. The number of transcripts present in the transcriptomic expression dataset is less, but these transcripts were experimentally validated in pilot studies so we used AtRTD2 transcriptomic expression dataset for further analysis. Although the sQTLs were randomly distributed across the genome, 1775 (~28%) of sQTLs localized on chromosome one, whereas chromosome two had the lowest number (956; ~15%) of SNPs linked to splicing patterns (Table 1 and Figure 1). The higher distribution of sQTLs on chromosome one is probably due to its bigger size as compared with other chromosomes (Rhee et al., 2003). To get a better understanding of the influence of the genetic variants (SNPs) on splicing patterns, sQTLs that were within 4 kb from their cognate gene were designated as *cis*-sQTL and every other sQTL outside this window, including those on a different chromosome, as *trans*-sQTLs. Subsequently, 356 *cis*-sQTLs (5% of the total mapped sQTLs) that were associated with the splicing of 301 genes and an extensively high frequency (95%) of *trans*-sQTLs were identified (Table 1 and Figure 2). Interestingly, an overrepresentation of *trans*-sQTLs (*trans*-sQTL hotspots) on chromosome one was observed, which indicates that the molecular factor(s) on this chromosome that may regulates the splicing of several transcripts are *trans*. The sQTLs were then mapped with a list of trait-associated SNPs available in GWAS catalog, and the exact match was found for 911 non-redundant SNPs associated with 757 genes (Supplementary Table 4). Among 911 sQTLs that overlapped with GWAS catalog SNPs, 61 are *cis*-sQTL and are associated with 48 different gene, while 850 are *trans*-sQTL and linked with 709 genes.

To complement the above analysis, we estimated the splicing ratios and AS categories of significant sQTL cognate genes (6129) containing 34,351 transcripts and 26968 AS events (Supplementary Datasets 2 and 3). The association of one of the top sQTL SNP that resides on chromosome 1 at position

1099063 with splicing ratios of *AT1G04170* (*EIF2 GAMMA; EUKARYOTIC TRANSLATION INITIATION FACTOR 2 GAMMA SUBUNIT*) showed that alteration in genotype from homozygote CC to heterozygote CT significantly modulates splicing ratios of the isoforms *AT1G04170_JC4* and *AT1G04170_P1* (Figure 3). Further analysis based on Percent Spliced in (PSI) values highlights the impact of sQTL at AS event level that reflects a significant change in A3 events (Figure 3). The overall AS estimates of sQTL cognate genes shows that IR (10630) was the dominant and mutually exclusive exons (MX) (60) was the least frequent alternative splicing event in diverse accessions of *A. thaliana* (Figure 4). Although less than IR events, the number of alternative 3' splice site (A3) events (8132) was significantly higher than the alternative 5' splice site (A5) events (4536). The number of Alternative first exon (AF) (1827) was close to skipped exons (ES) (1601) but way higher than alternative last exon (AL) (182). The AS events (26968) were subset into *cis* (1590) and *trans* (25378) sQTL cognate gene categories and observed a similar pattern of splicing events. Furthermore, predominance for IR and lower frequency for MX was also observed when analyzed AS genes that emerge as a result of the overlapping of sQTL with GWAS. We also highlighted the subclass of IR events known as Exitrons that are present in coding exons and possess the exonic potential to significantly modulate proteome diversity (Marquez et al., 2015). To illuminate their contribution towards IR events, we extracted exitrons (exon-like introns) by overlapping a list of publicly available 2459 exitrons with 3798 genes possessing 10630 IR event and found 913 common genes (Supplementary Table 5). This AS analysis based on sQTL cognate genes revealed a significant role of IR events in shaping transcriptome diversity and may influence a plant adaptation to complex environmental conditions.

sQTLs Are Enriched in Exonic Regions

In order to understand the biological significance of our results, we performed functional categorization of filtered SNPs (12,617,361) by classifying them into sQTLs (6,406) and non-sQTLs (12,610,954) SNPs. We were interested in better understanding their role in genome regulation, so we looked at the genomic distribution of sQTL/non-sQTL SNPs and characterized them as exonic SNPs and non-exonic SNPs. The sQTLs were enriched in exonic regions, which show their immense potential to modulate genomic architecture to generate phenotypic diversity. Among exonic variants, missense gene variants showed the highest frequency (1,883) followed by synonymous (1,764) and upstream gene (1,225) variant categories (Figure 5). Among non-exonic regions,

TABLE 1 | Genome-wide summary of sQTL mapped using 666 diverse ecotypes of *Arabidopsis thaliana*.

Chromosome	Total sQTLs	Cis-sQTL	Cis-sQTL linked genes	Trans-sQTL	Trans-sQTL linked genes
Chr1	1,775	116	93	5,177	1,741
Chr2	956	41	35	4,688	1,111
Chr3	1,266	45	46	4,773	1,083
Chr4	1,055	78	60	4,641	1,052
Chr5	1,354	76	76	5,112	1,537

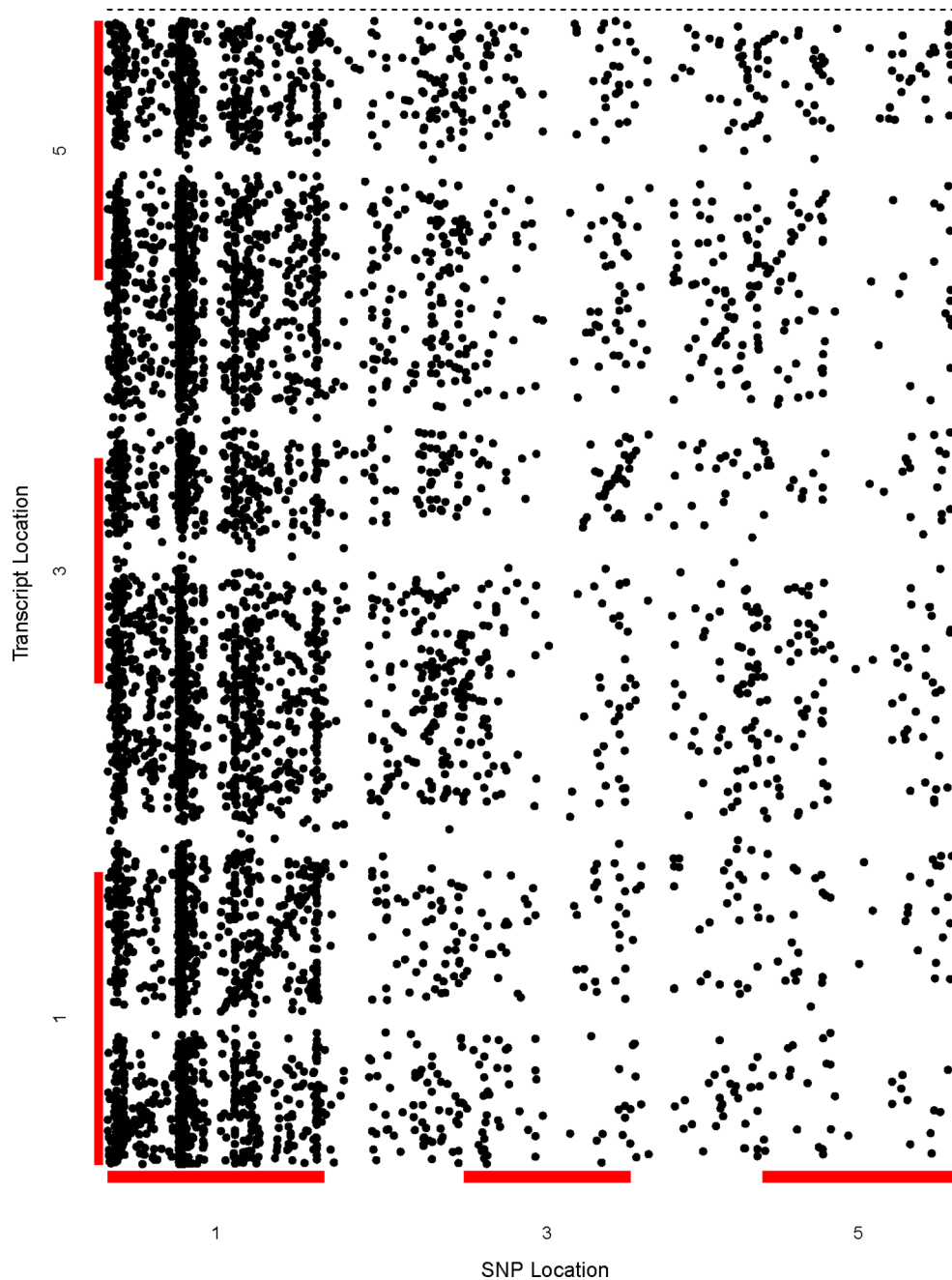


FIGURE 1 | Genetic analysis shows the genome-wide distribution of sQTLs and enrichment of *trans*-sQTL on chromosome 1. A comprehensive two-dimensional genome-wide sQTL map showing the distribution of all significant signals for unique sQTLs. Each dot represents an association between a genetic variant location (x-axis) and cognate transcript location (y-axis). The points along the diagonal correspond to *cis*-associations (window 4 kb) with the vertical band representing *trans*-sQTLs.

sQTLs that reside in the intronic regions were significantly higher (282) than intergenic regions (7). Although the non-sQTLs were also enriched in exonic regions, yet within exonic regions, they painted a different picture as compared to sQTLs as they showed a higher number of upstream gene variant regions, compared with the rest of exonic non-sQTLs. However, the proportion of both categories (intron, intergenic) of non-exonic non-sQTLs is almost

similar. Analysis of the functional context of sQTLs provided by SnpEff revealed a high impact on splice acceptor and stop gained functional categories, although their occurrence in sQTLs is low. The summary of all exonic and non-exonic SNPs for both functional classes (sQTLs and non-sQTLs) showed around 95% SNPs are enriched in variants belong to exonic regions and the majority of them represents the *trans* category, while only 5%

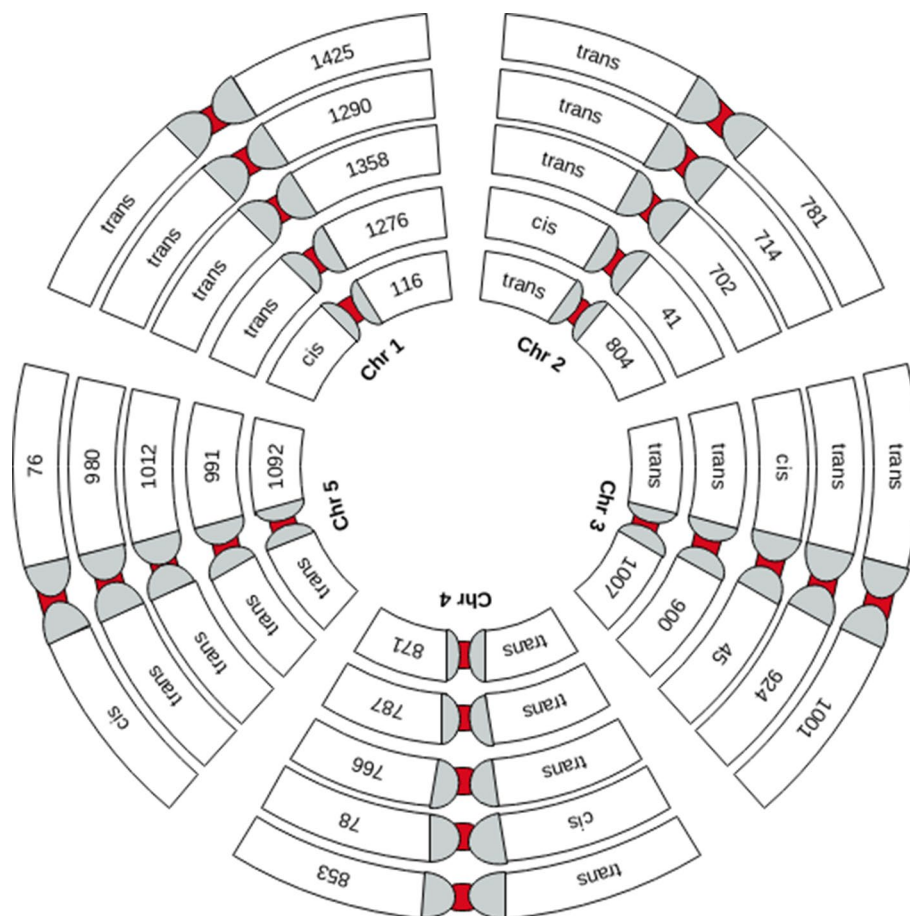


FIGURE 2 | Graphical representation of the chromosome-wide distribution of sQTLs. Each group of five chromosomes (Chr1 to 5 is depicted by small to bigger circles, respectively) shows *cis*-sQTLs on one of the five chromosomes and *trans*-sQTLs on other chromosomes.

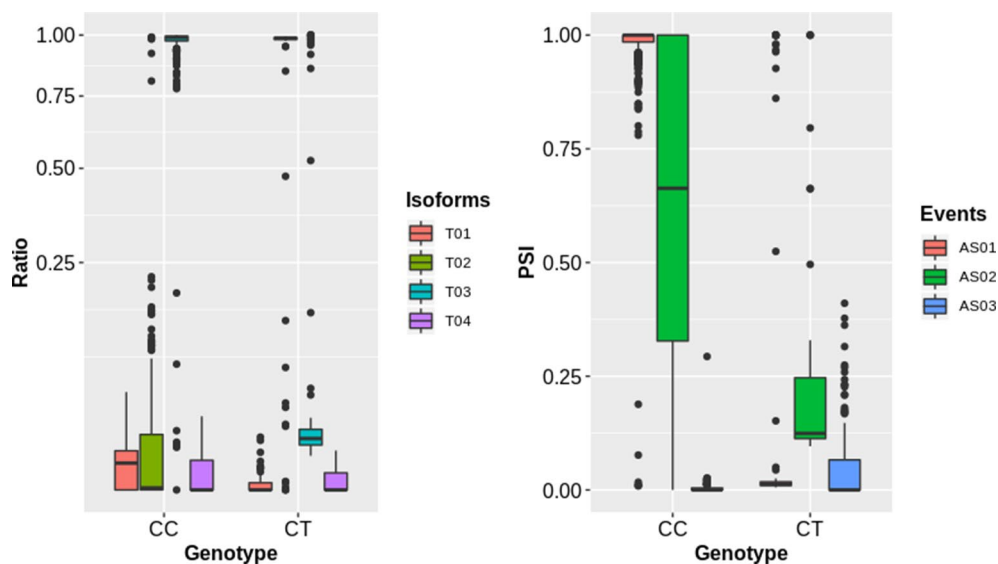


FIGURE 3 | The impact of sQTLs (snp_1_1099063) on splicing isoform ratios and AS events of the gene (AT1G04170). The left panel shows the impact of change in genotype on splicing ratios of transcripts (T01: AT1G04170_c1, T02: AT1G04170_Jc4, T03: AT1G04170_P1, T04: AT1G04170_P2) with sharp change for T01 and T03. The right panel shows the splicing events (AS01, AT1G04170;A3:Chr1:1097120-1097399:1097120-1097405:+, AS02, AT1G04170;A3:Chr1:1097286-1097396:1097286-1097399:+, AS03, AT1G04170;A5:Chr1:1097286-1097399:1097120-1097399:+) with significant change for AS01 and AS02 event.

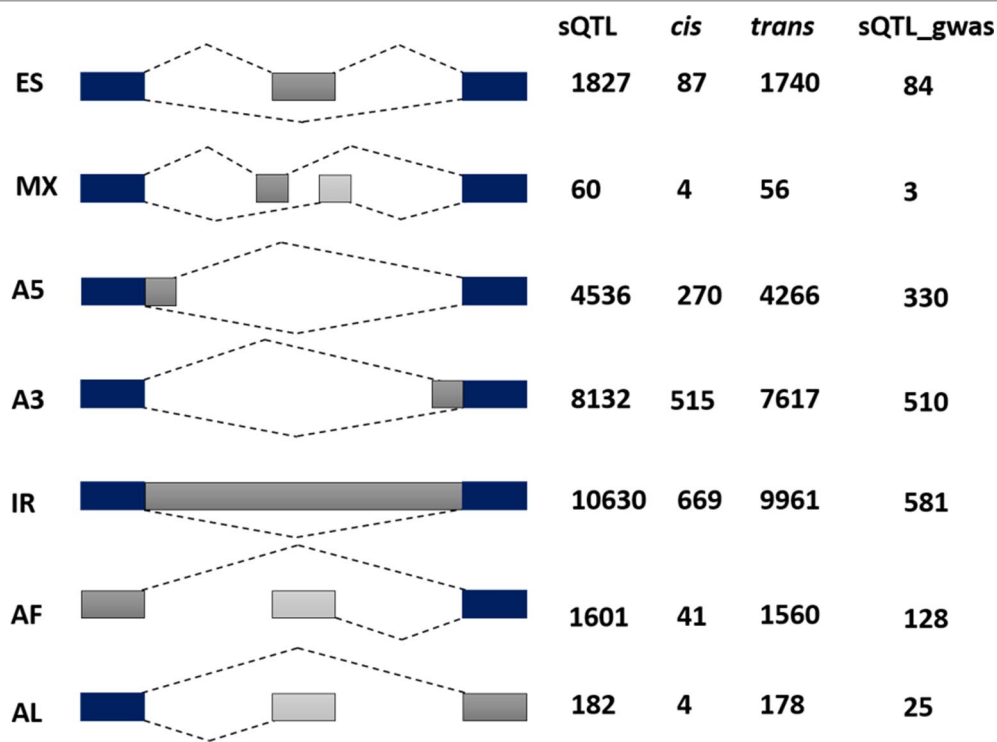


FIGURE 4 | Alternative splicing (AS) categorization of 6129 significant sQTL cognate genes possessing 26968 AS events. Among these 1590 are associated with *cis* and 25378 with the *trans* category. The last column reflects AS pattern of 757 sQTL-GWAS cognate genes possessing 1661 AS events. Among AS categories (exon skipping (ES), mutually exclusive exons (MX), Alternative 5'/3' splice-site (A5/A3), Intron retention (IR), Alternative First/Last Exons (AF/AL), IR are the most common, whereas mutually exclusive exons were the least frequent type of alternative splicing.

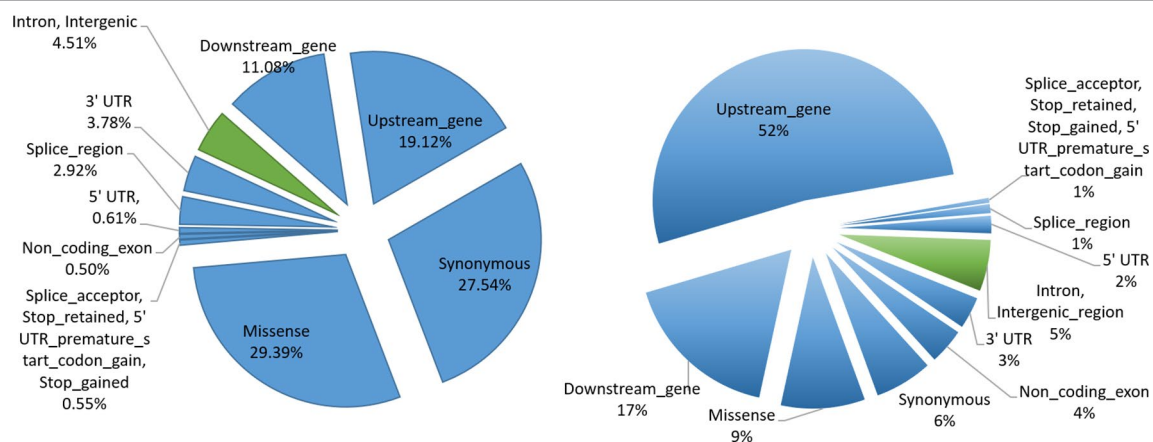


FIGURE 5 | sQTL functional characterization. Pie charts indicating proportions of SNPs annotated with each functional category. SNPs in exonic and non-exonic regions are indicated by bluish and greenish colors, respectively. The left shows the functional categorization of sQTL SNPs, whereas the right panel depicts the functional categorization of non-sQTL SNPs.

were found in intronic regions. Furthermore, we have analysed the distribution of *cis* and *trans*-sQTLs along with the exonic locations. Of the 356 *cis*-sQTLs, significantly more (342) reside in exonic locations (hypergeometric *p*-value ≤ 0.05). Similarly, significantly more (5,775 out of 6,050) *trans*-sQTLs localized in exonic regions (hypergeometric *p*-value ≤ 0.05).

Biological Significance of sQTL Cognate Genes

Functional annotation analysis of 6,129 sQTL cognate genes was performed using DAVID (Huang et al., 2009) (**Figure 6; Supplementary Table 6**). The statistically significant gene enrichment terms were filtered based on FDR ≤ 0.05 to illuminate

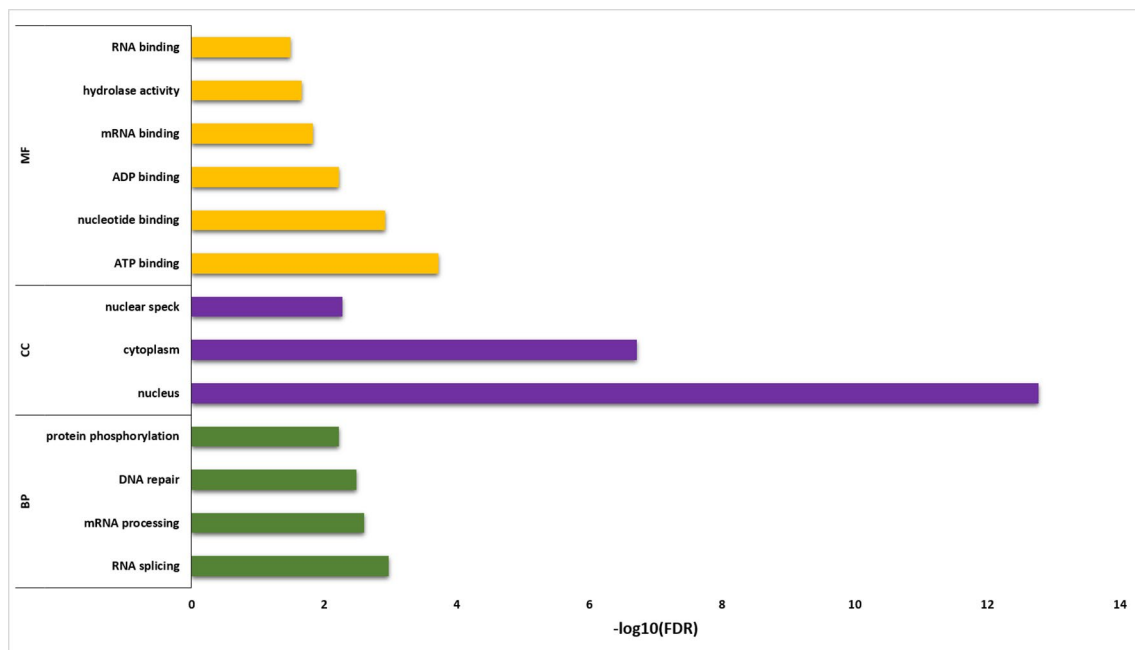


FIGURE 6 | Gene enrichment analysis for sQTL-associated genes. Y-axis shows the gene enrichment categories, whereas X-axis illustrates $-\log_{10}(\text{FDR})$. BP, CC, and MF represent a biological process, cellular component, and molecular function, respectively.

the momentous role of sQTL cognate genes in diverse biological processes, cellular components, and molecular function. The involvement of sQTL cognate genes in complex biological processes like RNA splicing/processing shows its tremendous potential to modulate the transcriptomic architecture and ability of sQTLs to affect genes associated with DNA repair speculates its critical involvement in genome stability. Furthermore, sQTL association with post-translational modifications in the shape of protein phosphorylation implies their proteome regulatory role that can generate phenotypic diversity.

The presence of sQTL cognate genes in vital cellular components (nucleus, cytoplasm, nuclear speck) and its association with significant binding (RNA, mRNA, ATP, ADP) and catalysis (hydrolase activity) molecular functions highlights its involvement in integral cellular processing that can help in a plant adaptation to the microclimatic conditions.

Genome Regulatory Role of sQTL

sQTL enrichment within annotated genome regulatory regions was analyzed as mapped sQTLs were spread across the genome and can be enriched among various genome regulatory elements. Moreover, transcription factors (TFs) are important genome regulators as they can mediate transcription by binding in the upstream region of their target genes (Jin et al., 2017). Therefore, the list of TFs from the Plant Transcription Factor Database (PlantTFDB) v4.0 (Jin et al., 2017) was downloaded and showed that sQTLs for 389 genes overlapped with 51 TF families.

Furthermore, the binding of regulatory proteins to *cis*-regulatory DNA elements (CREs) can orchestrate gene expression.

DNase I hypersensitive sites (DHSs) are significantly enriched in CREs that provide chromatin accessibility to regulatory proteins. The DHS sites and nucleosome positioning/occupancy for Arabidopsis were downloaded from PlantDHS database (Zhang et al., 2016) and revealed a significant overlap of sQTLs with CRE enriched regions. The leaf and flower tissue nucleosome positioning data of Arabidopsis was used to see the frequency of sQTLs that reside in nucleosome enriched regions and revealed that 462 sQTLs are flowering specific, 399 are leaf specific, and 4,962 are shared between both tissues. The list of 395 *A. thaliana* splicing-related genes from SRGD (splicing-related gene database) was downloaded to interrogate any overlap with non-redundant sQTL cognate genes and found 128 (7 *cis*- and 121 *trans*-associated) overlapping with splicing-related genes (Supplementary Figure 2; Supplementary Table 7). Among these splicing-related genes (128), the highest number was found on chromosome one (31) and least number on chromosome two (18), which also relates to chromosome size. The overlapped splicing-related genes (128) were associated with 2,397 sQTLs (~37% of overall sQTLs), which show their tremendous potential to serve as significant genome regulatory elements.

sQTLs Are Enriched Among Stress Responsive, Clock, and Flowering Genes

In order to understand the underlying dynamics of sQTLs with the spatial distribution of 666 accessions worldwide, we analyzed three highly significant gene functional categories (stress response, flowering, and circadian clock) among sQTL associated genes. The three categories are intimately associated with each other and confer

adaptation to different climatic regions. Towards this goal, a list of 3150 Arabidopsis stress-responsive genes was downloaded from STIFDB2 database (Naika et al., 2013). In total, 742 stress-responsive genes associated with significant sQTLs were identified, highlighting the potential of AS in the plant stress response mechanism. Next, we downloaded a list of 346 flowering genes (306 flowering time and 46 flowering development genes) from FLOR-ID database (Bouché et al., 2016) and identified 122 genes that were associated with sQTLs. Similarly, out of 28 core clock genes, 16 were found to be associated with sQTLs (**Supplementary Figure 3A**).

Interestingly, we found six common genes (**Table 2**; **Supplementary Figure 3B**) between the three groups (circadian, flowering, and stress) that were associated with sQTLs. Besides core clock components like *circadian clock-associated 1* (*CCA1*), *late elongated hypocotyl* (*LHY*), *timing of cab expression 1* (*TOC1*), and pseudo response regulator 7 (*PRR7*), we found *phytochrome interacting factor 5* (*PIF5*) and *b-box domain protein 19* (*BBX19*) genes, which are associated with light-sensing, flowering, and photomorphogenesis, respectively (Suárez-López et al., 2001; Hayama and Coupland, 2003; Somers et al., 2004; Wang et al., 2014; Greenham and McClung, 2015; Wang et al., 2015; Wang and Dehesh, 2015; Nasim et al., 2017;). Since circadian clock and *PIFs* (*PIF 4/5* and others) influence the timing of flowering by modulating the expression of flowering-related genes, all accessions were divided on the basis of the timing of flowering at 10°C and 16°C (**Supplementary Figure 4**). Clock genes not only control global transcription patterns but their transcript abundance is also modulated by AS (James et al., 2012). Therefore, the mean expression of the aforementioned six genes revealed whether the expression of these genes is associated with the flowering time at 10°C and 16°C. Interestingly, expression of *LHY*, *CCA1*, and *PIF5* is significantly higher for plants growing at 16°C and is accompanied by the low expression of *TOC1* and *BBX19* (flowering repressor) among accessions that flower between 51 and 60 days. The relationship between the expression of these genes and flowering between 61 and 110 days is not straightforward; however, the expression of *PIF5* tremendously increased among accessions that flower late and is accompanied by a lower level of the expression of *BBX19*, which represses flowering by sequestering the *Flowering Time* (*FT*) gene (Wang et al., 2014). On the contrary, the expression of *PIF5* is generally higher among accessions flowering between 51 and 110 days at 16°C; however, late-flowering (131–160 days) accessions show a dramatic increase in the expression of *LHY*. Since most of the late flowering accessions are from Sweden, photoperiod-dependent flowering regulation *via LHY* is more pronounced (Park et al., 2016). Overall, these results indicate that clock, *PIF5*,

and *BXX19* gene mediated flowering time is important for diverse accessions to occupy different geographical regions and many sQTLs mediate these responses.

DISCUSSION

In this study, we analyzed population-scale transcriptomic and genotypic data of highly diverse 666 *A. thaliana* accessions to comprehensively identify the genomic regions regulating splicing (sQTLs). We used the AtRTD2 database as well as our own genomic assemblies to map sQTLs and found a significant overlap between the two approaches. Since there is significant overlap and the AtRTD2 database is highly validated and non-redundant, we suggest following the transcriptomic approach for sQTL mapping in Arabidopsis and other species where accurate datasets are available. Furthermore, using available transcriptomic datasets and Salmon based approaches are rapid and give reliable results without creating own transcriptome assemblies against the genome. The sQTLs based on the transcriptomic approach are spread genome-wide; however, their frequency varies across chromosomes. The chromosomal distribution relative to the genomic location of their cognate genes (*cis* or *trans*) showed that chromosome one harbors the highest number of *trans*-sQTL hotspots. The chromosomal distribution relative to the genomic location of their cognate genes (*cis* or *trans*) showed that chromosome one harbors the highest number of *trans*-sQTL hotspots. Among the top five associations, two of them reside on chromosome one and belongs to *cis* (snp_1_1099063; gene *AT1G04170*) and *trans* (snp_1_10688832; gene *AT1G30320*) categories; this strengthens our observation and reflects the immense potential of sQTLs to contribute towards genome complexity and proteome diversity. In order to gain more insight into the biological role and genetic control of splicing, we co-localized the sQTLs with Arabidopsis GWAS hits to illuminate their association with different trait-associated loci and their potential to modulate plants phenotypic variability (Yoo et al., 2016). Functional analysis of sQTLs showed that these genetic variants are significantly localized within exonic regions. Among the exonic regions variants, we discovered a high proportion of missense variants, which can modulate the structure and function of proteins. However, functional annotation painted a different picture by showing the moderate effect of missense variants but the high impact of stop gained and splice acceptor variants, which can modulate AS patterns and proteome diversity (Wachsman et al., 2017).

Although all classes of AS contributed to the sQTLs, intron retention (IR) was prevalent than any other AS type. This is consistent with the previous observations of IR as the most common class of AS and a well-established mechanism for regulating gene expression in plants (Syed et al., 2012; Reddy et al., 2013). However, how IR contribute and/or modulates proteome complexity and the extent to which IR sQTLs influence phenotypic variability remains obscure (Chaudhary et al., 2019). Although IR is associated with nonsense-mediated decay (NMD), this is by no means the only consequence as transcripts can still escape NMD *via* sequestration in the nucleus and some are preferentially recruited to ribosomes (Palusa and Reddy, 2010; Kalyna et al., 2012; Marquez et al., 2012). While our analysis showed that the majority of sQTLs are associated

TABLE 2 | Association of six genes with sQTL-SNPs and GWAS-SNPs.

Genes ID	Gene name	Association with sQTL	Association with GWAS SNPs
AT5G02810	PRR7	344	18
AT3G59060	PIF5	44	10
AT1G01060	LHY	38	18
AT4G38960	BBX19	38	10
AT2G46830	CCA1	2	13
AT5G61380	TOC1	2	5

with IR events, most SNPs were localized within exonic regions. This is conceivable because only 356 *cis* sQTLs fall in this category and may have a subtle effect on IR in the *cis*-regulatory context. It is plausible that exonic variants cause changes in the pre-mRNA secondary structure, which impacts spliceosome recognition of exon–exon junctions (McManus and Graveley, 2011).

We interrogated the relationship between sQTL associated genes and annotated genome regulatory proteins and found a number of TF families linked to sQTLs. Although experimental validation is required to understand the regulation of TFs *via* AS, these results elaborate the potential of splicing as a mechanism for regulating gene expression (Nasim et al., 2017; Takata et al., 2017). Similarly, we also found a strong overlap between sQTLs, CREs, and nucleosome occupancy, especially among flower and leaf specific genes. Since nucleosome occupancy is much higher in exons, it helps to define intron–exon definition (Schwartz et al., 2009) to orchestrate an appropriate splicing response under variable environmental conditions. In addition, nucleosome occupancy has a strong influence on RNA polymerase II processing as its speed tends to be higher in regions with more open chromatin structure and Pol II speed regulates AS (Ullah et al., 2018; Godoy Herz et al., 2019). It is therefore not surprising that many sQTLs are enriched in the DHS site and have a strong association with regions with higher nucleosome occupancy among important life-history traits like leaf and flower. This data also indicates that many of the important traits like flowering are regulated *via* epigenetic means, and this is reminiscent of the downregulation of the *flowering* repressor locus *c* (*FLC*) that regulates flowering in a cold-dependent manner (Michaels and Amasino, 1999; Shindo et al., 2006).

Functional characterization of sQTLs among stress-responsive, circadian clock, and flowering genes revealed six common genes (*CCA1*, *LHY*, *TOC1*, *PRR7*, *PIF5*, and *BXX19*) that are shared in these categories (**Supplementary Figure 5, Supplementary Table 8**). Among these six genes, *PRR7* (*AT5G02810*) showed high significance in the sQTL analysis by accruing second rank (**Supplementary Dataset 2**). *PRR7* was impacted by the trans-QTL (*snp_5_626009*) present on chromosome 5 at position 626009, and this SNP modulated the splicing ratios and affected the AS patterns. We are aware that co-localisation of sQTLs in these categories would need further validation but we hope that it would provide an interesting starting point and a list of useful genes that may be regulated *via* alternative splicing. It is well known that the circadian clock plays a vital role in the normal functioning of plants and is intimately associated with carbon fixation during the day and starch mobilization to promote growth during the night (Dodd et al., 2005; Graf et al., 2010). Rhythmicity of clock components is not only associated with appropriate growth responses under variable and often stressful conditions but also promotes fitness and adaptive responses in plants (Dodd et al., 2005). Interestingly, down-regulation of *LHY* and *CCA1* around noon time among Arabidopsis hybrids and polyploids promotes heterosis *via* upregulation of chlorophyll, starch synthesis, and metabolism genes (Ni et al., 2009). Intriguingly, *1-aminocyclopropane-1-carboxylate synthase* (*ACS*, a rate-limiting enzyme in ethylene synthesis) is also downregulated by *CCA1* and phytochrome-interacting factor 5 (*PIF5*) during the

day and night, respectively, to promote heterosis in Arabidopsis (Song et al., 2018). Further, *LHY* and *CCA1* have been implicated in cold temperature acclimation responses and may also be important under higher temperature and stressful temperature conditions. *CCA1* and *LHY* are partially redundant but show different expression and splicing patterns under cold conditions (Calixto et al., 2018). Similarly, *PRR7* has been associated with temperature responses in Arabidopsis and also shows different splicing patterns under normal and cold conditions (James et al., 2012). Recent experimental and modeling data suggest that *TOC1*, in concert with *ABA* levels, plays the role of an environmental sensor that coordinates the pace of the central oscillator to affect downstream processes (Pokhilko et al., 2013). Mechanistic details of *TOC1* mediated drought responses were revealed by analyzing the relationship between *TOC1* and an *ABA*-related gene (*ABAR/CHLH/GUN5*) (Legnaioli et al., 2009; Castells et al., 2010). Under drought conditions and elevated *ABA* levels, *TOC1* binds the *ABAR* promoter and modulates its circadian expression, resulting in clock-dependent gating of *ABA* function and drought tolerance (Legnaioli et al., 2009). Recent findings revealed that *LHY* modulates the expression of many genes involved in *ABA* signaling pathway to fine-tune plant performance under drought and osmotic stress conditions. However, *LHY* also maintained seed germination and plant growth *via* alleviation of the inhibitory effect of *ABA* (Adams et al., 2018).

The timing of flowering is crucial for the survival and adaptation of diverse Arabidopsis accessions in different geographical regions of the world. It is well known that the timing of flowering has a huge impact on seed set, grain filling, maturation, and overall productivity. Intriguingly, plants can also speed up their flowering in response to various environmental fluctuations and stresses, presumably as a survival strategy to produce seeds as quickly as possible (Wang and Dehesh, 2015). It is not surprising that one of the common genes between clock, flowering, and stress-related genes is *BBX19*. This gene has been shown to work as a circadian clock output and downregulates *CONSTANS* (*CO*) and precisely times the expression of *flowering* locus *T* (*FT*) in a day-length dependent manner to orchestrate appropriate flowering time (Wang et al., 2015). Interestingly, E3 ubiquitin ligase *constitutive photomorphogenic 1* (*COP1*), *early flowering 3* (*ELF3*), *phytochrome-interacting factor 4* (*PIF4*), and *PIF5* also influence *BBX19* function to mediate photomorphogenic responses in Arabidopsis (Wang et al., 2015). Furthermore, expression of *BBX19* is significantly reduced as a result of high levels of *methylerythritol cyclodiphosphate* (*MEcPP*), which is a plastidial isoprenoid intermediate that also functions as a stress-response retrograde signal to orchestrate appropriate transcriptional response (Xiao et al., 2012; Wang and Dehesh, 2015). Therefore, the *BBX19* gene provides the role of a flowering checkpoint that links a stress-specific retrograde signal (*MEcPP*) *via* sequestering the active *CO* gene, which is essential for *FT* transcription to promote flowering (Wang and Dehesh, 2015). Higher expression of *BBX19* delays flowering; however, its expression is positively correlated with *PIF5* expression to promote hypocotyl growth. Taken together, *BBX19* plays a dual role to modulate plant growth and stress-responsive flowering (Wang et al., 2015). Therefore, it is not surprising that

many sQTL SNPs are associated with these genes and may be important for defining phenotypes and underlying genotypes among geographically diverse lines of Arabidopsis. Since *BBX19* controls flowering in a clock-regulated and stress-dependent manner, we propose that expression and sQTL patterns of *BBX19* may have a bearing on the flowering patterns of 666 accessions from different geographical regions of the world and play a role in adaptive responses. Recent evidence also shows that SNPs are present at almost every 200 bp in different ecotypes of Arabidopsis and can alter genomic architecture to affect splicing efficiency, gene expression pattern, and phenotypic diversity (Gan et al., 2011). We envisage that the presence of widespread variation in diverse ecotypes of Arabidopsis and our sQTL analysis would provide a solid platform for sQTL discovery and their influence on phenotypic traits in the future.

DATA AVAILABILITY STATEMENT

Publicly available datasets were analyzed in this study. Genotype data can be found here: <https://1001genomes.org/data/GMI-MPI/releases/v3.1/> and RNA-Seq data can be found here: <http://signal.salk.edu/1001.php>

AUTHOR CONTRIBUTIONS

NS and WK conceived the study, WK performed most of the analysis, and MH, AR, SC, and IJ contributed to it. All authors contributed toward preparing the manuscript. WK and MH contributed equally.

ACKNOWLEDGMENTS

We thank the funding agencies for research support: Leverhulme Trust [RPG-2016-014]; DOE Office of Science, Office of Biological and Environmental Research (BER) [DE-SC0010733]; National Science Foundation; and the US Department of Agriculture (ASNR). MH is supported by a University of Edinburgh Chancellor's fellowship and in part by Bill and Melinda Gates Foundation and with UK aid from the UK Government Department for International Development (Grant Agreement OPP1127286)

REFERENCES

- Adams, S., Grundy, J., Veflingstad, S. R., Dyer, N. P., Hannah, M. A., Ott, S., et al. (2018). Circadian control of abscisic acid biosynthesis and signalling pathways revealed by genome-wide analysis of LHY binding targets. *New Phytol.* 220 (3), 893–907. doi: 10.1111/nph.15415
- Alamancos, G. P., Pagès, A., Trincado, J. L., Bellora, N., and Eyra, E. (2015). Leveraging transcript quantification for fast computation of alternative splicing profiles. *RNA* 21, 1521–1531. doi: 10.1261/rna.051557.115
- Alamancos, G., Pagès, A., Trincado, J., and Bellora, N. (2014). SUPPA: a super-fast pipeline for alternative splicing analysis from RNA-Seq. *bioRxiv* 008763. doi: 10.1101/008763
- Alonso-Blanco, C., Andrade, J., Becker, C., Bemm, E., Bergelson, J., Borgwardt, K. M., et al. (2016). 1,135 Genomes reveal the global pattern of polymorphism in Arabidopsis thaliana. *Cell* 166, 481–491. doi: 10.1016/j.cell.2016.05.063

under the auspices of the Centre for Tropical Livestock Genetics and Health, established jointly by the University of Edinburgh, Scotland's Rural College, and the International Livestock Research Institute. The findings and conclusions contained within are those of the authors and do not necessarily reflect positions or policies of Bill & Melinda Gates Foundation nor the UK Government. The Roslin Institute receives institute strategic programme funds from Biotechnology and Biological Sciences Research Council (BBSRC).

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fpls.2019.01160/full#supplementary-material>

SUPPLEMENTARY FIGURE 1 | Complementarity of the genome and transcriptome-based approaches for sQTL analysis.

SUPPLEMENTARY FIGURE 2 | Summary of sQTL cognate genes overlapping with splicing related genes. The x-axis shows the type of splicing related genes and Y-axis exhibits the number of input and matched splicing related genes.

SUPPLEMENTARY FIGURE 3 | Functional characterization of genes associated with sQTL. (A) sQTL cognate genes were highly enriched in different core functional categories (e.g. stress response, flowering). (B) Six genes were shared between different functional categories.

SUPPLEMENTARY FIGURE 4 | Phenotypic association of six genes with flowering time. The x-axis shows days of flowering at 10 °C (A) and 16 °C (B) and Y-axis show the average/relative gene expression value for six genes across a diverse set of 666 A. thaliana accessions.

SUPPLEMENTARY FIGURE 5 | The impact of sQTLs on splicing isoform ratios and AS events of six genes. The left panel shows the impact of change in genotype on splicing ratios of transcripts and the right panel shows the splicing events. For a detailed description of all isoforms and transcripts (see **Supplementary Table 8**).

