

# PLANT RNA BIOLOGY



EDITED BY: Dóra Szakonyi, Ana Confraria, Concetta Valerio, Paula Duque  
and Dorothee Staiger  
PUBLISHED IN: Frontiers in Plant Science







# frontiers

## Frontiers Copyright Statement

© Copyright 2007-2019 Frontiers Media SA. All rights reserved.

All content included on this site, such as text, graphics, logos, button icons, images, video/audio clips, downloads, data compilations and software, is the property of or is licensed to Frontiers Media SA ("Frontiers") or its licensees and/or subcontractors. The copyright in the text of individual articles is the property of their respective authors, subject to a license granted to Frontiers.

The compilation of articles constituting this e-book, wherever published, as well as the compilation of all other content on this site, is the exclusive property of Frontiers. For the conditions for downloading and copying of e-books from Frontiers' website, please see the Terms for Website Use. If purchasing Frontiers e-books from other websites or sources, the conditions of the website concerned apply.

Images and graphics not forming part of user-contributed materials may not be downloaded or copied without permission.

Individual articles may be downloaded and reproduced in accordance with the principles of the CC-BY licence subject to any copyright or other notices. They may not be re-sold as an e-book.

As author or other contributor you grant a CC-BY licence to others to reproduce your articles, including any graphics and third-party materials supplied by you, in accordance with the Conditions for Website Use and subject to any copyright notices which you include in connection with your articles and materials.

All copyright, and all rights therein, are protected by national and international copyright laws.

The above represents a summary only. For the full conditions see the Conditions for Authors and the Conditions for Website Use.

ISSN 1664-8714  
ISBN 978-2-88963-025-7  
DOI 10.3389/978-2-88963-025-7

## 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](mailto:researchtopics@frontiersin.org)



# PLANT RNA BIOLOGY

Topic Editors:

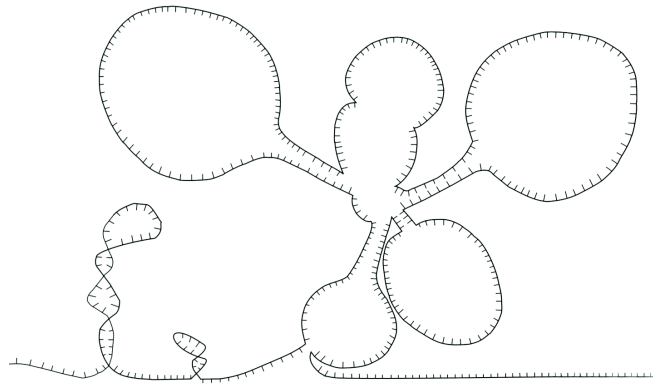
**Dóra Szakonyi**, Instituto Gulbenkian de Ciência, Portugal

**Ana Confraria**, Instituto Gulbenkian de Ciência, Portugal

**Concetta Valerio**, Instituto Gulbenkian de Ciência, Portugal

**Paula Duque**, Instituto Gulbenkian de Ciência, Portugal

**Dorothee Staiger**, Bielefeld University, Germany



Artwork by Dóra Szakonyi

Discoveries from the past decades revealed that RNA molecules are much more than inert intermediates between the coding DNA sequences and their functional products, proteins. Today, RNAs are recognized as active regulatory molecules influencing gene expression, chromatin organization and genome stability, thus impacting all aspects of plant life including development, growth, reproduction and stress tolerance.

Innovations in methodologies, the expanding application of next-generation sequencing technologies, and the creation of public datasets and databases have exposed a new universe of RNA-based mechanisms and led to the discovery of new families of non-coding RNAs, uncovered the large extent of alternative splicing events, and highlighted the potential roles of RNA modifications and RNA secondary structures. Furthermore, considerable advances have been made in identifying RNA-binding and processing factors involved in the synthesis and maturation of different forms of RNA molecules as well as in RNA processing, biochemical modifications or degradation.

This Research Topic showcases the broad biological significance of RNAs in plant systems and contains eight original research articles, one review and four mini-reviews, covering various RNA-based mechanisms in higher plants. Emerging new technologies and novel multidisciplinary approaches are empowering the



scientific community and will expectedly bring novel insights into our understanding of the mechanisms through which RNA is regulated and regulates biological processes in plant cells.