SUPPLEMENTARY DATASET 1 | RNA-Sequencing Analysis using genome as a reference

SUPPLEMENTARY DATASET 2 | (Splicing ratios of sQTL cognate Genes): <https://figshare.com/s/4dbd07a41f0fd4c6988c>

SUPPLEMENTARY DATASET 3 | (Percent Spliced In (PSI) of sQTL cognate Genes): <https://figshare.com/s/89a7f23e21d7de9643c3>

SUPPLEMENTAL DATASET 4 | (Stringtie merged assembly): <https://figshare.com/s/fd1a95af9e2b3025b8ff>

- Andrews, S. (2010). FastQC: a quality control tool for high throughput sequence data.
- Bouché, F., Lobet, G., Tocquin, P., and Périlleux, C. (2016). FLOR-ID: an interactive database of flowering-time gene networks in Arabidopsis thaliana. *Nucleic Acids Res.* 44, D1167–D1171. doi: 10.1093/nar/gkv1054
- Calixto, C. P. G., Guo, W., James, A. B., Tzioutziou, N. A., Entizne, J. C., Panter, P. E., et al. (2018). Rapid and dynamic alternative splicing impacts the Arabidopsis cold response transcriptome. *Plant Cell* 30, 1424–1444. doi: 10.1105/tpc.18.00177
- Castells, E., Portolés, S., Huang, W., and Mas, P. (2010). A functional connection between the clock component TOC1 and abscisic acid signaling pathways. *Plant Signal. Behav.* 5 (4), 409–411. doi: 10.4161/psb.5.4.11213
- Chaudhary, S., Jabre, I., Reddy, A. S. N., Staiger, D., and Syed, N. H. (2019). Perspective on alternative splicing and proteome complexity in plants. *Trends Plant Sci.* xx, 1–11. doi: 10.1016/j.tplants.2019.02.006
- Chen, Q., Han, Y., Liu, H., Wang, X., Sun, J., Zhao, B., et al. (2018). Genome-wide association analyses reveal the importance of alternative splicing in

- diversifying gene function and regulating phenotypic variation in maize. *Plant Cell* 30, 1404–1423. doi: 10.1105/tpc.18.00109
- Dobin, A., Davis, C. A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., et al. (2013). STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* 29, 15–21. doi: 10.1093/bioinformatics/bts635
- Dodd, A. N., Toth, R., Kevei, E., Webb, A. A. R., Salathia, N., Hibberd, J. M., et al. (2005). Plant circadian clocks increase photosynthesis, growth, survival, and competitive advantage. *Science* 330 (2005), 630–633. doi: 10.1126/science.1115581
- Feng, J., Li, J., Gao, Z., Lu, Y., Yu, J., Zheng, Q., et al. (2015). SKIP Confers Osmotic Tolerance during salt stress by controlling alternative gene splicing in Arabidopsis. *Mol. Plant* 8, 1038–1052. doi: 10.1016/j.molp.2015.01.011
- Filichkin, S. A., Cumbie, J. S., Dharmawardhana, P., Jaiswal, P., Chang, J. H., Palusa, S. G., et al. (2015). Environmental stresses modulate abundance and timing of alternatively spliced circadian transcripts in Arabidopsis. *Mol. Plant* 8, 207–227. doi: 10.1016/j.molp.2014.10.011
- Filichkin, S. A., Hamilton, M., Dharmawardhana, P. D., Singh, S. K., Sullivan, C., Ben-Hur, A., et al. (2018). Abiotic stresses modulate landscape of poplar transcriptome via alternative splicing, differential intron retention, and isoform ratio switching. *Front. Plant Sci.* 9, 1–17. doi: 10.3389/fpls.2018.00005
- Filichkin, S. A., Priest, H. D., Givan, S. A., Shen, R., Bryant, D. W., Fox, S. E., et al. (2010). Genome-wide mapping of alternative splicing in Arabidopsis thaliana. *Gen. Res.* 20 (1), 45–58. doi: 10.1101/gr.093302.109
- Frachon, L., Bartoli, C., Carrère, S., Bouchez, O., Chaubet, A., Gautier, M., et al. (2018). A Genomic Map of Climate Adaptation in *Arabidopsis thaliana* at a Micro-Geographic Scale. *Front. Plant Sci.* 9, 967. doi: 10.3389/fpls.2018.00967
- François, O., Blum, M. G. B., Jakobsson, M., and Rosenberg, N. A. (2008). Demographic history of European populations of *Arabidopsis thaliana*. *PLoS Genet.* 4 (5), e1000075. doi: 10.1371/journal.pgen.1000075
- Gan, X., Stegle, O., Behr, J., Steffen, J. G., and Drewe, P., Hildebrand, K. L., et al. (2011). Multiple reference genomes and transcriptomes for *Arabidopsis thaliana*. *Nature* 477 (7365), 419–423. doi: 10.1038/nature10414
- Godoy Herz, M. A., Kubaczka, M. G., Brzyżek, G., Servi, L., Krzyszton, M., Simpson, C., et al. (2019). Light regulates plant alternative splicing through the control of transcriptional elongation. *Mol. Cell* 73 (5), 1066–1074.e3. doi: 10.1016/j.molcel.2018.12.005
- Graf, A., Schlereth, A., Stitt, M., and Smith, A. M. (2010). Circadian control of carbohydrate availability for growth in Arabidopsis plants at night. *Proc. Natl. Acad. Sci.* 107 (20), 9458–9463. doi: 10.1073/pnas.0914299107
- Greenham, K., and McClung, C. R. (2015). Integrating circadian dynamics with physiological processes in plants. *Nat. Rev. Genet.* 16(10), 598–610. doi: 10.1038/nrg3976
- Han, S., Jung, H., Lee, K., Kim, H., and Kim, S. (2017). Genome wide discovery of genetic variants affecting alternative splicing patterns in human using bioinformatics method. *Genes Genomics* 39, 453–459. doi: 10.1007/s13258-016-0466-7
- Hartmann, L., Wiefner, T., and Wachter, A. (2018). Subcellular compartmentation of alternatively-spliced transcripts defines SERINE/ARGININE-RICH PROTEIN 30 expression. *Plant Physiol.* 176, 2886–2903. doi: 10.1104/pp.17.01260
- Hayama, R., and Coupland, G. (2003). Shedding light on the circadian clock and the photoperiodic control of flowering. *Curr. Opin. Plant Biol.* 6 (1), 13–19. doi: 10.1016/S1369-5266(02)00011-0
- Huang, D. W., Sherman, B. T., and Lempicki, R. A. (2009). Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat. Protoc.* 4, 44–57. doi: 10.1038/nprot.2008.211
- Huang, J., Gu, L., Zhang, Y., Yan, T., Kong, G., Kong, L., et al. (2017). An oomycete plant pathogen reprograms host pre-mRNA splicing to subvert immunity. *Nat. Commun.* 8, 2051. doi: 10.1038/s41467-017-02233-5
- Iida, K., and Go, M. (2006). Survey of conserved alternative splicing events of mRNAs encoding SR proteins in land plants. *Mol. Biol. Evol.* 23, 1085–1094. doi: 10.1093/molbev/msj118
- Jabre, I., Reddy, A. S. N., Kalyna, M., Chaudhary, S., Khokhar, W., Byrne, L. J., et al. (2019). Does co-transcriptional regulation of alternative splicing mediate plant stress responses? *Nucleic Acids Res.* 47 (6), 2716–2726. doi: 10.1093/nar/gkz121
- James, A. B., Syed, N. H., Bordage, S., Marshall, J., Nimmo, G. A., Jenkins, G. I., et al. (2012). Alternative splicing mediates responses of the Arabidopsis circadian clock to temperature changes. *Plant Cell.* 24 (3), 961–981. doi: 10.1105/tpc.111.093948
- Jiang, J., Liu, X., Liu, C., Liu, G., Li, S., and Wang, L. (2017). Integrating omics and alternative splicing reveals insights into grape response to high temperature. *Plant Physiol.* 173, 1502–1518. doi: 10.1104/pp.16.01305
- Jin, J., Tian, F., Yang, D.-C., Meng, Y.-Q., Kong, L., Luo, J., et al. (2017). PlantTFDB 4.0: toward a central hub for transcription factors and regulatory interactions in plants. *Nucleic Acids Res.* 45, D1040–D1045. doi: 10.1093/nar/gkw982
- Kalyna, M., Lopato, S., Voronin, V., and Barta, A. (2006). Evolutionary conservation and regulation of particular alternative splicing events in plant SR proteins. *Nucleic Acids Res.* 34 (16), 4395–4405. doi: 10.1093/nar/gkl570
- Kalyna, M., Simpson, C. G., Syed, N. H., Lewandowska, D., Marquez, Y., Kusenda, B., et al. (2012). Alternative splicing and nonsense-mediated decay modulate expression of important regulatory genes in Arabidopsis. *Nucleic Acids Res.* 40 (6), 2454–2469. doi: 10.1093/nar/gkr932
- Kawakatsu, T., Huang, S. C., Jupe, F., Sasaki, E., Schmitz, R. J., Urlich, M. A., et al. (2016). Epigenomic diversity in a global collection of *Arabidopsis thaliana* accessions. *Cell* 166, 492–505. doi: 10.1016/j.cell.2016.06.044
- Klepikova, A. V., Kasianov, A. S., Gerasimov, E. S., Logacheva, M. D., and Penin, A. A. (2016). A high resolution map of the *Arabidopsis thaliana* developmental transcriptome based on RNA-seq profiling. *Plant J.* 88, 1058–1070. doi: 10.1111/tjp.13312
- Krueger, F. (2015). Trim galore!: a wrapper tool around cutadapt and FastQC to consistently apply quality and adapter trimming to FastQ files.
- Lee, A., Lee, S. S., Jung, W. Y., Park, H. J., Lim, B. R., Kim, H. S., et al. (2016). The OsCYP19-4 gene is expressed as multiple alternatively spliced transcripts encoding Isoforms with distinct cellular localizations and PPLase activities under cold stress. *Int. J. Mol. Sci.* 17, 1154. doi: 10.3390/ijms17071154
- Lee, C.-R., Svardal, H., Farlow, A., Exposito-Alonso, M., Ding, W., Novikova, P., et al. (2017). On the post-glacial spread of human commensal *Arabidopsis thaliana*. *Nat. Commun.* 8, 14458. doi: 10.1038/ncomms14458
- Legnaioli, T., Cuevas, J., and Mas, P. (2009). TOC1 functions as a molecular switch connecting the circadian clock with plant responses to drought. *EMBO J.* 28 (23), 3745–3757. doi: 10.1038/emboj.2009.297
- Liu, J., Sun, N., Liu, M., Liu, J., Du, B., Wang, X., et al. (2013). An autoregulatory loop controlling Arabidopsis HsfA2 expression: role of heat shock-induced alternative splicing. *Plant Physiol.* 162, 512–521. doi: 10.1104/pp.112.205864
- Liu, Z., Qin, J., Tian, X., Xu, S., Wang, Y., Li, H., et al. (2018). Global profiling of alternative splicing landscape responsive to drought, heat and their combination in wheat (*Triticum aestivum* L.). *Plant Biotechnol. J.* 16, 714–726. doi: 10.1111/pbi.12822
- Lutz, U., Posé, D., Pfeifer, M., Gundlach, H., Hagmann, J., Wang, C., et al. (2015). Modulation of ambient temperature-dependent flowering in *Arabidopsis thaliana* by natural variation of FLOWERING LOCUS M. *PLoS Genet.* 11 (10), e1005588. doi: 10.1371/journal.pgen.1005588
- Marquez, Y., Brown, J. W. S., Simpson, C., and Barta, A., and Kalyna, M. (2012). Transcriptome survey reveals increased complexity of the alternative splicing landscape in Arabidopsis. *Genome Res.* 22 (6), 1184–1195. doi: 10.1101/gr.134106.111
- Marquez, Y., Höpfler, M., Ayatollahi, Z., Barta, A., and Kalyna, M. (2015). Unmasking alternative splicing inside protein-coding exons defines exons and their role in proteome plasticity. *Genome Res.* 25, 995–1007. doi: 10.1101/gr.186585.114
- McManus, C. J., and Graveley, B. R. (2011). RNA structure and the mechanisms of alternative splicing. *Current Opinion in Genetics & Development* 21 (4), 373–379. doi: 10.1016/j.gde.2011.04.001
- Michaels, S. D., and Amasino, R. M. (1999). FLOWERING LOCUS C encodes a novel MADS domain protein that acts as a repressor of flowering. *Plant Cell Online.* 11 (5), 949–956. doi: 10.1105/tpc.11.5.949
- Monlong, J., Calvo, M., Ferreira, P. G., and Guigó, R. (2014). Identification of genetic variants associated with alternative splicing using sQTLseeker. *Nat. Commun.* 5, 4698. doi: 10.1038/ncomms5698
- Naika, M., Shameer, K., Mathew, O. K., Gowda, R., and Sowdhamini, R. (2013). STIFDB2: an updated version of plant stress-responsive transcription factor database with additional stress signals, stress-responsive transcription factor binding sites and stress-responsive genes in Arabidopsis and rice. *Plant Cell Physiol.* 54, e8–e8. doi: 10.1093/pcp/pcs185
- Nasim, Z., Fahim, M., and Ahn, J. H. (2017). Possible role of MADS AFFECTING FLOWERING 3 and B-BOX DOMAIN PROTEIN 19 in flowering time

- regulation of Arabidopsis mutants with defects in nonsense-mediated mRNA decay. *Front. Plant Sci.* 8, 191. doi: 10.3389/fpls.2017.00191
- Ni, Z., Kim, E. D., Ha, M., Lackey, E., Liu, J., Zhang, Y., et al. (2009). Altered circadian rhythms regulate growth vigour in hybrids and allopolyploids. *Nature* 457 (7227), 327–331. doi: 10.1038/nature07523
- Pajoro, A., Severing, E., Angenent, G. C., and Immink, R. G. H. (2017). Histone H3 lysine 36 methylation affects temperature-induced alternative splicing and flowering in plants. *Genome Biol.* 18 (1), 102. doi: 10.1186/s13059-017-1235-x
- Palusa, S. G., Ali, G. S., and Reddy, A. S. N. (2007). Alternative splicing of pre-mRNAs of Arabidopsis serine/arginine-rich proteins: regulation by hormones and stresses. *Plant J.* 49, 1091–1107. doi: 10.1111/j.1365-313X.2006.03020.x
- Palusa, S. G., and Reddy, A. S. N. (2010). Extensive coupling of alternative splicing of pre-mRNAs of serine/arginine (SR) genes with nonsense-mediated decay. *New Phytol.* 185, 83–89. doi: 10.1111/j.1469-8137.2009.03065.x
- Park, E., Pan, Z., Zhang, Z., Lin, L., and Xing, Y. (2018). The expanding landscape of alternative splicing variation in human populations. *Am. J. Hum. Genet.* 102, 11–26. doi: 10.1016/j.ajhg.2017.11.002
- Park, M. J., Kwon, Y. J., Gil, K. E., and Park, C. M. (2016). LATE ELONGATED HYPOCOTYL regulates photoperiodic flowering via the circadian clock in Arabidopsis. *BMC Plant Biol.* 16, 1–12. doi: 10.1186/s12870-016-0810-8
- Patro, R., Duggal, G., Love, M. I., Irizarry, R. A., and Kingsford, C. (2017). Salmon provides fast and bias-aware quantification of transcript expression. *Nat. Methods* 14, 417–419. doi: 10.1038/nmeth.4197
- Pertea, M., Pertea, G. M., Antonescu, C. M., Chang, T.-C., Mendell, J. T., & Salzberg, S. L. (2015). StringTie enables improved reconstruction of a transcriptome from RNA-seq reads. *Nat. Biotechnol.* 33 (3), 290–295. doi: 10.1038/nbt.3122
- Pokhilko, A., Mas, P., and Millar, A. J. (2013). Modelling the widespread effects of TOC1 signalling on the plant circadian clock and its outputs. *BMC Syst. Biol.* 7 (1), 7–23. doi: 10.1186/1752-0509-7-23
- Qu, W., Gurdziel, K., Pique-Regi, R., and Ruden, D. M. (2017). Identification of splicing quantitative trait loci (sQTL) in *Drosophila melanogaster* with developmental lead (Pb2+) exposure. *Front. Genet.* 8, 145. doi: 10.3389/fgenet.2017.00145
- Reddy, A., Marquez, Y., Kalyna, M., and Barta, A. (2013). Complexity of the alternative splicing landscape in plants. *Plant Cell* 25 (10), 3657–3683. doi: 10.1105/tpc.113.117523
- Rhee, S. Y., Beavis, W., Berardini, T. Z., Chen, G., Dixon, D., Doyle, A., et al. (2003). The Arabidopsis Information Resource (TAIR): a model organism database providing a centralized, curated gateway to Arabidopsis biology, research materials and community. *Nucleic Acids Res.* 31, 224–228. doi: 10.1093/nar/gkg076
- Schwartz, S., Meshorer, E., and Ast, G. (2009). Chromatin organization marks exon-intron structure. *Nat. Struct. Mol. Biol.* 16 (9), 990–995. doi: 10.1038/nsmb.1659
- Seaton, D. D., Graf, A., Baerenfaller, K., Stitt, M., Millar, A. J., and Gruissem, W. (2018). Photoperiodic control of the Arabidopsis proteome reveals a translational coincidence mechanism. *Mol. Syst. Biol.* 14, e7962. doi: 10.15252/msb.20177962
- Shen, Y., Zhou, Z., Wang, Z., Li, W., Fang, C., Wu, M., et al. (2014). Global dissection of alternative splicing in paleopolyploid soybean. *Plant Cell* 26, 996–1008. doi: 10.1105/tpc.114.122739
- Shindo, C., Lister, C., Crevillen, P., Nordborg, M., and Dean, C. (2006). Variation in the epigenetic silencing of FLC contributes to natural variation in Arabidopsis vernalization response. *Genes Dev.* 20 (22), 3079–3083. doi: 10.1101/gad.405306
- Singh, R. K., and Cooper, T. A. (2012). Pre-mRNA splicing in disease and therapeutics. *Trends Mol. Med.* 18, 472–482. doi: 10.1016/j.molmed.2012.06.006
- Smith, C. C. R., Tittes, S., Mendieta, J. P., Collier-zans, E., Rowe, H. C., Rieseberg, L. H., et al. (2018). Genetics of alternative splicing evolution during sunflower domestication. *Proc. Natl. Acad. Sci.* 115 (26), 6768–6773. doi: 10.1073/pnas.1803361115
- Somers, D. E., Kim, W.-Y., and Geng, R. (2004). The F-Box protein ZEITLUPE confers dosage-dependent control on the circadian clock, photomorphogenesis, and flowering time. *Plant Cell Online* 16 (3), 769–782. doi: 10.1105/tpc.016808
- Song, Q., Ando, A., Xu, D., Fang, L., Zhang, T., Huq, E., et al. (2018). Diurnal down-regulation of ethylene biosynthesis mediates biomass heterosis. *Proc. Natl. Acad. Sci.* 115 (21), 5606–5611. doi: 10.1073/pnas.1722068115
- Suárez-López, P., Wheatley, K., Robson, F., Onouchi, H., Valverde, F., and Coupland, G. (2001). CONSTANS mediates between the circadian clock and the control of flowering in Arabidopsis. *Nature* 410 (6832), 1116–1120. doi: 10.1038/35074138
- Syed, N. H., Kalyna, M., Marquez, Y., Barta, A., and Brown, J. W. S. (2012). Alternative splicing in plants - coming of age. *Trends Plant Sci.* 17, 616–623. doi: 10.1016/j.tplants.2012.06.001
- Takata, A., Matsumoto, N., and Kato, T. (2017). Genome-wide identification of splicing QTLs in the human brain and their enrichment among schizophrenia-associated loci. *Nat. Commun.* 8, 14519. doi: 10.1038/ncomms14519
- Togninalli, M., Seren, Ü., Meng, D., Fitz, J., Nordborg, M., Weigel, D., et al. (2018). The AraGWAS Catalog: a curated and standardized Arabidopsis thaliana GWAS catalog. *Nucleic Acids Res.* 46, D1150–D1156. doi: 10.1093/nar/gkx954
- Ullah, F., Hamilton, M., Reddy, A. S. N., and Ben-Hur, A. (2018). Exploring the relationship between intron retention and chromatin accessibility in plants. *BMC Genomics* 19, 21. doi: 10.1186/s12864-017-4393-z
- Wachsman, G., Modliszewski, J. L., Valdes, M., and Benfey, P. N. (2017). A SIMPLE pipeline for mapping point mutations. *Plant Physiol.* 174, 1307–1313. doi: 10.1104/pp.17.00415
- Wang, C.-Q., Guthrie, C., Sarmast, M. K., and Dehesh, K. (2014). BBX19 interacts with CONSTANS to repress FLOWERING LOCUS T transcription, defining a flowering time checkpoint in Arabidopsis. *Plant Cell* 26 (9), 3589–3602. doi: 10.1105/tpc.114.130252
- Wang, C.-Q., Sarmast, M. K., Jiang, J., and Dehesh, K. (2015). The transcriptional regulator BBX19 promotes hypocotyl growth by facilitating COP1-mediated EARLY FLOWERING3 degradation in Arabidopsis. *Plant Cell*. doi: 10.1105/tpc.15.00044
- Wang, C., and Dehesh, K. (2015). From retrograde signaling to flowering time. *Plant Signal. Behav.* 10(6), e1022012. doi: 10.1080/15592324.2015.1022012
- Xiao, Y., Savchenko, T., Baidoo, E. E. K., Chehab, W. E., Hayden, D. M., Tolstikov, V., et al. (2012). Retrograde signaling by the plastidial metabolite MeCPP regulates expression of nuclear stress-response genes. *Cell* 149 (7), 1525–1535. doi: 10.1016/j.cell.2012.04.038
- Yang, Q., Hu, Y., Li, J., and Zhang, X. (2017). ulfasQTL: an ultra-fast method of composite splicing QTL analysis. *BMC Genomics* 18, 963. doi: 10.1186/s12864-016-3258-1
- Yoo, W., Kyung, S., Han, S., and Kim, S. (2016). Investigation of splicing quantitative trait loci in Arabidopsis thaliana. *Genomics Inform.* 14, 211–215. doi: 10.5808/GI.2016.14.4.211
- Zhang, P., Deng, H., Xiao, F. M., and Liu, Y. S. (2013). Alterations of alternative splicing patterns of Ser/Arg-Rich (SR) genes in response to hormones and stresses treatments in different ecotypes of rice (*Oryza sativa*). *J. Integr. Agric.* 12, 737–748. doi: 10.1016/S2095-3119(13)60260-9
- Zhang, R., Calixto, C. P. G., Marquez, Y., Venhuizen, P., Tzioutziou, N. A., Guo, W., et al. (2017). A high quality Arabidopsis transcriptome for accurate transcript-level analysis of alternative splicing. *Nucleic Acids Res.* 45, 5061–5073. doi: 10.1093/nar/gkx267
- Zhang, R., Calixto, C. P. G., Tzioutziou, N. A., James, A. B., Simpson, C. G., Guo, W., et al. (2015). AtRTD - a comprehensive reference transcript dataset resource for accurate quantification of transcript-specific expression in Arabidopsis thaliana. *New Phytol.* 208, 96–101. doi: 10.1111/nph.13545
- Zhang, T., Marand, A. P., and Jiang, J. (2016). PlantDHS: a database for DNase I hypersensitive sites in plants. *Nucleic Acids Res.* 44 (D1), D1148–53. doi: 10.1093/nar/gkv962
- Zhang, X., Cal, A. J., and Borevitz, J. O. (2011). Genetic architecture of regulatory variation in Arabidopsis thaliana. *Genome Res.* 21, 725–733. doi: 10.1101/gr.115337.110

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2019 Khokhar, Hassan, Reddy, Chaudhary, Jabre, Byrne and Syed. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



Alternative Splicing of Circadian Clock Genes Correlates With Temperature in Field-Grown Sugarcane

Luíza L. B. Dantas^{1†}, Cristiane P. G. Calixto², Maira M. Dourado¹, Monalisa S. Carneiro³, John W. S. Brown^{2,4*} and Carlos T. Hotta^{1*}

OPEN ACCESS

Edited by:

Ezequiel Petrillo,
CONICET Institute of Physiology,
Molecular Biology and Neurosciences
(IFIBYNE), Argentina

Reviewed by:

Naeem Hasan Syed,
Canterbury Christ Church University,
United Kingdom
Micaela Godoy Herz,
CONICET Institute of Physiology,
Molecular Biology and Neurosciences
(IFIBYNE), Argentina

*Correspondence:

John W. S. Brown
j.w.s.brown@dundee.ac.uk
Carlos T. Hotta
hotta@iq.usp.br

†Present address:

Luíza L. B. Dantas,
Max Planck Institute of Molecular Plant
Physiology, Potsdam-Golm, Germany

Specialty section:

This article was submitted to
Plant Physiology,
a section of the journal
Frontiers in Plant Science

Received: 29 July 2019

Accepted: 15 November 2019

Published: 23 December 2019

Citation:

Dantas LLB, Calixto CPG,
Dourado MM, Carneiro MS,
Brown JWS and Hotta CT (2019)
Alternative Splicing of Circadian Clock
Genes Correlates With Temperature in
Field-Grown Sugarcane.
Front. Plant Sci. 10:1614.
doi: 10.3389/fpls.2019.01614

¹ Departamento de Bioquímica, Instituto de Química, Universidade de São Paulo, São Paulo, Brazil, ² Division of Plant Sciences, School of Life Sciences, University of Dundee at the James Hutton Institute, Dundee, United Kingdom,

³ Departamento de Biotecnologia, Produção Vegetal e Animal, Centro de Ciências Agrícolas, Universidade Federal de São Carlos, Araras, Brazil, ⁴ Cell and Molecular Sciences, The James Hutton Institute, Dundee, United Kingdom

Alternative Splicing (AS) is a mechanism that generates different mature transcripts from precursor mRNAs (pre-mRNAs) of the same gene. In plants, a wide range of physiological and metabolic events are related to AS, as well as fast responses to changes in temperature. AS is present in around 60% of intron-containing genes in *Arabidopsis*, 46% in rice, and 38% in maize and it is widespread among the circadian clock genes. Little is known about how AS influences the circadian clock of C4 plants, like commercial sugarcane, a C4 crop with a complex hybrid genome. This work aims to test if the daily dynamics of AS forms of circadian clock genes are regulated by environmental factors, such as temperature, in the field. A systematic search for AS in five sugarcane clock genes, *ScLHY*, *ScPRR37*, *ScPRR73*, *ScPRR95*, and *ScTOC1* using different organs of sugarcane sampled during winter, with 4 months old plants, and during summer, with 9 months old plants, revealed temperature- and organ-dependent expression of at least one alternatively spliced isoform in all genes. Expression of AS isoforms varied according to the season. Our results suggest that AS events in circadian clock genes are correlated with temperature.

Keywords: alternative splicing, circadian clock, diel rhythms, field experiment, gene expression, sugarcane

INTRODUCTION

During gene expression in a eukaryotic cell, pre-mRNAs undergo splicing to remove introns and join exons in a mature transcript, generating an open reading frame (ORF) for protein synthesis. Splicing is largely co-transcriptional in yeast, *Drosophila melanogaster*, mammals and *Arabidopsis thaliana* (L.) Heyn. (Beyer and Osheim, 1988; Oesterreich et al., 2016; Saldi et al., 2016; Godoy Herz et al., 2019; Jabre et al., 2019). Alternative splicing (AS) is a mechanism that generates different RNA transcripts from a single gene. As a result of AS, the mature mRNAs represent another level of gene expression regulation at the post-transcriptional level by, for example, insertion of premature termination codons (PTC), which can target some AS isoforms to degradation by the nonsense-mediated mRNA decay (NMD) pathway (Filichkin and Mockler, 2012; Kalyna et al., 2012; Marquez et al., 2012). Alternatively, those transcripts carrying PTCs could produce truncated polypeptides missing functional domains and

motifs that can compete with the corresponding functional protein (Seo et al., 2011; Mastrangelo et al., 2012; Reddy et al., 2013). In addition, AS can increase protein by producing mRNAs from the same gene that encode protein variants with different diversity in function, localization, and stability (Syed et al., 2012; Chaudhary et al., 2019). AS is a ubiquitous process observed from *Drosophila* to humans and plants (Graveley, 2005; Filichkin et al., 2010; Marquez et al., 2012; Kornblihtt et al., 2013). In plants, a wide range of physiological and metabolic events and responses are related to AS. This mechanism is so widespread that it was reported in more than 60% of intron-containing genes in *Arabidopsis*, 46% in rice, and 38% in maize (Zhang et al., 2010; Marquez et al., 2012; Staiger and Brown, 2013a; Thatcher et al., 2014; Chamala et al., 2015; Filichkin et al., 2015b; Min et al., 2015). There is evidence of organ and tissue-specific alternative transcript forms and even alternative transcript isoforms in different subcellular locations (Nagashima et al., 2011; Kriechbaumer et al., 2012; Remy et al., 2013; Vaneechoutte et al., 2017). AS impacts development, from the early gametic cell specification to the seed maturation (Moll et al., 2008; Liu et al., 2009; Sugliani et al., 2010; Fouquet et al., 2011; Zhang et al., 2016; Szakonyi and Duque, 2018) and even flowering time and floral development (Zhang et al., 2011; Severing et al., 2012; Rosloski et al., 2013). Both biotic and abiotic stress responses are also closely related to AS (Staiger and Brown, 2013a; Shang et al., 2017; Wang et al., 2018). Plants under stress conditions change their AS patterns dramatically (Palusa et al., 2007; Staiger and Brown, 2013a; Ding et al., 2014; Filichkin et al., 2015a; Calixto et al., 2018; Calixto et al., 2019). Also, many circadian clock genes generate alternative transcript forms with PTCs under different environmental conditions (Filichkin et al., 2010; James et al., 2012c; James et al., 2012a; Jones et al., 2012; Filichkin et al., 2015a; Calixto et al., 2016). The presence of alternative transcripts in the circadian clock genes is highly conserved among different plant species, such as *Arabidopsis*, *Populus alba* L., *Brachypodium distachyon* (L.) P. Beauv., and rice (*Oryza sativa* L.)—all C3 plants (Filichkin and Mockler, 2012). Little is known about how AS influences the circadian clock of C4 plants.

The circadian clock is a 24 h endogenous timekeeper mechanism that anticipates the Earth's day/night and seasonal cycles (Hsu and Harmer, 2014; Millar, 2016; McClung, 2019). Like AS, the circadian clock is associated with growth, photosynthesis, and biomass in plants, so these two regulatory mechanisms may act together, or even regulate each other (Dodd et al., 2005; Lu et al., 2005; Harmer, 2009; Lai et al., 2012; Syed et al., 2012; Staiger and Brown, 2013a). The circadian clock consists of multiple interlocked transcription–translation feedback loops connected with input pathways that feed the circadian clock function with environmental cues, such as light and temperature, and with output pathways that are responsible for coordinating several major metabolic and physiological processes (Pokhilko et al., 2012; Haydon et al., 2013; Hsu and Harmer, 2014). In *Arabidopsis*, the main loop consists in three different components: *CIRCADIAN CLOCK ASSOCIATED 1* (*CCA1*), *LATE ELONGATED HYPOCOTYL* (*LHY*), expressed around dawn and *TIMING OF CHLOROPHYLL A/B BINDING PROTEIN 1* (*TOC1*), expressed around dusk (Alabadi et al.,

2001). Closely associated with this loop are the PSEUDO-RESPONSE REGULATORS 7, 3 and 9 (*PRR7*, *PRR3*, *PRR9*) (Locke et al., 2005; Zeilinger et al., 2006; Para et al., 2007; Nakamichi et al., 2010). The components of the central loop and the associated PRRs are conserved among other plant species, including crops like rice, maize (*Zea mays* L.), barley (*Hordeum vulgare* L.) and sugarcane (*Saccharum* hybrid) (Murakami et al., 2007; Khan et al., 2010; Hotta et al., 2013; Calixto et al., 2015). The sugarcane circadian clock, although sharing conserved components with other plants, may have a broader influence over sugarcane physiology, with 32% of sugarcane transcripts showing rhythms under circadian conditions (Hotta et al., 2013).

Sugarcane is a C4 grass that stores large amounts of sucrose in its stems, which can reach as much as 700 mM or 50% of the culm dry weight (Moore, 1995). Its genome is exceptionally complex, showing aneuploidy and a massive autopolyploidy that can range from six to fourteen copies of each chromosome (Garcia et al., 2013). The genome size of commercial modern sugarcane is estimated to be around 10 Gb (de Setta et al., 2014; Chan et al., 2018). Because modern sugarcane cultivars are interspecific hybrids progenies from *Saccharum officinarum* L. and *Saccharum spontaneum* L., about 80% of sugarcane chromosomes comes from *S. officinarum*, 10% comes from *S. spontaneum* and 10% are recombinants of these two species (D'Hont et al., 1996; Cuadrado et al., 2004; D'Hont, 2005). Sugarcane is a valuable commodity, responsible for 80% of sugar and 40% of ethanol worldwide (FAO, 2015). The remaining biomass from sugarcane can also be used for bioenergy production: the bagasse can be either burned to generate electricity or have its cell wall hydrolyzed to yield simple sugars, which can be fermented to produce second-generation biofuel (Amorim et al., 2011).

Although a great deal of data has been generated about the plant circadian clock, sugarcane, and AS, the majority of these studies have been performed under highly controlled experimental conditions. Such conditions are essential for reproducibility and, for the circadian clock, a constant environment is one way to demonstrate the inner mechanism generating self-sustained rhythms, as well as rhythmic responses. However, those conditions are far from the environment that crops face in nature, with fluctuations and complex interactions between abiotic and biotic variables (Annunziata et al., 2017; Annunziata et al., 2018; Shalit-Kaneh et al., 2018). In order to better understand the relationship between the circadian clock and AS and how this relationship impacts on crops, it is essential to expand experiments to field conditions. Indeed, essential in-field studies using *Arabidopsis* (Richards et al., 2012; Annunziata et al., 2018) and rice (Izawa et al., 2011a; Sato et al., 2011; Nagano et al., 2012) show that the complex natural cyclic environment has a broader impact on rhythmic gene expression. So far, no studies have approached the AS profile on circadian clock genes under such conditions.

In this study, we examined whether the daily dynamics of AS forms of circadian clock genes are regulated by environmental factors in the field. We used sugarcane organs extracted from

field-grown plants when individuals were 4-months-old, during the Brazilian winter, and 9-months-old, during the Brazilian summer. We investigated the AS profile of sugarcane circadian clock genes in this fluctuating natural environment. Data shows that there is at least one alternatively spliced form for each of the five circadian clock genes analyzed. During winter, when temperatures are lower, alternative transcripts are more highly expressed than in summer, with higher temperatures, which suggests that AS might be related to the fluctuating environmental temperature in the field. The different organs also showed different levels of AS and leaf has most of the diversity in AS events. Collectively, our data suggest temperature correlates with AS in the circadian clock of sugarcane plants grown in a natural environment, possibly as a mechanism of dynamic adjustment of the circadian clock.

MATERIAL AND METHODS

Field Conditions and Plant Harvesting

The sugarcane field where the experiment was conducted was located at the Federal University of São Carlos, campus Araras, in São Paulo state, Brazil (22°21'25" S, 47°23'3" W, at an altitude of 611 m). The soil of the site was classified as a Typic Eutroferic Red Latosol. Sugarcane tillers from the commercial variety SP80-3280 (*Saccharum* hybrid) were planted in soil in April/2012. Field design had 8 plots (**Figure S1A**). Each plot had 4 rows containing 20 tillers each. Only sugarcane plants from both central lines were used in order to avoid border effects. Sugarcane individuals were randomly picked from two plots in order to avoid variability of both the local environment and individual plants. Data on environmental conditions was acquired from a local weather station (**Figures S1B, C**). Leaves +1 (L1), a source organ and the first fully photosynthetically active leaf in sugarcane, were sampled from the selected individual plants during two different seasons, and therefore different developmental stages. In the first harvest, 4-months-old plants were sampled in August/2012, during winter; in the second harvest, 9-months-old plants were sampled in January/2013, during summer. In winter, dawn was at 6:30, and dusk was at 18:00 (11.5 h day/12.5 h night). In summer, dawn was at 5:45, and dusk was at 19:00 (13.25 h day/10.75 h night). To compare the rhythms of samples harvested in different seasons, the time of harvesting were normalized to a photoperiod of 12 h day/12 h night using the following equations: for times during the day: $ZT = 12 \cdot T \cdot P_d^{-1}$, where ZT is the normalized time, T is the time from dawn (in hours), and P_d is the length of the day (in hours); for times during the night: $ZT = 12 + 12 \cdot (T - P_d) \cdot P_n^{-1}$, where ZT is the normalized time, T is the time from dawn (in hours), P_d is the length of the day (in hours), and P_n is the length of the night (in hours). Because the 9-month-old plants had their culms fully developed, internodes 1 and 2 (I1) and internode 5 (I5) were also sampled. Both internodes are sink tissues with different profiles: internodes 1 and 2 mostly undergo intense cell division and elongation, whereas internode 5 undergoes sucrose storage. For every time point, 9 individuals were randomly selected in the assigned plots and harvested from the culm up. After that, those 9 individuals were separated into three pools of three individuals, each pool formed a biological replicate and then

their leaves +1 were extracted. For all harvests, plants were sampled every 2 h for 26 h, starting 2 h before dawn. In total, the time course consisted of 14 time points in each harvest/season. After every time point sampling, a process that took less than 30 min on average, tissue was immediately frozen in liquid nitrogen.

RNA Extraction

Sugarcane leaves previously frozen in liquid nitrogen were pulverized using dry ice and a grinder. Then, 100 mg of this ground tissue was used for total RNA extractions using Trizol (Life Technologies, Carlsbad, CA, USA), followed by treatment with DNase I (Life Technologies, Carlsbad, CA, USA) and cleaned with RNeasy Plant Mini Kit (QIAGEN, Valencia, CA, USA). The quality and quantity of each RNA sample were checked using an Agilent RNA 6000 Nano Kit Bioanalyzer (Agilent Technologies, Palo Alto, CA, USA). All RNA samples were stored at -80°C.

cDNA Synthesis

cDNA was synthesized using SuperScript III First-Strand Synthesis System for RT-PCR (Life Technologies, Carlsbad, CA, USA) starting from 5 µg of total RNA. For all reactions, both Oligo(dT) and Random Hexamers primers were used. All cDNA samples were stored at -20°C.

PCR Reactions

Primers used in PCR reactions were designed using the software PrimerQuest Tool (IDT) (<http://www.idtdna.com/primerquest/home/index>). Each pair of primers was gene-specific and amplified fragments ranging from 242 bp to 805 bp (**Table S1**). All PCR reactions were carried out using Go Taq DNA Polymerase (Promega, Madison, WI, USA) and following the manufacturer's protocol. Briefly, each 20-µl PCR reaction contained 2 µl of template, 10 µM of each primer, 4 µl of 5x Green Go Taq Buffer, 0.15 µl of Go Taq DNA Polymerase, 2 mM of dNTPs. PCR conditions were: an initial step at 94°C for 2 min, followed by 20 – 30 cycles of 94°C for 15 s, 50°C for 15 s, 72°C for 30 s, followed by a final extension of 72°C for 5 min. PCR reactions using primers amplifying control genes *ScGAPDH* and *ScPP2AA2* were performed for all cDNA samples. Reactions containing negative control using RNA as template and positive control using genomic DNA as template were carried out. All PCR-amplified fragments were analyzed by taking 10 µl of reaction and run on an electrophoresis gel of 1.5% agarose (Life Technologies, Carlsbad, CA, USA) and 1x TBE (50 mM Tris-HCl pH 8, 50 mM Boric Acid, 1mM EDTA).

High-Resolution RT-PCR

High-Resolution RT-PCR (HR RT-PCR) reactions were performed based on Simpson et al. (2007) and Simpson et al. (2019). For all reactions, the forward primer was labeled with 6-carboxyfluorescein (FAM). Reactions consisted of a final volume of 20 µl which had 2 µl of cDNA, 10 µM of each primer, 2 µl of 10x PCR Reaction Buffer with MgCl₂ (Roche Life Science, Indianapolis, IN, USA), 0.15 µl Taq DNA Polymerase (Roche Life Science, Indianapolis, IN, USA), and 2 mM of dNTPs. The PCR detailed program was: an initial step at 94°C for 2 min,

followed by 22–26 cycles of 94°C for 15 s, 50°C for 15 s, 70°C for 30 s, followed by a final extension of 70°C for 5 min. Once PCR reactions were complete, 1 µl of each reaction was added to a mix containing 9 µl of Hi-Di Formamide (Applied Biosystems, Life Technologies, Carlsbad, CA, USA) and 0.5 µl of GeneScan 500 LIZ Size Standard (Applied Biosystems, Carlsbad, CA, USA). The RT-PCR products were separated on an ABI 3730 Automatic DNA Sequencer (Applied Biosystems, Life Technologies, Carlsbad, CA, USA). The results were analyzed using GeneMapper fragment analysis software (Applied Biosystems, Carlsbad, CA, USA). LOESS (locally estimated scatterplot smoothing) regression was used to detect trends in the data. The maximum value of the LOESS curve between ZT0 and ZT22 was considered the peak of the rhythm. The code to fully reproduce our analysis is available on GitHub (<https://github.com/LabHotta/AlternativeSplicing>) and archived on Zenodo (<http://doi.org/10.5281/zenodo.3509232>).

Cloning and Sequencing

In order to identify alternatively spliced forms, as well as differentially expressed alleles, RT-PCR fragments were cloned and sequenced. For this, PCR fragments were purified using the QIAquick PCR Purification Kit (QIAGEN, Valencia, CA, USA). Each purified fragment was cloned into pGEM-T Easy Vector (Promega, Madison, WI, USA) following the manufacturer's protocol. Briefly, each reaction contained 3 µl of purified PCR product, 5 µl of Rapid Ligation Buffer, T4 DNA Ligase, 1 µl of pGEM-T Easy Vector (50 ng) and 1 µl T4 DNA Ligase (3 Weiss units/µl). Ligation reactions were incubated overnight at 4°C. Two micro-liter of each ligation reactions were used for heat-shock transformation of 50 µl of JM109 High-Efficiency Competent Cells (Promega, Madison, WI, USA), following manufacturer's instructions. Transformed cells were plated on LB/ampicillin/IPTG/X-gal media and incubated overnight at 37°C. Random colonies were selected to use in plasmid extraction using QIAprep Miniprep Kit (QIAGEN, Valencia, CA, USA). Positive plasmids were confirmed by PCR reactions, digestions using restriction enzymes *Pst*I and *Nco*I (Promega, Madison, WI, USA) and Sanger sequencing. In order to identify alternative splicing events and single-nucleotide polymorphisms, results were compared to sugarcane genomic sequences, and sugarcane transcripts from Sugarcane Assembled Sequences (SAS) from SUCEST (<http://sucest-fun.org/>) using Clustal Omega (<https://www.ebi.ac.uk/Tools/msa/clustalo/>). *In silico* translation was carried out using ExPaSy translation tool (<http://web.expasy.org/translate/>).

RESULTS

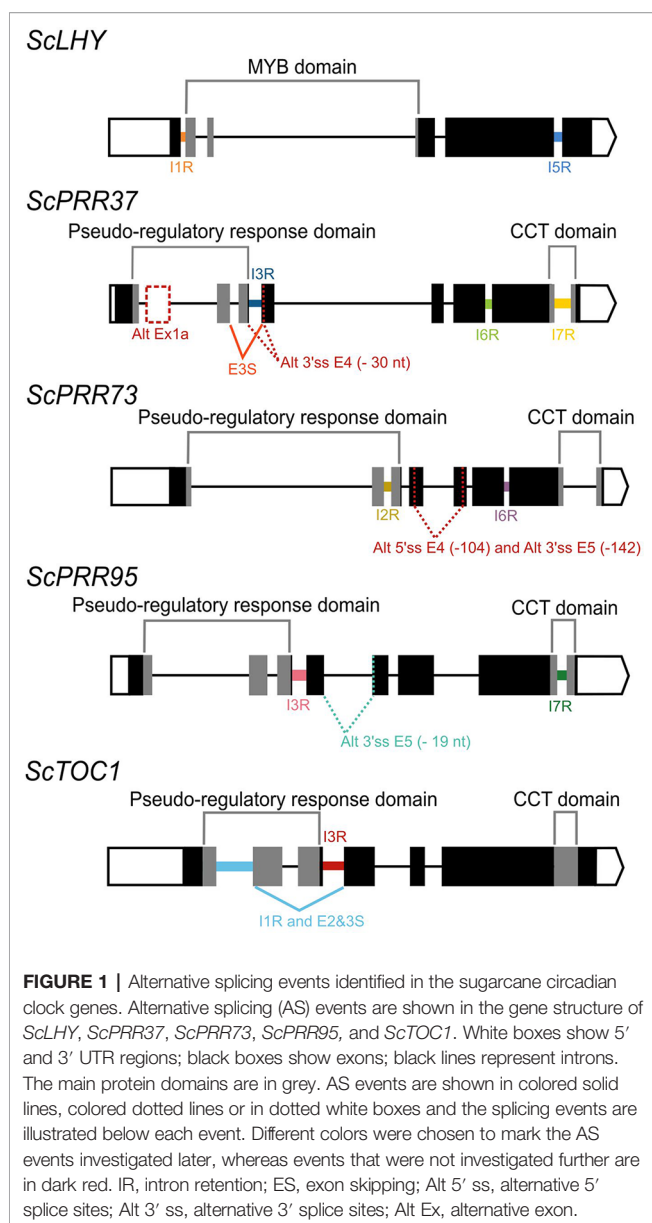
Identification of Alternatively Spliced Forms of Sugarcane Circadian Clock Genes

In *Arabidopsis*, alternative splicing (AS) events of the circadian clock genes are well known (James et al., 2012b). In order to investigate AS of sugarcane circadian clock genes, it is important

to first determine and annotate their genomic sequences/structures. We first described the gene structure of previously identified homologs *ScLHY*, *ScPRR73*, *ScPRR95* and *ScTOC1* (Hotta et al., 2013). These homologs were identified by Hotta et al. (2013) using BLAST searches of *Arabidopsis* and rice circadian clock coding DNA sequence (CDS), on the sugarcane Expressed Sequence Tags (ESTs) collection database at SUCEST (<http://sucest-fun.org/>). This study identified only CDS sequences of sugarcane circadian clock genes but not their genomic sequences due to the relatively few genomic sequences publicly available at that time for the sugarcane cultivar of interest, plus a high percentage of incomplete genomic sequences (Vicentini et al., 2012; de Setta et al., 2014). Here we used the SUCEST sequences (Hotta et al., 2013) to search for their genomic sequences in an unpublished sugarcane genomic database (Souza et al., 2019). We identified genomic contigs for *ScLHY* (previously identified as *ScCCA1*), *ScTOC1*, *ScPRR73* (previously identified as *ScPRR3*) and *ScPRR95* (previously identified as *ScPRR59*) (Table S2). A new homolog, *ScPRR37*, was identified here for the first time in sugarcane (Table S2, Figure 1). The sugarcane genes were annotated by comparison with genomic sequences of circadian clock homologs from barley and sorghum (Calixto et al., 2015). The exon/intron structures are shown in Figure 1; introns contained the canonical GT-AG dinucleotides at the exon-intron boundaries. The putative CDS of each gene analyzed was translated *in silico* in order to confirm intact open reading frames (ORFs).

To identify AS transcripts, RT-PCR was performed on cDNA synthesized from L1 RNA, collected at three different time points, which corresponded to the highest gene expression for each clock gene, based on the findings from Hotta et al. (2013). RT-PCR used pairs of primers to generate overlapping amplicons to cover the full length of the transcripts for each sugarcane circadian clock genes, *ScLHY*, *ScPRR37*, *ScPRR73*, *ScPRR95*, and *ScTOC1*. RT-PCR products were cloned and sequenced, which confirmed the annotation and identified AS events. Confirmation of different transcript isoforms was carried out by using two different approaches combined: RT-PCR product size and fragment cloning followed by sequencing.

We found evidence of AS events in all five genes analyzed (Table 1). Intron retention (IR) was the most frequent AS event identified, resulting in the inclusion of premature termination codons (PTC) in all cases (Figure 1). All five genes had at least two IR events, and *ScTOC1* had retention of the first intron (I1R) combined with the skipping of exons 2 and 3 (E23S). *ScLHY* had retention of introns 1 (I1R) and 5 (I5R). *ScPRR37* had retention of introns 3 (I3R), 6 (I6R) and 7 (I7R). *ScPRR73* had retention of introns 2 (I2R) and 6 (I6R). *ScPRR95* had two introns retained: intron 3 (I3R) and intron 7 (I7R) (Figure 1). All the other introns in the genes were efficiently spliced. Exon skipping (ES) was detected in *ScPRR37* (E3S) and *ScTOC1*, the latter having two exons skipped at once (E23S), combined with I1R (Figure 1). Alternative 3' splice sites (Alt 3' ss) were found in both *ScPRR37* and *ScPRR95* in exon 4 (E4) and exon 5 (E5), respectively (Figure 1). In *ScPRR73*, we found an AS event between exons



4 and 5 involving an alternative 5' splice site (Alt 5'ss) in exon 4 (E4) being spliced to an Alt3'ss in exon 5. This AS event removes 104 nt and 142 nt from exons 4 and 5, respectively (total of 246 nt), but remains in frame and potentially produces a protein missing 82 amino acids compared to the wild-type protein.

It is possible that some of the alternative sequences we found are not AS events but different haplotypes since sugarcane is highly polyploid and aneuploid. To exclude this possibility, we obtained the sequences for circadian clock genes from four different sugarcane sequencing projects (Riaño-Pachón and Mattiello, 2017; Garsmeur et al., 2018; Zhang et al., 2018; Souza et al., 2019). We found 11 sequences for the 5' portion of the *ScLHY*: 4 from *S. spontaneum* (Zhang et al., 2018), 6 from the commercial *Saccharum* hybrid SP80-3280 (Riaño-Pachón and Mattiello, 2017; Souza et al., 2019) and 1 from the

TABLE 1 | AS events found in the sugarcane circadian clock genes analyzed.

Gene	AS event*	Effect on transcript
<i>ScLHY</i>	Intron 1 retained (I1R)	Inserts a PTC on the transcript
	Intron 5 retained (I5R)	Creates a PTC on Exon 6
<i>ScPRR37</i>	Alternative exon 2 (Ex2a)	Inserts PTCs in the transcript
	Exon 3 skipped (E3S)	Inserts PTCs on Exon 4
	Intron 3 retained (I3R)	Creates PTC on Exon 3
	Intron 6 retained (I6R)	Inserts a PTC on the transcript
	Intron 7 retained (I7R)	Creates PTCs on Exon 8
	Alternative 3' splice site in exon 4	Protein stays in frame, removes 30 nucleotides from the PRR domain
<i>ScPRR73</i>	Intron 2 retained (I2R)	Inserts a PTC on the transcript
	Alternative 5' splice site in exon 4 and alternative 3' splice site in exon 5	Protein stays in frame, removes 245 nucleotides from transcript
<i>ScPRR95</i>	Intron 6 retained (I6R)	Inserts a PTC on the transcript
	Alternative 3' splice site in exon 5	Inserts a PTC on Exon 5
<i>ScTOC1</i>	Intron 7 retained (I7R)	Inserts a PTC on the transcript
	Intron 1 retained (I1R) combined with Exons 2 and 3 skipped (E23S)	Inserts a PTC on the transcript
	Exon 3 skipped (E3S)	Inserts a PTC on the transcript

*All AS events were confirmed by Sanger sequencing.

commercial *Saccharum* hybrid R570 (Garsmeur et al., 2018). All *ScLHY* sequences had the complete intron 1 (Figure S2). Similarly, all 16 sequences that had the end portion of *ScLHY* contained the complete intron 5 (Figure S2). All other detected AS events had 7 to 15 sequences supporting the conclusion that these alternative sequences are not the result of different sugarcane haplotypes (Figures S3–S6).

Expressed AS Forms in Different Seasons in Sugarcane Leaves

We used HR RT-PCR (Simpson et al., 2019) to examine the daily dynamics of the expressed isoforms of the sugarcane circadian clock genes in two different seasons, winter and summer, using field-grown plants that were 4 and 9 months old, respectively. Briefly, the HR RT-PCR system uses fluorescently labelled primers to amplify across an AS event, followed by fragment analysis in an automatic DNA sequencer that quantifies the relative levels of RT-PCR products and thereby splicing ratios that reflect different splice site choices. The primers used for the HR RT-PCR assays had the same sequence of those used to amplify each gene region on RT-PCR experiments (Table S1). Leaf +1 (L1), internode 1 and 2 (I1), and internode5 (I5) samples, harvested every 2 h during 26 h, starting from 2 h before dawn were used. As reference genes to normalize data, we used *ScGAPDH* and *ScPP2AA2* (Iskandar et al., 2004; James et al., 2012a). As the experiments were done in different seasons, we have normalized the time of sampling to fit in a 12 h day/12 h night photoperiod such that ZT00 is set to dawn, and ZT12 is set to dusk.

From the five sugarcane circadian clock genes analyzed, HR RT-PCR experiments detected high levels of AS in *ScLHY*, *ScPRR37*, and *ScPRR73* (Figure 2), but not in *ScPRR95* and *ScTOC1*, in L1 (Figure S7). In general, the AS isoform peaked earlier than the FS form in 8 of the 9 conditions assayed (Figures

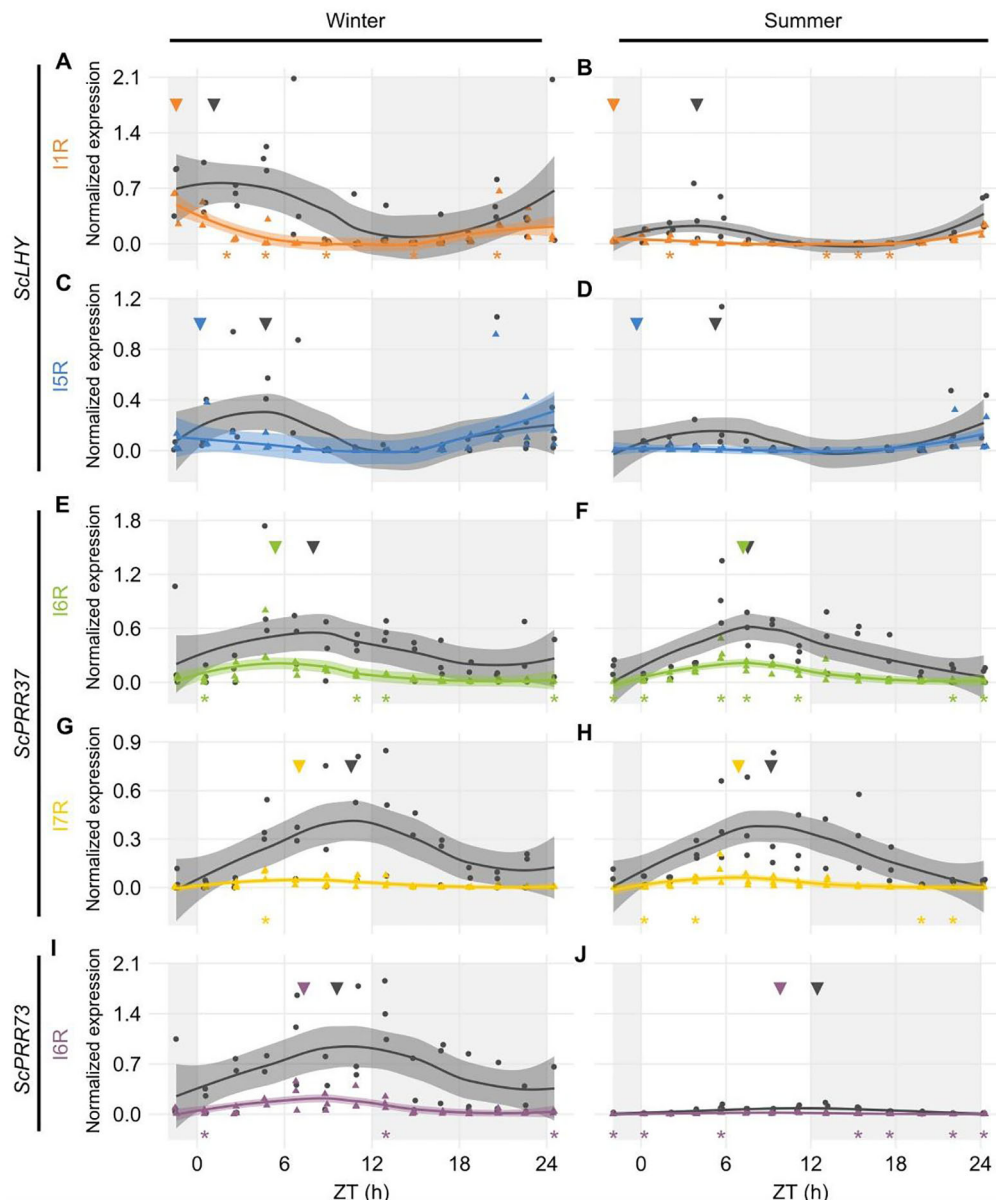


FIGURE 2 | Diel expression profile of fully spliced and alternative transcript isoforms in different seasons. Biological replicates (circles and triangles) and their LOESS curve (continuous lines \pm SE) of fully spliced (FS, black) and alternative transcript forms (AS, colored) for the winter samples (4-month-old plants, left) the summer samples (9-month-old plants, right). (A, B) *ScLHY* gene expression shows levels of I1R (orange) and (C, D) I5R events (blue). (E, F) *ScPRR37* gene expression shows levels of I6R (green), and (G, H) I6R (yellow); (I, J) *ScPRR73* gene expression shows levels of I2R (purple). Inverted triangles show the time of the maximum value of the LOESS curve. The light-gray boxes represent the night period. Statistical significance was analyzed by paired Student's t-test, * $p < 0.05$.

2A–I), apart from *ScLHY*, where the FS forms peaked at the same normalized time in winter and summer samples.

ScLHY had confirmed events of intron 1 and 5 retention (I1R and I5R) in both harvests (Figures 2A–D). The peak of expression of the *ScLHY* alternative isoforms did not match the peak of expression of the fully spliced functional (FS) isoform. *ScLHY* I1R peaked close to dawn in winter and summer samples, while the FS for this region peaked one hour after dawn in winter plants (ZT01) and five hours after dawn (ZT05) in summer plants (Figures 2A, B). *ScLHY* I5R peaked at

dawn for plants in winter samples, while it was not considered expressed for plants in summer samples (Figures 2C, D). The FS form for this region peaked between ZT04–05 for plants in both winter and summer samples (Figure 2).

ScPRR37 I6R had a peak at ZT05 in winter samples, but at ZT07 in summer samples (Figures 2E, F). The corresponding FS isoform had a peak at ZT08 in both conditions. In turn, *ScPRR37* I7R had a peak at ZT07 in both winter and summer plants, but the corresponding FS isoform had a peak at ZT09 and ZT11 in winter and summer samples, respectively (Figures 2G, H).

ScPRR37 I3R and E3S were not detectable using HR RT-PCR (Figures S7A, B). Both the *ScPRR73* I6R and its FS isoform only had high levels in winter samples, with a peak at ZT07 (Figures 2I, J). The corresponding FS isoform had a peak at ZT10.

Expressed AS Forms in the Different Source-Sink Sugarcane Organs

We extended the investigation to two other sugarcane organs, internodes 1 and 2 (I1) and internode 5 (I5). These internodes have different physiology: I1 has a high cellular and metabolic activity, whereas I5 is the first internode to actively accumulate sucrose. Only the plants harvested in the summer (9-month-old) had developed internodes that could be harvested. We only measured AS forms in genes *ScLHY* and *ScPRR37*, which were the homologs featuring the highest AS transcript expression in L1 (Figure 2). The main difference in expression levels between organs was observed in *ScPRR37* I6R, that had significantly higher

levels in leaves during the day compared to the internodes (One-way ANOVA with post-hoc Tukey HSD test, $*p < 0.05$, Figure S8).

In both internodes, *ScLHY* showed detectable levels of both AS events observed in L1. *ScLHY* I1R-containing transcript levels were very low with a peak 2 h before dawn (ZT22), with the FS isoform peaking between ZT03-04 (Figures 3A, B). *ScLHY* I5R was also identified in both internodes, at higher levels than *ScLHY* I1R. The AS isoform peaked at ZT22, while the FS form peaked between ZT03-04 (Figures 3C, D). During the end of the night, *ScLHY* I1R levels were significantly higher in the leaves compared to the internodes (One-way ANOVA with post-hoc Tukey HSD test, $*p < 0.05$, Figure S8).

The *ScPRR37* homolog featured variable levels of FS and AS isoforms when compared to both internodes. In I1, *ScPRR37* I6R and its FS isoform did not have a clear rhythm, with the I6R showing low levels of expression, and the FS isoform showing two peaks. In I5, *ScPRR37* I6R and its FS isoform had higher

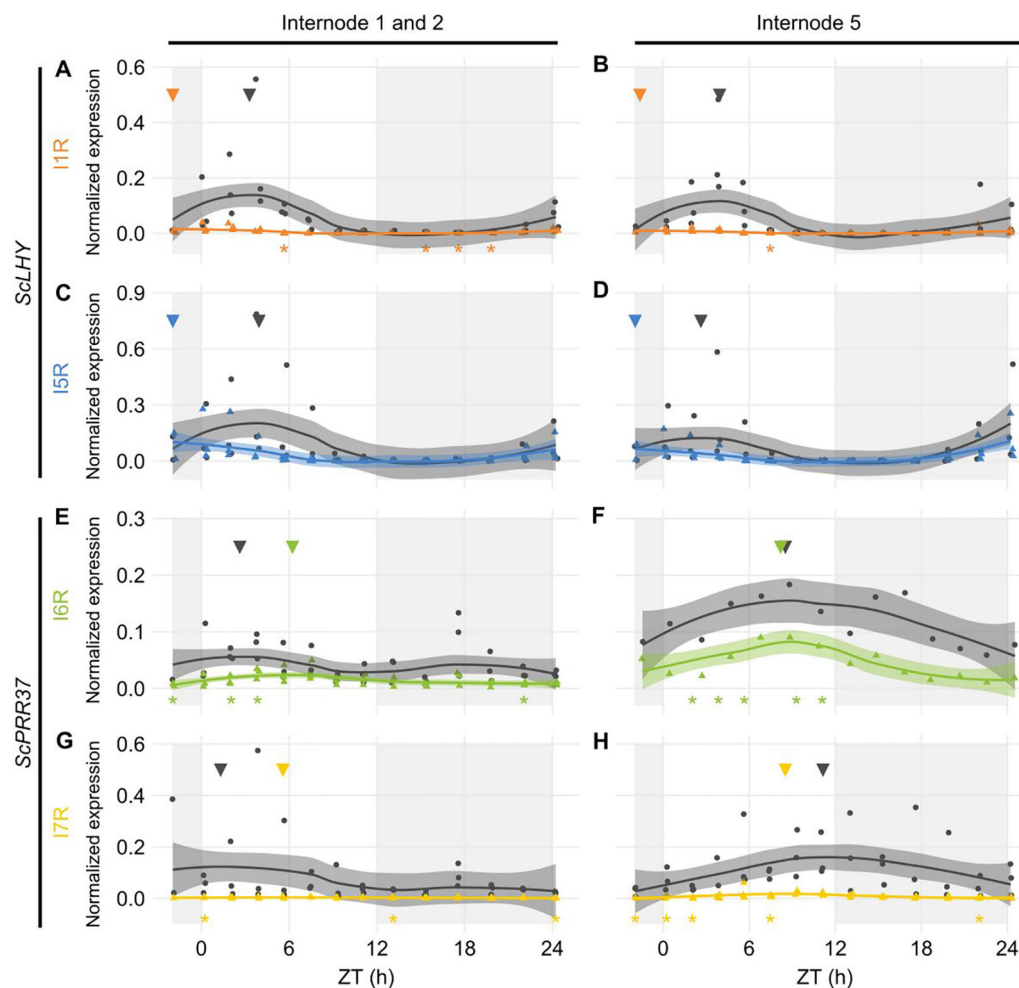


FIGURE 3 | Diel expression profile of fully spliced and alternative spliced transcript isoforms in internodes. Biological replicates (circles and triangles) and their LOESS curve (continuous lines \pm SE) of fully spliced (FS, black) and alternative spliced (AS, colored) transcript forms of sugarcane circadian clock genes during the summer harvest in internode 1 and 2 (left) and internode 5 (right). (A, B) *ScLHY* gene expression shows levels of I1R (orange) and (C, D) I5R (blue); (E, F) *ScPRR37* gene expression shows levels of I6R (green) and (G, H) I7R (yellow). Inverted triangles show the time of the maximum value of the LOESS curve. The light-gray boxes represent the night period. Statistical significance was analyzed by paired Student's t-test, $*p < 0.05$.

levels and a peak between ZT08-09 (**Figures 3E, F**). In both internodes, the I7R isoform was expressed at very low levels, but the FS isoform for that transcript region peaked at ZT01 in I1 and ZT11 in I5 (**Figures 3G, H**).

Alternative Splicing Events Are Dependent on the Time of the Day and Temperature

After the identification of rhythms in FS and AS transcripts, we tried to identify rhythms in their relative levels by examining the log of the splicing ratio of the AS to the FS transcripts [$\log(\text{AS}/\text{FS})$] from the HR RT-PCR data. *ScLHY*, *ScPRR37*, and *ScPRR73* showed evidence of splicing rhythms (**Figure 4**), but only *ScLHY* had more than a 10-fold difference between the maximum and

the minimum $\log(\text{AS}/\text{FS})$ (77-fold) (**Figure 4A**). The AS : FS ratios of all the time-point samples and organs were grouped, as they showed similar rhythmic patterns. In general, all AS events of a gene showed a similar rhythmic pattern. The only exception was E3S in *ScPRR37* (**Figure 4D**), that had a different phase from the other AS events found in *ScPRR37* (**Figure 4B**). The AS events observed in *ScLHY* and E3S in *ScPRR37* had a peak at the end of the night, between ZT20 and ZT24, and a trough between ZT06 and ZT08. In contrast, *ScPRR73* AS events and the remaining *ScPRR37* AS events had a peak between ZT05 and ZT06, and a trough between ZT16 and ZT18.

The rhythmic changes in the $\log(\text{AS}/\text{FS})$ values could be explained by changes in the expression of putative regulatory

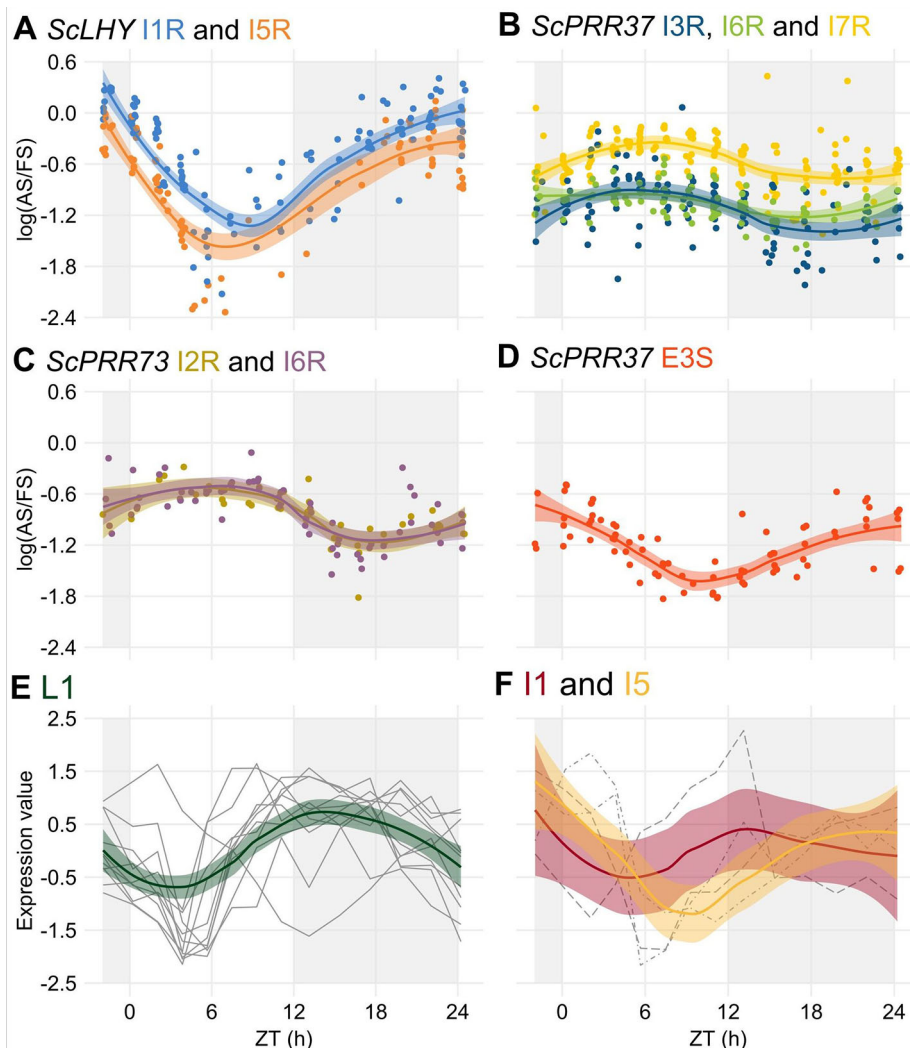


FIGURE 4 | Alternative splicing is rhythmic in sugarcane in field conditions. **(A–D)** The logarithm of the ratio of the expression levels of an AS isoform to its FS isoform, annotated as the $\log(\text{AS}/\text{FS})$, was plotted against the normalized time of the day (ZT). **(A)** *ScLHY* I1R (orange) and I5R (blue); *ScPRR37* I3R (dark blue), **(B)** I6R (green) and I7R (light yellow); **(C)** *ScPRR73* I2R (gold) and I6R (purple); and **(D)** *ScPRR37* E3S (red). **(E–F)** Normalized expression levels of rhythmic splicing-related transcripts taken from oligo array data (Dantas et al., 2019) in **(E)** leaves +1 (L1, green), and internodes 1 and 2 (I1, red) and **(F)** internode 5 (I5, yellow). Individual expression profiles were drawn in gray. LOESS regression was used to draw the trends in the data in all panels (continuous line \pm SE). The light-gray boxes represent the night period.

genes, such as splicing factors or spliceosomal protein genes. In a previous work, we have identified 6,705 rhythmic transcripts in L1, 3,755 in I1 and 3,242 in I5 in field-grown sugarcane (Dantas et al., 2019). Fourteen spliceosome-related transcripts in the oligo array were expressed in all three organs (**Table S2**): 9 transcripts were rhythmic only in L1, one was rhythmic in L1 and I1, one was rhythmic in L1 and I5, and one was rhythmic in I1 and I5 (**Table S3**). In L1, 9 of the 11 rhythmic transcripts peaked between ZT09 and ZT13, with a trough between ZT03 and ZT5 (**Figure 4E**). In the internodes, most of the transcripts peaked at ZT00 (**Figure 4F**).

To test if the temperature was a factor in the regulation of AS, we correlated temperature information with $\log(\text{AS}/\text{FS})$ values. Only *ScLHY* AS events showed a significant negative correlation (**Figure 5**). The negative correlation was found in both *ScLHY* AS events, at both harvests/seasons, and all organs. This suggests that the AS regulation of *ScLHY* genes are temperature-dependent.

DISCUSSION

In this paper, pioneer information on AS events in the sugarcane circadian clock core genes *ScLHY*, *ScPRR37*, *ScPRR73*, *ScPRR95*, and *ScTOC1* in field-grown sugarcane plants are presented (**Figure 1**). As for *Arabidopsis* and barley in previous studies, AS is widespread among circadian clock genes (Filichkin et al., 2010; Filichkin and Mockler, 2012; James et al., 2012a; Calixto et al., 2016). The circadian clock homolog *CCA1/LHY* in rice also displays the conserved IIR AS event, suggesting conserved patterns in AS events between the two species (Filichkin et al., 2015a). In barley, AS events were described for the homologs *HvLHY*, *HvPRR37*, *HvPRR73*, and *HvGI*. There are conserved AS isoforms expressed in both *Arabidopsis* and barley for *HvLHY* and *HvPRR37* (Calixto et al., 2016). The conservation of expression of AS forms across different plant species highlights the role that AS plays in gene expression of circadian clock genes.

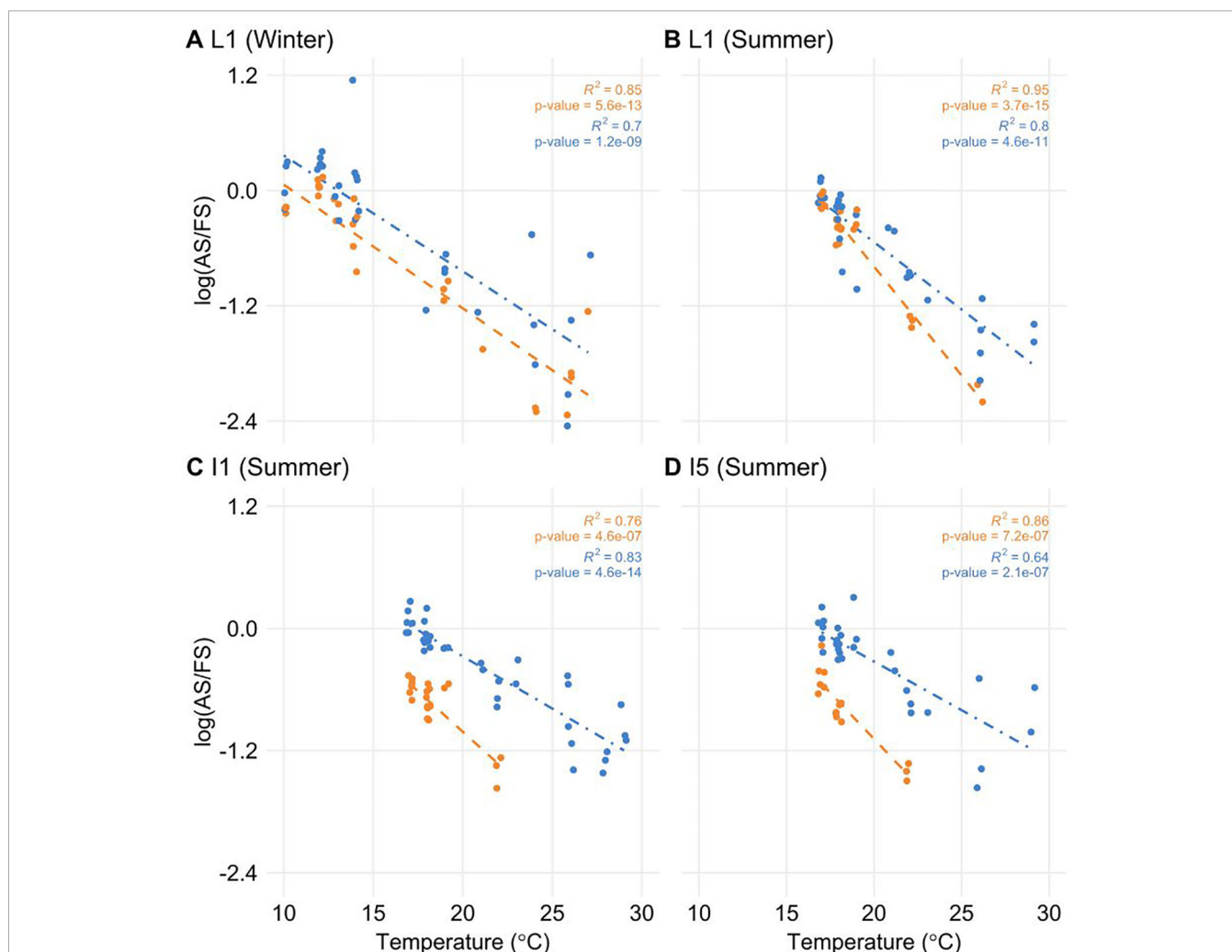


FIGURE 5 | The proportion of alternative spliced and fully spliced forms has a negative correlation with ambient temperature. The logarithm of the ratio of the expression levels of AS and FS isoforms, annotated as the $\log(\text{AS}/\text{FS})$, was plotted against the ambient temperature for *ScLHY* I1R (orange) and I5R (blue). Regression lines were added for each group of ratios. R^2 and P -value were calculated for each regression. Negative correlations were significant in leaf +1 (L1) in the (A) winter (4-month-old plants), and (B) summer, (C) internode 1 and 2 (I1), and (D) internode 5 (I5), both in the summer (9-month-old plants).

Our data in sugarcane identified at least one AS event in each of the clock gene homologs analyzed. Given the ploidy of sugarcane, it was necessary to demonstrate that the alternative transcript sequences did not originate from haplotypes but were *bona fide* AS events. By examination of existing sequenced genomic data, we conclude that they are AS events. The most frequent AS event in the sugarcane circadian clock genes was intron retention (IR), which was detected in all five genes. The presence of transcripts containing retained introns might also be an indication of partially spliced transcripts but the other introns in these genes were efficiently removed. The presence of retained introns in transcript isoforms can lead to post-transcriptional regulation of gene expression (Reddy et al., 2013). Since retained introns usually insert premature termination codons (PTCs) in the transcript, they could be substrates for degradation *via* the nonsense-mediated decay (NMD) pathway. However, in plants, transcripts with detained introns appear to avoid NMD since their abundance is unaffected in NMD mutants (Kalyna et al., 2012; Marquez et al., 2012). Furthermore, such transcripts have been shown to remain in the nucleus and thus avoid the NMD machinery (Göhring et al., 2014). Nuclear intron detention is now recognized as an important post-transcriptional regulatory mechanism (Jacob and Smith, 2017). Intron retention transcripts with PTCs can also potentially give rise to C-terminally truncated and dysfunctional proteins (Seo et al., 2011; Mastrangelo et al., 2012; Reddy et al., 2013). Other AS events in sugarcane circadian clock genes that were in frame and did not insert a PTC in the transcript were the alternative 5' splice site in *ScPRR37* exon 3 (Alt 5'ss E3) and the *ScPRR73* alternative splice site between exons 4 and 5: Alt 5'ss E4 (-104) and Alt 3' ss E5 (-142) (Table S2). In both cases, there is the removal of nucleotides from the coding sequence. *ScPRR37* Alt 5'ss E3 removes 30 nucleotides (10 amino acids) from the PRR domain. The resulting sequence is likely to be translated into a defective protein. In *ScPRR73*, the combination of Alt 5'ss E4 (-104) and Alt 3' ss E5 (-142) removes 246 nucleotides (82 amino acids) from the coding sequence but still leaves an ORF. However, it is possible that the loss of sequence could affect the normal function of the *ScPRR73* CTT domain. Therefore, the AS events identified in the sugarcane core circadian clock genes either produce transcripts that are likely to be kept in the nucleus and degraded or that are translated into incomplete proteins that are likely to be functionally defective. Thus, AS has an important role in regulating expression and production of core clock proteins. This might have a direct impact on the clock-dependent plant metabolism and physiology.

The levels of alternative transcript isoforms can vary under stress conditions (Staiger and Brown, 2013b; Shang et al., 2017; Calixto et al., 2018), at specific developmental stages (Szakonyi and Duque, 2018) or in different cell tissues (Shen et al., 2014; Thatcher et al., 2014). In previous work using microarrays, we found that rhythmic expression at the gene level was very organ-specific in sugarcane (Dantas et al., 2019). Our data shows differences in the AS in leaf and internodes at the transcript level. In L1, circadian clock transcripts undergo AS at higher relative levels than in I1 and I5 at the end of the night (Figures 2

and 3). In addition, the splicing ratios, $\log(\text{AS}/\text{FS})$, had rhythms in *ScLHY*, *ScPRR37*, and *ScPRR73*. In *ScPRR37*, the intron retention events peaked in the middle of the day, while the exon skipping event peaked at the end of the night, suggesting that the temporal regulation of these two types of AS events are independent of each other (Figures 4B–D). Although their expression profiles differ, the consequences of *ScPRR37* I3R and E3S events are likely to be similar. I3R introduces PTCs after exon 3 and E3S removes exon 3 (part of the PRR domain). *Arabidopsis* *PRR7* also has two mutually exclusive AS events in a similar region of the gene: retention of intron 3 (I3R) and skipping of exon 4 (E4S), both of which give nonproductive mRNAs (James et al., 2012a). The switch from intron retention (mainly during the day) to exon skipping (mainly during the night) in *ScPRR37* may reflect rhythmic changes in specific splicing factors. The splicing ratio rhythms of *ScLHY* and *ScPRR73* showed the same pattern, irrespective of the sampling season or organ (Figures 4A–C). This means that these splicing ratios are not organ-specific and are environmentally and circadian clock-regulated in order to have the same distribution during the day and the night regardless of their duration. To try and relate the rhythmic changes in splicing ratios of clock genes to the expression of splicing factor or spliceosomal protein genes, we examined the expression of the splicing-related genes that were rhythmic in L1 in our previous microarray study (Figure 4E) (Dantas et al., 2019). The gene-level expression of the majority of these genes peak around dusk (between ZT11 and ZT13), which does not coincide with the peaks in splicing ratios observed in circadian clock genes. The list of splicing-related genes in the microarray analysis was not extensive (Table S1) and many splicing factors are alternatively spliced to regulate the level of productive, protein-coding transcripts (Reddy et al., 2013; Staiger and Brown, 2013a). Transcript level RNA-seq will be required to more accurately measure protein-coding transcripts of splicing factors to identify candidate regulators of AS of the core clock genes analyzed here.

Another interesting observation is the noticeable variation in the AS transcripts across the different organs analyzed (Figure 3). The differences in transcript expression between source and sink tissues might reflect their metabolic differences. While L1 is a fully photosynthetically active leaf in sugarcane, therefore undergoing photosynthesis, both internodes sink in the assimilated carbon for different purposes: cell division and elongation in I1 and sucrose storage in I5. In *Arabidopsis*, sucrose has been shown to decrease *PRR7* levels, which decrease *ScLHY* levels as a consequence (Haydon et al., 2013; Frank et al., 2018). Thus, differences in *ScPRR37* I6R and *ScLHY* I1R levels (Figures 3 and S8) could be due to the sucrose that is stored in the internodes. In turn, these differences in the levels of circadian clock genes might affect circadian clock outputs. We have found that transcriptional rhythms are very organ-specific in sugarcane (Dantas et al., 2019).

In *Arabidopsis*, the circadian clock is associated with photosynthesis (Dodd et al., 2009), cell division (Fung-Uceda et al., 2018), and sugar accumulation (Graf et al., 2010; Graf and Smith, 2011; Ko et al., 2016). Taken together, these data suggest

that the circadian clock might regulate the different sugarcane organs in distinctive ways. It is already known that metabolites can feedback to regulate the circadian clock, as data show in *Arabidopsis* that photosynthetic sugars regulate clock functioning (Haydon et al., 2013). Considering that in *Arabidopsis* different tissues are enriched with different levels of circadian clock transcripts (Endo et al., 2014), a similar phenomenon might occur in sugarcane and explain the different organ transcript expressions and profiles, as well as helping to keep each organ different metabolic and physiologic profiles.

We found that the splicing ratios from *ScLHY* AS transcripts are negatively correlated with temperature in all organs. Previous studies on the presence of AS in circadian clock genes in *Arabidopsis* and barley revealed that, under controlled conditions, the AS status of circadian clock genes is regulated by temperature, especially *LHY* and its paralog *CCA1* (James et al., 2012a; James et al., 2012b; Park et al., 2012; Seo et al., 2012; Kwon et al., 2014; Filichkin et al., 2015a; Calixto et al., 2016; Marshall et al., 2016; Calixto et al., 2018). Lower temperatures led to increased abundance of alternative non-productive forms of *LHY*, *PRR7* and *PRR5* in *Arabidopsis* (James et al., 2012a). This could affect the expression of fully spliced transcripts of the circadian clock genes, promoting a functional modulation in the circadian clock central oscillator, which might be reflected by altered temporal control of the clock outputs. In *Arabidopsis*, cold temperatures reduced the amplitude of *CCA1/LHY*, as well as disrupted the circadian clock function (Bieniawska et al., 2008). Considering the natural environment context, where plants like sugarcane face fluctuations in temperature on a daily and yearly basis, the continuous temperature-regulation of AS of the circadian clock network could have a more profound impact on metabolism and, ultimately, on crop yield.

The data in our work shows that from winter to summer, as the temperature increases (Figure S1B), the expression of alternative forms of the circadian clock genes decreases, noticeably for *ScLHY* (Figures 2A–D). In *Arabidopsis*, there is evidence linking the circadian clock with sugar accumulation as starch through *CCA1* and *LHY* (Ni et al., 2009; Miller et al., 2012; Ng et al., 2014) and in field-grown maize, a C4 plant like sugarcane, two *CCA1* homologs are associated to photosynthesis and, therefore, sugar accumulation (Ko et al., 2016). All this evidence allows us to speculate that there might be differences in the sugar accumulation by the field-grown sugarcane from winter to summer, but a metabolomic analysis focused on sucrose and hexoses content would be necessary to bring evidence to support such speculation.

Recent studies featuring experiments conducted in field conditions in *Arabidopsis*, rice and tomato highlight the differences in gene expression, circadian regulation and plant metabolism compared to experiments conducted inside growth chambers (Annunziata et al., 2017; Annunziata et al., 2018; Izawa et al., 2011b; Higashi et al., 2016; Shalit-Kaneh et al., 2018). Because AS has an impact on the regulation of gene expression, which impacts circadian regulation and plant metabolism, it is important to start investigating the dynamic adjustment of AS in

response to a fluctuating environment. RNA-seq data from time-series experiments revealed that the AS status of *Arabidopsis* transcriptome is widely responsive to changes in temperature (Calixto et al., 2018). By progressively lowering temperature, rapid changes in the spliced forms of transcripts were detected, which suggests that AS might also act to regulate low-temperature responses and how plants tolerate such stress (Calixto et al., 2018). The *Arabidopsis* circadian clock also acts in regulating plant abiotic stress tolerance (Grundy et al., 2015). This also suggests that both AS and the circadian clock might act in synergy to help plants to cope with temperature changes in both the short and long term. In the field, this regulation might be even more important, due to the unexpected fluctuations in light, temperature, and humidity in which plants are exposed.

Our data show that AS occurs in sugarcane circadian clock genes and that the different transcript isoforms show a dynamic expression profile in sugarcane grown under field conditions. Furthermore, *ScLHY* AS regulation correlates with temperature in sugarcane circadian clock genes. Thus, the changes in expression of alternative isoforms of *ScLHY* transcripts observed across winter and summer might illustrate the combined effect of both the circadian clock and AS regulation and AS in *ScLHY* might be a key mechanism that allows the continuous dynamic adjustment of the circadian clock by temperature in sugarcane. It is important to start further studies on the impact of the seasonal variation on the AS isoforms of the circadian clock gene expression and, ultimately, sugarcane metabolism and yield.

DATA AVAILABILITY STATEMENT

All datasets generated for this study are included in GitHub (<https://github.com/LabHotta/AlternativeSplicing>) and archived on Zenodo (<http://doi.org/10.5281/zenodo.3509232>).

AUTHOR CONTRIBUTIONS

CH and JB designed this research. LD and CH harvested the biological material and carried out BLAST analyses. LD processed all samples and carried out cloning and HR RT-PCRs. LD, CC, JB, and CH participated in the interpretation of genomic annotation and HR RT-PCR data. MD contributed with cloning. MC contributed with the plants and space for the field experiment. LD and CH drafted the manuscript. All authors participated in its correction and have read and approved the final manuscript. CH and JB acquired the funding.

FUNDING

The present study was supported by the São Paulo Research Foundation (FAPESP) [grant nos. 11/00818-8 and 15/06260-0; BIOEN Program], and by the Serrapilheira Institute (grant no. Serra-1708-16001). LD was supported by FAPESP scholarships

[grants 11/08897-4 and 15/10220-3]. CC and JB were supported by funding from the Biotechnology and Biological Sciences Research Council (BBSRC) [BB/K006568/1 and BB/N022807/1] and the Scottish Government Rural and Environment Science and Analytical Services division (RESAS) [to JB].