**Citation:** Szakonyi, D., Confraria, A., Valerio, C., Duque, P., Staiger, D., eds. (2019). Plant RNA Biology: Lausanne: Frontiers Media. doi: 10.3389/978-2-88963-025-7



# Table of Contents

## **06 Editorial: Plant RNA Biology**

Dóra Szakonyi, Ana Confraria, Concetta Valerio, Paula Duque and Dorothee Staiger

## **RNA STRUCTURE**

### **09 New Era of Studying RNA Secondary Structure and its Influence on Gene Regulation in Plants**

Xiaofei Yang, Minglei Yang, Hongjing Deng and Yiliang Ding

## **SPlicing**

### **16 A Role of U12 Intron in Proper Pre-mRNA Splicing of Plant Cap Binding Protein 20 Genes**

Marcin Pieczynski, Katarzyna Kruska, Dawid Bielewicz, Jakub Dolata, Michal Szczesniak, Wojciech Karlowski, Artur Jarmolowski and Zofia Szweykowska-Kulinska

### **30 Ectopic Transplastomic Expression of a Synthetic MatK Gene Leads to Cotyledon-Specific Leaf Variegation**

Yujiao Qu, Julia Legen, Jürgen Arndt, Stephanie Henkel, Galina Hoppe, Christopher Thieme, Giovanna Ranzini, Jose M. Muino, Andreas Weihe, Uwe Ohler, Gert Weber, Oren Ostersetzer and Christian Schmitz-Linneweber

### **44 Alternative Splicing as a Regulator of Early Plant Development**

Dóra Szakonyi and Paula Duque

## **RNA STABILITY**

### **53 Functional Characterization of SMG7 Paralogs in Arabidopsis thaliana**

Claudio Capitaio, Neha Shukla, Aneta Wandrolova, Ortrun Mittelsten Scheid and Karel Riha

## **NON-CODING RNAs**

### **63 Under a New Light: Regulation of Light-Dependent Pathways by Non-coding RNAs**

Camila Sánchez-Retuerta, Paula Suárez-López and Rossana Henriques

### **80 Transposon-Derived Non-coding RNAs and Their Function in Plants**

Jungnam Cho

### **86 Cold-Dependent Expression and Alternative Splicing of Arabidopsis Long Non-coding RNAs**

Cristiane P. G. Calixto, Nikoleta A. Tzioutziou, Allan B. James, Csaba Hornyik, Wenbin Guo, Runxuan Zhang, Hugh G. Nimmo and John W. S. Brown

### **102 Nuclear Speckle RNA Binding Proteins Remodel Alternative Splicing and the Non-coding Arabidopsis Transcriptome to Regulate a Cross-Talk Between Auxin and Immune Responses**

Jérémie Bazin, Natali Romero, Richard Rigo, Celine Charon, Thomas Blein, Federico Ariel and Martin Crespi



- 115** *Regulation of Plant Microprocessor Function in Shaping microRNA Landscape*  
Jakub Dolata, Michał Taube, Mateusz Bajczyk, Artur Jarmolowski,  
Zofia Szweykowska-Kulinska and Dawid Bielewicz
- 123** *miRVIT: A Novel miRNA Database and its Application to Uncover Vitis Responses to Flavescence dorée Infection*  
Walter Chitarra, Chiara Pagliarani, Simona Abbà, Paolo Boccacci,  
Giancarlo Birello, Marika Rossi, Sabrina Palmano, Cristina Marzachi,  
Irene Perrone and Giorgio Gambino
- 139** *The MicroRNA319d/TCP10 Node Regulates the Common Bean – Rhizobia Nitrogen-Fixing Symbiosis*  
José Á. Martín-Rodríguez, Alfonso Leija, Damien Formey and  
Georgina Hernández
- 152** *Respective Contributions of URT1 and HESO1 to the Uridylation of 5' Fragments Produced From RISC-Cleaved mRNAs*  
Hélène Zuber, Hélène Scheer, Anne-Caroline Joly and Dominique Gagliardi





# Editorial: Plant RNA Biology

**Dóra Szakonyi<sup>1\*†</sup>, Ana Confraria<sup>1\*†</sup>, Concetta Valerio<sup>1</sup>, Paula Duque<sup>1</sup> and Dorothee Staiger<sup>2</sup>**

<sup>1</sup> Instituto Gulbenkian de Ciência, Oeiras, Portugal, <sup>2</sup> RNA Biology and Molecular Physiology, Faculty of Biology, Bielefeld University, Bielefeld, Germany