REFERENCES

- Alabadi, D., Oyama, T., Yanovsky, M. J., Harmon, F. G., Más, P., and Kay, S. A. (2001). Reciprocal regulation between TOC1 and LHY/CCA1 within the *Arabidopsis* circadian clock. *Science* 293, 880–883. doi: 10.1126/science.1061320
- Amorim, H. V., Lopes, M. L., De Castro Oliveira, J. V., Buckeridge, M. S., and Goldman, G. H. (2011). Scientific challenges of bioethanol production in Brazil. *Appl. Microbiol. Biotechnol.* 91, 1267–1275. doi: 10.1007/s00253-011-3437-6
- Annunziata, M. G., Apelt, F., Carillo, P., Krause, U., Feil, R., Mengin, V., et al. (2017). Getting back to nature: a reality check for experiments in controlled environments. *J. Exp. Bot.* 68, 4463–4477. doi: 10.1093/jxb/erx220
- Annunziata, M. G., Apelt, F., Carillo, P., Krause, U., Feil, R., Koehl, K., et al. (2018). Response of *Arabidopsis* primary metabolism and circadian clock to low night temperature in a natural light environment. *J. Exp. Bot.* 69, 4881–4895. doi: 10.1093/jxb/ery276
- Beyer, A. L., and Osheim, Y. N. (1988). Splice site selection, rate of splicing, and alternative splicing on nascent transcripts. *Genes Dev.* 2, 754–765. doi: 10.1101/gad.2.6.754
- Bieniaszka, Z., Espinoza, C., Schlereth, A., Sulpice, R., Hinch, D. K., and Hannah, M. A. (2008). Disruption of the *Arabidopsis* circadian clock is responsible for extensive variation in the cold-responsive transcriptome. *Plant Physiol* 147 (1), 263–279. doi: 10.1104/pp.108.118059
- Calixto, C. P. G., Waugh, R., and Brown, J. W. S. (2015). Evolutionary relationships among Barley and *Arabidopsis* core circadian clock and clock-associated genes. *J. Mol. Evol.* 80, 108–119. doi: 10.1007/s00239-015-9665-0
- Calixto, C. P. G., Simpson, C. G., Waugh, R., and Brown, J. W. S. (2016). Alternative splicing of Barley clock genes in response to low temperature. *PLoS One* 11, e0168028. doi: 10.1371/journal.pone.0168028
- Calixto, C. P. G., Guo, W., James, A. B., Tzioutziou, N. A., Entizne, J. C., Panter, P. E., et al. (2018). Rapid and dynamic alternative splicing impacts the *Arabidopsis* cold response transcriptome. *Plant Cell* 30, 1424–1444. doi: 10.1105/tpc.18.00177
- Calixto, C. P. G., Tzioutziou, N. A., James, A. B., Hornyik, C., Guo, W., Zhang, R., et al. (2019). Cold-dependent expression and alternative splicing of *Arabidopsis* long non-coding RNAs. *Front. Plant Sci.* 10, 235. doi: 10.3389/fpls.2019.00235
- Chamala, S., Feng, G., Chavarro, C., and Barbazuk, W. B. (2015). Genome-wide identification of evolutionarily conserved alternative splicing events in flowering plants. *Front. Bioeng. Biotechnol.* 3, 33. doi: 10.3389/fbioe.2015.00033
- Chan, A., Charron, C., van der Vossen, E., D'Hont, A., Town, C., Hervouet, C., et al. (2018). A mosaic monoploid reference sequence for the highly complex genome of sugarcane. *Nat. Commun.* 9, 1–10. doi: 10.1038/s41467-018-05051-5
- Chaudhary, S., Jabre, I., Reddy, A. S. N., Staiger, D., and Syed, N. H. (2019). Perspective on alternative splicing and proteome complexity in plants. *Trends In Plant Sci.* 496–506. doi: 10.1016/j.tplants.2019.02.006
- Cuadrado, A., Acevedo, R., Moreno Díaz De La Espina, S., Jouve, N., and De La Torre, C. (2004). Genome remodelling in three modern *S. officinarum* x *S. spontaneum* sugarcane cultivars. *J. Exp. Bot.* 55, 847–854. doi: 10.1093/jxb/erh093
- Dantas, L. L. B., de Lima, N. O., Nishiyama, M. Y., Carneiro, M. S., Souza, G. M., and Hotta, C. T. (2019). Rhythms of transcription in field-grown sugarcane are highly organ specific. *BioRx* doi: 10.1101/607002
- de Setta, N., Monteiro-Vitorello, C., Metcalfe, C., Cruz, G. M., Del Bem, L., Vicentini, R., et al. (2014). Building the sugarcane genome for biotechnology and identifying evolutionary trends. *BMC Genomics* 15, 540. doi: 10.1186/1471-2164-15-540
- D'Hont, A., Grivet, L., Feldmann, P., Rao, S., Berding, N., and Glaszmann, J. C. (1996). Characterization of the double genome structure of sugarcane cultivars (*Saccharum* ssp.) by molecular cytogenetics. *Mol. Gen. Genet.* 250, 405–413. doi: 10.1007/BF02174028
- D'Hont, A. (2005). Unraveling the genome structure of polyploids using FISH and GISH; examples of sugarcane and banana. *Cytogenetic Genome Res.* 109, 27–33. doi: 10.1159/000082378
- Ding, F., Cui, P., Wang, Z., Zhang, S., Ali, S., and Xiong, L. (2014). Genome-wide analysis of alternative splicing of pre-mRNA under salt stress in *Arabidopsis*. *BMC Genomics* 15, 431. doi: 10.1186/1471-2164-15-431
- Dodd, A. N., Love, J., and Webb, A. A. R. (2005). The plant clock shows its metal: circadian regulation of cytosolic free Ca²⁺. *Trends in Plant Sci.* 10, 15–21. doi: 10.1016/j.tplants.2004.12.001
- Dodd, A. N. (2009). Plant circadian clocks increase photosynthesis, growth, survival. *Science* 630, 630–633. doi: 10.1126/science.1115581
- Endo, M., Shimizu, H., Nohales, M., Araki, T., and Kay, S. (2014). Tissue-specific clocks in *Arabidopsis* show asymmetric coupling. *Nature* 515, 419–422. doi: 10.1038/nature13919
- FAO. (2015). FAOSTAT. Food and Agriculture Organization of the United Nations, Rome, Italy. <http://faostat.fao.org/default.aspx>
- Filichkin, S., and Mockler, T. C. (2012). Unproductive alternative splicing and nonsense mRNAs: a widespread phenomenon among plant circadian clock genes. *Biol. Direct* 7, 20. doi: 10.1186/1745-6150-7-20
- Filichkin, S. A., Priest, H. D., Givan, S. A., Shen, R., Bryant, D. W., Fox, S. E., et al. (2010). Genome-wide mapping of alternative splicing in *Arabidopsis thaliana*. *Genome Res.* 20 (1), 45–58. doi: 10.1101/gr.093302.109.2008
- Filichkin, S. A., Cumbie, J. S., Dharmawardhana, P., Jaiswal, P., Chang, J. H., Palusa, S. G., et al. (2015a). Environmental stresses modulate abundance and timing of alternatively spliced circadian transcripts in *Arabidopsis*. *Mol. Plant* 8, 207–227. doi: 10.1016/j.molp.2014.10.011
- Filichkin, S., Priest, H. D., Megraw, M., and Mockler, T. C. (2015b). Alternative splicing in plants: directing traffic at the crossroads of adaptation and environmental stress. *Curr. Opin. In Plant Biol.* 24, 125–135. doi: 10.1016/j.pbi.2015.02.008
- Fouquet, R., Martin, F., Fajardo, D. S., Gault, C. M., Gómez, E., Tseung, C.-W., et al. (2011). Maize rough endosperm3 encodes an RNA splicing factor required for endosperm cell differentiation and has a nonautonomous effect on embryo development. *Plant Cell* 23, 4280–4297. doi: 10.1105/tpc.111.092163
- Frank, A., Mattioli, C. C., Viana, A. J. C., Hearn, T. J., Kusakina, J., Belbin, F. E., et al. (2018). Circadian entrainment in *Arabidopsis* by the sugar-responsive transcription factor bZIP63. *Curr. Biol.* 28, 2597–2606.e6. doi: 10.1016/j.cub.2018.05.092
- Fung-Uceda, J., Lee, K., Seo, P. J., Polyn, S., De Veylder, L., and Mas, P. (2018). The circadian clock sets the time of DNA replication licensing to regulate growth in *Arabidopsis*. *Dev. Cell* 45, 101–113.e4. doi: 10.1016/j.devcel.2018.02.022
- Göhring, J., Jacak, J., and Barta, A. (2014). Imaging of endogenous messenger RNA splice variants in living cells reveals nuclear retention of transcripts inaccessible to nonsense-mediated decay in *Arabidopsis*. *Plant Cell* 26, 754–764. doi: 10.1105/tpc.113.118075
- Garcia, A. A. F., Mollinari, M., Marconi, T. G., Serang, O. R., Silva, R. R., Vieira, M. L. C., et al. (2013). SNP genotyping allows an in-depth characterisation of the genome of sugarcane and other complex autopolyploids. *Sci. Rep.* 3, 3399. doi: 10.1038/srep03399
- Garsmeur, O., Droc, G., Antonise, R., Grimwood, J., Potier, B., Aitken, K., et al. (2018). A mosaic monoploid reference sequence for the highly complex genome of sugarcane. *Nat. Commun.* 9, 2638. doi: 10.1038/s41467-018-05051-5
- Godoy Herz, M. A., Kubaczka, M. G., Brzyżek, G., Servi, L., Krzysztan, M., Simpson, C., et al. (2019). Light regulates plant alternative splicing through the

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fpls.2019.01614/full#supplementary-material>.

- control of transcriptional elongation. *Mol. Cell* 73 (5), 1066–1074. doi: 10.1016/j.molcel.2018.12.005
- Graf, A., and Smith, A. M. (2011). Starch and the clock: the dark side of plant productivity. *Trends Plant Sci.* 16, 169–175. doi: 10.1016/j.tplants.2010.12.003
- Graf, A., Schlereth, A., Stitt, M., and Smith, A. M. (2010). Circadian control of carbohydrate availability for growth in *Arabidopsis* plants at night. *Proc. Natl. Acad. Sci. U.S.A.* 107, 9458–9463. doi: 10.1073/pnas.0914299107
- Graveley, B. R. (2005). Mutually exclusive splicing of the insect Dscam Pre-mRNA directed by competing intronic RNA secondary structures. *Cell* 123, 65–73. doi: 10.1016/j.cell.2005.07.028
- Grundy, J., Stoker, C., and Carré, I. A. (2015). Circadian regulation of abiotic stress tolerance in plants. *Front. In Plant Sci.* 6, 648. doi: 10.3389/fpls.2015.00648
- Harmer, S. L. (2009). The circadian system in higher plants. *Annu. Rev. Plant Biol.* 60, 357–377. doi: 10.1146/annurev.arplant.043008.092054
- Haydon, M. J., Mielczarek, O., Robertson, F. C., Hubbard, K. E., and Webb, A. A. R. (2013). Photosynthetic entrainment of the *Arabidopsis thaliana* circadian clock. *Nature* 502, 689–692. doi: 10.1038/nature12603
- Higashi, T., Tanigaki, Y., Takayama, K., Nagano, A. J., Honjo, M. N., and Fukuda, H. (2016). Detection of diurnal variation of tomato transcriptome through the molecular timetable method in a sunlight-type plant factory. *Front. in Plant Sci.* 7, 1–9. doi: 10.3389/fpls.2016.00087
- Hotta, C. T., Nishiyama, M. Y., and Souza, G. M. (2013). Circadian rhythms of sense and antisense transcription in sugarcane, a highly polyploid crop. *PLoS One* 8, 1–16. doi: 10.1371/journal.pone.0071847
- Hsu, P. Y., and Harmer, S. L. (2014). Wheels within wheels: the plant circadian system. *Trends in Plant Sci.* 19, 240–249. doi: 10.1016/j.tplants.2013.11.007
- Iskandar, H. M., Simpson, R. S., Casu, R. E., Bonnett, G. D., Maclean, D. J., and Manners, J. M. (2004). Comparison of reference genes for quantitative real-time polymerase chain reaction analysis of gene expression in sugarcane. *Plant Mol. Biol. Rep.* 22, 325–337. doi: 10.1007/BF02772676
- Izawa, T., Mihara, M., Suzuki, Y., Gupta, M., Itoh, H., Nagano, A. J., et al. (2011a). Os-GIGANTEA confers robust diurnal rhythms on the global transcriptome of rice in the field. *Plant Cell* 23, 1741–1755. doi: 10.1105/tpc.111.083238
- Izawa, T., Mihara, M., Suzuki, Y., Gupta, M., Itoh, H., Nagano, A. J., et al. (2011b). Os-GIGANTEA confers robust diurnal rhythms on the global transcriptome of rice in the field. *Plant Cell* 23, 1741–1755. doi: 10.1105/tpc.111.083238
- Jabre, I., Reddy, A. S. N., Kalyna, M., Chaudhary, S., Khokhar, W., Byrne, L. J., et al. (2019). Does co-transcriptional regulation of alternative splicing mediate plant stress responses? *Nucleic Acids Res.* 47, 2716–2726. doi: 10.1093/nar/gkz121
- Jacob, A. G., and Smith, C. W. J. (2017). Intron retention as a component of regulated gene expression programs. *Hum. Genet.* 136, 1043–1057. doi: 10.1007/s00439-017-1791-x
- James, A. B., Syed, N. H., Bordage, S., Marshall, J., Nimmo, G. A., Jenkins, G. I., et al. (2012a). Alternative splicing mediates responses of the *Arabidopsis* circadian clock to temperature changes. *Plant Cell* 24, 961–981. doi: 10.1105/tpc.111.093948
- James, A. B., Syed, N. H., Bordage, S., Marshall, J., Nimmo, G. A., Jenkins, G. I., et al. (2012b). Alternative Splicing Mediates Responses of the *Arabidopsis* circadian clock to temperature changes. *Plant Cell* 24, 961–981. doi: 10.1105/tpc.111.093948
- James, A. B., Syed, N. H., Brown, J. W. S., and Nimmo, H. G. (2012c). Thermoplasticity in the plant circadian clock: how plants tell the temperature. *Plant Signaling Behav.* 7, 1219–1223. doi: 10.4161/psb.21491
- Jones, M. A., Williams, B. A., McNicol, J., Simpson, C. G., Brown, J. W. S., and Harmer, S. L. (2012). Mutation of *Arabidopsis* SPLICEOSOMAL TIMEKEEPER LOCUS1 causes circadian clock defects. *Plant Cell* 24, 4066–4082. doi: 10.1105/tpc.112.104828
- Kalyna, M., Simpson, C. G., Syed, N. H., Lewandowska, D., Marquez, Y., Kusenda, B., et al. (2012). Alternative splicing and nonsense-mediated decay modulate expression of important regulatory genes in *Arabidopsis*. *Nucleic Acids Res.* 40, 2454–2469. doi: 10.1093/nar/gkr932
- Khan, S., Rowe, S. C., and Harmon, F. G. (2010). Coordination of the maize transcriptome by a conserved circadian clock. *BMC Plant Biol.* 10, 126. doi: 10.1186/1471-2229-10-126
- Ko, D. K., Rohozinski, D., Song, Q., Taylor, S. H., Juenger, T. E., Harmon, F. G., et al. (2016). Temporal shift of circadian-mediated gene expression and carbon fixation contributes to biomass heterosis in Maize hybrids. *PLoS Genet.* 12, e1006197. doi: 10.1371/journal.pgen.1006197
- Kornblihtt, A. R., Schor, I. E., Alló, M., Dujardin, G., Petrillo, E., and Muñoz, M. J. (2013). Alternative splicing: a pivotal step between eukaryotic transcription and translation. *Nat. Rev. Mol. Cell Biol.* 14, 153–165. doi: 10.1038/nrm3525
- Kriechbaumer, V., Wang, P., Hawes, C., and Abell, B. M. (2012). Alternative splicing of the auxin biosynthesis gene YUCCA4 determines its subcellular compartmentation. *Plant J.* 70, 292–302. doi: 10.1111/j.1365-313X.2011.04866.x
- Kwon, Y.-J., Park, M.-J., Kim, S.-G., Baldwin, I. T., and Park, C.-M. (2014). Alternative splicing and nonsense-mediated decay of circadian clock genes under environmental stress conditions in *Arabidopsis*. *BMC Plant Biol.* 14, 136. doi: 10.1186/1471-2229-14-136
- Lai, A. G., Doherty, C. J., Mueller-Roeber, B., Kay, S. A., Schippers, J. H. M., and Dijkwel, P. P. (2012). CIRCADIAN CLOCK-ASSOCIATED 1 regulates ROS homeostasis and oxidative stress responses. *Proc. Natl. Acad. Sci.* 109, 17129–17134. doi: 10.1073/pnas.1209148109
- Liu, M., Yuan, L., Liu, N. Y., Shi, D. Q., Liu, J., and Yang, W. C. (2009). GAMETOPHYTIC FACTOR 1, involved in pre-mRNA splicing, is essential for megagametogenesis and embryogenesis in *Arabidopsis*. *J. Integr. Plant Biol.* 51, 261–271. doi: 10.1111/j.1744-7909.2008.00783.x
- Locke, J. C. W., Millar, A. J., and Turner, M. S. (2005). Modelling genetic networks with noisy and varied experimental data: the circadian clock in *Arabidopsis thaliana*. *J. Theor. Biol.* 234, 383–393. doi: 10.1016/j.jtbi.2004.11.038
- Lu, Y., Gehan, J. P., and Sharkey, T. D. (2005). Daylength and circadian effects on starch degradation and maltose metabolism. *Plant Physiol.* 138, 2280–2291. doi: 10.1104/pp.105.061903
- Marquez, Y., Brown, J. W. S., 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. doi: 10.1101/gr.134106.111
- Marshall, C. M., Tartaglio, V., Duarte, M., and Harmon, F. G. (2016). The *Arabidopsis* sickle mutant exhibits altered circadian clock responses to cool temperatures and temperature-dependent alternative splicing. *Plant Cell* 28, 2560–2575. doi: 10.1105/tpc.16.00223
- Mastrangelo, A. M., Marone, D., Laidò, G., De Leonardi, A. M., and De Vita, P. (2012). Alternative splicing: enhancing ability to cope with stress via transcriptome plasticity. *Plant Sci.* 185–186, 40–49. doi: 10.1016/j.plantsci.2011.09.006
- McClung, C. (2019). The plant circadian oscillator. *Biology* 8, 14. doi: 10.3390/biology8010014
- McCormick, R. F., Truong, S. K., Sreedasyam, A., Jenkins, J., Shu, S., Sims, D., et al. (2018). The Sorghum bicolor reference genome: improved assembly, gene annotations, a transcriptome atlas, and signatures of genome organization. *Plant J.* 93, 338–354. doi: 10.1111/tpj.13781
- Millar, A. J. (2016). The intracellular dynamics of circadian clocks reach for the light of ecology and evolution. *Annu. Rev. Plant Biol.* 67, annurev-arplant-043014-115619. doi: 10.1146/annurev-arplant-043014-115619
- Miller, M., Zhang, C., and Chen, Z. J. (2012). Ploidy and hybridity effects on growth vigor and gene expression in *Arabidopsis thaliana* hybrids and their parents. *G3 (Bethesda)* 2, 505–513. doi: 10.1534/g3.112.002162
- Min, X. J., Powell, B., Braessler, J., Meinken, J., Yu, F., and Sablok, G. (2015). Genome-wide cataloging and analysis of alternatively spliced genes in cereal crops. *BMC Genomics* 16, 721. doi: 10.1186/s12864-015-1914-5
- Moll, C., Von Lyncker, L., Zimmermann, S., Kägi, C., Baumann, N., Twell, D., et al. (2008). CLO/GFA1 and ATO are novel regulators of gametic cell fate in plants. *Plant J.* 56, 913–921. doi: 10.1111/j.1365-313X.2008.03650.x
- Moore, P. (1995). Temporal and spatial regulation of sucrose accumulation in the sugarcane stem. *Aust. J. Plant Physiol.* 22, 661. doi: 10.1071/PP950661
- Murakami, M., Tago, Y., Yamashino, T., and Mizuno, T. (2007). Comparative overviews of clock-associated genes of *Arabidopsis thaliana* and *Oryza sativa*. *Plant Cell Physiol.* 48, 110–121. doi: 10.1093/pcp/pcl043

- Nagano, A. J., Sato, Y., Mihara, M., Antonio, B. A., Motoyama, R., Itoh, H., et al. (2012). Deciphering and prediction of transcriptome dynamics under fluctuating field conditions. *Cell* 151, 1358–1369. doi: 10.1016/j.cell.2012.10.048
- Nagashima, Y., Mishiba, K.-I., Suzuki, E., Shimada, Y., Iwata, Y., and Koizumi, N. (2011). *Arabidopsis* IRE1 catalyzes unconventional splicing of bZIP60 mRNA to produce the active transcription factor. *Sci. Rep.* 1, 29. doi: 10.1038/srep00029
- Nakamichi, N., Kiba, T., Henriques, R., Mizuno, T., Chua, N.-H., and Sakakibara, H. (2010). PSEUDO-RESPONSE REGULATORS 9, 7, and 5 are transcriptional repressors in the *Arabidopsis* circadian clock. *Plant Cell* 22, 594–605. doi: 10.1105/tpc.109.072892
- Ng, D. W.-K., Miller, M., Yu, H. H., Huang, T.-Y., Kim, E.-D., Lu, J., et al. (2014). A role for CHH methylation in the parent-of-origin effect on altered circadian rhythms and biomass heterosis in *Arabidopsis* intraspecific hybrids. *Plant Cell* 26, 2430–2440. doi: 10.1105/tpc.113.115980
- Ni, Z., Kim, E.-D., Ha, M., Lackey, E., Liu, J., Zhang, Y., et al. (2009). Altered circadian rhythms regulate growth vigour in hybrids and allopolyploids. *Nature* 457, 327–331. doi: 10.1038/nature07523
- Oesterreich, F. C., Herzel, L., Strauber, K., Hujer, K., Howard, J., and Neugebauer, K. M. (2016). Splicing of nascent RNA coincides with intron exit from RNA polymerase II. *Cell* 165, 372–381. doi: 10.1016/j.cell.2016.02.045
- Palusa, S. G., Ali, G. S., and Reddy, A. S. N. (2007). Alternative splicing of pre-mRNAs of *Arabidopsis* serine/arginine-rich proteins: regulation by hormones and stresses. *Plant J.* 49, 1091–1107. doi: 10.1111/j.1365-313X.2006.03020.x
- Para, A., Farré, E. M., Imaizumi, T., Pruneda-Paz, J. L., Harmon, F. G., and Kay, S. A. (2007). PRR3 Is a vascular regulator of TOC1 stability in the *Arabidopsis* circadian clock. *Plant Cell* 19, 3462–3473. doi: 10.1105/tpc.107.054775
- Park, M.-J., Seo, P. J., and Park, C.-M. (2012). CCA1 alternative splicing as a way of linking the circadian clock to temperature response in *Arabidopsis*. *Plant Signal Behav.* 7, 1194–1196. doi: 10.4161/psb.21300
- Pokhilko, A., Fernández, A. P., Edwards, K. D., Southern, M. M., Halliday, K. J., and Millar, A. J. (2012). The clock gene circuit in *Arabidopsis* includes a repressilator with additional feedback loops. *Mol. Syst. Biol.* 8, 574. doi: 10.1038/msb.2012.6
- Reddy, A. S. N., Marquez, Y., Kalyna, M., and Barta, A. (2013). Complexity of the alternative splicing landscape in plants. *Plant Cell* 25, 3657–3683. doi: 10.1105/tpc.113.117523
- Remy, E., Cabrito, T. R., Baster, P., Batista, R. A., Teixeira, M. C., Friml, J., et al. (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. doi: 10.1105/tpc.113.110353
- Riaño-Pachón, D. M., and Mattiello, L. (2017). Draft genome sequencing of the sugarcane hybrid SP80-3280. *F1000Res* 6, 861. doi: 10.12688/f1000research.11859.2
- Richards, C. L., Rosas, U., Banta, J., Bhambhra, N., and Purugganan, M. D. (2012). Genome-wide patterns of *Arabidopsis* gene expression in nature. *PLoS Genet.* 8 (4), 1–14. doi: 10.1371/journal.pgen.1002662
- Rosloski, S. M., Singh, A., Jali, S. S., Balasubramanian, S., Weigel, D., and Grbic, V. (2013). Functional analysis of splice variant expression of MADS AFFECTING FLOWERING 2 of *Arabidopsis thaliana*. *Plant Mol. Biol.* 81, 57–69. doi: 10.1007/s11103-012-9982-2
- Saldi, T., Cortazar, M. A., Sheridan, R. M., and Bentley, D. L. (2016). Coupling of RNA polymerase II transcription elongation with pre-mRNA splicing. *J. Mol. Biol.* 428, 2623–2635. doi: 10.1016/j.jmb.2016.04.017
- Sato, Y., Antonio, B., Namiki, N., Motoyama, R., Sugimoto, K., Takehisa, H., et al. (2011). Field transcriptome revealed critical developmental and physiological transitions involved in the expression of growth potential in japonica rice. *BMC Plant Biol.* 11, 10. doi: 10.1186/1471-2229-11-10
- Seo, P. J., Hong, S. Y., Kim, S. G., and Park, C. M. (2011). Competitive inhibition of transcription factors by small interfering peptides. *Trends In Plant Sci.* 16, 541–549. doi: 10.1016/j.tplants.2011.06.001
- Seo, P. J., Park, M.-J., Lim, M.-H., Kim, S.-G., Lee, M., Baldwin, I. T., et al. (2012). A self-regulatory circuit of CIRCADIAN CLOCK-ASSOCIATED1 underlies the circadian clock regulation of temperature responses in *Arabidopsis*. *Plant Cell* 24, 2427–2442. doi: 10.1105/tpc.112.098723
- Severing, E. I., van Dijk, A. D. J., Morabito, G., Busscher-Lange, J., Immink, R. G. H., and van Ham, R. C. H. J. (2012). Predicting the impact of alternative splicing on plant MADS domain protein function. *PLoS One* 7 (1), 1–13. doi: 10.1371/journal.pone.0030524
- Shalit-Kaneh, A., Kumimoto, R. W., Filkov, V., and Harmer, S. L. (2018). Multiple feedback loops of the *Arabidopsis* circadian clock provide rhythmic robustness across environmental conditions. *Proc. Natl. Acad. Sci. U.S.A.* 115, 7147–7152. doi: 10.1073/pnas.1805524115
- Shang, X., Cao, Y., and Ma, L. (2017). Alternative splicing in plant genes: A means of regulating the environmental fitness of plants. *Int. J. Mol. Sci.* 18 (2), 1–18. doi: 10.3390/ijms18020432
- Shen, Y., Zhou, Z., Wang, Z., Li, W., Fang, C., Wu, M., et al. (2014). Global dissection of alternative splicing in paleopolyploid soybean. *Plant Cell* 26, 996–1008. doi: 10.1105/tpc.114.122739
- Simpson, C. G., Fuller, J., Maronova, M., Kalyna, M., Davidson, D., McNicol, J., et al. (2007). Monitoring changes in alternative precursor messenger RNA splicing in multiple gene transcripts. *Plant J.* 53, 1035–1048. doi: 10.1111/j.1365-313X.2007.03392.x
- Simpson, C. G., Fuller, J., Rapazote-Flores, P., Mayer, C. D., Calixto, C. P. G., Milne, L., et al. (2019). High-resolution RT-PCR analysis of alternative barley transcripts. *Methods In Mol. Biol.* 1900, 269–281. doi: 10.1007/978-1-4939-8944-7_17
- Souza, G. M., Van Sluys, M.-A., Lembke, C. G., Lee, H., Margarido, G. R. A., Hotta, C. T., et al. (2019). Assembly of the 373k gene space of the polyploid sugarcane genome reveals reservoirs of functional diversity in the world's leading biomass crop. *GigaScience* 8 (12), giz129. doi: 10.1093/gigascience/giz129
- Staiger, D., and Brown, J. W. S. (2013a). Alternative splicing at the intersection of biological timing, development, and stress responses. *Plant Cell* 25, 3640–3656. doi: 10.1105/tpc.113.113803
- Staiger, D., and Brown, J. W. S. (2013b). Alternative splicing at the intersection of biological timing, development, and stress responses. *Plant Cell* 25, 3640–3656. doi: 10.1105/tpc.113.113803
- Sugliani, M., Brambilla, V., Clercx, E. J. M., Koornneef, M., and Soppe, W. J. J. (2010). The conserved splicing factor SUA controls alternative splicing of the developmental regulator ABI3 in *Arabidopsis*. *Plant Cell* 22, 1936–1946. doi: 10.1105/tpc.110.074674
- Syed, N. H., Kalyna, M., Marquez, Y., Barta, A., and Brown, J. W. S. (2012). Alternative splicing in plants - coming of age. *Trends In Plant Sci.* 17, 616–623. doi: 10.1016/j.tplants.2012.06.001
- Szakonyi, D., and Duque, P. (2018). Alternative splicing as a regulator of early plant development. *Front. In Plant Sci.* 9, 1–9. doi: 10.3389/fpls.2018.01174
- Thatcher, S. R., Zhou, W., Leonard, A., Wang, B.-B., Beatty, M., Zastrow-Hayes, G., et al. (2014). Genome-wide analysis of alternative splicing in ze mays: landscape and genetic regulation. *Plant Cell Online* 26, 3472–3487. doi: 10.1105/tpc.114.130773
- Vaneechoutte, D., Estrada, A. R., Lin, Y. C., Loraine, A. E., and Vandepoele, K. (2017). Genome-wide characterization of differential transcript usage in *Arabidopsis thaliana*. *Plant J.* 92, 1218–1231. doi: 10.1111/tpj.13746
- Vicentini, R., Del Bem, L. E. V., Van Sluys, M. A., Nogueira, F. T. S., and Vincentz, M. (2012). Gene content analysis of sugarcane public ESTs reveals thousands of missing coding-genes and an unexpected pool of grasses conserved ncRNAs. *Trop. Plant Biol.* 5, 199–205. doi: 10.1007/s12042-012-9103-z
- Wang, X., Yang, M., Ren, D., Terzaghi, W., Deng, X. W., and He, G. (2018). Cis-regulated alternative splicing divergence and its potential contribution to environmental responses in *Arabidopsis*. *Plant J.* 97, 555–570. doi: 10.1111/tpj.14142
- Zeilinger, M. N., Farré, E. M., Taylor, S. R., Kay, S., and Doyle, F. J. (2006). A novel computational model of the circadian clock in *Arabidopsis* that incorporates PRR7 and PRR9. *Mol. Syst. Biol.* 2, 58. doi: 10.1038/msb4100101
- Zhang, X., Fang, X., Li, R., Li, S., Zhang, G., Tao, Y., et al. (2010). Deep RNA sequencing at single base-pair resolution reveals high complexity of the rice transcriptome. *Genome Res.* 20, 646–654. doi: 10.1101/gr.100677.109

- Zhang, Z., Zhang, S., Zhang, Y., Wang, X., Li, D., Li, Q., et al. (2011). Arabidopsis floral initiator SKB1 confers high salt tolerance by regulating transcription and pre-mRNA splicing through altering histone H4R3 and small nuclear ribonucleoprotein LSM4 methylation. *Plant Cell* 23, 396–411. doi: 10.1105/tpc.110.081356
- Zhang, Q., Zhang, X., Wang, S., Tan, C., Zhou, G., and Li, C. (2016). Involvement of alternative splicing in Barley seed germination. *PloS One* 11, e0152824. doi: 10.1371/journal.pone.0152824
- Zhang, J., Zhang, X., Tang, H., Zhang, Q., Hua, X., Ma, X., et al. (2018). Allele-defined genome of the autopolyploid sugarcane *Saccharum spontaneum* L. *Nat. Genet.* 50, 1565. doi: 10.1038/s41588-018-0237-2

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2019 Dantas, Calixto, Dourado, Carneiro, Brown and Hotta. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



Thermo-Sensitive Alternative Splicing of *FLOWERING LOCUS M* Is Modulated by Cyclin-Dependent Kinase G2

Candida Nibau^{1*}, Marçal Gallemí², Despoina Dadarou^{1†}, John H. Doonan¹ and Nicola Cavallari^{2,3*}

OPEN ACCESS

Edited by:

Kranthi Kiran Mandadi,
Texas A&M University,
United States

Reviewed by:

Lee Jeong Hwan,
Chonbuk National University,
South Korea
Dong-Hwan Kim,
Chung-Ang University,
South Korea

*Correspondence:

Nicola Cavallari
nicola.cavallari@ist.ac.at
Candida Nibau
csn@aber.ac.uk

†Present address:

Despoina Dadarou,
School of Life Sciences,
University of Warwick, Coventry,
United Kingdom

Specialty section:

This article was submitted to
Plant Physiology,
a section of the journal
Frontiers in Plant Science

Received: 29 July 2019

Accepted: 29 November 2019

Published: 22 January 2020

Citation:

Nibau C, Gallemí M, Dadarou D,
Doonan JH and Cavallari N (2020)
Thermo-Sensitive Alternative
Splicing of *FLOWERING*
LOCUS M Is Modulated by
Cyclin-Dependent Kinase G2.
Front. Plant Sci. 10:1680.
doi: 10.3389/fpls.2019.01680

¹ Institute of Biological, Environmental, and Rural Sciences, Aberystwyth University, Aberystwyth, United Kingdom, ² Institute of Science and Technology Austria, Klosterneuburg, Austria, ³ Max F. Perutz Laboratories, Medical University of Vienna, Vienna, Austria

The ability to sense environmental temperature and to coordinate growth and development accordingly, is critical to the reproductive success of plants. Flowering time is regulated at the level of gene expression by a complex network of factors that integrate environmental and developmental cues. One of the main players, involved in modulating flowering time in response to changes in ambient temperature is *FLOWERING LOCUS M* (*FLM*). *FLM* transcripts can undergo extensive alternative splicing producing multiple variants, of which *FLM-β* and *FLM-δ* are the most representative. While *FLM-β* codes for the flowering repressor *FLM* protein, translation of *FLM-δ* has the opposite effect on flowering. Here we show that the cyclin-dependent kinase G2 (*CDKG2*), together with its cognate cyclin, *CYCLYN L1* (*CYCL1*) affects the alternative splicing of *FLM*, balancing the levels of *FLM-β* and *FLM-δ* across the ambient temperature range. In the absence of the *CDKG2/CYCL1* complex, *FLM-β* expression is reduced while *FLM-δ* is increased in a temperature dependent manner and these changes are associated with an early flowering phenotype in the *cdkg2* mutant lines. In addition, we found that transcript variants retaining the full *FLM* intron 1 are sequestered in the cell nucleus. Strikingly, *FLM* intron 1 splicing is also regulated by *CDKG2/CYCL1*. Our results provide evidence that temperature and CDKs regulate the alternative splicing of *FLM*, contributing to flowering time definition.

Keywords: alternative splicing, cyclin-dependent kinase, temperature, *FLOWERING LOCUS M*, flowering time, *Arabidopsis thaliana*

INTRODUCTION

Reproductive success in a constantly changing environment is a major challenge for all kingdoms of life and often involves adjusting behavior or development according to prevailing conditions. Higher plants, for example, can maximize their reproductive chances by timing their flowering to suit geographic location and seasonal weather patterns (Amasino, 2010).

While light is considered a master input in the transition to flowering (Liu, 2001; Weller et al., 2001; Searle and Coupland, 2004; Mattson and Erwin, 2005), temperature is also critical and can promote or delay flowering according to the species (Balasubramanian et al., 2006; Nakano et al., 2013; Romera-Branchat et al., 2014). Moreover, heat pulses can have mixed effects (Bouché et al., 2015) while prolonged cold exposure (a process called vernalization) has been long known to accelerate flowering in many species adapted to post-winter flowering (Chouard, 1960).

The switch from the vegetative to reproductive phase is coordinated by a large number of genes scattered across several pathways but a few *loci* can have a wide effect (Srikanth and Schmid, 2011; Strange et al., 2011; Song et al., 2013; Susila et al., 2018). *FLOWERING LOCUS C (FLC)*, for example, has a major role in vernalization, ensuring that flowering does not occur until after winter, and variation at this *locus* has been implicated in determining fitness across geographic locations and plant lineages (Nordborg and Bergelson, 1999; Caicedo et al., 2004; Reeves et al., 2007). In this case, transition to flowering demands antisense-mediated chromatin silencing at the *FLC* locus and involves a complex regulatory array composed of *FLC* activators, like *FRIGIDA*, as well as repressors, the autonomous pathway, and cold (Whittaker and Dean, 2017).

Many *Arabidopsis* ecotypes require several weeks of vernalization to accelerate flowering (Lempe et al., 2005; Shindo et al., 2006), while commonly used accessions flower without prior exposure to cold (Johanson et al., 2000). Consequently, species-specific mechanisms regulate flowering time at ambient temperature and few have been reported at a detailed molecular level (Samach and Wigge, 2005; Lee et al., 2008; Wigge, 2013).

Unlike mammals (Vriens et al., 2014), plants lack a clear class of thermoreceptors but phytochrome B (phyB) and phototropins have been shown to fulfill dual roles as both thermo- and light-sensors (Jung et al., 2016; Legris et al., 2016; Fujii et al., 2017). In addition, many of the factors involved in temperature sensitive decisions in plants are still unknown and, therefore, our understanding of the molecular correlations in the temperature sensing/response pathways remains sketchy.

Increasing evidence points to a role of messenger RNA (mRNA) splicing as an endogenous “molecular thermometer” in temperature adaptations (Capovilla et al., 2015). Splicing is the removal of intronic (mostly non-coding) sequences from a primary transcript (pre-mRNA) to form the mature messenger RNA (mRNA) by a multi-megadalton complex called the spliceosome (Will and Luhrmann, 2011; Lee and Rio, 2015). Combinatorial recognition and usage of distinct nucleotide sequences (splice sites) on the pre-mRNA can lead to the assembly of multiple different transcripts in a process called alternative splicing (AS).

In higher plants, AS affects 60–70% of intron-containing genes (Chamala et al., 2015; Zhang et al., 2015). Hence, AS is a transcriptome-wide mechanism that has the potential to profoundly affect the level of gene expression in response to environmental stimuli. Indeed, generation of nonproductive transcript isoforms, targeted for non-sense-mediated mRNA

decay (NMD) (Kalyna et al., 2012), or translation of protein variants with altered amino acid sequence and function can quickly modify the cell proteome (Marquez et al., 2015).

Several reports exploring different temperature ranges and environments have shown that AS plays a critical role in response to extreme (Mastrangelo et al., 2012; Leviatan et al., 2013; Staiger and Brown, 2013; Hartmann et al., 2016; Klepikova et al., 2016; Calixto et al., 2018; Laloum et al., 2018) as well as to very small variations in ambient temperature (Streitner et al., 2013; Capovilla et al., 2015; Pajoro et al., 2017; Verhage et al., 2017; Capovilla et al., 2018; James et al., 2018).

In *Arabidopsis*, ambient temperature modulation of flowering time involves the FLC-related MADS-box transcription factor *FLOWERING LOCUS M (FLM, MAF1)*. Loss-of-function mutations in *FLM* reduce the temperature dependency of flowering suggesting its role as a repressor (Scortecci et al., 2001; Werner et al., 2005). Indeed, *FLM* modulates flowering time over a wide temperature range (from 5 to 23°C) and can bind, like *FLC* (Lee et al., 2007), to the *SHORT VEGETATIVE PHASE* protein (SVP), to form a potent *FLM*-SVP repressor complex (Lee et al., 2013; Posé et al., 2013). Besides *FLM*, flowering time is also regulated by the other *FLC*-clade proteins, *MAF2*–*MAF5* (Ratcliffe, 2003; Li et al., 2008; Gu et al., 2013; Lee et al., 2013; Airolidi et al., 2015; Theißen et al., 2018).

Temperature information on *FLM* gene expression is integrated at the post-transcriptional level by the interplay of AS events leading to the production of several *FLM* mRNA forms (Capovilla et al., 2017). In the reference accession, Columbia-0 (Col-0), two of these variants, *FLM-β* and *FLM-δ*, are the predominant transcripts, which result from the alternative usage of the mutually exclusive exons 2 (*FLM-β*) and 3 (*FLM-δ*) (Lee et al., 2013; Posé et al., 2013).

The resulting proteins *FLM-β* and *FLM-δ* have been implicated in repressing or promoting flowering respectively. In particular, *FLM-β* was found to bind both SVP and to promoter regions of regulated target genes (Posé et al., 2013). Hence, *FLM-β* has been recognized as the real protagonist in the temperature dependent repressor complex while the function of *FLM-δ* as flowering promoter has been much debated (Capovilla et al., 2017). The effect of other *FLM* isoforms in flowering is also poorly understood.

Recently, specific splicing factors have been reported to modulate flowering time by affecting the balance between *FLM-β* and *FLM-δ*, like the U2 auxiliary factors *ATU2AF65A* and *ATU2AF65B*, the glycine rich proteins *ATGRP7* and *ATGRP8* and Splicing Factor 1 *ATSF1* (Lee et al., 2017; Park et al., 2019; Steffen et al., 2019).

Cyclin-dependent kinases (CDKs), an evolutionarily conserved group of serine/threonine kinases initially implicated in cell cycle control (Strausfeld et al., 1996; Rane et al., 1999; Sherr and Roberts, 1999), are involved in pre-mRNA processing through interaction with the spliceosome components (Ko et al., 2001; Hu et al., 2003; Loyer et al., 2005; Even et al., 2006; Cheng et al., 2012). In plants, CDKs regulate a myriad of developmental processes including flowering time. CDKC, for instance is part of the positive transcription elongation factor b (P-TEFb) that

phosphorylates the C-terminal domain of RNA polymerase II (PolII) (Cui et al., 2007), modulates the localization of spliceosome components (Kitsios et al., 2008) and can regulate flowering time through promoting expression of an *FLC* antisense transcript called *COOLAIR* (Wang et al., 2014). The CDKG group is the most closely related to mammalian CDKs (Menges et al., 2005; Umeda, 2005) that are involved in mRNA processing (Bartkowiak et al., 2010; Chen et al., 2006; Loyer et al., 2005; Even et al., 2006) and have been also shown to regulate splicing (Huang et al., 2013; Cavallari et al., 2018), meiosis (Zheng et al., 2014) and flowering responses (Ma et al., 2015).

In *Arabidopsis*, two closely related genes, *CDKG1* and *CDKG2*, encode for the catalytic subunit of the kinase, which physically interacts with the regulatory subunit, *CYCLIN L1* (*CYCL1*) (Van Leene et al., 2010). *CDKG1*, *CDKG2*, and *CYCL1* have a role in mRNA splicing (Xu et al., 2012; Huang et al., 2013) forming part of an ambient temperature responsive AS cascade targeting genes involved in splicing (Cavallari et al., 2018). Moreover, *CDKG1* is required for chromosome pairing and recombination at high ambient temperature (Zheng et al., 2014) while *CDKG2* was reported as a negative regulator of flowering (Ma et al., 2015) although the molecular pathway involved was not identified.

Here, we show that the early flowering phenotype in *cdkg2-1* as well as in the double *cdkg2-1;cycl1-1* mutant lines is maintained across the ambient temperature range and under different light conditions (both long and short day). Early flowering is associated with impaired AS of *FLM* transcripts as mutants in both the kinase and cyclin genes showed differential integration of temperature cues into *FLM* mRNA. Specifically, *CDKG2* and *CYCL1*, but not *CDKG1*, are required for balancing *FLM-β* and *FLM-δ* levels across the ambient temperature range. Moreover, lack of *CDKG2* and *CYCL1* also affect the correct processing of the alternative introns 1 and 4 in *FLM* mRNAs. In addition, we report that mRNA variants retaining *FLM* intron 1 are sequestered in the cell nucleus.

Taken together our data provide evidence that the temperature pathways and the *CDKG2/CYCL1* complex converge on the regulation of *FLM* AS to fine tune the flowering process.

MATERIALS AND METHODS

Plant Materials and Growth Conditions

The wild type Columbia (Col-0) and mutant stocks *cdkg1-1* (SALK_075762), *cdkg2-1* (SALK_012428), and *cycl1-1* (SAIL_285_G10) used in this study were obtained from the Nottingham Arabidopsis Stock Centre and have previously been described (Zheng et al., 2014; Ma et al., 2015; Cavallari et al., 2018).

For analysis of the splicing events, plants were grown in Petriplates containing plant medium (0.5x MS salts and vitamins, pH 5.8, 0.7% plant agar) for 2 weeks at 23°C under either long day (LD) conditions (16 h light, 8 h dark) or short day (SD) conditions (8 h light, 16 h dark). Plants in LD or SD were then

transferred to 15, 23, or 27°C for 2 days and collected for mRNA isolation. For the experiments listed above, Philips GreenPower LED production modules were used to provide a combination of red (660 nm)/far red (720 nm)/blue (455 nm), light with a photon density of about 140 $\mu\text{molm}^{-2}\text{s}^{-1}$ \pm 20%.

For the flowering experiments, seeds were sown in pots containing soil mix (80% Levington F2 and 20% sand) and placed at 15, 23, or 27°C either in LD (16 h light, 8 h dark) or SD (8 h light, 16 h dark). The light was provided by Sylvania 840 lamps and the light intensity 150 $\mu\text{molm}^{-2}\text{s}^{-1}$ for LD and 250 $\mu\text{molm}^{-2}\text{s}^{-1}$ for SD. Flowering was scored by counting the number of rosette leaves at bolting for each genotype.

Ribonucleic Acid Extraction, Real Time, and Quantitative Polymerase Chain Reaction

Total RNA (3–5 seedlings per sample) was extracted from whole rosettes using the RNeasy Plant Mini kit (Qiagen). One microgram of total mRNA was used to generate cDNA using iScript™ cDNA Synthesis Kit (Bio-Rad). The primers used for the analysis of the AS of the different genes are listed in **Supplementary Table 2**. Three hundred eighty-four-well plates (Roche) were loaded using a JANUS Automated Workstation (PerkinElmer) with a 5 μl reaction containing 2.5 μl Luna® Universal qPCR Master Mix (New England Biolabs). Quantitative PCRs (qPCRs) were performed using the LightCycler 480 (Roche). Samples ($n \geq 3$) were measured in technical triplicates and expression of *PP2AA3* (AT1G13320) was used as a reference (Czechowski, 2005). Data were analyzed using the LightCycler® 480 Software (Roche).

Construct Generation and Plant Transformation

For transient expression in *Nicotiana benthamiana* leaves, the *CDKG2-GFP* (Cavallari et al., 2018) and *RSP34-RFP* (Lorković et al., 2004) cassettes were cloned into the pEAQ-HT-DEST2 vector (Sainsbury et al., 2009) and transformed into *Agrobacterium tumefaciens* strain LBA4404. Leaf infiltration was performed as described (Sainsbury et al., 2009). After 5 days, leaves were harvested for confocal imaging using a Leica TCS SP5 II confocal laser scanning microscope (CLSM) controlled by Leica LAS-AF software.

Protoplast Isolation and Subsequent Cell Fractionation

Mesophyll protoplasts were isolated from 3-week-old Col-0 plants as described by Wu et al. (Wu et al., 2009). Subsequent cell fractions were prepared as described by Goehring et al. (Goehring et al., 2014) with slight modifications. Briefly, 2×10^6 *Arabidopsis thaliana* mesophyll protoplasts were resuspended in 1 ml NIB lysis buffer [10 mM 2-(N-morpholino) ethanesulfonic acid-potassium hydroxide pH 5.5, 200 mM sucrose, 2.5 mM ethylenediaminetetraacetic acid, 2.5 mM dithiothreitol, 0.1 mM spermine, 10 mM NaCl, 0.2% Triton X-100, 1 U/ μl RNasin (Promega)] and lysed using a 25 G gauge needle (6 to 10 passages). Complete lysis was confirmed by light microscopy.

For the total fraction, 100 μ l of lysed cells were immediately resuspended in 1 ml TRIzol (Ambion) and kept on ice until the remaining fractions were processed. The lysate was pelleted for 10 min at 500 g and 1 ml of supernatant, which represents the cytoplasmic fraction, was removed, and centrifuged for another 15 min at 10,000 g. Eight hundred μ l of supernatant was resuspended in 8 ml TRIzol and the pellet, which represents the nuclear fraction, resuspended in 4 ml NRBT (20 mM Tris-HCl pH 7.5, 25% glycerol, 2.5 mM MgCl₂, 0.2% Triton X-100), centrifuged at 500 g for 10 min and washed three times. After washing, the nuclear pellet was resuspended in 500 μ l NRB2 (20 mM Tris-HCl pH 7.5, 250 mM sucrose, 10 mM MgCl₂, 0.5% Triton X-100, 5 mM β -mercaptoethanol) and carefully overlaid on top of 500 μ l NRB3 (20 mM Tris-HCl pH 7.5, 1.7 M sucrose, 10 mM MgCl₂, 0.5% Triton X-100, 5 mM β -mercaptoethanol) and centrifuged at 16,000 g for 45 min. Finally, the nuclear pellet was resuspended in 1 ml TRIzol and RNA was isolated following the manufacturer's instructions. Samples for protein analysis [total (whole protoplasts), cytoplasmic, and nuclear] were also kept.

Protein Extraction and Western Blotting

Protein samples from the fractionation experiments (total, cytoplasmic, and nuclear) were resuspended in sample loading buffer and heated up at 65°C for 10 min before loading in a 10–20% polyacrylamide gradient gel (Bio-Rad) and transferred to polyvinylidene fluoride membranes. Membranes were probed with anti-H3 antibody (Abcam 1791) or anti-alcohol dehydrogenase (Agrisera AS10 685) at a dilution of 1:5,000 and the secondary antibody used was goat anti-rabbit immunoglobulin G coupled to unmodified horseradish peroxidase (Sigma) at a 1:10,000 dilution. Detection was done using the ECL Western Blotting Detection Reagent (Amersham) and signal detected with Image Quant LAS4000 (GE).

Statistical Analysis

Statistical analyses were performed using PRISM 8 (GraphPad Software) or Excel (Microsoft Office, Microsoft). P-values were calculated using an unpaired, two-tailed Student's t-test (**p < 0.001; *p < 0.01; *p < 0.05; ns, not significant). Unless otherwise indicated in the figure legend, data represent mean \pm standard deviation.

RESULTS

CKDG2 Regulates the Alternative Splicing of the Flowering Regulator *FLM*

It has been previously shown that the CDKG group of kinases and their cognate cyclin, CYCLIN L1 (CYCL1), are important regulators of temperature dependent AS in *Arabidopsis* (Huang et al., 2013; Cavallari et al., 2018). Moreover, plants lacking CDKG2 display an early flowering phenotype when grown at ambient temperature (Ma et al., 2015). This led us to hypothesize that the early flowering phenotype in *cdkg2-1* mutant lines could be maintained along the ambient temperature range as a result of defective AS in genes involved in the temperature transduction

pathway. To test this possibility, we grew wild type, single *cdkg2-1* and *cycl1-1*, and the double *cdkg2-1;cycl1-1* mutant lines at 23°C under a LD light regime. Under these conditions, both the single *cdkg2-1* and *cycl1-1* and the double *cdkg2-1;cycl1-1* mutants flowered significantly earlier than the wild type (Figures 1A, B). On the contrary, no flowering phenotype was observed in the *cdkg1-1* mutant line (Supplementary Figure 1A).

Subsequently, we conducted a reverse transcriptase-polymerase chain reaction (RT-PCR) screen in *cdkg* mutant lines to test AS and expression levels of a small panel of genes including splicing factors, clock genes, and flowering regulators (Supplementary Table 1). Remarkably, among the investigated targets we found that in the single *cdkg2-1*, *cycl1-1* and in the double *cdkg2-1;cycl1-1* mutant lines, the processing of *FLM* (*MAF1*), a master regulator of the ambient temperature flowering pathway, was altered in terms of the relative levels of *FLM*- β and *FLM*- δ transcripts (Supplementary Figure 1B). In contrast the AS of *MAF2* (Airoldi et al., 2015), a close *FLM* paralogue was not affected in the different lines (Supplementary Figure 1C). In addition, no differences in the splicing of *FLM* or *MAF2* were observed in the single *cdkg1-1* mutant (Supplementary Figures 1B, C). The double *cdkg1-1;cdkg2-1* loss of function line could not be assessed as this genotype cannot be recovered and is assumed to be lethal (Zheng et al., 2014).

In order to investigate the changes in AS in more detail we quantified the levels of *FLM*- β and *FLM*- δ transcripts in the different mutants grown at 23°C by quantitative RT-PCR (RT-qPCR, see Figure 1C and Supplemental Figure 2A for *FLM* gene structure, AS events, and primer position). As observed by RT-PCR, lower levels of *FLM*- β (coding for the flowering repressor isoform) and increased expression of *FLM*- δ were observed in the single *cdkg2-1* and *cycl1-1* and in the double *cdkg2-1;cycl1-1* mutant lines (Figures 1D, E). Specifically, *FLM*- β expression was severely reduced in *cdkg2-1* (0.52 ± 0.10 fold) and in *cycl1-1* (0.70 ± 0.09 fold) and further impaired in the double *cdkg2-1;cycl1-1* (0.36 ± 0.03 fold) mutant lines in comparison to Col-0 (Figure 1D). The levels of *FLM*- δ were instead found significantly higher (up to 1.8 fold) in both *cdkg2-1* and *cdkg2-1;cycl1-1* mutants compared to wild type (Figure 1E) suggesting that CDKG2 together with CYCL1 maintains the balance between these two mutually exclusive isoforms. No significant change in *FLM*- β and *FLM*- δ expression were found between Col-0 and the *cdkg1-1* mutant lines (Supplementary Figures 3A, B).

Analysis of other flowering regulators involved in the temperature pathway showed that while there were no differences in the expression levels for *FLC* and the *TEMPRANILLO* genes (*TEM1* and *TEM2*; Supplementary Figures 4A, B), total *SVP* transcripts were reduced in all the mutant lines in comparison to Col-0 (Supplementary Figure 4C). This was mostly due to a reduction in the expression of one of two major *SVP* isoforms, *SVP2*, in mutant lines as determined by RT-PCR. Moreover, the lack of CDKG2 did not affect the AS of *FLM* regulatory genes like *ATU2AF65A* (Cavallari et al., 2018), *ATSF1*, or *ATGRP7* (Supplementary Figure 4D).

Consistent with its role in splicing, the CDKG2-GFP protein localizes to the nucleus of plant cells where it co-localizes with

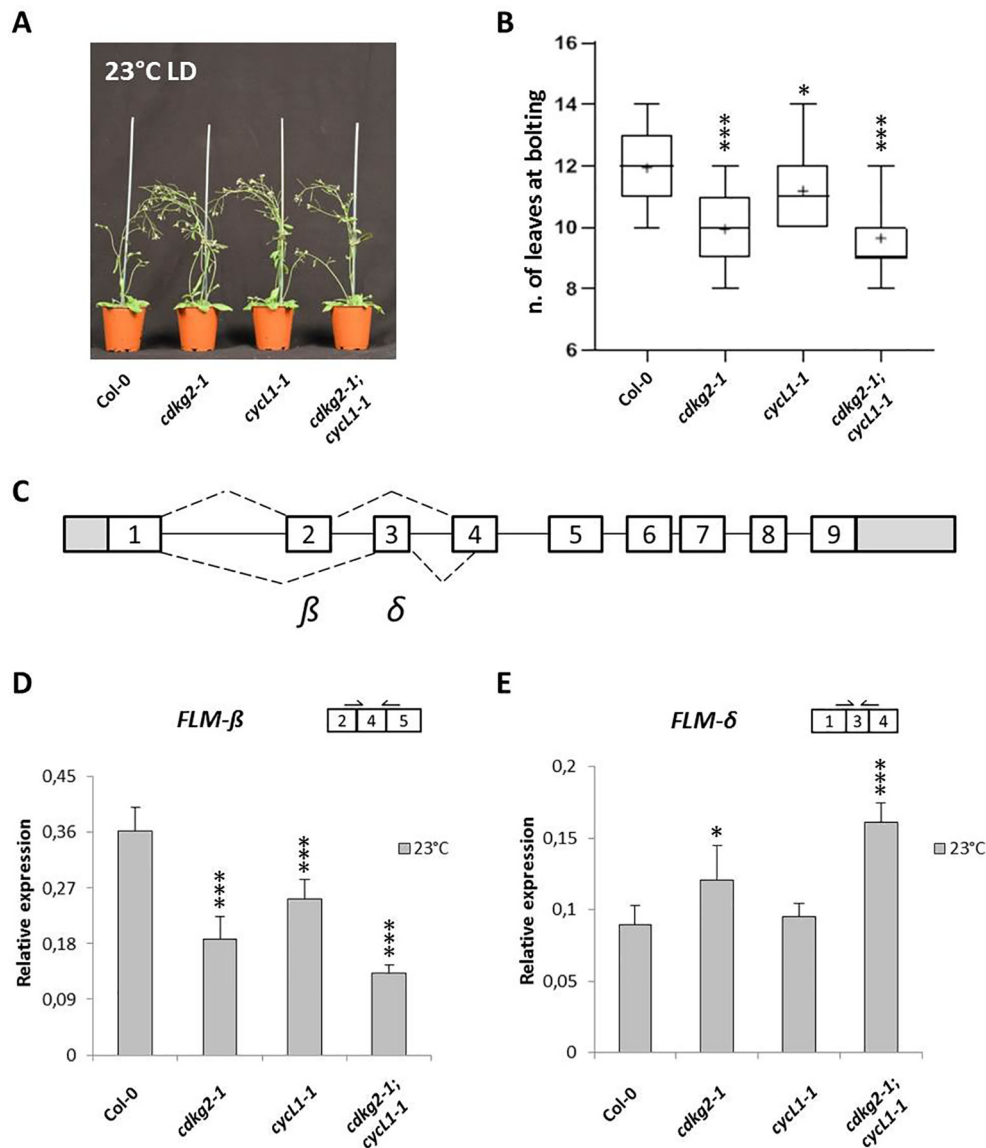


FIGURE 1 | Lack of CDKG2 and CYCL1 is associated to early flowering and alters the alternative splicing of *FLM*. **(A)** Flowering phenotype of Col-0, *cdkg2-1*, *cycl1-1*, and *cdkg2-1; cycl1-1* mutants grown at 23°C under long day (LD) conditions. **(B)** Flowering time of the plants shown in **(A)** quantified by counting the number of rosette leaves present at bolting ($n \geq 30$). Boxes represent 2nd and 3rd quartiles, bars minimum to maximum values, and crosses average of the groups. **(C)** Schematic representation of *FLM* locus and messenger RNA (mRNA) variants, including exons (boxes) and introns (lines). White boxes correspond to coding exons, gray boxes correspond to non-coding exon sequences (UTRs). Dotted lines represent alternative splicing (AS) events. The major isoforms produced are also indicated (β and δ). **(D)** and **(E)** Relative expression levels of *FLM- β* **(D)** and *FLM- δ* **(E)** mRNA as quantified by real-time quantitative PCR in the different lines grown at 23°C under LD conditions ($n \geq 5$). Student's t-test comparing *cdkg2-1*, *cycl1-1*, or *cdkg2-1; cycl1-1* to Col-0, *** $p < 0.001$, and * $p < 0.05$.

the spliceosome component RSp34-RFP (Supplementary Figures 5A, B).

Altered *FLM* Splicing in *cdkg2* Mutants Is Associated With Early Flowering Across Different Temperatures

We have previously shown that the CDKG group is actively involved in maintaining plant homeostasis along the ambient temperature range (Zheng et al., 2014; Cavallari et al., 2018). This

led us to test the possibility that the lack of CDKG2 or of its co-factor CYCL1 may regulate flowering along the ambient temperature range. For this, plants were grown at both 15 and 27°C under LD conditions (LD, 16 h light/8 h dark). As observed at 23°C, the early flowering phenotype of the single *cdkg2-1* and double *cdkg2-1; cycl1-1* was maintained at the different temperatures tested (Figures 2A, B) albeit with some small differences. At 15°C the double mutant lines flowered slightly earlier than the single *cdkg2-1* and *cycl1-1* lines (Figures 2A, B).

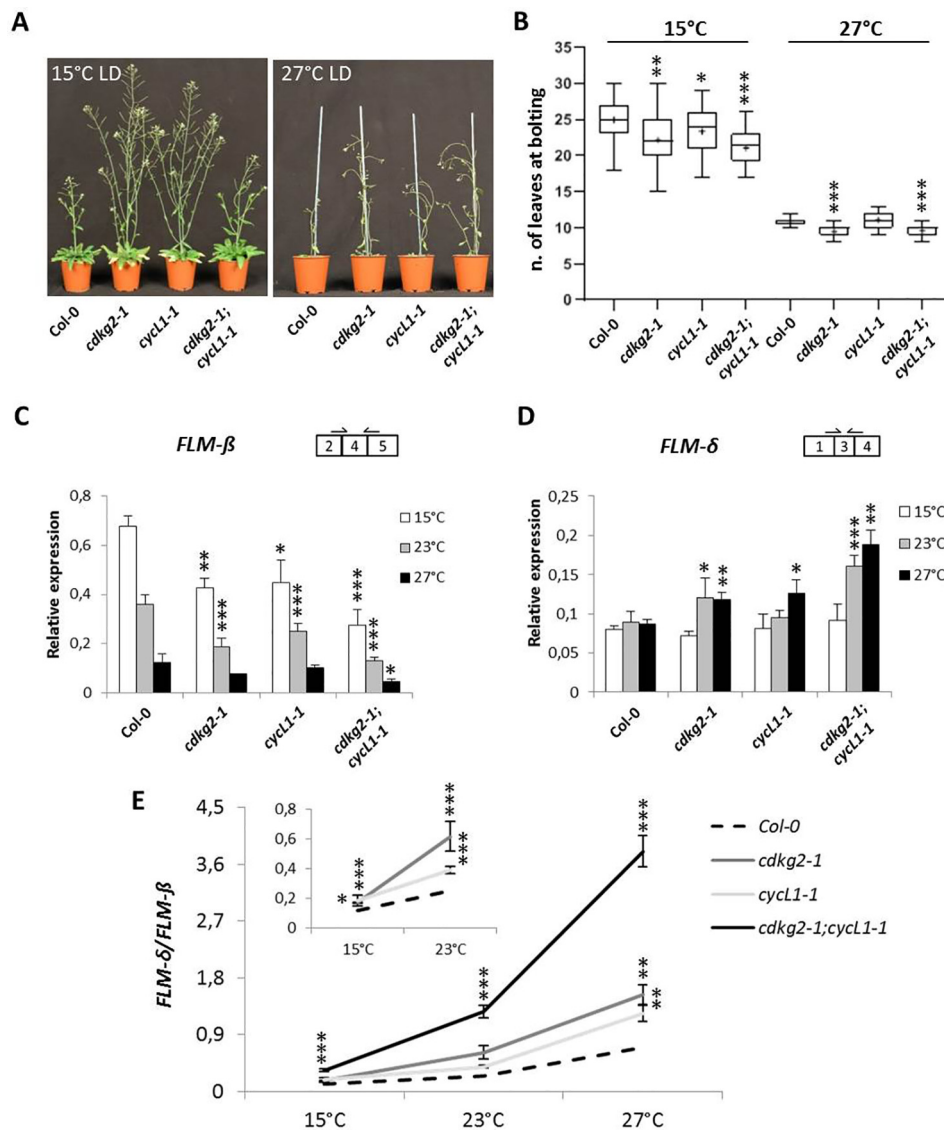


FIGURE 2 | Lack of CDKG2 and CYCL1 unbalances the alternative splicing of *FLM* across the ambient temperature range. **(A)** Flowering phenotype of Col-0, *cdkg2-1*, *cycl1-1*, and *cdkg2-1; cycl1-1* mutants grown under long day (LD) conditions at 15 and 27°C as indicated. **(B)** Flowering time of the plants shown in **(A)** quantified by counting the number of rosette leaves present at bolting ($n \geq 27$ at 15°C and $n \geq 23$ at 27°C). Boxes represent 2nd and 3rd quartiles, bars minimum to maximum values, and crosses average of the groups. **(C)** and **(D)** Relative expression levels of *FLM-β* **(C)** and *FLM-δ* **(D)** messenger RNA (mRNA) as quantified by real-time quantitative PCR in the different lines, grown at 15, 27, and 23°C for comparison under LD conditions as indicated ($n \geq 3$). **(E)** Ratio of *FLM-δ/FLM-β* mRNA in *cdkg2-1*, *cycl1-1*, and *cdkg2-1; cycl1-1* in comparison to Col-0 at the respective temperature (LD, long day). In the inset, detail of Col-0, *cdkg2-1*, and *cycl1-1* for statistic display. Student's t-test *** $p < 0.001$, ** $p < 0.01$, and * $p < 0.05$.

At 27°C, both the single *cdkg2-1* and the double mutant line were flowering significantly earlier than the wild type (**Figures 2A, B**) while no significant differences in flowering time were seen for the *cycl1-1* mutant.

To determine the effect of temperature on *FLM* splicing in the different mutant backgrounds, we quantified the levels of *FLM-β* and *FLM-δ* by RT-qPCR in 2-week old seedlings grown under LD conditions by shifting growth temperature from 23°C either to 15°C or to 27°C for 48 h before sampling.

In the wild type, *FLM-β* levels displayed temperature sensitivity as previously reported (Posé et al., 2013) with transcript levels raising at 15°C and decreasing at 27°C (**Figure 2C**) while *FLM-δ* expression remained relatively stable in wild type (**Figure 2D**). Strikingly, we found a more pronounced reduction in *FLM-β* levels along the temperature range in mutant lines and a significant increase in *FLM-δ* at 23 and 27°C (**Figures 2C, D**) in comparison to wild type. The detrimental effect of temperature increases on splicing in the mutant lines

became more evident when the ratio between *FLM-β* and *FLM-δ* (*FLM-δ/FLM-β*) was calculated at each temperature point (Figure 2E). While in Col-0 the ratio increased with the temperature (5.9 ± 1.2 fold from 15 to 27°C) this increase was higher in the mutant lines (11.5 ± 0.7 fold in *cdkg2-1;cycl1-1*).

We also examined the relative levels of *SVP* expression in the various mutant backgrounds at different temperatures and found that the double *cdkg2-1;cycl1-1* mutant had constitutively lower *SVP* expression than the wt control across the temperature range (Supplementary Figure 6A).

Although *FLM* is known to influence flowering particularly at lower temperatures (Lutz et al., 2015), the *cdkg2-1* and the *cdkg2-1;cycl1-1* double mutants flowered earlier than Col-0 also at 27°C suggesting the involvement of additional regulatory mechanisms. Expression of *FLC* was reported to have a strong impact on flowering time particularly at high temperatures (Balasubramanian et al., 2006). However, we observed no significant changes in *FLC* expression between Col-0 and the mutant lines at 27°C suggesting that the early flowering phenotype of the *cdkg2* mutants is not due to altered *FLC* expression (Supplementary Figure 6B).

CDKG2/CYCL1 Has a Wide Effect on FLM Transcript Processing

In order to determine if other major splicing events in *FLM* were affected by the lack of CDKG2 and CYCL1, we analyzed expression of the mRNAs that retain intron 4, namely splicing variants *ASF7* or *ASF10* (Capovilla et al., 2017). Retention of the in frame intron 4, either in combination with exon 2 or 3, could translate for proteins with characteristics similar to *FLM-β* or *FLM-δ* respectively (see Figure 3A and Supplementary Figure 2A for splicing scheme). Remarkably, we found reduced levels of intron 4 retention for *ASF7* transcripts in the *cdkg2-1* and *cdkg2-1;cycl1-1* mutant lines at 23 and 27°C while *ASF10* was mildly but significantly affected in the single *cycl1-1* and in the double mutant albeit at different temperatures (Figures 3B, C).

Variations in *FLM* intron 1 sequence were previously shown to fine tune flowering time and to be involved in adaptation to temperature (Lutz et al., 2015; Lutz et al., 2017) and based on database annotations (Araport11) there are several potential intron 1 retention *FLM* mRNAs (AT1G77080.6, AT1G77080.7, AT1G77080.9, AT1G77080.10). These alternative *FLM* transcripts could thus affect *FLM* expression. We observed that *FLM* intron 1 retention (*FLMi1*) was not affected by temperature in Col-0 while single and double mutant lines showed remarkably higher retention levels at 23 and 27°C (up to 2.6 ± 0.34 fold, Figure 3D).

Taken together, these data suggested that the lack of CDKG2 and CYCL1 affected not only the balance between *FLM-β* and *FLM-δ* but also the processing of other *FLM* transcripts spanning from exon 1 to intron 4 along the ambient temperature range.

The observed differences in the relative *FLM* isoform abundance and how these may impact on the expression of *FLM*, prompted us to evaluate the total levels of *FLM* by measuring *FLM* exon 1 (*FLMex1*) containing transcripts by RT-qPCR. Total levels of *FLM* mRNA decreased along the temperature range in Col-0 and were further reduced in the

mutant lines both at 15 and 23°C but not at 27°C (Figure 3E). These observations suggest that the lower *FLM* levels observed in the *cdkg2-1* and *cycl1-1* mutants may reflect intrinsic differences in *FLM* isoform stability, although we cannot completely exclude a concomitant reduction in transcription at specific temperatures.

Lack of CDKG2 and CYCL1 Promotes Flowering and Alters FLM Alternative Splicing Independently of the Photoperiod

Since the photoperiod also has a strong effect on flowering time, we assessed the flowering phenotypes of the single and double *cdkg2-1* and *cycl1-1* mutants under SD conditions. For this we grew plants at 15, 23, and 27°C in SD (8 h light/16 h dark) which is considered a non-inductive condition for *Arabidopsis* (Balasubramanian and Weigel, 2006). We hypothesize if mainly the temperature pathway was affected then early flowering should be maintained independently of day length.

The double *cdkg2-1;cycl1-1* mutant lines still flowered earlier than wild type plants in SD conditions at all temperatures while in *cdkg2-1* this effect was present at 23°C and at 27°C (Figures 4A–F).

As we observed under LD conditions, plants grown in SD showed small decreases in *FLM-β* with increased temperature (Figure 5A). Increases in *FLM-δ* transcripts were significant only in the double *cdkg2-1;cycl1-1* mutant (Figure 5B). In addition, *FLMi1* levels were also increased in mutant lines while *FLMex1* was lower mainly at 23°C (Figures 5C, D). Since, no differences in flowering were observed in SD conditions (23°C) and expression of *FLM-β* and *FLM-δ* was similar between Col-0 and the *cdkg2-1* mutant line (Supplementary Figures 7A–C), we decided not to further test this mutant in the present investigation.

The data profile obtained for *ASF7* and *ASF10* under SD conditions was comparable to that seen under LD conditions. Indeed, *ASF7* levels decreased with temperature and this effect was more accentuated in the mutants at higher temperatures. The effect on *ASF10* was less pronounced (Supplementary Figures 8A, B). Notably, the expression profile of *SVP* was not affected in the mutant lines in SD conditions (Supplementary Figure 8C).

Taken together, these results suggest that the temperature-dependent effect of the CDKG2/CYCL1 complex on the AS of *FLM* is independent of the photoperiod.

FLMi1 Transcripts Accumulate in the Cell Nucleus

Intron retention events in plants can promote mRNA sequestration in the nucleus (Gohring et al., 2014) so that the affected mRNAs are unlikely to be translated into proteins in the cytoplasm. Hence, a possible consequence of the significant increase in *FLM* intron 1 containing transcripts in the mutant lines could be the increase of the nuclear *FLM* mRNA pool. This could represent an interesting, yet previously unknown, mechanism of *FLM* regulation based on CDKG2 activity and controlling *FLM* nuclear export.

The accumulation of intron 1 containing transcripts in the *cdkg2-1* and *cycl1-1* single and double mutants could be the consequence of either an increase in *FLM* pre-mRNA (unprocessed transcripts) or of a specific CDKG2 effect on

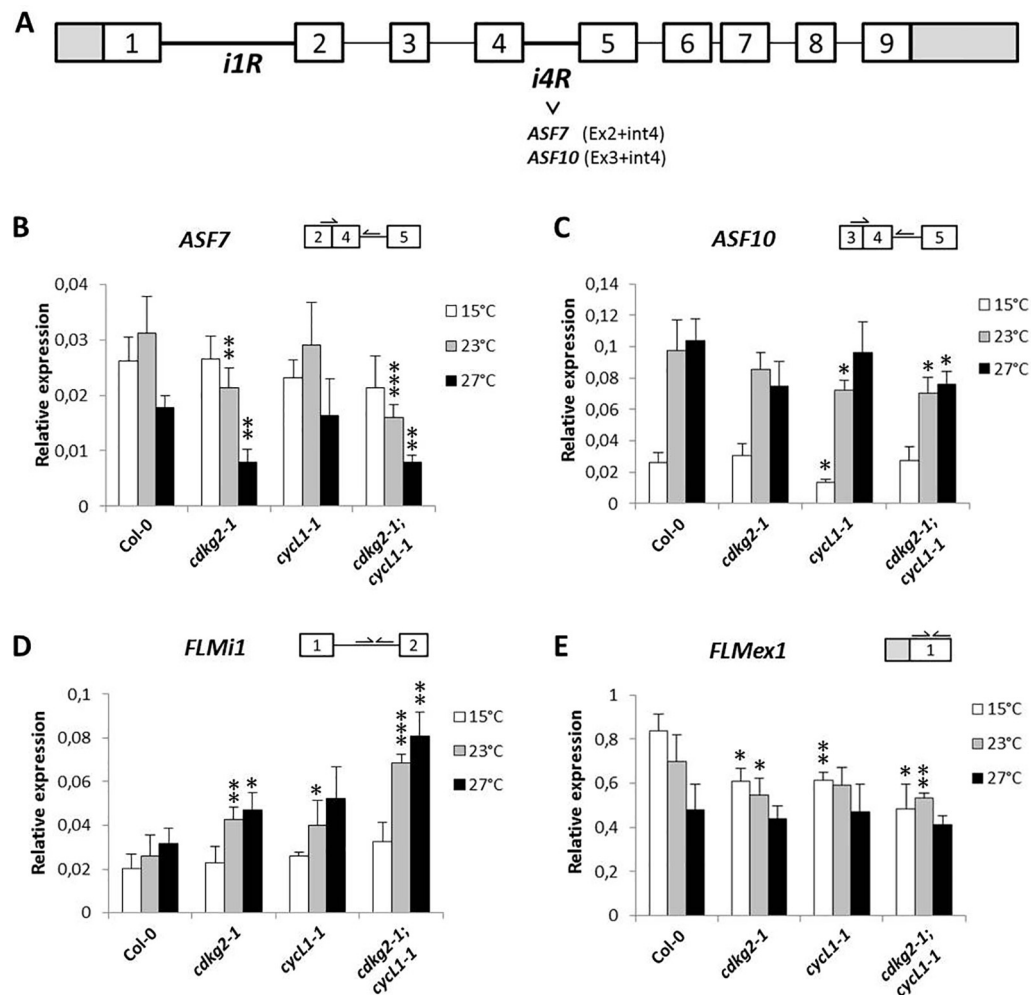


FIGURE 3 | CDKG2 and CYCL1 also regulate the abundance of other *FLM* alternative splicing isoforms. **(A)** Schematic representation of *FLM* locus and messenger RNA (mRNA) variants, including exons (boxes) and introns (lines). White boxes correspond to coding exons, gray boxes correspond to non-coding exon sequences (UTRs). iR, intron retention. **(B–E)** Relative expression levels of *ASF7* **(B)**, *ASF10* **(C)**, *FLMi1* **(D)**, and *FLMex1* **(E)** mRNA as quantified by real-time quantitative PCR in the different lines, grown at 15, 23, and 27°C under long day conditions as indicated. Student's t-test comparing *cdkg2-1*, *cycl1-1*, or *cdkg2-1; cycl1-1* to Col-0 at the respective temperature, $n \geq 3$, *** $p < 0.001$, ** $p < 0.01$, and * $p < 0.05$.

intron 1 AS. To distinguish between these two possibilities, we amplified only processed messengers by RT-PCR by positioning the primers at the end of *FLM* intron 1 and at the exon 4/exon 5 junction (*FLMi1e2F* and *FLMe5-4R*; **Supplementary Table 2**). Interestingly, the transcripts we found had size corresponding to *FLMi1* mRNAs that contain both intron 2 and intron 3 (and relative exons) or only intron 2 (**Supplementary Figure 9A**). Moreover, we observed that these isoforms were more abundant in the double *cdkg2-1; cycl1-1* mutant than in Col-0, confirming that we see increased *FLM* intron 1 retention in the absence of CDKG2/CYCL1 (**Supplementary Figure 9A**).

These findings prompted us to fractionate protoplast cell mRNA and assess sub-cellular localization of specific transcripts. Strikingly, nuclear and cytoplasmic fractions showed that while *FLM-β* and *FLM-δ* forms are present in the cytoplasm (as expected, being

protein coding isoforms) *FLMi1* was retained in the nucleus (**Figure 6A**). The purity of the fractions was confirmed by RT-PCR for *SEF Factor* (AT5G37955) and by Western blot.

In summary, the data presented here show that the CDKG2/CYCL1 complex affects the temperature-dependent splicing of *FLM*. Finally, the nuclear retention of *FLM* intron1 containing transcripts could provide a new layer of *FLM* regulation across the temperature range.

DISCUSSION

The identification of key components in ambient temperature sensing/response in plants is crucial not least in times of global

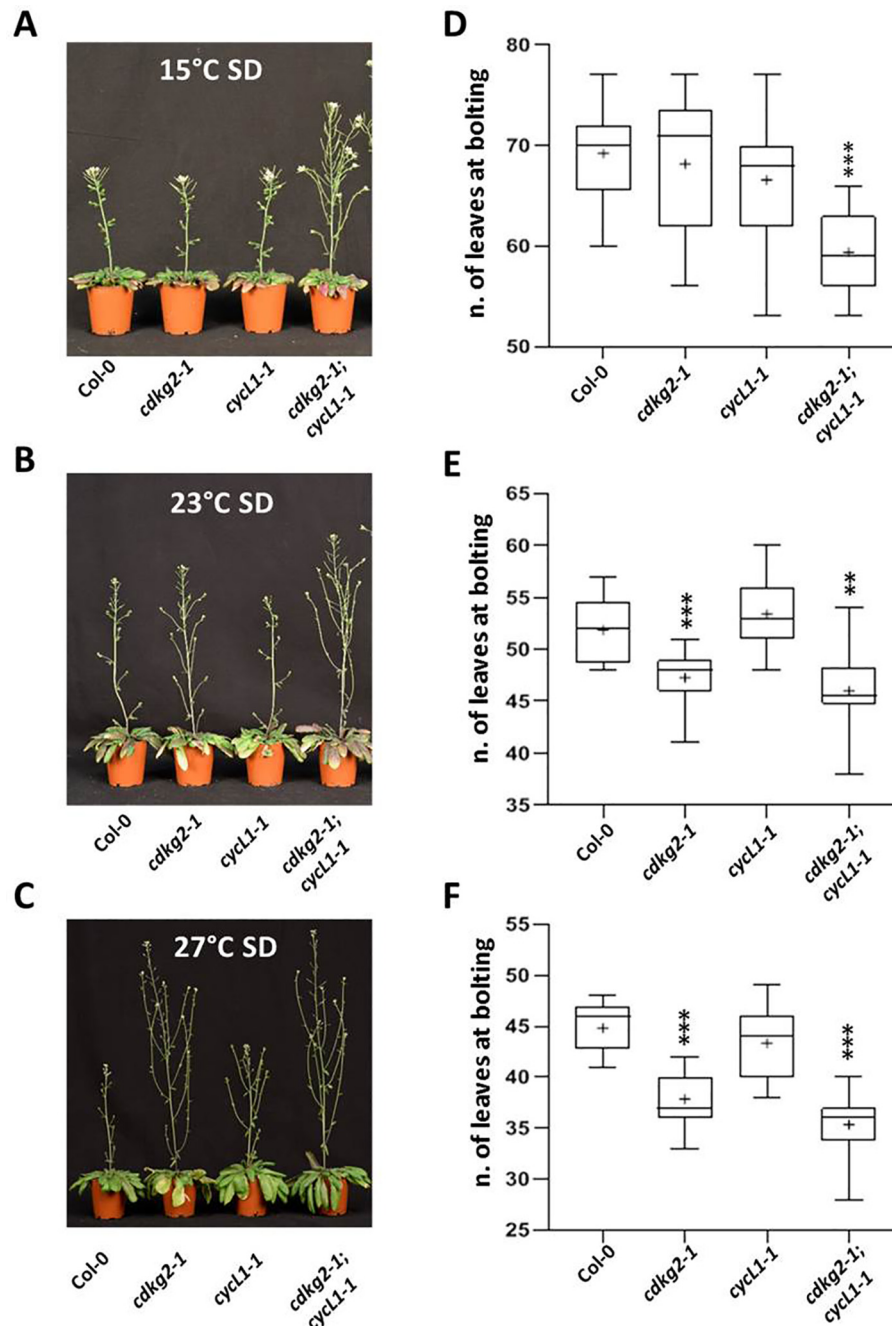


FIGURE 4 | Lack of CDKG2/CYCL1 complex promotes flowering across the ambient temperature range in short day conditions. **(A–C)** Flowering phenotype of *Col-0*, *cdkg2-1*, *cycl1-1*, and *cdkg2-1; cycl1-1* mutants grown under SD conditions at 15°C **(A)**, at 23°C **(B)**, and at 27°C **(C)**. **(D–F)** Flowering time of the plants shown in **(A)**, **(B)**, and **(C)** quantified by counting the number of rosette leaves present at bolting ($n \geq 15$, $n \geq 10$, and $n \geq 14$ respectively). Boxes represent 2nd and 3rd quartiles, bars minimum to maximum values, and crosses average of the groups. Student's t-test comparing *cdkg2-1*, *cycl1-1*, or *cdkg2-1; cycl1-1* to *Col-0* at the respective temperature, *** $p < 0.001$ and ** $p < 0.01$.

warming where increased temperature variation could produce ecological changes that will negatively impact on the present agricultural system (Wheeler and von Braun, 2013; Moore and Lobell, 2015; Jagadish et al., 2016). Hence, investigation and analysis

of the molecular circuits involved in the temperature transduction pathways in plants is now of considerable importance.

While animals have developed specialized receptor classes for specific environmental variables (Terakita and Nagata, 2014;

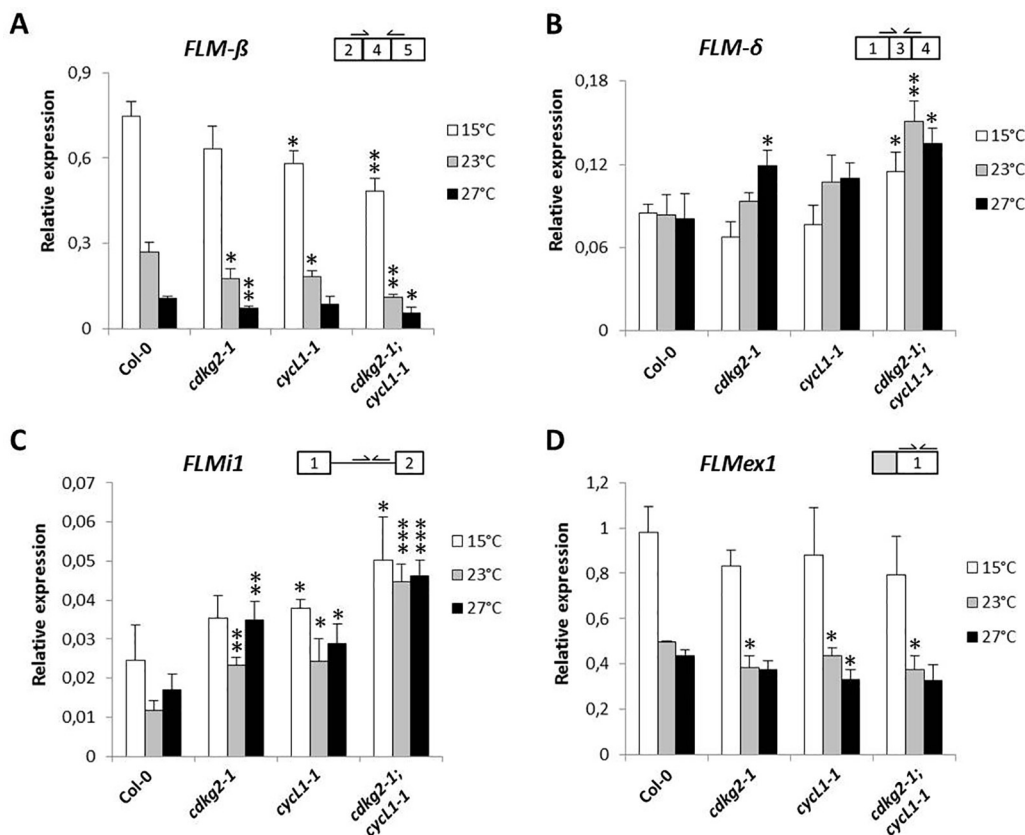


FIGURE 5 | Lack of CDKG2 and CYCL1 unbalances the alternative splicing of *FLM* across the ambient temperature range in short day (SD) conditions. (A–D)

Relative expression levels of *FLM-β* (A), *FLM-δ* (B), *FLMi1* (C), and *FLMex1* (D) messenger RNA as quantified by real-time quantitative PCR in the different lines, Col-0, *cdkg2-1*, *cycl1-1*, *cdkg2-1; cycl1-1* grown at 15, 23, and 27°C under SD conditions (8 h light, 16 h dark) as indicated. Student's t-test comparing *cdkg2-1*, *cycl1-1*, or *cdkg2-1; cycl1-1* to Col-0 at the respective temperature, n = 3, ***p < 0.001, **p < 0.01, and *p < 0.05.

Vriens et al., 2014), the sensors so far identified in plants belong to diverse gene families and can have wider roles in both sensing and integrating environmental cues (Paik and Huq, 2019). The CDKG group of kinases, for example, has an important role in inherently temperature sensitive processes like meiosis and flowering (Zheng et al., 2014; Ma et al., 2015).

Recently we found that CDKGs can also integrate ambient temperature inputs by modulating an alternative mRNA splicing cascade (Cavallari et al., 2018) raising the question as to whether the role of CDKs in the aforementioned developmental processes could be acting through AS.

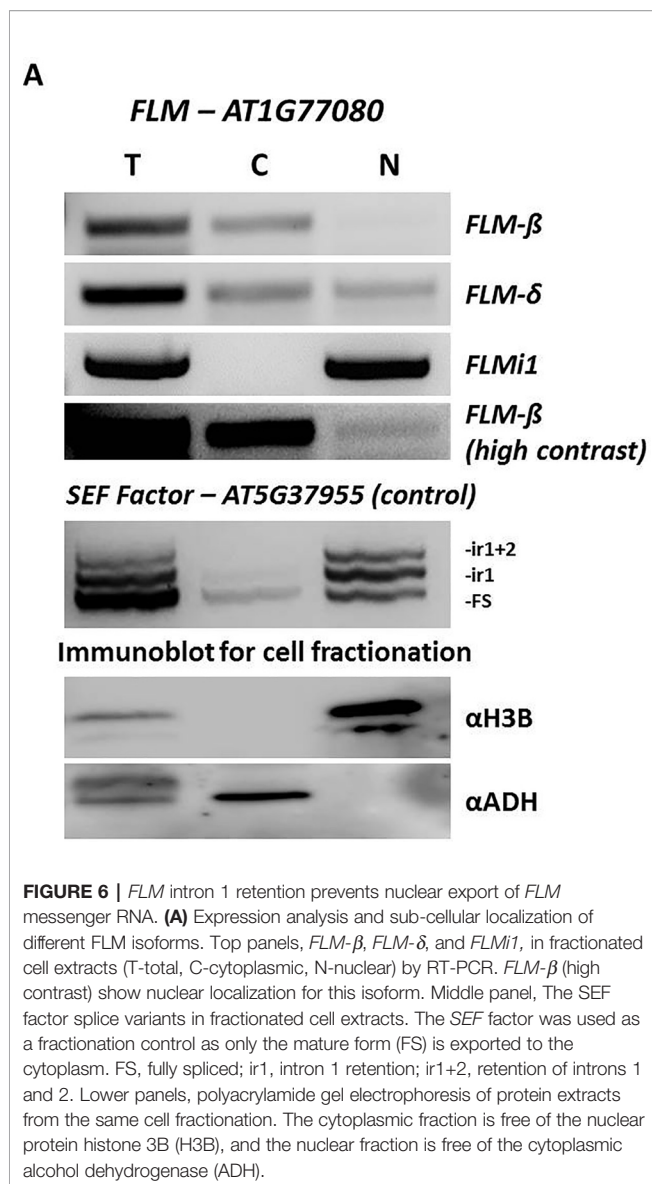
In the current report, we demonstrate that the CDKG2/CYCL1 modulates AS of the flowering regulator *FLM*, possibly providing an additional mechanism fine-tuning flowering time across the ambient temperature range.

FLM mRNA processing responds strongly to ambient temperature coding for some known (i.e., *FLM-β* and *FLM-δ*) as well as putative isoforms (i.e., *ASF7* and *ASF10*) (Posé et al., 2013; Capovilla et al., 2017). While the repressive role of *FLM-β* in flowering time regulation is well accepted there is still debate about the function of *FLM-δ*. In addition, functional

characterization of *ASF7* and *ASF10* proteins (with predictably similar functions as *FLM-β* and *FLM-δ*) is still missing. Indeed, *ASF7* and *ASF10* transcripts contain the in-frame *FLM* intron 4 which belongs to the exitron class (Marquez et al., 2015). Exitrons define a particular intron group associated with translation of alternative protein variants, suggesting that *ASF7* and *ASF10* might code for alternative proteins with different (and as yet unknown) functions.

Besides the strong temperature regulation of *FLM* AS, we found that the absence of CDKG2 and CYCL1 resulted in changes in the abundance of *FLM-β* and *FLM-δ* and, to a minor extent, of *ASF7* and *ASF10* across the temperature range (Figures 2C, D and 3B, C) and under LD and SD conditions (Figures 5A, B and Supplementary Figures 8A, B). While temperature increases affects levels of the active floral repressor *FLM-β*, CDKG2 acted against the temperature signal to dampen the shift on the production of its non-repressive counterpart *FLM-δ*.

Moreover, CDKG2 and CYCL1 control the levels of *FLM* intron 1 retention and this new regulatory mechanism may influence the *FLM* intracellular mRNA trafficking (Reed, 2003). The nuclear retained *FLMi1* mRNAs could potentially be further



processed, as was recently shown for the splicing factor *SR30* (Hartmann et al., 2018), and be stored or released from the cell nucleus in response to changing environmental conditions to promote or delay transition to flowering respectively. Indeed, the two *FLMi1* isoforms found by RT-PCR (**Supplementary Figure 6A**) retaining intron 2 and intron 3 may be spliced either into *FLM-β* or *FLM-δ* variants.

Hence, modulation of CDKG2 kinase activity is likely to impact on flowering time definition changing the AS of *FLM*, either by altering the ratio of *FLM-β* and *FLM-δ* as reported for other splicing factors (Lee et al., 2017; Park et al., 2019; Steffen et al., 2019) or by promoting retention of *FLM* intron 1. Indeed, the predicted increase in nuclear retention for *FLMi1* isoforms would provide a new additional, elegant, and rapid signaling module to adjust flowering time in response to changes in ambient temperature. Furthermore, the observation that the

effect on AS in mutant lines was greater at higher temperatures (**Figure 2E**) suggests that CDKGs may contribute to temperature compensation during mRNA processing, a feature which is very important for other cellular mechanisms like the circadian clock (Avello et al., 2019). Consistent with this idea, *FLM-δ* and *FLMi1* expression became temperature dependent in *cdkg2-1* mutant lines, contrary to Col-0 where these isoforms were stably expressed (**Figures 2D** and **3D**).

Previously we showed that CDKG1 affected the splicing of *ATU2AF65A* (Cavallari et al., 2018) and recently, loss of this fundamental spliceosome component has been reported to regulate flowering time in *Arabidopsis* by altering the expression patterns of several flowering related genes including *FLM* (Park et al., 2019). The observations that CDKG2 and *CYCL1* control the AS of both *CDKG1* and *FLM* along the ambient temperature range, place this complex at the top of a signal transduction cascade translating environmental signals into developmental changes by regulating the AS of key regulatory genes in the temperature pathway.

Indeed, our data suggest a model whereby interplay between temperature and CDKs can modulate flowering time *via* AS of key floral regulators. We speculate that the flowering phenotype observed in *cdkg2* mutant lines may go beyond just a direct action on *FLM* considering that additional flowering genes are affected at the expression or AS levels (like *SVP*). A deeper understanding of the genetic interactions between CDKG related functions and the flowering time pathway could provide insights into the role of AS in regulating flowering and, particularly, the role it might play in temperature compensation.

However, whether temperature related differences in AS pertains to mRNA secondary structure modifications, as in the yeast model (Meyer et al., 2011) or to a sensor mediated signaling cascade, the molecular mechanisms ruling temperature dependent mRNA processing are yet to be fully elucidated.

The complexity and plasticity of the environmental sensing landscape in plants is only just emerging (Legris et al., 2016; Fujii et al., 2017; Casal and Qüesta, 2018; Dickinson et al., 2018; Wang et al., 2018; Han et al., 2019; Paik and Huq, 2019) and our results highlight the capacity of AS to bridge the interactions between environmental input pathways, specifically temperature, and central regulatory mechanisms, such as the cyclin dependent protein kinases, to control gene expression.

DATA AVAILABILITY STATEMENT

All datasets generated for this study are included in the article/Supplementary Material.

AUTHOR CONTRIBUTIONS

NC and CN conceived the project and designed research. NC, CN, MG, and DD performed research. NC and CN analyzed data. NC, CN, and JD wrote the paper. JD supervised the project and obtained funding.

FUNDING

CN, DD, and JD were funded by the BBSRC (grant number BB/M009459/1). NC was funded by the VIPS Program of the Austrian Federal Ministry of Science and Research and the City of Vienna.

ACKNOWLEDGMENTS

We would like to thank Prof. Eva Benkova (IST AUSTRIA, Klosterneuburg, Austria) and Prof. Andrea Barta (MFPL,

Vienna, Austria) for constant support and Dr. Mariya Kalyna (BOKU, Vienna, Austria) and Dr. Gergely Molnar (BOKU, Tulln an der Donau, Austria) for helpful suggestions.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fpls.2019.01680/full#supplementary-material>

REFERENCES

- Airoldi, C. A., McKay, M., and Davies, B. (2015). MAF2 Is regulated by temperature-dependent splicing and represses flowering at low temperatures in parallel with FLM. *PLoS One* 10, e0126516. doi: 10.1371/journal.pone.0126516
- Amasino, R. (2010). Seasonal and developmental timing of flowering. *Plant J.* 61, 1001–1013. doi: 10.1111/j.1365-3113.2010.04148.x
- Avello, P. A., Davis, S. J., Ronald, J., and Pitchford, J. W. (2019). Heat the clock: entrainment and compensation in *Arabidopsis* circadian rhythms. *J. Circadian Rhythms* 17, 5. doi: 10.5334/jcr.179
- Balasubramanian, S., and Weigel, D. (2006). Temperature induced flowering in *Arabidopsis thaliana*. *Plant Signaling Behav.* 1, 227–228. doi: 10.4161/psb.1.53452
- Balasubramanian, S., Sureshkumar, S., Lempe, J., and Weigel, D. (2006). Potent induction of *Arabidopsis thaliana* flowering by elevated growth temperature. *PLoS Genet.* 2, e106. doi: 10.1371/journal.pgen.0020106
- Bartkowiak, B., Liu, P., Phatnani, H. P., Fuda, N. J., Cooper, J. J., Price, D. H., et al. (2010). CDK12 is a transcription elongation-associated CTD kinase, the metazoan ortholog of yeast Ctk1. *Genes Dev.* 24, 2303–2316. doi: 10.1101/gad.1968210
- Bouché, F., Detry, N., and Périlleux, C. (2015). Heat can erase epigenetic marks of vernalization in *Arabidopsis*. *Plant Signaling Behav.* 10, e990799. doi: 10.4161/15592324.2014.990799
- Caicedo, A. L., Stinchcombe, J. R., Olsen, K. M., Schmitt, J., and Purugganan, M. D. (2004). Epistatic interaction between *Arabidopsis* FRI and FLC flowering time genes generates a latitudinal cline in a life history trait. *Proc. Natl. Acad. Sci.* 101, 15670–15675. doi: 10.1073/pnas.0406232101
- Calixto, C. P. G., Guo, W., James, A. B., Tzioutziou, N. A., Entizne, J. C., Panter, P. E., et al. (2018). Rapid and dynamic alternative splicing impacts the *Arabidopsis* cold response transcriptome. *Plant Cell tpc* 30, 1424–1444. doi: 10.1105/tpc.18.00177
- Capovilla, G., Pajaro, A., Immink, R. G., and Schmid, M. (2015). Role of alternative pre-mRNA splicing in temperature signaling. *Curr. Opin. In Plant Biol.* 27, 97–103. doi: 10.1016/j.pbi.2015.06.016
- Capovilla, G., Symeonidi, E., Wu, R., and Schmid, M. (2017). Contribution of major FLM isoforms to temperature-dependent flowering in *Arabidopsis thaliana*. *J. Exp. Bot.* 68, 5117–5127. doi: 10.1093/jxb/erx328
- Capovilla, G., Delhomme, N., Collani, S., Shutava, I., Bezrukov, I., Symeonidi, E., et al. (2018). PORCUPINE regulates development in response to temperature through alternative splicing. *Nat. Plants* 4, 534–539. doi: 10.1038/s41477-018-0176-z
- Casal, J. J., and Qüesta, J. I. (2018). Light and temperature cues: multitasking receptors and transcriptional integrators. *New Phytol.* 217, 1029–1034. doi: 10.1111/nph.14890
- Cavallari, N., Nibau, C., Fuchs, A., Dadarou, D., Barta, A., and Doonan, J. H. (2018). The cyclin-dependent kinase G group defines a thermo-sensitive alternative splicing circuit modulating the expression of *Arabidopsis* *ATU2AF65A*. *Plant J.* 94, 1010–1022. doi: 10.1111/tpj.13914
- Chamala, S., Feng, G., Chavarro, C., and Barbazuk, W. B. (2015). Genome-wide identification of evolutionarily conserved alternative splicing events in flowering plants. *Front. In Bioeng. Biotechnol.* 3, 33. doi: 10.3389/fbioe.2015.00033
- Chen, H.-H., Wang, Y.-C., and Fann, M.-J. (2006). Identification and characterization of the CDK12/Cyclin L1 complex involved in alternative splicing regulation. *Mol. Cell. Biol.* 26, 2736–2745. doi: 10.1128/MCB.26.7.2736-2745.2006
- Cheng, S.-W. G., Kuzys, M. A., Moradian, A., Ichu, T.-A., Chang, V. C.-D., Tien, J. F., et al. (2012). Interaction of Cyclin-Dependent Kinase 12/CrkRS with Cyclin K1 Is Required for the Phosphorylation of the C-Terminal Domain of RNA Polymerase II. *Mol. Cell. Biol.* 32, 4691–4704. doi: 10.1128/MCB.06267-11
- Chouard, P. (1960). Vernalization and its relations to dormancy. *Annu. Rev. Plant Physiol.* 11, 191–238. doi: 10.1146/annurev.pp.11.060160.001203
- Cui, X., Fan, B., Scholz, J., and Chen, Z. (2007). Roles of *Arabidopsis* cyclin-dependent kinase C complexes in cauliflower mosaic virus infection, plant growth, and development. *The Plant Cell Online* 19, 1388–1402. doi: 10.1105/tpc.107.051375
- Czechowski, T. (2005). Genome-wide identification and testing of superior reference genes for transcript normalization in *Arabidopsis*. *Plant Physiol.* 139, 5–17. doi: 10.1104/pp.105.063743
- Dickinson, P. J., Kumar, M., Martinho, C., Yoo, S. J., Lan, H., Artavanis, G., et al. (2018). Chloroplast signaling gates thermotolerance in *Arabidopsis*. *Cell Rep.* 22, 1657–1665. doi: 10.1016/j.celrep.2018.01.054
- Even, Y., Durieux, S., Escande, M.-L., Lozano, J. C., Peaucellier, G., Weil, D., et al. (2006). CDC2L5, a Cdk-like kinase with RS domain, interacts with the ASF/SF2-associated protein p32 and affects splicing *in vivo*. *J. Cell. Biochem.* 99, 890–904. doi: 10.1002/jcb.20986
- Fujii, Y., Tanaka, H., Konno, N., Ogasawara, Y., Hamashima, N., Tamura, S., et al. (2017). Phototropin perceives temperature based on the lifetime of its photoactivated state. *Proc. Natl. Acad. Sci.* 114, 9206–9211. doi: 10.1073/pnas.1704462114
- Gohring, J., Jacak, J., and Barta, A. (2014). Imaging of endogenous messenger RNA splice variants in living cells reveals nuclear retention of transcripts inaccessible to nonsense-mediated decay in *Arabidopsis*. *Plant Cell* 26, 754–764. doi: 10.1105/tpc.113.118075
- Gu, X., Le, C., Wang, Y., Li, Z., Jiang, D., Wang, Y., et al. (2013). *Arabidopsis* FLC clade members form flowering-repressor complexes coordinating responses to endogenous and environmental cues. *Nat. Commun.* 4, 1947. doi: 10.1038/ncomms2947
- Han, S.-H., Park, Y.-J., and Park, C.-M. (2019). Light priming of thermotolerance development in plants. *Plant Signaling Behav.* 14, 1554469. doi: 10.1080/15592324.2018.1554469
- Hartmann, L., Drewe-Boß, P., Wießner, T., Wagner, G., Geue, S., Lee, H.-C., et al. (2016). Alternative splicing substantially diversifies the transcriptome during early photomorphogenesis and correlates with the energy availability in *Arabidopsis*. *Plant Cell* 28, 2715–2734. doi: 10.1105/tpc.16.00508
- Hartmann, L., Wießner, T., and Wachter, A. (2018). Subcellular compartmentation of alternatively spliced transcripts defines *SERINE/ARGININE-RICH PROTEIN30* expression. *Plant Physiol.* 176, 2886–2903. doi: 10.1104/pp.17.01260
- Hu, D., Mayeda, A., Trembley, J. H., Lahti, J. M., and Kidd, V. J. (2003). CDK11 complexes promote pre-mRNA splicing. *J. Biol. Chem.* 278, 8623–8629. doi: 10.1074/jbc.M210057200

- Huang, X.-Y., Niu, J., Sun, M.-X., Zhu, J., Gao, J.-F., Yang, J., et al. (2013). CYCLIN-DEPENDENT KINASE G1 is associated with the spliceosome to regulate CALLOSE SYNTHASE5 splicing and pollen wall formation in *Arabidopsis*. *Plant Cell* 25, 637–648. doi: 10.1105/tpc.112.107896
- Jagadish, S. V. K., Bahuguna, R. N., Djanaguiraman, M., Gamuyao, R., Prasad, P. V. V., and Craufurd, P. Q. (2016). Implications of high temperature and elevated CO₂ on flowering time in plants. *Front. Plant Sci.* 7, 913. doi: 10.3389/fpls.2016.00913
- James, A. B., Calixto, C. P. G., Tzioutziou, N. A., Guo, W., Zhang, R., Simpson, C. G., et al. (2018). How does temperature affect splicing events? Isoform switching of splicing factors regulates splicing of *LATE ELONGATED HYPOCOTYL (LHY)*: Splicing factor temperature signaling to the clock. *Plant Cell Environ.* 41, 1539–1550. doi: 10.1111/pce.13193
- Johanson, U., West, J., Lister, C., Michaels, S., Amasino, R., and Dean, C. (2000). Molecular analysis of FRIGIDA, a major determinant of natural variation in *Arabidopsis* flowering time. *Science* 290, 344–347. doi: 10.1126/science.290.5490.344
- Jung, J.-H., Domijan, M., Klose, C., Biswas, S., Ezer, D., Gao, M., et al. (2016). Phytochromes function as thermosensors in *Arabidopsis*. *Science* 354, 886–889. doi: 10.1126/science.aaf6005
- Kalyna, M., Simpson, C. G., Syed, N. H., Lewandowska, D., Marquez, Y., Kusenda, B., et al. (2012). Alternative splicing and nonsense-mediated decay modulate expression of important regulatory genes in *Arabidopsis*. *Nucleic Acids Res.* 40, 2454–2469. doi: 10.1093/nar/gkr932
- Kitsios, G., Alexiou, K. G., Bush, M., Shaw, P., and Doonan, J. H. (2008). A cyclin-dependent protein kinase, CDK2, colocalizes with and modulates the distribution of spliceosomal components in *Arabidopsis*. *Plant J.* 54, 220–235. doi: 10.1111/j.1365-313X.2008.03414.x
- Klepikova, A. V., Kasianov, A. S., Gerasimov, E. S., Logacheva, M. D., and Penin, A. A. (2016). A high resolution map of the *Arabidopsis thaliana* developmental transcriptome based on RNA-seq profiling. *Plant J.* 88, 1058–1070. doi: 10.1111/tpl.13312
- Ko, T. K., Kelly, E., and Pines, J. (2001). CrkRS: a novel conserved Cdc2-related protein kinase that colocalizes with SC35 speckles. *J. Cell Sci.* 114, 2591–2603.
- Laloum, T., Martin, G., and Duque, P. (2018). Alternative splicing control of abiotic stress responses. *Trends In Plant Sci.* 23, 140–150. doi: 10.1016/j.tplants.2017.09.019
- Lee, Y., and Rio, D. C. (2015). Mechanisms and regulation of alternative pre-mRNA splicing. *Annu. Rev. Biochem.* 84, 291–323. doi: 10.1146/annurev-biochem-060614-034316
- Lee, J. H., Yoo, S. J., Park, S. H., Hwang, I., Lee, J. S., and Ahn, J. H. (2007). Role of SVP in the control of flowering time by ambient temperature in *Arabidopsis*. *Genes Dev.* 21, 397–402. doi: 10.1101/gad.1518407
- Lee, J. H., Lee, J. S., and Ahn, J. H. (2008). Ambient temperature signaling in plants: an emerging field in the regulation of flowering time. *J. Plant Biol.* 51, 321–326. doi: 10.1007/BF03036133
- Lee, J. H., Ryu, H. S., Chung, K. S., Posé, D., Kim, S., Schmid, M., et al. (2013). Regulation of temperature-responsive flowering by MADS-Box Transcription Factor Repressors. *Science* 342, 628–632. doi: 10.1126/science.1241097
- Lee, K. C., Jang, Y. H., Kim, S.-K., Park, H.-Y., Thu, M. P., Lee, J. H., et al. (2017). RRM domain of *Arabidopsis* splicing factor SF1 is important for pre-mRNA splicing of a specific set of genes. *Plant Cell Rep.* 36, 1083–1095. doi: 10.1007/s00299-017-2140-1
- Legris, M., Klose, C., Burgie, E. S., Costigliolo, C., Neme, M., Hiltbrunner, A., et al. (2016). Phytochrome B integrates light and temperature signals in *Arabidopsis*. *Science* 354 (6314), 897–900. doi: 10.1126/science.aaf5656
- Lempe, J., Balasubramanian, S., Sureshkumar, S., Singh, A., Schmid, M., and Weigel, D. (2005). Diversity of flowering responses in wild *Arabidopsis thaliana* strains. *PLoS Genet.* 1, e6. doi: 10.1371/journal.pgen.0010006
- Leviatan, N., Alkan, N., Leshkowitz, D., and Fluhr, R. (2013). Genome-wide survey of cold stress regulated alternative splicing in *Arabidopsis thaliana* with tiling microarray. *PLoS One* 8, e66511. doi: 10.1371/journal.pone.0066511
- Li, D., Liu, C., Shen, L., Wu, Y., Chen, H., Robertson, M., et al. (2008). A repressor complex governs the integration of flowering signals in *Arabidopsis*. *Dev. Cell* 15, 110–120. doi: 10.1016/j.devcel.2008.05.002
- Liu, J. (2001). Isolation of a CONSTANS ortholog from *pharbitis nil* and its role in flowering. *Plant Physiol.* 125, 1821–1830. doi: 10.1104/pp.125.41821
- Lorković, Z. J., Hilscher, J., and Barta, A. (2004). Use of fluorescent protein tags to study nuclear organization of the spliceosomal machinery in transiently transformed living plant cells. *Mol. Biol. Cell* 15, 3233–3243. doi: 10.1091/mbc.e04-01-0055
- Loyer, P., Trembley, J., Katona, R., Kidd, V., and Lahti, J. (2005). Role of CDK/cyclin complexes in transcription and RNA splicing. *Cell. Signalling* 17, 1033–1051. doi: 10.1016/j.cellsig.2005.02.005
- Lutz, U., Posé, D., Pfeifer, M., Gundlach, H., Hagmann, J., Wang, C., et al. (2015). Modulation of ambient temperature-dependent flowering in *Arabidopsis thaliana* by natural variation of FLOWERING LOCUS M. *PLoS Genet.* 11, e1005588. doi: 10.1371/journal.pgen.1005588
- Lutz, U., Nussbaumer, T., Spannagl, M., Diener, J., Mayer, K. F., and Schwechheimer, C. (2017). Natural haplotypes of FLM non-coding sequences fine-tune flowering time in ambient spring temperatures in *Arabidopsis*. *Elife*, e22114. doi: 10.7554/eLife.22114
- Ma, X., Qiao, Z., Chen, D., Yang, W., Zhou, R., Zhang, W., et al. (2015). CYCLIN-DEPENDENT KINASE G2 regulates salinity stress response and salt mediated flowering in *Arabidopsis thaliana*. *Plant Mol. Biol.* 88, 287–299. doi: 10.1007/s11103-015-0324-z
- Marquez, Y., Höpfner, M., Ayatollahi, Z., Barta, A., and Kalyna, M. (2015). Unmasking alternative splicing inside protein-coding exons defines exons and their role in proteome plasticity. *Genome Res.* 25, 995–1007. doi: 10.1101/gr.186585.114
- Mastrangelo, A. M., Marone, D., Laidò, G., De Leonardi, A. M., and De Vita, P. (2012). Alternative splicing: enhancing ability to cope with stress via transcriptome plasticity. *Plant Sci.* 185–186, 40–49. doi: 10.1016/j.plantsci.2011.09.006
- Mattson, N. S., and Erwin, J. E. (2005). The impact of photoperiod and irradiance on flowering of several herbaceous ornamentals. *Sci. Hortic.* 104, 275–292. doi: 10.1016/j.scienta.2004.08.018
- Menges, M., De Jager, S. M., Gruissem, W., and Murray, J. A. H. (2005). Global analysis of the core cell cycle regulators of *Arabidopsis* identifies novel genes, reveals multiple and highly specific profiles of expression and provides a coherent model for plant cell cycle control: global analysis of plant core cell cycle regulators. *Plant J.* 41, 546–566. doi: 10.1111/j.1365-313X.2004.02319.x
- Meyer, M., Plass, M., Pérez-Valle, J., Eyraes, E., and Vilardell, J. (2011). Deciphering 3' ss selection in the yeast genome reveals an RNA thermosensor that mediates alternative splicing. *Mol. Cell* 43, 1033–1039. doi: 10.1016/j.molcel.2011.07.030
- Moore, F. C., and Lobell, D. B. (2015). The fingerprint of climate trends on European crop yields. *Proc. Natl. Acad. Sci. U.S.A.* 112, 2670–2675. doi: 10.1073/pnas.1409606112
- Nakano, Y., Higuchi, Y., Sumitomo, K., and Hisamatsu, T. (2013). Flowering retardation by high temperature in chrysanthemums: involvement of FLOWERING LOCUS T-like 3 gene repression. *J. Exp. Bot.* 64, 909–920. doi: 10.1093/jxb/ers370
- Nordborg, M., and Bergelson, J. (1999). The effect of seed and rosette cold treatment on germination and flowering time in some *Arabidopsis thaliana* (Brassicaceae) ecotypes. *Am. J. Bot.* 86, 470–475. doi: 10.2307/2656807
- Paik, I., and Huq, E. (2019). Plant photoreceptors: Multi-functional sensory proteins and their signaling networks. *Semin. In Cell Dev. Biol.* 92, 114–121. doi: 10.1016/j.semcdb.2019.03.007
- Pajoro, A., Severing, E., Angenent, G. C., and Immink, R. G. H. (2017). Histone H3 lysine 36 methylation affects temperature-induced alternative splicing and flowering in plants. *Genome Biol.* 18, 102. doi: 10.1186/s13059-017-1235-x
- Park, H.-Y., Lee, H. T., Lee, J. H., and Kim, J.-K. (2019). *Arabidopsis* U2AF65 regulates flowering time and the growth of pollen tubes. *Front. Plant Sci.* 10, 569. doi: 10.3389/fpls.2019.00569
- Posé, D., Verhage, L., Ott, F., Yant, L., Mathieu, J., Angenent, G. C., et al. (2013). Temperature-dependent regulation of flowering by antagonistic FLM variants. *Nature* 503, 414–417. doi: 10.1038/nature12633
- Rane, S. G., Dubus, P., Mettus, R. V., Galbreath, E. J., Boden, G., Reddy, E. P., et al. (1999). Loss of Cdk4 expression causes insulin-deficient diabetes and Cdk4 activation results in β -islet cell hyperplasia. *Nat. Genet.* 22, 44–52. doi: 10.1038/8751
- Ratcliffe, O. J. (2003). Analysis of the *Arabidopsis* MADS AFFECTING FLOWERING gene family: MAF2 prevents vernalization by short periods of cold. *Plant Cell Online* 15, 1159–1169. doi: 10.1105/tpc.009506
- Reed, R. (2003). Coupling transcription, splicing and mRNA export. *Curr. Opin. In Cell Biol.* 15, 326–331. doi: 10.1016/S0955-0674(03)00048-6

- Reeves, P. A., He, Y., Schmitz, R. J., Amasino, R. M., Panella, L. W., and Richards, C. M. (2007). Evolutionary conservation of the *FLOWERING LOCUS C*-mediated vernalization response: evidence from the sugar beet (*Beta vulgaris*). *Genetics* 176, 295–307. doi: 10.1534/genetics.106.069336
- Romera-Branchat, M., Andrés, F., and Coupland, G. (2014). Flowering responses to seasonal cues: what's new? *Curr. Opin. In Plant Biol.* 21, 120–127. doi: 10.1016/j.pbi.2014.07.006
- Sainsbury, F., Thuenemann, E. C., and Lomonosoff, G. P. (2009). pEAQ: versatile expression vectors for easy and quick transient expression of heterologous proteins in plants. *Plant Biotechnol. J.* 7, 682–693. doi: 10.1111/j.1467-7652.2009.00434.x
- Samach, A., and Wigge, P. A. (2005). Ambient temperature perception in plants. *Curr. Opin. In Plant Biol.* 8, 483–486. doi: 10.1016/j.pbi.2005.07.011
- Scortecchi, K. C., Michaels, S. D., and Amasino, R. M. (2001). Identification of a MADS-box gene, *FLOWERING LOCUS M*, that represses flowering. *Plant J.* 26, 229–236. doi: 10.1046/j.1365-313x.2001.01024.x
- Searle, I., and Coupland, G. (2004). Induction of flowering by seasonal changes in photoperiod. *EMBO J.* 23, 1217–1222. doi: 10.1038/sj.emboj.7600117
- Sherr, C. J., and Roberts, J. M. (1999). CDK inhibitors: positive and negative regulators of G1-phase progression. *Genes Dev.* 13, 1501–1512.
- Shindo, C., Lister, C., Crevillen, P., Nordborg, M., and Dean, C. (2006). Variation in the epigenetic silencing of *FLC* contributes to natural variation in *Arabidopsis* vernalization response. *Genes Dev.* 20, 3079–3083. doi: 10.1101/gad.405306
- Song, Y. H., Ito, S., and Imaizumi, T. (2013). Flowering time regulation: photoperiod- and temperature-sensing in leaves. *Trends In Plant Sci.* 18, 575–583. doi: 10.1016/j.tplants.2013.05.003
- Srikanth, A., and Schmid, M. (2011). Regulation of flowering time: all roads lead to Rome. *Cell. Mol. Life Sci.* 68, 2013–2037. doi: 10.1007/s00018-011-0673-y
- Staiger, D., and Brown, J. W. S. (2013). Alternative splicing at the intersection of biological timing, development, and stress responses. *Plant Cell* 25, 3640–3656. doi: 10.1105/tpc.113.113803
- Steffen, A., Elgner, M., and Staiger, D. (2019). Regulation of flowering time by the RNA-binding proteins *AtGRP7* and *AtGRP8*. *Plant Cell Physiol.* 60, 2040–2050. doi: 10.1093/pcp/pcz124
- Strange, A., Li, P., Lister, C., Anderson, J., Warthmann, N., Shindo, C., et al. (2011). Major-effect alleles at relatively few loci underlie distinct vernalization and flowering variation in *Arabidopsis* accessions. *PLoS One* 6, e19949. doi: 10.1371/journal.pone.0019949
- Strausfeld, U. P., Howell, M., Descombes, P., Chevalier, S., Rempel, R. E., Adamczewski, J., et al. (1996). Both cyclin A and cyclin E have S-phase promoting (SPF) activity in *Xenopus* egg extracts. *J. Cell Sci.* 109, 1555–1563.
- Streitner, C., Simpson, C. G., Shaw, P., Danisman, S., Brown, J. W. S., and Staiger, D. (2013). Small changes in ambient temperature affect alternative splicing in *Arabidopsis thaliana*. *Plant Signaling Behav.* 8, e24638. doi: 10.4161/psb.24638
- Susila, H., Nasim, Z., and Ahn, J. (2018). Ambient temperature-responsive mechanisms coordinate regulation of flowering Time. *Int. J. Mol. Sci.* 19, 3196. doi: 10.3390/ijms19103196
- Terakita, A., and Nagata, T. (2014). Functional properties of opsins and their contribution to light-sensing physiology. *Zool. Sci.* 31, 653–659. doi: 10.2108/zs140094
- Theissen, G., Rümpler, F., and Gramzow, L. (2018). Array of MADS-Box Genes: Facilitator for Rapid Adaptation? *Trends In Plant Sci.* 23, 563–576. doi: 10.1016/j.tplants.2018.04.008
- Umeda, M. (2005). Control of cell division and transcription by cyclin-dependent kinase-activating kinases in Plants. *Plant Cell Physiol.* 46, 1437–1442. doi: 10.1093/pcp/pci170
- Van Leene, J., Hollunder, J., Eeckhout, D., Persiau, G., Van De Slijke, E., Stals, H., et al. (2010). Targeted interactomics reveals a complex core cell cycle machinery in *Arabidopsis thaliana*. *Mol. Syst. Biol.* 6, 397. doi: 10.1038/msb.2010.53
- Verhage, L., Severing, E. I., Bucher, J., Lammers, M., Busscher-Lange, J., Bonnema, G., et al. (2017). Splicing-related genes are alternatively spliced upon changes in ambient temperatures in plants. *PLoS One* 12, e0172950. doi: 10.1371/journal.pone.0172950
- Vriens, J., Nilius, B., and Voets, T. (2014). Peripheral thermosensation in mammals. *Nat. Rev. Neurosci.* 15, 573–589. doi: 10.1038/nrn3784
- Wang, Z.-W., Wu, Z., Raitskin, O., Sun, Q., and Dean, C. (2014). Antisense-mediated *FLC* transcriptional repression requires the P-TEFb transcription elongation factor. *Proc. Natl. Acad. Sci.* 111, 7468–7473. doi: 10.1073/pnas.1406635111
- Wang, Q., Zuo, S., Wang, X., Liu, Q., Gu, L., Oka, Y., et al. (2018). Beyond the photocycle — how cryptochromes regulate photoresponses in plants? *Curr. Opin. In Plant Biol.* 45, 120–126. doi: 10.1016/j.pbi.2018.05.014
- Weller, J. L., Beauchamp, N., Kerckhoffs, L. H. J., Platten, J. D., and Reid, J. B. (2001). Interaction of phytochromes A and B in the control of de-etiolation and flowering in pea: Phytochromes A and B in pea. *Plant J.* 26, 283–294. doi: 10.1046/j.1365-313X.2001.01027.x
- Werner, J. D., Borevitz, J. O., Warthmann, N., Trainer, G. T., Ecker, J. R., Chory, J., et al. (2005). Quantitative trait locus mapping and DNA array hybridization identify an *FLM* deletion as a cause for natural flowering-time variation. *Proc. Natl. Acad. Sci.* 102, 2460–2465. doi: 10.1073/pnas.0409474102
- Wheeler, T., and von Braun, J. (2013). Climate change impacts on global food security. *Science* 341, 508–513. doi: 10.1126/science.1239402
- Whittaker, C., and Dean, C. (2017). The *FLC* locus: a platform for discoveries in epigenetics and adaptation. *Annu. Rev. Cell Dev. Biol.* 33, 555–575. doi: 10.1146/annurev-cellbio-100616-060546
- Wigge, P. A. (2013). Ambient temperature signalling in plants. *Curr. Opin. In Plant Biol.* 16, 661–666. doi: 10.1016/j.pbi.2013.08.004
- Will, C. L., and Luhrmann, R. (2011). Spliceosome structure and function. *Cold Spring Harbor Perspect. In Biol.* 3, a003707–a003707. doi: 10.1101/cshperspect.a003707
- Wu, F.-H., Shen, S.-C., Lee, L.-Y., Lee, S.-H., Chan, M.-T., and Lin, C.-S. (2009). Tape-*Arabidopsis* Sandwich - a simpler *Arabidopsis* protoplast isolation method. *Plant Methods* 5, 16. doi: 10.1186/1746-4811-5-16
- Xu, F., Xu, S., Wiermer, M., Zhang, Y., and Li, X. (2012). The cyclin L homolog *MOS12* and the *MOS4*-associated complex are required for the proper splicing of plant resistance genes: *MOS12* and *MAC* in the splicing of R genes. *Plant J.* 70, 916–928. doi: 10.1111/j.1365-313X.2012.04906.x
- Zhang, R., Calixto, C. P. G., Tzioutziou, N. A., James, A. B., Simpson, C. G., Guo, W., et al. (2015). *AtRTD* - a comprehensive reference transcript dataset resource for accurate quantification of transcript-specific expression in *Arabidopsis thaliana*. *New Phytol.* 208, 96–101. doi: 10.1111/nph.13545
- Zheng, T., Nibau, C., Phillips, D. W., Jenkins, G., Armstrong, S. J., and Doonan, J. H. (2014). CDKG1 protein kinase is essential for synapsis and male meiosis at high ambient temperature in *Arabidopsis thaliana*. *Proc. Natl. Acad. Sci.* 111, 2182–2187. doi: 10.1073/pnas.1318460111

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2020 Nibau, Gallemí, Dadarou, Doonan and Cavallari. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



Alternative Splicing and DNA Damage Response in Plants

Barbara Anna Nimeth[†], Stefan Riegler^{†‡} and Maria Kalyna^{*†}

Department of Applied Genetics and Cell Biology, BOKU—University of Natural Resources and Life Sciences, Vienna, Austria

OPEN ACCESS

Edited by:

Paula Casati,
CONICET Center for Photosynthetic
and Biochemical Studies (CEFOBI),
Argentina

Reviewed by:

Aditya Banerjee,
St. Xavier's College, India
Rohini Garg,
Shiv Nadar University, India

*Correspondence:

Maria Kalyna
mariya.kalyna@boku.ac.at

†ORCID ID:

Barbara Anna Nimeth
orcid.org/0000-0001-8692-8413
Stefan Riegler
orcid.org/0000-0003-3413-1343
Maria Kalyna
orcid.org/0000-0003-4702-7625

‡Present Address:

Stefan Riegler,
Institute of Science and Technology
Austria, Klosterneuburg, Austria

Specialty section:

This article was submitted to
Plant Physiology,
a section of the journal
Frontiers in Plant Science

Received: 29 July 2019

Accepted: 21 January 2020

Published: 19 February 2020

Citation:

Nimeth BA, Riegler S and Kalyna M
(2020) Alternative Splicing and DNA
Damage Response in Plants.
Front. Plant Sci. 11:91.
doi: 10.3389/fpls.2020.00091

Plants are exposed to a variety of abiotic and biotic stresses that may result in DNA damage. Endogenous processes - such as DNA replication, DNA recombination, respiration, or photosynthesis - are also a threat to DNA integrity. It is therefore essential to understand the strategies plants have developed for DNA damage detection, signaling, and repair. Alternative splicing (AS) is a key post-transcriptional process with a role in regulation of gene expression. Recent studies demonstrate that the majority of intron-containing genes in plants are alternatively spliced, highlighting the importance of AS in plant development and stress response. Not only does AS ensure a versatile proteome and influence the abundance and availability of proteins greatly, it has also emerged as an important player in the DNA damage response (DDR) in animals. Despite extensive studies of DDR carried out in plants, its regulation at the level of AS has not been comprehensively addressed. Here, we provide some insights into the interplay between AS and DDR in plants.

Keywords: alternative splicing, DNA repair, DNA damage response, Arabidopsis, plant, stress, splicing factor

DNA DAMAGE RESPONSE IN PLANTS

The genomic integrity of living cells is perpetually challenged by a variety of environmental and internal cellular factors. Environmental stresses, such as drought, salinity, ultraviolet (UV), ionizing radiation, xenobiotic toxicity, heavy metals, and mutagenic chemicals damage DNA and affect its stability (Hu et al., 2016; Nisa et al., 2019). Cellular replication, recombination errors, and reactive oxygen species resulting as a byproduct of metabolism also cause DNA damage. A cell's reaction to genotoxic stress, referred to as DNA damage response (DDR), starts with cell cycle arrest and, in the case of plants, endoreplication (De Veylder et al., 2011). To ensure the repair of a variety of different types of DNA lesions, several DNA repair mechanisms are active and constitute the DNA repair phase of DDR. Should the repair of DNA damage not be sufficient, programmed cell death eliminates the damaged cell and ensures homeostasis (Manova and Gruszka, 2015; Kim et al., 2019). Due to their sessile nature, plants find themselves at increased risk to detrimental environmental factors. It has also been shown that light and temperature conditions affect DNA repair mechanisms such as homologous recombination and photoreactivation (Li et al., 2002; Boyko et al., 2005).

The repair of UV-induced lesions by photoreactivation appears to be an ancient conserved DNA damage repair mechanism. It relies on the activity of photolyase, utilizing the energy of UV-A or blue light to reverse UV damage in the DNA (Manova and Gruszka, 2015; Kavakli et al., 2017; Zhang et al., 2017a). Another mechanism of UV damage repair is nucleotide excision repair (NER), which identifies,

removes, and repairs the damaged base(s) using the other DNA strands as a template. In addition to UV lesions, NER repairs bulky adducts that change the DNA conformation. Global genomic repair (GGR) and transcription-coupled repair (TCR), although differing in their mode of damage recognition, share similarities in their mechanisms of action (Hanawalt, 2002). The DNA glycosylases, which initiate base excision repair (BER) at damaged sites, facilitate the repair of a variety of DNA lesions (Wallace, 2014). There is evidence for BER being active in chloroplasts to counter the effects of reactive oxygen species production during photosynthesis (Gutman and Niyogi, 2009). The mismatch repair (MMR) pathway is responsible for the repair of replication errors, such as mismatches and indels, UV, and oxidative damage (Li et al., 2016; Liu et al., 2017; Belfield et al., 2018). Double-strand breaks (DSBs) are repaired *via* non-homologous end joining (NHEJ) and homologous recombination (HR). While HR requires homologous sequences to ensure efficient repair, NHEJ joins DSBs without considering sequence context and is, thus, an error prone mechanism, which can result in mutations and DNA changes (Manova and Gruszka, 2015).

Two protein kinases, ATM (ATAXIA-TELANGIECTASIA MUTATED) and ATR (ATAXIA TELANGIECTASIA-MUTATED AND RAD3-RELATED), initiate eukaryotic DDR. Once activated, they signal *via* checkpoint kinases 1 and 2 (CHK1 and CHK2), respectively. Human homologs of CHK1 and CHK2 activate p53, which in turn controls cell cycle arrest, DNA damage repair, and programmed cell death. While the downstream processes of ATM, ATR, and p53 have been studied extensively, data on their upstream activation and regulation remains scarce. Neither orthologs of CHK1 and CHK2, nor of p53, have been identified in plants so far. However, a functional homolog of p53, SUPPRESSOR OF GAMMA RESPONSE 1 (SOG1), transcriptionally regulating DDR downstream of ATM and ATR was found (Preuss and Britt, 2003; Yoshiyama et al., 2009; Yoshiyama, 2016). Indeed, SOG1 was identified as a master regulator transcription factor of the plant DDR, influencing expression of genes related to the cell cycle and DNA repair (Ogita et al., 2018). About 300 direct targets of SOG1 were identified, including transcription factors, DNA repair genes, and regulators of the cell cycle (Bourbousse et al., 2018).

A recent research update highlights the growing interest in DDR in plants but also serves to show that a role for alternative splicing (AS) remains to be established (Gimenez and Manzano-Agugliaro, 2017).

OVERVIEW OF ALTERNATIVE SPLICING

Most messenger RNAs in higher eukaryotes are synthesized as precursors, which contain intervening sequences, known as introns. To provide a template for protein synthesis, messenger RNA (mRNA) introns have to be removed and exons joined in a process termed pre-mRNA splicing. However, exons and introns or their parts can be differentially included in mRNA by AS. AS produces transcript and protein variants from a single gene with different fates and functions, and is a fundamental aspect of RNA

biology that has a key role in our understanding of gene expression regulation. Up to 95% of human and 70% of plant multi-exonic genes are alternatively spliced (Pan et al., 2008; Wang et al., 2008; Marquez et al., 2012; Chamala et al., 2015; Zhang et al., 2017b). Further studies report that about 50% of the genes in soybeans, 46% in rice, 40% in maize, and over 60% in tomatoes and barley undergo AS (Thatcher et al., 2014; Chamala et al., 2015; Clark et al., 2019; Rapazote-Flores et al., 2019), emphasizing its importance in crop plant development and environmental response. AS has a broad role in many aspects of plant biology, but its role in responding to DNA damage is mostly unknown and requires further investigation.

Pre-mRNA splicing requires the core splicing signals, which consist of the 5' and 3' splice sites and a branch site (Wang and Burge, 2008). However, multiple additional features, such as intronic and exonic splicing regulatory *cis*-elements (splicing enhancers and silencers), length of introns and exons, and differential guanine-cytosine content between exons and introns, affect the recognition and selection of the core splicing signals (Braunschweig et al., 2013). The secondary structure of the pre-mRNA can alter access to splicing signals and binding sites for splicing factors (SFs) or change the distance between these elements (Shepard and Hertel, 2008). Differential DNA methylation, histone modifications, and nucleosome positioning modulate RNA polymerase II elongation speed and recruitment of SFs, thus also resulting in alternative splice site selection [for a recent review see (Jabre et al., 2019)].

Common types of AS events include exon skipping, usage of alternative 5' and 3' splice sites, mutually exclusive exons, and intron retention. Exon skipping is the predominant event in animals, whereas it is infrequent in plants (Marquez et al., 2012; Braunschweig et al., 2013). Intron retention is widespread both in plants and animals (Marquez et al., 2012; Braunschweig et al., 2014). Interestingly, intron retention transcripts are often not substrates for nonsense-mediated mRNA decay due to their nuclear localization (James et al., 2012; Kalyna et al., 2012; Leviatan et al., 2013; Gohring et al., 2014). Retention of introns may regulate protein abundance during developmental transitions and in response to stress (including DNA damage). When transcripts with retained introns are recognized as incompletely processed they remain in the nucleus until a change in the cellular environment results in post-transcriptional splicing (Yap et al., 2012; Boothby et al., 2013; Boutz et al., 2015; Brown et al., 2015). Microexons (ultra-short exons of 3-30 nucleotides) found in hundreds of animal genes, and recently identified exons (alternatively spliced internal regions of protein-coding exons), which occur in ~7% of Arabidopsis and 4% of human protein-coding genes, complement the repertoire of AS events (Marquez et al., 2012; Irimia et al., 2014; Marquez et al., 2015; Staiger and Simpson, 2015; Sibley et al., 2016; Ustianenko et al., 2017; Zhang et al., 2017b).

Hundreds of proteins participate in the splicing process (Chen and Moore, 2015). However, the modulation of splice site recognition is mainly governed by two families of SFs - serine/arginine-rich (SR) proteins and heterogeneous nuclear

ribonucleoproteins (hnRNPs) - through binding to regulatory *cis*-elements in the pre-mRNA (Barta et al., 2010; Manley and Krainer, 2010; Yeap et al., 2014; Howard and Sanford, 2015). SR proteins and hnRNPs act as activators and repressors of splice site selection, respectively, however, the effect often depends on their binding position. Expression levels, localization, and post-translational modifications (PTMs) (phosphorylation, acetylation, ubiquitination, and sumoylation) of SFs in a particular cell are one of the components of the splicing code, which governs the AS outcomes (Barash et al., 2010; Baralle and Baralle, 2018). Interestingly, SR proteins and hnRNPs participate in multiple cellular processes, such as mRNA export, RNA stability and quality control, and translation.

ALTERNATIVE SPLICING AND DNA DAMAGE RESPONSE, INSIGHTS FROM STUDIES IN ANIMALS

It is becoming clear that RNA-binding proteins and AS are important in DDR. One of the first pieces of evidence that SFs may have a role in DDR came from a study which demonstrated that the depletion of a canonical human SR protein, SRSF1 (SF2/ASF), resulted in increased DSB formation and genome instability (Li and Manley, 2005). Several studies in animals have unexpectedly identified SFs and other RNA processing proteins associated with response to irradiation and DNA damaging chemicals. For example, genome-wide siRNA knockdown of multiple genes have shown that splicing and RNA processing factors are the most enriched functional category within factors whose depletion mediates DNA damage (Paulsen et al., 2009; Lackner et al., 2011). Studies of individual SFs, including SR proteins, have demonstrated changes in their expression levels, AS profiles, phosphorylation state, and subcellular distribution in response to DNA damage (Matsuoka et al., 2007; Busa et al., 2010; Sakashita and Endo, 2010; Ip et al., 2011; Adamson et al., 2012; Leva et al., 2012). The importance of AS and splicing factors in DDR in animals has been reviewed extensively (Naro et al., 2015; Shkreta and Chabot, 2015; Giono et al., 2016; Kai, 2016; Mikolaskova et al., 2018).

The interplay between DDR and AS occurs at multiple levels (Figure 1). One of the most rapid responses to stress and DNA damage is the change in activity of already translated proteins by PTMs. Multiple SFs have been identified in DDR-regulated phosphoproteomes (Bennetzen et al., 2010; Bensimon et al., 2010; Beli et al., 2012). The kinases ATM and ATR are directly activated by DNA lesions and phosphorylate hundreds of proteins in response to ionizing radiation, including several hnRNPs and SR proteins (Matsuoka et al., 2007). Studies using the treatment of mammalian cells with several genotoxic agents revealed reduced SR protein phosphorylation levels affecting their accumulation in nuclear granules. These studies also found differential AS of genes involved in DNA repair, cell cycle control, and apoptosis (Bennetzen et al., 2010; Leva et al., 2012; Shkreta et al., 2016). Remarkably, detained introns, a recently identified subgroup of

retained introns, are enriched in genes involved in DDR. Moreover, DNA damage and the activity of certain Clk kinases, which maintain the hyperphosphorylated status of SR proteins, can modulate splicing of detained introns (Boutz et al., 2015). Changes in the activity of SR proteins also have been associated with their acetylation state in response to cisplatin-induced DNA damage (Edmond et al., 2011; Nakka et al., 2015). Interestingly, acetyltransferases can indirectly impact the translocation of SR proteins *via* the modification of SR protein kinases (Edmond et al., 2011). Recent studies also demonstrated the acetylation of hnRNPs in response to DNA damage (Magni et al., 2019; Siam et al., 2019). Ubiquitination, besides its regulatory activity during spliceosome assembly, affects SFs upon DNA damage (Lu and Legerski, 2007). Genotoxic agents cause deubiquitylation and sumoylation of hnRNPs (Vassileva and Matunis, 2004).

As localization and shuttling of SFs is highly dependent on their phosphorylation state, it is not surprising that DNA damage-induced nuclear translocation of SR protein kinases results in the hyperphosphorylation and subsequent nuclear accumulation of certain SR proteins (Edmond et al., 2011). UV irradiation also affects the redistribution of SFs into the cytoplasm, therefore impacting AS (van der Houven van Oordt et al., 2000; Llorian et al., 2005; Guil et al., 2006). The DNA damage-induced relocalization of SFs appears to be dependent on cell type and genotoxic treatment (Tissier et al., 2010; Wong et al., 2013).

In plants, members of different Arabidopsis SR protein subfamilies localize into distinct populations of nuclear speckles (Lorkovic et al., 2008), with their localization dependent on their phosphorylation status (Ali et al., 2003; Tillemans et al., 2005). Different classes of kinases (such as SR protein kinases, PRP4 kinases, Cdc2-like or LAMMER-type kinases, and mitogen-activated protein kinases) phosphorylate plant SFs, including SR proteins and hnRNPs (Golovkin and Reddy, 1999; Savaldi-Goldstein et al., 2000; Feilner et al., 2005; de la Fuente van Bentem et al., 2006; de la Fuente van Bentem et al., 2008; Kanno et al., 2018), suggesting that DNA damage in plants could lead to altered SF activities and changes in AS. However, to which extent this occurs, which SFs are affected and the roles of different PTMs remain the subject of further studies.

In addition to the post-translational regulation of SFs during DDR, their activity can be altered by changes in their AS. Studies in animal cells have illustrated the impact DNA damage has on the AS of SF genes (Solier et al., 2010; Ip et al., 2011; Leva et al., 2012). Munoz and colleagues describe a mechanism by which AS is regulated during DDR (Munoz et al., 2009; Munoz et al., 2017). The hyperphosphorylation of the C-terminal domain of RNA polymerase II (RNAPII) is associated with a decrease in RNAPII elongation speed. This slowing down of RNAPII favors the selection of weaker splice sites as the time window for their recognition by the splicing machinery is extended before stronger downstream sites are synthesized. The hyperphosphorylation and slowdown of RNAPII in response to UV exposure leads to differential exon skipping events in multiple genes associated with apoptosis, cell cycle, and cancer (Munoz et al., 2009; Munoz et al., 2017). These findings raise questions regarding the mechanisms and PTMs affecting RNAPII

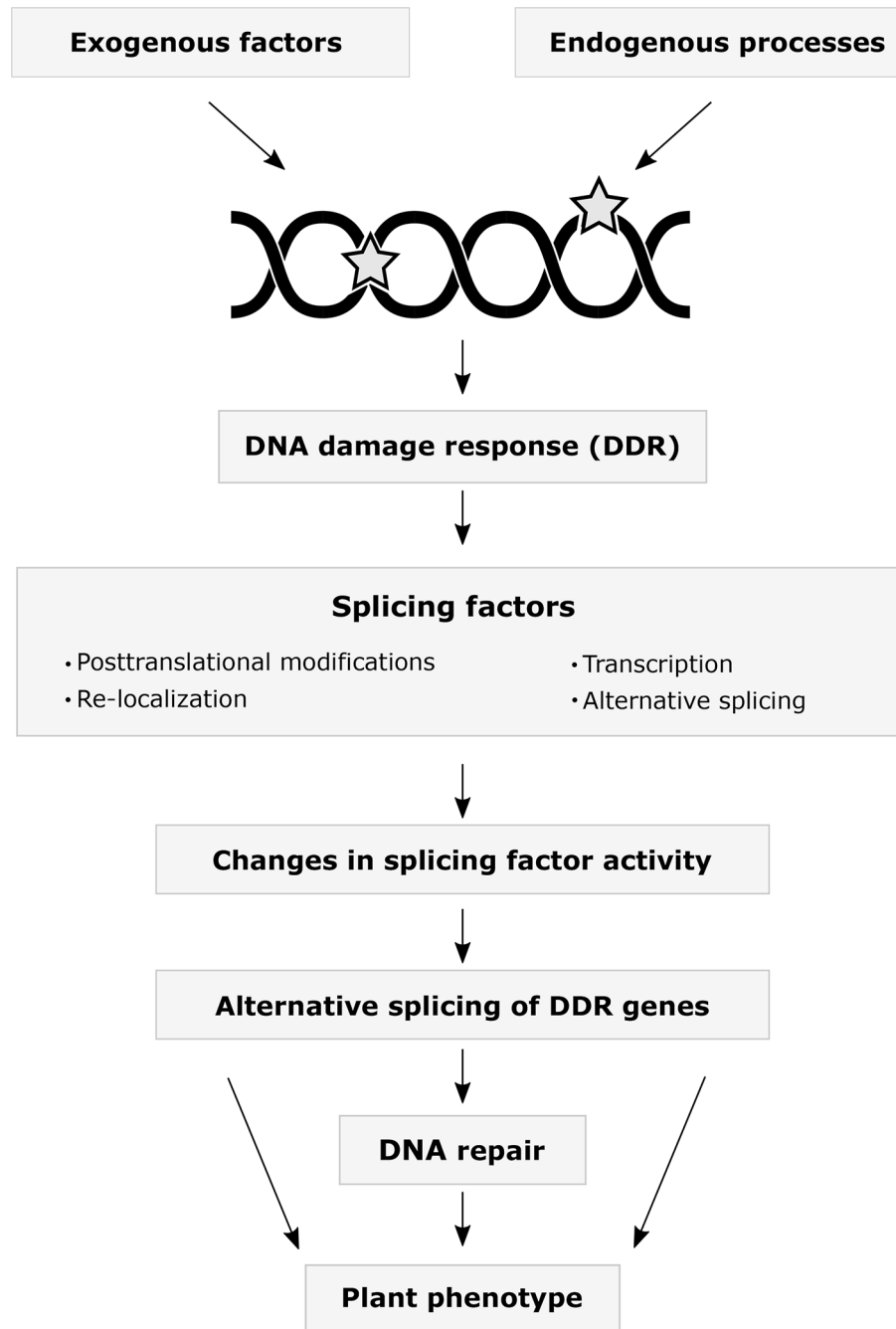


FIGURE 1 | The interplay between the DNA damage response and alternative splicing. A variety of exogenous environmental stress factors and endogenous cellular processes may result in DNA damage. Numerous studies on animals have demonstrated that splicing factors change their expression levels, alternative splicing patterns, post-translational modification states, and subcellular localization in response to DNA damage. Altered expression and activities of splicing factors may regulate DNA repair by modulating alternative splicing of DDR genes. Current data indicates that many plant DDR genes undergo alternative splicing. Which plant splicing factors are involved in the DDR, how they are regulated, what are their target genes, and how the splicing changes are translated into the plant phenotype remains to be addressed in the future.

TABLE 1 | Overview of alternative splicing in genes involved in DNA damage response.

	Gene name	Gene ID	Alternative splicing			Gene name	Gene ID	Alternative splicing			
			AtRTD2 ¹	Reference				AtRTD2 ¹	Reference		
A	Base excision repair (BER)					C	Homologous recombination (HR)				
	OGG1	At1g21710	+		MRE11		At5g54260	+			
	FPG	At1g52500	+	2,3	RAD50		At2g31970	+			
	NTH1	At2g31450	+		NBS1		At3g02680	+			
	NTH2	At1g05900	+		COM1		At3g52115	+			
	DME	At5g04560	+		RECQ4A		At1g10930	+			
	ROS1	At2g36490	+		RAD51		At5g20850	+			
	UNG	At3g18630	–		RAD51B		At2g28560	+			
	DML3	At4g34060	+		RAD51C		At2g45280	+	7		
	MBD4L	At3g07930	+	4	RAD51D		At1g07745	+			
	ARP	At2g41460	+		XRCC2		At5g64520	+			
	APE1L	At3g48425	+		XRCC3		At5g54750	+	7		
	APE2	At4g36050	+		BRCA2A		At4g00020	+			
	ZDP	At3g14890	+		BRCA2B		At5g01630	–			
	TDP1	At5g15170	+		RAD54		At3g19210	+			
	XRCC1	At1g80420	+		SRS2		At4g25120	+			
	SAV6	At5g26680	+		FANCM		At1g35530	+			
	PARP1	At2g31320	+		EME1A		At2g21800	+			
	PARP2	At4g02390	+		EME1B		At2g22140	+			
		Pol δ	See section E				MUS81	At4g30870	+		
	Pol ε	See section E			GEN1	At1g01880	+				
	LIG1	See section E			SEND1	At3g48900	+				
B	Nucleotide excision repair (NER)						TOP3α	At5g63920	+		
	RAD4	At5g16630	+			RMI1	At5g63540	+			
	RAD23A	At1g16190	+			Pol δ		See section E			
	RAD23B	At1g79650	+	5		PCNA		See section E			
	RAD24C	At3g02540	+	5		RFC		See section E			
	RAD23D	At5g38470	+	5	D	DNA mismatch repair (MMR)					
	CEN2	At4g37010	+			MSH2	At3g18524	–			
	DDB1A	At4g05420	+			MSH3	At4g25540	–			
	DDB1B	At4g21100	–			MSH6	At4g02070	+			
	DDB2	At5g58760	+			MSH7	At3g24495	+			
	CSA	At1g27840	+			MLH1	At4g09140	+			
		At1g19750	+			RFC		See section E			
	CHR8	At2g18760	+			PCNA		See section E			
	CHR24	At5g63950	+			EXO1		See section E			
	XPB1	At5g41370	+			RPA		See section F			
	XPB2	At5g41360	+		POL δ		See section E				
	UVH6	At1g03190	+		E	Components involved in metabolic pathways					
	TFIIH1	At1g55750	+			EXO1	At1g29630	+			
		At1g61420	+			PCNA	At1g07370	–			
							At2g29570	–			
	GTF2H2	At1g05055	–			Pol δ	At1g09815	–			
	TFIIH3	At1g18340	+				At2g42120	+			
	TFIIH4	At4g17020	+				At5g63960	+			
	TTDA	At1g12400	+				At1g08260	–			
		At1g62886	–			Pol ε	At2g27120	+			
	CDKD;1	At1g73690	–				At5g22110	+			
	CDKD;2	At1g66750	+			At1g21690	+				
	CDKD;3	At1g18040	+		RFC	At1g63160	–				
	CYCH;1	At5g27620	+			At1g77470	+				
	MAT1	At4g30820	+			At5g22010	+				
	UVH3	At3g28030	+			At5g27740	–				
	UVH1	At5g41150	+	6		LIG1	At1g08130	–			
	ERCC1	At3g05210	+				At1g49250	–			
		RPA	See section F		F	Replication protein A (RPAs)					
		PCNA	See section E			RPA1	At2g06510	+			
		RFC	See section E				At5g08020	+			
		Pol δ	See section E				At5g45400	–			
		Pol ε	See section E				At5g61000	–			
		LIG1	See section E				At4g19130	–			

(Continued)

TABLE 1 | Continued

Gene name	Gene ID	Alternative splicing		Gene name	Gene ID	Alternative splicing	
		AtRTD2 ¹	Reference			AtRTD2 ¹	Reference
				RPA2	At2g24490	–	
					At3g02920	+	
				RPA3	At3g52630	+	
					At4g18590	+	

elongation speed and the subsequent changes in splicing outcomes during DDR in plants. Which plant SFs are alternatively spliced during DDR, how their transcript isoforms differ in their function, and how their AS influences DDR itself also remains to be addressed in the future.

ALTERNATIVE SPLICING, A NEW PLAYER IN THE PLANT DNA DAMAGE RESPONSE?

Despite extensive studies of DDR and AS in animals, comparatively little is known about this relationship in plants. The PubMed search with the terms “Splicing” and “DNA damage” or “DNA repair” returns a handful of papers in the plant field, which is in stark contrast to about 700 non-plant papers. The first papers describing AS of the Arabidopsis DNA damage/repair gene At-FPG/At-MMH DNA glycosylase were published about 20 years ago (Ohtsubo et al., 1998; Murphy and Gao, 2001). Since then, several key DNA repair genes have been reported to undergo AS, supporting the importance of AS in DDR in plants (Table 1). For example, genes encoding At-RAD1/UHV1 (homologous to yeast RAD1 and human XPF DNA repair endonuclease) and AtPOLK polymerase generate AS isoforms in a tissue-specific pattern (Vonarx et al., 2002; Garcia-Ortiz et al., 2004; Garcia-Ortiz et al., 2007). Two Arabidopsis translesion synthesis DNA polymerases, AtREV and AtPOLH, are regulated by AS, and complementation analysis of AtPOLH AS isoforms in *Rad30*-deficient yeast showed that the AtPOLH C-terminus is required for functional activity (Santiago et al., 2009). Several studies also reveal differential AS in DNA repair genes in crop plants, such as rice class II DNA photolyase (Hirouchi et al., 2003), endonuclease OsMUS81 (Mimida et al., 2007), and checkpoint protein OsRad9 (Li et al., 2017).

To estimate the extent of AS in DNA repair genes at the genome-wide level, we queried the Arabidopsis reference transcript dataset (AtRTD2), which contains 82,190 transcripts from 34,212 genes (Zhang et al., 2017b), with a list of 102 Arabidopsis DNA repair genes (Spampinato, 2017). Only nine genes from this list have previously been reported to be alternatively spliced. Remarkably, this survey revealed that more than 80% of these genes show evidence of AS in the AtRTD2 (Table 1). Further, key regulators of DDR in plants, SOG1, ATM, and ATR (not in the Spampinato, 2017 list), also undergo AS. Although this brief survey deals with a subset of

DDR genes, it clearly illustrates a hidden potential for AS and regulation of DDR in plants. Plant mechanisms and SFs involved in DDR regulation remain to be investigated.

CONCLUSIONS

The cellular response to DNA damage must be tightly regulated. Numerous studies on animals reveal interactions between DDR and AS at multiple levels and demonstrate that AS has an important role in DDR. In plants, initial studies show that AS has a function in plant DDR, but many questions remain to be addressed. How is the expression and activity of plant SFs regulated in DDR, what are their target genes, and do RNAPII processivity or changes in chromatin structure convey DDR into differential splicing outcomes in plants? Comprehensive transcriptome analyses will identify genes that show differences in AS patterns in response to genotoxic stress. Moreover, SFs, RNA processing factors, and DNA repair genes that undergo changes in AS may be detected and help determine the complex interplay between DDR and AS in plants. Finally, the major stress factors restrict plant growth and decrease yield in crop plants. Recent studies report extensive AS in crop species, emphasizing the need for further investigations to establish AS involvement in the response mechanisms to stress exposure and DNA damage.

AUTHOR CONTRIBUTIONS

MK designed the project. BN performed the survey of alternative splicing of Arabidopsis DNA repair genes and prepared the table and figure. The manuscript was written by BN, SR, and MK.

FUNDING

This work is supported by the Austrian Science Fund (FWF) (P26333 to MK).

ACKNOWLEDGMENTS

Authors thank Peter Venhuizen and Craig Simpson for their comments on the manuscript.

REFERENCES

- Adamson, B., Smogorzewska, A., Sigoillot, F. D., King, R. W., and Elledge, S. J. (2012). A genome-wide homologous recombination screen identifies the RNA-binding protein RBMX as a component of the DNA-damage response. *Nat. Cell Biol.* 14 (3), 318–328. doi: 10.1038/ncb2426
- 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 (6), 883–893. doi: 10.1046/j.1365-3113X.2003.01932.x
- Baralle, M., and Baralle, F. E. (2018). The splicing code. *Biosystems* 164, 39–48. doi: 10.1016/j.biosystems.2017.11.002
- Barash, Y., Calarco, J. A., Gao, W., Pan, Q., Wang, X., Shai, O., et al. (2010). Deciphering the splicing code. *Nature* 465 (7294), 53–59. doi: 10.1038/nature09000
- 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. *Plant Cell* 22 (9), 2926–2929. doi: 10.1105/tpc.110.078352
- Belfield, E. J., Ding, Z. J., Jamieson, F. J. C., Visscher, A. M., Zheng, S. J., Mithani, A., et al. (2018). DNA mismatch repair preferentially protects genes from mutation. *Genome Res.* 28 (1), 66–74. doi: 10.1101/gr.219303.116
- Beli, P., Lukashchuk, N., Wagner, S. A., Weinert, B. T., Olsen, J. V., Baskcomb, L., et al. (2012). Proteomic investigations reveal a role for RNA processing factor THRAP3 in the DNA damage response. *Mol. Cell* 46 (2), 212–225. doi: 10.1016/j.molcel.2012.01.026
- Bennetzen, M. V., Larsen, D. H., Bunkenborg, J., Bartek, J., Lukas, J., and Andersen, J. S. (2010). Site-specific phosphorylation dynamics of the nuclear proteome during the DNA damage response. *Mol. Cell Proteomics* 9 (6), 1314–1323. doi: 10.1074/mcp.M900616-MCP200
- Bensimon, A., Schmidt, A., Ziv, Y., Elkon, R., Wang, S. Y., Chen, D. J., et al. (2010). ATM-dependent and -independent dynamics of the nuclear phosphoproteome after DNA damage. *Sci. Signal* 3 (151), rs3. doi: 10.1126/scisignal.2001034
- Boothby, T. C., Zipper, R. S., van der Wee, C. M., and Wolniak, S. M. (2013). Removal of retained introns regulates translation in the rapidly developing gametophyte of *Marsilea vestita*. *Dev. Cell* 24 (5), 517–529. doi: 10.1016/j.devcel.2013.01.015
- Bourbousse, C., Vegesna, N., and Law, J. A. (2018). SOG1 activator and MYB3R repressors regulate a complex DNA damage network in Arabidopsis. *Proc. Natl. Acad. Sci. U. S. A.* 115 (52), E12453–e12462. doi: 10.1073/pnas.1810582115
- Boutz, P. L., Bhutkar, A., and Sharp, P. A. (2015). Detained introns are a novel, widespread class of post-transcriptionally spliced introns. *Genes Dev.* 29 (1), 63–80. doi: 10.1101/gad.247361.114
- Boyko, A., Filkowski, J., and Kovalchuk, I. (2005). Homologous recombination in plants is temperature and day-length dependent. *Mutat. Res.* 572 (1–2), 73–83. doi: 10.1016/j.mrfmmm.2004.12.011
- Braunschweig, U., Gueroussov, S., Plocik, A. M., Graveley, B. R., and Blencowe, B. J. (2013). Dynamic integration of splicing within gene regulatory pathways. *Cell* 152 (6), 1252–1269. doi: 10.1016/j.cell.2013.02.034
- Braunschweig, U., Barbosa-Morais, N. L., Pan, Q., Nachman, E. N., Alipanahi, B., Gonatopoulos-Pournatzis, T., et al. (2014). Widespread intron retention in mammals functionally tunes transcriptomes. *Genome Res.* 24 (11), 1774–1786. doi: 10.1101/gr.177790.114
- Brown, J. W., Simpson, C. G., Marquez, Y., Gadd, G. M., Barta, A., and Kalyna, M. (2015). Lost in translation: pitfalls in deciphering plant alternative splicing transcripts. *Plant Cell* 27 (8), 2083–2087. doi: 10.1105/tpc.15.00572
- Busa, R., Geremia, R., and Sette, C. (2010). Genotoxic stress causes the accumulation of the splicing regulator Sam68 in nuclear foci of transcriptionally active chromatin. *Nucleic Acids Res.* 38 (9), 3005–3018. doi: 10.1093/nar/gkq004
- Chamala, S., Feng, G., Chavarro, C., and Barbazuk, W. B. (2015). Genome-wide identification of evolutionarily conserved alternative splicing events in flowering plants. *Front. Bioeng. Biotechnol.* 3, 33. doi: 10.3389/fbioe.2015.00033
- Chen, W., and Moore, M. J. (2015). Spliceosomes. *Curr. Biol.* 25 (5), R181–R183. doi: 10.1016/j.cub.2014.11.059
- Clark, S., Yu, F., Gu, L., and Min, X. J. (2019). Expanding alternative splicing identification by integrating multiple sources of transcription data in tomato. *Front. Plant Sci.* 10, 689. doi: 10.3389/fpls.2019.00689
- de la Fuente van Bentem, S., Anrather, D., Roitinger, E., Djamei, A., Hufnagel, T., Barta, A., et al. (2006). Phosphoproteomics reveals extensive in vivo phosphorylation of Arabidopsis proteins involved in RNA metabolism. *Nucleic Acids Res.* 34 (11), 3267–3278. doi: 10.1093/nar/gkl429
- de la Fuente van Bentem, S., Anrather, D., Dohnal, I., Roitinger, E., Csaszar, E., Joore, J., et al. (2008). Site-specific phosphorylation profiling of Arabidopsis proteins by mass spectrometry and peptide chip analysis. *J. Proteome Res.* 7 (6), 2458–2470. doi: 10.1021/pr8000173
- De Veylder, L., Larkin, J. C., and Schnittger, A. (2011). Molecular control and function of endoreplication in development and physiology. *Trends Plant Sci.* 16 (11), 624–634. doi: 10.1016/j.tplants.2011.07.001
- Edmond, V., Moysan, E., Khochbin, S., Matthias, P., Brambilla, C., Brambilla, E., et al. (2011). Acetylation and phosphorylation of SRSF2 control cell fate decision in response to cisplatin. *EMBO J.* 30 (3), 510–523. doi: 10.1038/emboj.2010.333
- Feilner, T., Hultschig, C., Lee, J., Meyer, S., Immink, R. G., Koenig, A., et al. (2005). High throughput identification of potential Arabidopsis mitogen-activated protein kinases substrates. *Mol. Cell Proteomics* 4 (10), 1558–1568. doi: 10.1074/mcp.M500007-MCP200
- Garcia-Ortiz, M. V., Ariza, R. R., Hoffman, P. D., Hays, J. B., and Roldan-Arjona, T. (2004). Arabidopsis thaliana AtPOLK encodes a DinB-like DNA polymerase that extends mispaired primer termini and is highly expressed in a variety of tissues. *Plant J.* 39 (1), 84–97. doi: 10.1111/j.1365-3113X.2004.02112.x
- Garcia-Ortiz, M. V., Roldan-Arjona, T., and Ariza, R. R. (2007). The noncatalytic C-terminus of AtPOLK Y-family DNA polymerase affects synthesis fidelity, mismatch extension and translesion replication. *FEBS J.* 274 (13), 3340–3350. doi: 10.1111/j.1742-4658.2007.05868.x
- Gimenez, E., and Manzano-Agugliaro, F. (2017). DNA damage repair system in plants: a worldwide research update. *Genes (Basel)* 8 (11), e299. doi: 10.3390/genes8110299
- Giono, L. E., Nieto Moreno, N., Cambindo Botto, A. E., Dujardin, G., Munoz, M. J., and Kornblitt, A. R. (2016). The RNA response to DNA damage. *J. Mol. Biol.* 428 (12), 2636–2651. doi: 10.1016/j.jmb.2016.03.004
- Gohring, J., Jacak, J., and Barta, A. (2014). Imaging of endogenous messenger RNA splice variants in living cells reveals nuclear retention of transcripts inaccessible to nonsense-mediated decay in Arabidopsis. *Plant Cell* 26 (2), 754–764. doi: 10.1105/tpc.113.118075
- Golovkin, M., and Reddy, A. S. (1999). An SC35-like protein and a novel serine/arginine-rich protein interact with Arabidopsis UI-70K protein. *J. Biol. Chem.* 274 (51), 36428–36438. doi: 10.1074/jbc.274.51.36428
- Guil, S., Long, J. C., and Caceres, J. F. (2006). hnRNP A1 relocalization to the stress granules reflects a role in the stress response. *Mol. Cell Biol.* 26 (15), 5744–5758. doi: 10.1128/MCB.00224-06
- Gutman, B. L., and Niyogi, K. K. (2009). Evidence for base excision repair of oxidative DNA damage in chloroplasts of Arabidopsis thaliana. *J. Biol. Chem.* 284 (25), 17006–17012. doi: 10.1074/jbc.M109.008342
- Hanawalt, P. C. (2002). Subpathways of nucleotide excision repair and their regulation. *Oncogene* 21 (58), 8949–8956. doi: 10.1038/sj.onc.1206096
- Hirouchi, T., Nakajima, S., Najrana, T., Tanaka, M., Matsunaga, T., Hidema, J., et al. (2003). A gene for a Class II DNA photolyase from *Oryza sativa*: cloning of the cDNA by dilution-amplification. *Mol. Genet. Genomics* 269 (4), 508–516. doi: 10.1007/s00438-003-0856-9
- Howard, J. M., and Sanford, J. R. (2015). The RNAissance family: SR proteins as multifaceted regulators of gene expression. *Wiley Interdiscip. Rev. RNA* 6 (1), 93–110. doi: 10.1002/wrna.1260
- Hu, Z., Cools, T., and De Veylder, L. (2016). Mechanisms used by plants to cope with DNA damage. *Annu. Rev. Plant Biol.* 67, 439–462. doi: 10.1146/annurev-arplant-043015-111902
- Ip, J. Y., Schmidt, D., Pan, Q., Ramani, A. K., Fraser, A. G., Odom, D. T., et al. (2011). Global impact of RNA polymerase II elongation inhibition on alternative splicing regulation. *Genome Res.* 21 (3), 390–401. doi: 10.1101/gr.111070.110
- Irimia, M., Weatheritt, R. J., Ellis, J. D., Parikshak, N. N., Gonatopoulos-Pournatzis, T., Babor, M., et al. (2014). A highly conserved program of neuronal microexons is misregulated in autistic brains. *Cell* 159 (7), 1511–1523. doi: 10.1016/j.cell.2014.11.035
- Ishikawa, Y., Endo, M., Abe, K., Osakabe, K., Nakajima, N., Saji, H., et al. (2004). Isolation of four RAD23 Genes from Arabidopsis thaliana and detection of alternative splicing variants. *Plant Biotechnol.* 21 (1), 65–71. doi: 10.5511/plantbiotechnology.21.65
- Jabre, I., Reddy, A. S. N., Kalyna, M., Chaudhary, S., Khokhar, W., Byrne, L. J., et al. (2019). Does co-transcriptional regulation of alternative splicing mediate plant stress responses? *Nucleic Acids Res.* 47 (6), 2716–2726. doi: 10.1093/nar/gkz121

- James, A. B., Syed, N. H., Bordage, S., Marshall, J., Nimmo, G. A., Jenkins, G. I., et al. (2012). Alternative splicing mediates responses of the Arabidopsis circadian clock to temperature changes. *Plant Cell* 24 (3), 961–981. doi: 10.1105/tpc.111.093948
- Kai, M. (2016). Roles of RNA-binding proteins in DNA damage response. *Int. J. Mol. Sci.* 17 (3), 310. doi: 10.3390/ijms17030310
- Kalyna, M., Simpson, C. G., Syed, N. H., Lewandowska, D., Marquez, Y., Kusenda, B., et al. (2012). Alternative splicing and nonsense-mediated decay modulate expression of important regulatory genes in Arabidopsis. *Nucleic Acids Res.* 40 (6), 2454–2469. doi: 10.1093/nar/gkr932
- Kanno, T., Venhuizen, P., Wen, T. N., Lin, W. D., Chiou, P., Kalyna, M., et al. (2018). PRP4KA, a Putative Spliceosomal Protein Kinase, is important for alternative splicing and development in Arabidopsis thaliana. *Genetics* 210 (4), 1267–1285. doi: 10.1534/genetics.118.301515
- Kavakli, I. H., Baris, I., Tardu, M., Gul, S., Oner, H., Cal, S., et al. (2017). The photolyase/cryptochrome family of proteins as DNA repair enzymes and transcriptional repressors. *Photochem. Photobiol.* 93 (1), 93–103. doi: 10.1111/php.12669
- Kim, J. H., Ryu, T. H., Lee, S. S., Lee, S., and Chung, B. Y. (2019). Ionizing radiation manifesting DNA damage response in plants: an overview of DNA damage signaling and repair mechanisms in plants. *Plant Sci.* 278, 44–53. doi: 10.1016/j.plantsci.2018.10.013
- Lackner, D. H., Durocher, D., and Karlseder, J. (2011). A siRNA-based screen for genes involved in chromosome end protection. *PLoS One* 6 (6), e21407. doi: 10.1371/journal.pone.0021407
- Leva, V., Giuliano, S., Bardoni, A., Camerini, S., Crescenzi, M., Lisa, A., et al. (2012). Phosphorylation of SRSF1 is modulated by replicational stress. *Nucleic Acids Res.* 40 (3), 1106–1117. doi: 10.1093/nar/gkr837
- Leviatan, N., Alkan, N., Leshkowitz, D., and Fluhr, R. (2013). Genome-wide survey of cold stress regulated alternative splicing in Arabidopsis thaliana with tiling microarray. *PLoS One* 8 (6), e66511. doi: 10.1371/journal.pone.0066511
- Li, X., and Manley, J. L. (2005). Inactivation of the SR protein splicing factor ASF/SF2 results in genomic instability. *Cell* 122 (3), 365–378. doi: 10.1016/j.cell.2005.06.008
- Li, S., Paulsson, M., and Bjorn, L. O. (2002). Temperature-dependent formation and photorepair of DNA damage induced by UV-B radiation in suspension-cultured tobacco cells. *J. Photochem. Photobiol. B.* 66 (1), 67–72. doi: 10.1016/S1011-1344(01)00277-9
- Li, Z., Pearlman, A. H., and Hsieh, P. (2016). DNA mismatch repair and the DNA damage response. *DNA Repair (Amst)* 38, 94–101. doi: 10.1016/j.dnarep.2015.11.019
- Li, R., Wang, W., Li, F., Wang, Q., Wang, S., Xu, Y., et al. (2017). Response of alternative splice isoforms of OsRad9 gene from *Oryza sativa* to environmental stress. *Z. Naturforsch. C.* 72 (7–8), 325–334. doi: 10.1515/znc-2016-0257
- Liu, D., Keijzers, G., and Rasmussen, L. J. (2017). DNA mismatch repair and its many roles in eukaryotic cells. *Mutat. Res.* 773, 174–187. doi: 10.1016/j.mrrev.2017.07.001
- Llorian, M., Beullens, M., Lesage, B., Nicolaescu, E., Beke, L., Landuyt, W., et al. (2005). Nucleocytoplasmic shuttling of the splicing factor SIPP1. *J. Biol. Chem.* 280 (46), 38862–38869. doi: 10.1074/jbc.M509185200
- Lorkovic, Z. J., Hilscher, J., and Barta, A. (2008). Co-localisation studies of Arabidopsis SR splicing factors reveal different types of speckles in plant cell nuclei. *Exp. Cell Res.* 314 (17), 3175–3186. doi: 10.1016/j.yexcr.2008.06.020
- Lu, X., and Legerski, R. J. (2007). The Prp19/Pso4 core complex undergoes ubiquitylation and structural alterations in response to DNA damage. *Biochem. Biophys. Res. Commun.* 354 (4), 968–974. doi: 10.1016/j.bbrc.2007.01.097
- Magni, M., Buscemi, G., Maita, L., Peng, L., Chan, S. Y., Montecucco, A., et al. (2019). TSPYL2 is a novel regulator of SIRT1 and p300 activity in response to DNA damage. *Cell Death Differ.* 26 (5), 918–931. doi: 10.1038/s41418-018-0168-6
- Manley, J. L., and Krainer, A. R. (2010). A rational nomenclature for serine/arginine-rich protein splicing factors (SR proteins). *Genes Dev.* 24 (11), 1073–1074. doi: 10.1101/gad.1934910
- Manova, V., and Gruszka, D. (2015). DNA damage and repair in plants - from models to crops. *Front. Plant Sci.* 6, 885. doi: 10.3389/fpls.2015.00885
- 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 (6), 1184–1195. doi: 10.1101/gr.134106.111
- Marquez, Y., Hopfler, M., Ayatollahi, Z., Barta, A., and Kalyna, M. (2015). Unmasking alternative splicing inside protein-coding exons defines exons and their role in proteome plasticity. *Genome Res.* 25 (7), 995–1007. doi: 10.1101/gr.186585.114
- Matsuoka, S., Ballif, B. A., Smogorzewska, A., McDonald, J. E. R. 3rd, Hurov, K. E., Luo, J., et al. (2007). ATM and ATR substrate analysis reveals extensive protein networks responsive to DNA damage. *Science* 316 (5828), 1160–1166. doi: 10.1126/science.1140321
- Mikolaskova, B., Jurcik, M., Cipakova, I., Kretova, M., Chovanec, M., and Cipak, L. (2018). Maintenance of genome stability: the unifying role of interconnections between the DNA damage response and RNA-processing pathways. *Curr. Genet.* 64 (5), 971–983. doi: 10.1007/s00294-018-0819-7
- Mimida, N., Kitamoto, H., Osakabe, K., Nakashima, M., Ito, Y., Heyer, W. D., et al. (2007). Two alternatively spliced transcripts generated from OsMUS81, a rice homolog of yeast MUS81, are up-regulated by DNA-damaging treatments. *Plant Cell Physiol.* 48 (4), 648–654. doi: 10.1093/pcp/pcm029
- Munoz, M. J., Perez Santangelo, M. S., Paronetto, M. P., de la Mata, M., Pelisch, F., Boireau, S., et al. (2009). DNA damage regulates alternative splicing through inhibition of RNA polymerase II elongation. *Cell* 137 (4), 708–720. doi: 10.1016/j.cell.2009.03.010
- Munoz, M. J., Nieto Moreno, N., Giono, L. E., Cambindo Botto, A. E., Dujardin, G., Bastianello, G., et al. (2017). Major roles for pyrimidine dimers, nucleotide excision repair, and ATR in the alternative splicing response to UV irradiation. *Cell Rep.* 18 (12), 2868–2879. doi: 10.1016/j.celrep.2017.02.066
- Murphy, T. M., and Gao, M. J. (2001). Multiple forms of formamidopyrimidine-DNA glycosylase produced by alternative splicing in Arabidopsis thaliana. *J. Photochem. Photobiol. B.* 61 (3), 87–93. doi: 10.1016/S1011-1344(01)00172-5
- Nakka, K. K., Chaudhary, N., Joshi, S., Bhat, J., Singh, K., Chatterjee, S., et al. (2015). Nuclear matrix-associated protein SMAR1 regulates alternative splicing via HDAC6-mediated deacetylation of Sam68. *Proc. Natl. Acad. Sci. U. S. A.* 112 (26), E3374–E3383. doi: 10.1073/pnas.1418603112
- Naro, C., Bielli, P., Pagliarini, V., and Sette, C. (2015). The interplay between DNA damage response and RNA processing: the unexpected role of splicing factors as gatekeepers of genome stability. *Front. Genet.* 6, 142. doi: 10.3389/fgene.2015.00142
- Nisa, M. U., Huang, Y., Benhamed, M., and Raynaud, C. (2019). The plant DNA damage response: signaling pathways leading to growth inhibition and putative role in response to stress conditions. *Front. Plant Sci.* 10, 653. doi: 10.3389/fpls.2019.00653
- Nota, F., Cambiagno, D. A., Ribone, P., and Alvarez, M. E. (2015). Expression and function of AtMBD4L, the single gene encoding the nuclear DNA glycosylase MBD4L in Arabidopsis. *Plant Sci.* 235, 122–129. doi: 10.1016/j.plantsci.2015.03.011
- Ogita, N., Okushima, Y., Tokizawa, M., Yamamoto, Y. Y., Tanaka, M., Seki, M., et al. (2018). Identifying the target genes of SUPPRESSOR OF GAMMA RESPONSE 1, a master transcription factor controlling DNA damage response in Arabidopsis. *Plant J.* 94 (3), 439–453. doi: 10.1111/tpj.13866
- Ohtsubo, T., Matsuda, O., Iba, K., Terashima, I., Sekiguchi, M., and Nakabeppu, Y. (1998). Molecular cloning of AtMMH, an Arabidopsis thaliana ortholog of the Escherichia coli mutM gene, and analysis of functional domains of its product. *Mol. Gen. Genet.* 259 (6), 577–590. doi: 10.1007/s004380050851
- Osakabe, K., Yoshioka, T., Ichikawa, H., and Toki, S. (2002). Molecular cloning and characterization of RAD51-like genes from Arabidopsis thaliana. *Plant Mol. Biol.* 50 (1), 71–81. doi: 10.1023/A:1016047231597
- 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 (12), 1413–1415. doi: 10.1038/ng.259
- Paulsen, R. D., Soni, D. V., Wollman, R., Hahn, A. T., Yee, M. C., Guan, A., et al. (2009). A genome-wide siRNA screen reveals diverse cellular processes and pathways that mediate genome stability. *Mol. Cell* 35 (2), 228–239. doi: 10.1016/j.molcel.2009.06.021
- Preuss, S. B., and Britt, A. B. (2003). A DNA-damage-induced cell cycle checkpoint in Arabidopsis. *Genetics* 164 (1), 323–334.
- Rapazote-Flores, P., Bayer, M., Milne, L., Mayer, C.-D., Fuller, J., Guo, W., et al. (2019). BaRTv1.0: an improved barley reference transcript dataset to determine accurate changes in the barley transcriptome using RNA-seq. *BMC Genomics.* 20 (1), 968. doi: 10.1186/s12864-019-6243-7
- Sakashita, E., and Endo, H. (2010). SR and SR-related proteins redistribute to segregated fibrillar components of nucleoli in a response to DNA damage. *Nucleus* 1 (4), 367–380. doi: 10.4161/nucl.1.4.12683

- Santiago, M. J., Alejandro-Durán, E., and Ruiz-Rubio, M. (2009). Alternative splicing of two translesion synthesis DNA polymerases from *Arabidopsis thaliana*. *Plant Sci.* 176 (4), 591–596. doi: 10.1016/j.plantsci.2009.01.018
- Savaldi-Goldstein, S., Sessa, G., and Fluhr, R. (2000). The ethylene-inducible PK12 kinase mediates the phosphorylation of SR splicing factors. *Plant J.* 21 (1), 91–96. doi: 10.1046/j.1365-3113x.2000.00657.x
- Shepard, P. J., and Hertel, K. J. (2008). Conserved RNA secondary structures promote alternative splicing. *RNA* 14 (8), 1463–1469. doi: 10.1261/rna.1069408
- Shkreta, L., and Chabot, B. (2015). The RNA splicing response to DNA damage. *Biomolecules* 5 (4), 2935–2977. doi: 10.3390/biom5042935
- Shkreta, L., Toutant, J., Durand, M., Manley, J. L., and Chabot, B. (2016). SRSF10 connects DNA damage to the alternative splicing of transcripts encoding apoptosis, cell-cycle control, and DNA repair factors. *Cell Rep.* 17 (8), 1990–2003. doi: 10.1016/j.celrep.2016.10.071
- Siam, A., Baker, M., Amit, L., Regev, G., Rabner, A., Najar, R. A., et al. (2019). Regulation of alternative splicing by p300-mediated acetylation of splicing factors. *RNA* 25 (7), 813–824. doi: 10.1261/rna.069856.118
- Sibley, C. R., Blazquez, L., and Ule, J. (2016). Lessons from non-canonical splicing. *Nat. Rev. Genet.* 17 (7), 407–421. doi: 10.1038/nrg.2016.46
- Solier, S., Barb, J., Zeeberg, B. R., Varma, S., Ryan, M. C., Kohn, K. W., et al. (2010). Genome-wide analysis of novel splice variants induced by topoisomerase I poisoning shows preferential occurrence in genes encoding splicing factors. *Cancer Res.* 70 (20), 8055–8065. doi: 10.1158/0008-5472.CAN-10-2491
- Spampinato, C. P. (2017). Protecting DNA from errors and damage: an overview of DNA repair mechanisms in plants compared to mammals. *Cell Mol. Life Sci.* 74 (9), 1693–1709. doi: 10.1007/s00018-016-2436-2
- Staiger, D., and Simpson, G. G. (2015). Enter exons. *Genome Biol.* 16 (1), 136. doi: 10.1186/s13059-015-0704-3
- Thatcher, S. R., Zhou, W., Leonard, A., Wang, B. B., Beatty, M., Zastrow-Hayes, G., et al. (2014). Genome-wide analysis of alternative splicing in *Zea mays*: landscape and genetic regulation. *Plant Cell* 26 (9), 3472–3487. doi: 10.1105/tpc.114.130773
- Tillemans, V., Dispa, L., Remacle, C., Collinge, M., and Motte, P. (2005). Functional distribution and dynamics of *Arabidopsis* SR splicing factors in living plant cells. *Plant J.* 41 (4), 567–582. doi: 10.1111/j.1365-3113X.2004.02321.x
- Tissier, A., Janel-Bintz, R., Coulon, S., Klaile, E., Kannouche, P., Fuchs, R. P., et al. (2010). Crosstalk between replicative and translesional DNA polymerases: PDIP38 interacts directly with Poleta. *DNA Repair (Amst)* 9 (8), 922–928. doi: 10.1016/j.dnarep.2010.04.010
- Ustianenko, D., Weyn-Vanhenhenryck, S. M., and Zhang, C. (2017). Microexons: discovery, regulation, and function. *Wiley Interdiscip. Rev. RNA* 8 (4), e1418. doi: 10.1002/wrna.1418
- van der Houven van Oordt, W., Diaz-Meco, M. T., Lozano, J., Krainer, A. R., Moscat, J., and Caceres, J. F. (2000). The MKK(3/6)-p38-signaling cascade alters the subcellular distribution of hnRNP A1 and modulates alternative splicing regulation. *J. Cell Biol.* 149 (2), 307–316. doi: 10.1083/jcb.149.2.307
- Vassileva, M. T., and Matunis, M. J. (2004). SUMO modification of heterogeneous nuclear ribonucleoproteins. *Mol. Cell Biol.* 24 (9), 3623–3632. doi: 10.1128/MCB.24.9.3623-3632.2004
- Vonarx, E. J., Howlett, N. G., Schiestl, R. H., and Kunz, B. A. (2002). Detection of *Arabidopsis thaliana* ATRAD1 cDNA variants and assessment of function by expression in a yeast rad1 mutant. *Gene* 296 (1–2), 1–9. doi: 10.1016/S0378-1119(02)00869-7
- Wallace, S. S. (2014). Base excision repair: a critical player in many games. *DNA Repair (Amst)* 19, 14–26. doi: 10.1016/j.dnarep.2014.03.030
- Wang, Z., and Burge, C. B. (2008). Splicing regulation: from a parts list of regulatory elements to an integrated splicing code. *RNA* 14 (5), 802–813. doi: 10.1261/rna.876308
- Wang, E. T., Sandberg, R., Luo, S., Khrebukova, I., Zhang, L., Mayr, C., et al. (2008). Alternative isoform regulation in human tissue transcriptomes. *Nature* 456 (7221), 470–476. doi: 10.1038/nature07509
- Wong, A., Zhang, S., Mordue, D., Wu, J. M., Zhang, Z., Darzynkiewicz, Z., et al. (2013). PDIP38 is translocated to the spliceosomes/nuclear speckles in response to UV-induced DNA damage and is required for UV-induced alternative splicing of MDM2. *Cell Cycle* 12 (19), 3184–3193. doi: 10.4161/cc.26221
- Yap, K., Lim, Z. Q., Khandelia, P., Friedman, B., and Makeyev, E. V. (2012). Coordinated regulation of neuronal mRNA steady-state levels through developmentally controlled intron retention. *Genes Dev.* 26 (11), 1209–1223. doi: 10.1101/gad.188037.112
- Yeap, W. C., Namasivayam, P., and Ho, C. L. (2014). HnRNP-like proteins as post-transcriptional regulators. *Plant Sci.* 227, 90–100. doi: 10.1016/j.plantsci.2014.07.005
- Yoshiyama, K., Conklin, P. A., Huefner, N. D., and Britt, A. B. (2009). Suppressor of gamma response 1 (SOG1) encodes a putative transcription factor governing multiple responses to DNA damage. *Proc. Natl. Acad. Sci. U. S. A.* 106 (31), 12843–12848. doi: 10.1073/pnas.0810304106
- Yoshiyama, K. O. (2016). SOG1: a master regulator of the DNA damage response in plants. *Genes Genet. Syst.* 90 (4), 209–216. doi: 10.1266/ggs.15-00011
- Zhang, M., Wang, L., and Zhong, D. (2017a). Photolyase: dynamics and mechanisms of repair of sun-induced DNA damage. *Photochem. Photobiol.* 93 (1), 78–92. doi: 10.1111/php.12695
- Zhang, R., Calixto, C. P. G., Marquez, Y., Venhuizen, P., Tzioutziou, N. A., Guo, W., et al. (2017b). A high quality *Arabidopsis* transcriptome for accurate transcript-level analysis of alternative splicing. *Nucleic Acids Res.* 45 (9), 5061–5073. doi: 10.1093/nar/gkx267

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2020 Nimeth, Riegler and Kalyna. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



Current Challenges in Studying Alternative Splicing in Plants: The Case of *Physcomitrella patens* SR Proteins

José Pedro Melo¹, Maria Kalyna^{2*†} and Paula Duque^{1*†}

¹ Instituto Gulbenkian de Ciência (IGC), Oeiras, Portugal, ² Department of Applied Genetics and Cell Biology, BOKU – University of Natural Resources and Life Sciences, Vienna, Austria

OPEN ACCESS

Edited by:

Michael J. Haydon,
The University of Melbourne, Australia

Reviewed by:

Stefan A. Rensing,
University of Marburg, Germany
Igor Fesenko,
Institute of Bioorganic Chemistry
(RAS), Russia

*Correspondence:

Maria Kalyna
mariya.kalyna@boku.ac.at
Paula Duque
duquep@igc.gulbenkian.pt

†ORCID:

Maria Kalyna
orcid.org/0000-0003-4702-7625
Paula Duque
orcid.org/0000-0002-4910-2900

Specialty section:

This article was submitted to
Plant Physiology,
a section of the journal
Frontiers in Plant Science

Received: 18 November 2019

Accepted: 26 February 2020

Published: 24 March 2020

Citation:

Melo JP, Kalyna M and Duque P
(2020) Current Challenges in Studying
Alternative Splicing in Plants:
The Case of *Physcomitrella patens*
SR Proteins. *Front. Plant Sci.* 11:286.
doi: 10.3389/fpls.2020.00286

To colonize different terrestrial habitats, early land plants had to overcome the challenge of coping with harsh new environments. Alternative splicing – an RNA processing mechanism through which splice sites are differentially recognized, originating multiple transcripts and potentially different proteins from the same gene – can be key for plant stress tolerance. Serine/arginine-rich (SR) proteins constitute an evolutionarily conserved family of major alternative splicing regulators that in plants subdivides into six subfamilies. Despite being well studied in animals and a few plant species, such as the model angiosperm *Arabidopsis thaliana* and the crop *Oryza sativa*, little is known of these splicing factors in early land plants. Establishing the whole complement of SR proteins in different species is essential to understand the functional and evolutionary significance of alternative splicing. An *in silico* search for SR proteins in the extant moss *Physcomitrella patens* revealed inconsistencies both in the published data and available databases, likely arising from automatic annotation lacking adequate manual curation. These misannotations interfere with the description not only of the number and subfamily classification of *Physcomitrella* SR proteins but also of their domain architecture, potentially hindering the elucidation of their molecular functions. We therefore advise caution when looking into *P. patens* genomic resources. Our systematic survey nonetheless confidently identified 16 *P. patens* SR proteins that fall into the six described subfamilies and represent counterparts of well-established members in *Arabidopsis* and rice. Intensified research efforts should disclose whether SR proteins were already determining alternative splicing modulation and stress tolerance in early land plants.

Keywords: *Physcomitrella patens*, *Arabidopsis thaliana*, RNA splicing, alternative splicing, SR proteins, gene annotation, stress, evolution

THE ADAPTATION OF PLANTS TO LIFE ON LAND

About 500 million years ago the first plants colonized land, resulting in one of the most important events in the history of life on Earth (Morris et al., 2018). The evolution and diversification of land plants shaped the biosphere and created the conditions that allowed the colonization of land by metazoans and the subsequent formation of complex terrestrial habitats (Floyd and Bowman, 2007).

However, terrestrial and aquatic environments differ greatly, mainly regarding water, temperature and radiation exposure. Thus, plants had to develop a number of regulatory cellular and physiological traits to cope with these adverse conditions (Kenrick and Crane, 1997). Some of these features are already present in charophycean algae, such as a three-dimensional, predominant haploid gametophyte, but others evolved during or after colonization. Furthermore, some appeared *de novo* while others evolved from existing traits. Some appeared only once, others originated multiple times, whereas others were gained and then lost in some lineages. Key innovations include the internalization of vital functions and organs and the development of impermeable exterior surfaces, leading to the appearance of specialized sexual organs (gametangia), vascular tissues, stomata, symbiosis with fungi, branched shoots, leaves, roots, seeds and flowers (Kenrick and Crane, 1997; Heckman et al., 2001; Floyd and Bowman, 2007; Rensing et al., 2008; Morris et al., 2018; One Thousand Plant Transcriptomes Initiative, 2019). At the cellular level, plants evolved better osmoregulation, desiccation, freezing and heat resistance as well as enhanced DNA repair mechanisms (Gensel, 2008; Ju et al., 2015; de Vries and Archibald, 2018).

Most of these traits evolved prior to the appearance of flowering plants, highlighting the importance of establishing and studying model species that represent all major plant groups. The extant moss *Physcomitrella patens* occupies a key position in the evolution of plants, between aquatic green algae and vascular plants, and presents characteristics of both, rendering it an invaluable tool to study the onset of terrestrial-based plant life, as well as the appearance and diversification of important traits for modern day crops (Rensing et al., 2008; Smidkova et al., 2010).

ALTERNATIVE SPLICING AS A MEANS OF ADAPTING TO A CHANGING ENVIRONMENT

A characteristic of eukaryotic protein-coding genes is the presence of segments of non-coding DNA, called introns, interspaced with coding DNA segments, the exons. When the precursor mRNA (pre-mRNA) is processed, specific splice sites are recognized, the introns are removed and the exons joined together (or spliced). However, splice sites can be differentially recognized, leading to the inclusion or removal of different segments of RNA, resulting in multiple transcripts, and potentially proteins, originating from the same gene. This mechanism is termed alternative splicing and represents an effective means of both increasing transcriptome and proteome diversity and regulating gene expression by affecting the stability of the transcripts (Shang et al., 2017). Alternative splicing can often generate transcripts with premature termination codons, which are targeted to degradation by nonsense-mediated mRNA decay (NMD). Coupling of NMD with alternative splicing can represent an important means of regulating the abundance of certain proteins, such as splicing regulators, in a homeostatic feedback loop (Nasif et al., 2018).

A large proportion of genes is known to be alternatively spliced, with recent studies indicating that up to 70% of plant

multi-exon genes undergo alternative splicing (Marquez et al., 2012; Chamala et al., 2015; Zhang et al., 2017). This number has been increasing over time as genome annotations improve and the use of next generation sequencing provides not only more but also deeper data (Syed et al., 2012). Estimates of alternative splicing rates in *P. patens* show a similar percentage (58%) of genes being alternatively spliced (Zimmer et al., 2013; Lang et al., 2016), although the actual number may be higher.

Alternative splicing is involved in numerous biological processes. In plants, it is especially important in the response to external cues, particularly environmental stresses (Staiger and Brown, 2013), and is hence likely to have played an important role in the process of land colonization (Mastrangelo et al., 2012). However, this hypothesis remains to be adequately tested. In *P. patens*, alternative splicing has been shown to be regulated by light (Wu et al., 2014) and to improve tolerance to heat stress (Chang et al., 2014). These findings are consistent with this posttranscriptional mechanism having helped early land plants cope with adverse terrestrial conditions. On the other hand, though further studies are needed to support this notion, it has also been reported that alternative splicing does not significantly affect proteome diversity in *P. patens* (Fesenko et al., 2017), suggesting that this mechanism could increase stress tolerance through the modulation of transcript stability and thereby protein abundance. It has also been shown that about 32% of alternatively spliced genes in *P. patens* are targeted to NMD, pointing to an important role of this mRNA degradation mechanism in gene regulation (Lloyd et al., 2018). The scarcity of available data and the existence of conflicting evidence underscore the need for thorough bioinformatics and evolutionary studies to address the role of alternative splicing in plant conquest of land.

SR PROTEINS: A HIGHLY CONSERVED FAMILY OF KEY ALTERNATIVE SPLICING REGULATORS

Serine/arginine-rich (SR) proteins represent an important family of RNA-binding proteins that is highly conserved among eukaryotes. Their most well-known role is in the regulation of pre-mRNA splicing, being involved in the recognition of splice sites and recruitment and assembly of the spliceosome. Most importantly, these proteins influence the recognition of splice sites by core spliceosomal components and are thus major modulators of alternative splicing. Nonetheless, these are not the only known functions of SR proteins. Several studies in both animal and plant systems have also unveiled a myriad of other functions, such as genome maintenance, mRNA stability and export, and oncogenic transformation (Huang and Steitz, 2005; Tillemans et al., 2006; Long and Cáceres, 2009; Zhong et al., 2009; Twyffels et al., 2011).

Their multifunctional roles and the historical timing of their discovery has led to several distinct SR protein classifications, an issue that was addressed almost a decade ago by Manley and Krainer (2010), who proposed a standardization of the definition and nomenclature of SR proteins. This was mostly based on

mammalian proteins, and adapting the system to plants proved to be an arduous task due to the higher number and diversity of family members. In addition, some plant-specific SR proteins do not present orthologs in mammals and exhibit unique features that did not fall under Manley and Krainer's definition. For this reason, in the same year, Barta et al. (2010) proposed an updated nomenclature for plant SR proteins to facilitate the assignment and comparison of these proteins across plant species.

According to the established definition for plants, SR proteins are characterized by the presence of one or two N-terminal RNA Recognition Motifs (RRMs) and a C-terminal arginine/serine-rich (RS) region of at least 50 amino acids with a minimum of 20% RS or SR dipeptides (Barta et al., 2010 and **Figure 1**). Based on this definition, the genomes of the dicotyledonous model plant species *Arabidopsis thaliana* and the monocot crop *Oryza sativa* (rice) encode 18 and 22 SR proteins, respectively, which are classified into six subfamilies according to their protein domain architecture. Three of these subfamilies have orthologs in mammals - the SR subfamily is orthologous to the mammalian SRSF1/SF2/ASF, the RSZ subfamily to SRSF7/9G8 and the SC subfamily to SRSF2/SC35. The remaining three subfamilies, however, are plant-specific, presenting no clear orthologs in animals. There are also a few proteins that contain an RRM and an RS region, but are nevertheless no longer considered SR proteins due to the presence of an additional N-terminal RS region and are therefore named SR-like.

One peculiarity of plant SR genes is that many occur as duplicated pairs of paralogs, thus explaining why there are almost twice as many as in humans (Kalyna and Barta, 2004). Genome duplications, either whole-genome or large segmental duplications, are very common throughout the evolution of several plant species and lineages. However, most of what is known has been studied in angiosperms, such as *Arabidopsis* and rice, with only occasional reports providing insight into earlier plants (Adams and Wendel, 2005; Soltis et al., 2015; Panchy et al., 2016; Clark and Donoghue, 2018; Wu et al., 2020). Interestingly, extant mosses such as *P. patens*, but not hornworts and liverworts, also underwent genome duplications. More specifically, *P. patens* appears to have undergone two whole-genome duplication events, one 40–48 million years ago (Mya) and the other 27–35 Mya, giving rise to the current 27 chromosomes and ~33,000 genes, a similar number to *A. thaliana* (Rensing et al., 2007, 2013; Lang et al., 2018). Another remarkable feature of plant SR genes is that they undergo highly conserved alternative splicing events in their longest introns, which in some cases has been maintained from *P. patens* or the single cell green alga *Chlamydomonas reinhardtii* to dicots throughout ~1.1 billion years of evolution (Iida and Go, 2006; Kalyna et al., 2006). Interestingly, alternative splice sites in SR genes of the RS and RS2Z subfamilies are embedded in ultraconserved regions preserved for >400 million years of land plant evolution (Kalyna et al., 2006). A similar feature has also been observed for mouse and human SR genes (Lareau et al., 2007), pointing to the importance of this mode of regulation. Establishing the whole complement of SR proteins in *P. patens* would prove invaluable to understand both the general regulation of these RNA-binding proteins and their impact on alternative splicing in an early land plant.

HOW MANY SR PROTEINS DOES THE *PHYSCOMITRELLA PATENS* GENOME ENCODE?

A few individual *P. patens* SR proteins have been previously reported and studied by different research groups (Iida and Go, 2006; Kalyna et al., 2006; Richardson et al., 2011; Califice et al., 2012; Rauch et al., 2014; Fesenko et al., 2017; Lloyd et al., 2018), but the gene family has hitherto not been addressed as a whole. Furthermore, not all studies report the same complement, subfamily classification or coding sequences, underscoring the need for revisiting these analyses. We conducted a comprehensive *in silico* search for SR genes in *P. patens* to determine their number and whether they fall under the current definition of this family as well as to establish reference protein sequences and domain architectures for each *Physcomitrella* SR protein.

A BLASTp search against the *P. patens* proteome using each *A. thaliana* SR protein sequence in the NCBI¹, CoGe² and Plaza³ databases yielded a list of 18 candidate *P. patens* SR proteins. However, after analyzing the sequences and metadata, including intron/exon positions, genome location, syntenicity and expression data, in JBrowse, manual curation of the retrieved sequences resulted in the identification of only 16 *bona fide* *P. patens* SR proteins (**Figure 1**).

INCONSISTENCIES WITHIN THE AVAILABLE DATABASES

Our initial BLAST search identified three *P. patens* proteins belonging to the RSZ subfamily. However, despite being annotated as two different genes, Pp3c11_26740V3.1 and Pp3c11_26750V3.1 occupy the same position in the genome (**Supplementary Figure S1**). We considered Pp3c11_26750V3.1 as the correct gene, as it is the longest annotation of the two and is unequivocally supported by EST, cDNA and RNA-seq data.

Another identified inconsistency related to Pp3c1_31300V3.1 from the SC subfamily. This gene includes a very short open reading frame (ORF) that does not support a full-length protein and is preceded by an unusually long 5' UTR (**Supplementary Figure S2**). It is positioned next to another member of the SC subfamily, Pp3c1_31280V3.1, with which it shares a similar sequence.

Noteworthy, some of the retrieved genes do not present annotated untranslated regions (UTRs), and the annotated exon and intron positions are not supported by the EST, cDNA, and RNA-seq data available in NCBI, CoGe or Phytozome⁴. One such example is the Pp3c16_1000V3.1 gene model, which does not include an annotated 5'UTR, and whose annotated first exon is not supported by experimental evidence. Moreover, although the second exon is supported by EST,

¹<https://www.ncbi.nlm.nih.gov>

²<https://genomevolution.org/coge/>

³<https://bioinformatics.psb.ugent.be/plaza>

⁴<https://phytozome.jgi.doe.gov/pz/portal.html>

Subfamily		<i>Physcomitrella patens</i>		<i>Arabidopsis thaliana</i>	
SR		Pp-SR36	Pp3c17_17620	At-SR30	At1g09140
		Pp-SR39	Pp3c2_2470	At-SR34	At1g02840
		Pp-SR40	Pp3c1_35700	At-SR34a	At3g49430
				At-SR34b	At4g02430
RSZ		Pp-RSZ23	Pp3c7_5100	At-RSZ21	At1g23860
		Pp-RSZ25	Pp3c11_26750*	At-RSZ22	At4g31580
				At-RSZ22a	At2g24590
SC		Pp-SC37	Pp3c1_31280		
		Pp-SC39	Pp3c11_21430	At-SC35	At5g64200
		Pp-SC41	Pp3c7_7590		
SCL		Pp-SCL29	Pp3c5_15600	At-SCL28	At5g18810
		Pp-SCL30	Pp3c5_15610	At-SCL30	At3g55460
		Pp-SCL33	Pp3c6_9110	At-SCL30a	At3g13570
		Pp-SCL42	Pp3s152_10	At-SCL33	At1g55310
RS2Z		Pp-RS2Z27	Pp3c20_7750*	At-RS2Z32	At3g53500
		Pp-RS2Z37	Pp3c14_23400*	At-RS2Z33	At2g37340
		Pp-RS2Z38	Pp3c17_20550		
RS				At-RS31	At3g61860
				At-RS31a	At2g46610
		Pp-RS27	Pp3c16_1000*	At-RS40	At4g25500
				At-RS41	At5g52040

FIGURE 1 | Schematic representation of the *Arabidopsis thaliana* and *Physcomitrella patens* SR protein families. All SR proteins are characterized by at least one RNA Recognition Motif (RRM) and an arginine/serine-rich (RS) region. Proteins in the RSZ subfamily include a glycine-rich region (G) after the first RRM. RSZ and RS2Z proteins also contain one and two zinc knuckles (Zn), respectively, with RS2Z members harboring an SP-rich extension after the RS region. The SCL subfamily is characterized by an N-terminal charged extension. The light-green shaded box indicates the three plant-specific SR protein subfamilies. Genes with incorrect reference models in the publicly available databases are marked with an asterisk (*).

cDNA and RNA-seq data, its sequence harbors no AUG. The first AUG is found only in the third exon. As such, the correct protein for this member of the RS subfamily appears to start in the third annotated exon (**Supplementary Figure S3**), with the second exon likely being included in the 5'UTR. This is likely due to an automatic annotation of the genome lacking proper manual curation. Another such example is Pp3c20_7750V3.1 from the RS2Z subfamily, whose reference gene model also lacks annotated UTRs, despite the fact that both alternative transcripts have annotated UTRs (**Supplementary Figure S4**).

An additional misannotation occurs in Pp3c14_23400V3.1, another RS2Z subfamily member, for which the first exon is not fully supported by EST, cDNA and RNA-seq data (**Supplementary Figure S5**). The real transcript appears to begin only after the annotated methionine, which would lead to translation starting in the second exon, giving rise to a full-length protein with all the canonical domains. Furthermore, this protein would present a molecular weight of 37 kD and a similar sequence to its paralog Pp3c17_20550V3.1.

After manually checking and curating the sequences, the remaining 16 SR proteins were aligned and compared with *A. thaliana* SR proteins. Importantly, all were unambiguously assigned to one of the six previously described subfamilies, despite the varying number of members within each subfamily.

A protein domain search was then performed using MyHits⁵ to confirm that all *P. patens* SR proteins presented the characteristics of their assigned subfamily.

IS THE PHYSCOMITRELLA PATENS SR PROTEIN FAMILY STRUCTURALLY SIMILAR TO THAT OF VASCULAR PLANTS?

A comparison between the *A. thaliana* and *P. patens* SR protein families is shown in **Figure 1**, while **Figure 2** and **Supplementary Table S1** summarize the number of SR proteins within each subfamily among six representative species from different phylogenetic groups. The SR subfamily, comprised of four members in both *Arabidopsis* and rice, includes three proteins in *P. patens*, one in *Marchantia polymorpha* and two in both *C. reinhardtii* and *Chara braunii*. SR subfamily members all harbor two RRM, sharing a conserved SWQDLKD motif in the second RRM as well as the characteristic RS region. All *P. patens* proteins from this subfamily contain a glycine-rich region separating the two RRM, a trait shared by three SR subfamily *Arabidopsis* proteins but not present in At-SR30. The RSZ subfamily comprises three members in *A. thaliana* and

⁵https://myhits.isb-sib.ch/cgi-bin/motif_scan

	Subfamily						Total #	Reference/Database
	SR	RSZ	SC	SCL	RS2Z	RS		
<i>Chlamydomonas reinhardtii</i>	2	2	0	3	0	2	9	Bowman <i>et al.</i> 2017, Phytozome, Plaza, CoGe
<i>Chara braunii</i> *	1	2	1	2	2	1	9	Nishiyama <i>et al.</i> 2018, OrcAE
<i>Marchantia polymorpha</i>	1	1	1	1	1	1	6	Bowman <i>et al.</i> 2017, Phytozome, Plaza, CoGe
<i>Physcomitrella patens</i>	3	2	3	4	3	1	16	This study
<i>Arabidopsis thaliana</i>	4	3	1	4	2	4	18	Barta <i>et al.</i> 2010
<i>Oryza sativa</i>	4	3	3	6	4	2	22	Barta <i>et al.</i> 2010

FIGURE 2 | SR protein numbers in six representative plant species. The column on the right indicates the sources of the data. Gray shading indicates SR proteins requiring further annotation. For *Chara braunii*, marked with an asterisk, refer to **Supplementary Table S1**. Values in bold indicate *Physcomitrella patens* SR proteins manually verified and curated in this study. The green shaded subfamilies are plant-specific. Lines on the left are a schematic representation of the phylogenetic relationships among the species.

O. sativa, only two in *P. patens*, *C. reinhardtii* and *C. braunii*, and one in *M. polymorpha*. It is characterized by the presence of a zinc knuckle between the RRM domain and the RS region, which is preceded by a glycine-rich region. Strikingly, the SC subfamily, while comprising only one member in *A. thaliana*, includes three in rice and four in *P. patens*, with all of these proteins containing one RRM followed by the RS region. *C. braunii* and *M. polymorpha* also have only one SC subfamily member, while the *C. reinhardtii* genome does not appear to code for any. The plant-specific SCL subfamily is present in all species analyzed, with *P. patens* and *Arabidopsis* possessing four members, rice six, and *C. reinhardtii*, *C. braunii* and *M. polymorpha* containing three, two and one, respectively. This subfamily is characterized by an N-terminal charged extension, in addition to the canonical RRM and RS region. The RS2Z subfamily is also plant-specific and is characterized by two zinc knuckles between the RRM and the RS region, as well as a C-terminal SP-rich region. Contrarily to the RSZ subfamily, there is no glycine-rich region after the RRM, with *Arabidopsis* and *C. braunii* containing two RS2Z members, rice four, *M. polymorpha* one, *P. patens* three and *C. reinhardtii* none. Finally, the plant-specific RS subfamily includes four members in *Arabidopsis*, two in *O. sativa* and *C. reinhardtii*, but apparently only one in *C. braunii*, *M. polymorpha*, and *P. patens*. It is characterized by two RRMs, with the second containing no SWQDLKD motif, and an RS region. The different number of SR proteins within each subfamily among different plant species could have played a role in the adaptation to different habitats and perhaps even speciation, particularly if novel and distinct functions were acquired. Future comparative studies across species should reveal the functions of SR orthologs or, at the very least, provide insight into the evolutionary history of each lineage.

CONCLUDING REMARKS AND OUTLOOK

Colonizing new habitats requires adaption to new environmental conditions, which can only be achieved by evolving traits that allow an organism to cope with such conditions. Given the documented role of alternative splicing in allowing plants to

cope with adverse environments and its prevalence across the plant kingdom, it seems likely that this posttranscriptional mechanism was important in plant adaptation to terrestrial habitats. Naturally, it follows that if early land plants already modulated alternative splicing through SR proteins, then these RNA-binding factors could have provided an advantage in the adaptation to land. As mentioned above, comparative studies across species representative of the major plant groups should help understand how certain traits appeared and evolved. A number of comparative studies have already been conducted (Kalyna and Barta, 2004; Iida and Go, 2006; Kalyna *et al.*, 2006; Richardson *et al.*, 2011; Califice *et al.*, 2012; Rauch *et al.*, 2014), but as more genomes and transcriptomes become sequenced and annotations improve, more information can be extracted from these analyses, thus requiring continuous updates.

However, comparative studies require solid and reliable data that are easily accessible in public databases. In fact, an indispensable prerequisite to the comparison of the number of SR proteins across species is the availability of well-annotated genomes, transcriptomes, and proteomes. Recent years have seen substantial advances in this regard, but there are still numerous misannotations, including in protein sequences and domain composition, that significantly hamper gene classification and functional studies. In the case of our survey of SR proteins in the moss *P. patens*, we were able to unambiguously classify all SR genes into the previously established subfamilies, thus helping validate the definition proposed by Barta *et al.* (2010). Nevertheless, our analysis shows that care should be taken when working with the *P. patens* databases, as the genome is clearly not perfectly curated and annotated. A recent paper on *M. polymorpha* reported 17 *P. patens* SR proteins (Bowman *et al.*, 2017), as they did not consider that Pp3c11_26740V3.1 and Pp3c11_26750V3.1 occupy the same position in the genome, likely because that information is not readily available in any database. Another study identifies 18 SR proteins in *P. patens*, though they do not discriminate between SR and SR-like proteins, thus overestimating the number of SR proteins (Fesenko *et al.*, 2017). Furthermore, the authors place Pp3c7_5100V3.1 in the RS2Z subfamily despite the fact that it only contains one zinc knuckle, which underlines the confusion in nomenclature regarding this family. Earlier studies comparing SR proteins in

plants (Califice et al., 2012; Rauch et al., 2014) and 27 eukaryotes (Richardson et al., 2011) report only 10 members of this gene family in *P. patens*, though also finding representatives for all the subfamilies. The tendency to report more SR genes in recent years reflects the improvement in the *P. patens* genome annotation. However, the gene models currently available in the several public databases, which are based on the current genome annotation, are not yet perfect, as in some cases they are not supported by expression data.

Regarding SR-like proteins, although they were formally excluded from the SR family when the nomenclature was updated a decade ago (Barta et al., 2010; Manley and Krainer, 2010), many studies still refer to them as such, as they also harbor both RRM and RS regions. *A. thaliana* expresses two SR-like proteins, but their number in *P. patens* remains uncertain. Our preliminary analysis identifies three (data not shown), but Fesenko and coworkers reported four (Fesenko et al., 2017). This discrepancy needs to be addressed in future studies.

Our analysis of the current *P. patens* genome annotation substantiates the existence of 16 canonical SR proteins distributed across the six previously described families. Conservation of the number of subfamilies and the protein structure organization of SR proteins from multicellular green algae to bryophytes to angiosperms, including the three plant-specific subfamilies, supports an ancient diversification of this family, likely predating the colonization of land habitats. Of note is the absence of the SC and the RS2Z subfamilies in *C. reinhardtii*. The SC subfamily, being also present in other eukaryotes including animals, could have been lost in this lineage. On the other hand, it is also possible that the plant-specific RS2Z subfamily had not evolved yet, appearing only in more complex algae, such as *C. braunii*. A number of evolutionary innovations previously assigned only to land plants have recently been identified in *C. braunii* (Nishiyama et al., 2018). It is tempting to speculate that the appearance of the RS2Z subfamily was one such innovation.

Experimental evidence for the conservation of function between *P. patens* and other plant model species is required

to elucidate to which extent SR proteins already played a key regulatory role in both constitutive and alternative splicing when plants colonized terrestrial habitats. Additionally, a detailed functional analysis of *P. patens* SR proteins will help identify other mechanisms and signaling pathways in which these proteins could be involved. It is interesting to note that SR protein family composition varies greatly among plant lineages, both in total number and in subfamily size, suggesting a role for SR gene loss and gain in plant evolution and adaptation. Comprehensive comparative studies will ultimately shed new light on land plant evolution and elucidate the role of SR proteins in stress tolerance of early land plants.

AUTHOR CONTRIBUTIONS

JM, MK, and PD designed the study and wrote the manuscript. JM conducted the *in silico* analyses and prepared the figures.

FUNDING

JM and PD were funded by the Fundação para a Ciência e a Tecnologia (FCT) through grants PD/BD/138327/2018 (awarded to JM in the frame of the Plants for Life Ph.D. Program), PTDC/BIA-FBT/31018/2017 and PTDC/BIA-BID/30608/2017. Funding from the GREEN-it research unit (UID/Multi/04551/2019) is also acknowledged. MK was supported by the Austrian Science Fund (FWF) through grant P26333.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fpls.2020.00286/full#supplementary-material>

REFERENCES

- Adams, K. L., and Wendel, J. F. (2005). Polyploidy and genome evolution in plants. *Curr. Opin. Plant Biol.* 8, 135–141. doi: 10.1016/j.pbi.2005.01.001
- Barta, A., Kalyna, M., and Reddy, A. S. N. (2010). Implementing a rational and consistent nomenclature for serine/arginine-rich protein splicing factors (SR Proteins) in Plants. *Plant Cell* 22, 2926–2929. doi: 10.1105/tpc.110.078352
- Bowman, J. L., Kohchi, T., Yamato, K. T., Jenkins, J., Shu, S., Ishizaki, K., et al. (2017). Insights into land plant evolution garnered from the *Marchantia polymorpha* genome. *Cell* 171, 287–304.e15. doi: 10.1016/j.cell.2017.09.030
- Califice, S., Baurain, D., Hanikenne, M., and Motte, P. (2012). A single ancient origin for prototypical serine/arginine-rich splicing factors. *Plant Physiol.* 158, 546–560. doi: 10.1104/pp.111.189019
- Chamala, S., Feng, G., Chavarro, C., and Barbazuk, W. B. (2015). Genome-wide identification of evolutionarily conserved alternative splicing events in flowering plants. *Front. Bioeng. Biotechnol.* 3:33. doi: 10.3389/fbioe.2015.00033
- Chang, C. Y., Lin, W. D., and Tu, S. L. (2014). Genome-wide analysis of heat-sensitive alternative splicing in *physcomitrella patens*. *Plant Physiol.* 165, 826–840. doi: 10.1104/pp.113.230540
- Clark, J. W., and Donoghue, P. C. J. (2018). Whole-genome duplication and plant macroevolution. *Trends Plant Sci.* 23, 933–945. doi: 10.1016/j.tplants.2018.07.006
- de Vries, J., and Archibald, J. M. (2018). Plant evolution: landmarks on the path to terrestrial life. *New Phytol.* 217, 1428–1434. doi: 10.1111/nph.14975
- Fesenko, I., Khazigaleeva, R., Kirov, I., Kniazev, A., Glushenko, O., Babalyan, K., et al. (2017). Alternative splicing shapes transcriptome but not proteome diversity in *Physcomitrella patens*. *Sci. Rep.* 7:2698. doi: 10.1038/s41598-017-02970-z
- Floyd, S. K., and Bowman, J. L. (2007). The ancestral developmental tool kit of land plants. *Int. J. Plant Sci.* 168, 1–35. doi: 10.1086/509079
- Gensel, P. G. (2008). The earliest land plants. *Ann. Rev. Ecol. Evol. Syst.* 39, 459–477. doi: 10.1146/annurev.ecolsys.39.110707.173526
- Heckman, D. S., Geiser, D. M., Eidell, B. R., Stauffer, R. L., Kardos, N. L., and Hedges, S. B. (2001). Molecular evidence for the early colonization of land by fungi and plants. *Science* 293, 1129–1133. doi: 10.1126/science.1061457
- Huang, Y., and Steitz, J. A. (2005). SRprises along a messenger's journey. *Mol. Cell* 17, 613–615. doi: 10.1016/j.molcel.2005.02.020

- Iida, K., and Go, M. (2006). Survey of conserved alternative splicing events of mRNAs encoding SR proteins in land plants. *Mol. Biol. Evol.* 23, 1085–1094. doi: 10.1093/molbev/msj118
- Ju, C., Van De Poel, B., Cooper, E. D., Thierier, J. H., Gibbons, T. R., Delwiche, C. F., et al. (2015). Conservation of ethylene as a plant hormone over 450 million years of evolution. *Nat. Plants* 1:14004. doi: 10.1038/nplants.2014.4
- 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. doi: 10.1042/BST0320561
- Kalyna, M., Lopato, S., Voronin, V., and Barta, A. (2006). Evolutionary conservation and regulation of particular alternative splicing events in plant SR proteins. *Nucleic Acids Res.* 34, 4395–4405. doi: 10.1093/nar/gkl570
- Kenrick, P., and Crane, P. R. (1997). The origin and early evolution of plants on land. *Nature* 389, 33–39. doi: 10.1038/37918
- Lang, D., Ullrich, K. K., Murat, F., Fuchs, J., Jenkins, J., Haas, F. B., et al. (2018). The *Physcomitrella patens* chromosome-scale assembly reveals moss genome structure and evolution. *Plant J.* 93, 515–533. doi: 10.1111/tpj.13801
- Lang, D., Van Gessel, N., Ullrich, K. K., and Reski, R. (2016). The genome of the model moss *Physcomitrella patens*. *Genom. Evol. Charophyt. Bryophyt. Lycophyt. Ferns* 78, 97–140. doi: 10.1016/bs.abr.2016.01.004
- Lareau, L. F., Inada, M., Green, R. E., Wengrod, J. C., and Brenner, S. E. (2007). Unproductive splicing of SR genes associated with highly conserved and ultraconserved DNA elements. *Nature* 446, 926–929. doi: 10.1038/nature05676
- Lloyd, J. P. B., Lang, D., Zimmer, A. D., Causier, B., Reski, R., and Davies, B. (2018). The loss of SMG1 causes defects in quality control pathways in *Physcomitrella patens*. *Nucleic Acids Res.* 46, 5822–5836. doi: 10.1093/nar/gky225
- 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. doi: 10.1042/BJ20081501
- Manley, J. L., and Krainer, A. R. (2010). A rational nomenclature for serine/arginine-rich protein splicing factors (SR proteins). *Genes Dev.* 24, 1073–1074. doi: 10.1101/gad.1934910
- Marquez, Y., Brown, J. W. S., 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. doi: 10.1101/gr.134106.111
- Mastrangelo, A. M., Marone, D., Laido, G., De Leonardi, A. M., and De Vita, P. (2012). Alternative splicing: enhancing ability to cope with stress via transcriptome plasticity. *Plant Sci.* 185, 40–49. doi: 10.1016/j.plantsci.2011.09.006
- Morris, J. L., Puttick, M. N., Clark, J. W., Edwards, D., Kenrick, P., Pressel, S., et al. (2018). The timescale of early land plant evolution. *Proc. Natl. Acad. Sci. U.S.A.* 115, E2274–E2283. doi: 10.1073/pnas.1719588115
- Nasif, S., Contu, L., and Muhlemann, O. (2018). Beyond quality control: the role of nonsense-mediated mRNA decay (n.d.) in regulating gene expression. *Semin. Cell Dev. Biol.* 75, 78–87. doi: 10.1016/j.semcdb.2017.08.053
- Nishiyama, T., Sakayama, H., De Vries, J., Buschmann, H., Saint-Marcoux, D., Ullrich, K. K., et al. (2018). The chara genome: secondary complexity and implications for plant terrestrialization. *Cell* 174:448. doi: 10.1016/j.cell.2018.06.033
- One Thousand Plant Transcriptomes Initiative. (2019). One thousand plant transcriptomes and the phylogenomics of green plants. *Nature* 574, 679–685. doi: 10.1038/s41586-019-1693-2
- Panchy, N., Lehti-Shiu, M., and Shiu, S. H. (2016). Evolution of gene duplication in plants. *Plant Physiol.* 171, 2294–2316. doi: 10.1104/pp.16.00523
- Rauch, H. B., Patrick, T. L., Klusman, K. M., Battistuzzi, F. U., Mei, W., Brendel, V. P., et al. (2014). Discovery and expression analysis of alternative splicing events conserved among plant SR proteins. *Mol. Biol. Evol.* 31, 605–613. doi: 10.1093/molbev/mst238
- Rensing, S. A., Beike, A. K., and Lang, D. (2013). “Evolutionary importance of generative polyploidy for genome evolution of haploid-dominant land plants,” in *Plant Genome Diversity*, eds J. Greilhuber, J. Dolezel, and J. Wendel. (Vienna: Springer).
- Rensing, S. A., Ick, J., Fawcett, J. A., Lang, D., Zimmer, A., Van De Peer, Y., et al. (2007). An ancient genome duplication contributed to the abundance of metabolic genes in the moss *Physcomitrella patens*. *BMC Evol. Biol.* 7:130. doi: 10.1186/1471-2148-7-130
- Rensing, S. A., Lang, D., Zimmer, A. D., Terry, A., Salamov, A., Shapiro, H., et al. (2008). The *Physcomitrella* genome reveals evolutionary insights into the conquest of land by plants. *Science* 319, 64–69. doi: 10.1126/science.1150646
- Richardson, D. N., Rogers, M. F., Labadorf, A., Ben-Hur, A., Guo, H., Paterson, A. H., et al. (2011). Comparative analysis of serine/arginine-rich proteins across 27 eukaryotes: insights into sub-family classification and extent of alternative splicing. *PLoS One* 6:e24542. doi: 10.1371/journal.pone.0024542
- Shang, X., Cao, Y., and Ma, L. (2017). Alternative splicing in plant genes: a means of regulating the environmental fitness of plants. *Int. J. Mol. Sci.* 18:E432. doi: 10.3390/ijms18020432
- Smidkova, M., Hola, M., and Angelis, K. J. (2010). Efficient biolistic transformation of the moss *Physcomitrella patens*. *Biol. Plant.* 54, 777–780. doi: 10.1007/s10535-010-0141-9
- Soltis, P. S., Marchant, D. B., Van De Peer, Y., and Soltis, D. E. (2015). Polyploidy and genome evolution in plants. *Curr. Opin. Genet. Dev.* 35, 119–125. doi: 10.1016/j.gde.2015.11.003
- Staiger, D., and Brown, J. W. S. (2013). Alternative splicing at the intersection of biological timing, development, and stress responses. *Plant Cell* 25, 3640–3656. doi: 10.1105/tpc.113.113803
- 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. doi: 10.1016/j.tplants.2012.06.001
- Tillemans, V., Leponce, I., Rausin, G., Dispa, L., and Motte, P. (2006). Insights into nuclear organization in plants as revealed by the dynamic distribution of *Arabidopsis* SR splicing factors. *Plant Cell* 18, 3218–3234. doi: 10.1105/tpc.106.044529
- Twyffels, L., Gueydan, C., and Krays, V. (2011). Shuttling SR proteins: more than splicing factors. *FEBS J.* 278, 3246–3255. doi: 10.1111/j.1742-4658.2011.08274.x
- Wu, H. P., Su, Y. S., Chen, H. C., Chen, Y. R., Wu, C. C., Lin, W. D., et al. (2014). Genome-wide analysis of light-regulated alternative splicing mediated by photoreceptors in *Physcomitrella patens*. *Genome Biol.* 15:R10. doi: 10.1186/gb-2014-15-1-r10
- Wu, S., Han, B., and Jiao, Y. (2020). Genetic contribution of paleopolyploidy to adaptive evolution in angiosperms. *Mol. Plant* 13, 59–71. doi: 10.1016/j.molp.2019.10.012
- Zhang, R., Calixto, C. P. G., Marquez, Y., Venhuizen, P., Tzioutziou, N. A., Guo, W., et al. (2017). A high quality *Arabidopsis* transcriptome for accurate transcript-level analysis of alternative splicing. *Nucleic Acids Res.* 45, 5061–5073. doi: 10.1093/nar/gkx267
- Zhong, X. Y., Wang, P., Han, J., Rosenfeld, M. G., and Fu, X. D. (2009). SR proteins in vertical integration of gene expression from transcription to RNA processing to translation. *Mol. Cell* 35, 1–10. doi: 10.1016/j.molcel.2009.06.016
- Zimmer, A. D., Lang, D., Buchta, K., Rombauts, S., Nishiyama, T., Hasebe, M., et al. (2013). Reannotation and extended community resources for the genome of the non-seed plant *Physcomitrella patens* provide insights into the evolution of plant gene structures and functions. *BMC Genom.* 14:498. doi: 10.1186/1471-2164-14-498

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2020 Melo, Kalyna and Duque. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

Advantages of publishing in Frontiers



OPEN ACCESS

Articles are free to read
for greatest visibility
and readership



FAST PUBLICATION

Around 90 days
from submission
to decision



HIGH QUALITY PEER-REVIEW

Rigorous, collaborative,
and constructive
peer-review



TRANSPARENT PEER-REVIEW

Editors and reviewers
acknowledged by name
on published articles

Frontiers

Avenue du Tribunal-Fédéral 34
1005 Lausanne | Switzerland

Visit us: www.frontiersin.org

Contact us: info@frontiersin.org | +41 21 510 17 00



REPRODUCIBILITY OF RESEARCH

Support open data
and methods to enhance
research reproducibility



DIGITAL PUBLISHING

Articles designed
for optimal readership
across devices



FOLLOW US

[@frontiersin](https://twitter.com/frontiersin)



IMPACT METRICS

Advanced article metrics
track visibility across
digital media



EXTENSIVE PROMOTION

Marketing
and promotion
of impactful research



LOOP RESEARCH NETWORK

Our network
increases your
article's readership