**Keywords:** posttranscriptional regulation, RNA processing, RNA stability, non-coding RNAs, RNA structure, transcriptomics

## Editorial on the Research Topic

### Plant RNA Biology

Initially regarded as mere intermediates between coding DNA sequences and proteins, RNAs have been ascribed a plethora of novel roles in the past decades. Today, they are recognized as active regulatory molecules, influencing processes such as gene expression, chromatin organization or genome stability that affect all aspects of a plant's life. This Research Topic aims at highlighting the broad biological significance of RNAs in plant systems and contains eight original research articles, one review and four mini-reviews, covering various RNA-based mechanisms in higher plants.

As they are being synthesized, RNAs fold into secondary structures, which can influence RNA metabolism at multiple levels, including transcript processing and stability, or protein translation. Yang et al. not only bring attention to the significance of RNA structures, but also summarize the currently available methods to study the folding of specific RNA molecules or profile genome-scale RNA structures.

After transcription, RNAs require proper maturation through a multi-step process integrating capping, splicing, and polyadenylation. Pieczynski et al. describe the importance of accurate splicing of the precursor mRNA (pre-mRNA) encoding the small subunit of the Arabidopsis nuclear cap-binding complex, *CBP20*, which contains seven introns. Of these, intron 4 belongs to the U12 class of introns and separates the gene into an N-terminal half encoding the RNA-binding moiety and a C-terminal half encoding a nuclear localization signal. Using intron swapping and mutagenesis, the authors demonstrate that the U12 intron is crucial for correct *CBP20* splicing. Mitochondrial and chloroplastial transcripts also undergo processing, which in the chloroplast is largely carried out by imported nuclear-encoded RNA-binding proteins and RNA-binding factors encoded by the organellar genome. Chloroplasts of land plants retain a single plastid-encoded splicing factor, intron maturase MatK. Qu et al. ectopically expressed MatK in tobacco chloroplasts and observed variegation of the cotyledons, indicating that adequate MatK levels are required for proper chloroplast development. Interestingly, the splicing pattern of selected plastid genes was not strongly affected in these lines, but rather translation was inhibited, leading to reprogramming of chloroplast gene expression.

Alternative splicing, which results from differential splicing of the same pre-mRNA, represents a major source of transcriptome and proteome diversity and an important regulation mechanism for development and stress responses in plants. Szakonyi and Duque scrutinize the role of alternative splicing during early plant development. They review the current knowledge on how alternative splicing changes during seed maturation, germination, and seedling formation in the dark and light. Furthermore, they compile the splicing factors with crucial roles during these initial steps of development and the alternatively-spliced genes with putative functional roles for different splice forms.

## OPEN ACCESS

### Edited by:

Diane C. Bassham,  
Iowa State University, United States

### Reviewed by:

Marie-Theres Hauser,  
University of Natural Resources and  
Life Sciences Vienna, Austria

### \*Correspondence:

Dóra Szakonyi  
dszakonyi@igc.gulbenkian.pt  
Ana Confraria  
aagusto@igc.gulbenkian.pt

<sup>†</sup>These authors have contributed  
equally to this work

### Specialty section:

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

**Received:** 02 May 2019

**Accepted:** 21 June 2019

**Published:** 09 July 2019

### Citation:

Szakonyi D, Confraria A, Valerio C,  
Duque P and Staiger D (2019)  
Editorial: Plant RNA Biology.  
Front. Plant Sci. 10:887.  
doi: 10.3389/fpls.2019.00887

Alternative splicing or mutations can introduce premature termination codons (PTCs) that target mRNAs to degradation by an evolutionarily-conserved process termed nonsense-mediated decay (NMD). The work presented by Capitao et al. focuses on SMG7 proteins, conserved components of the NMD machinery in eukaryotes. Plants lost several core NMD genes during evolution, but SMG7 genes expanded and acquired novel functions. In contrast to previous findings, the authors report that SMG7 is not an essential gene in Arabidopsis and fulfills an additional, NMD-unrelated function in meiosis. Moreover, SMG7-like (SMG7L), a close paralog arising from gene duplication in dicots, is not functionally redundant with SMG7 though its biological role remains unclear.

Research in the last couple of decades has also uncovered several different types of non-coding RNAs, grouped according to size into long non-coding (lnc) RNAs and a highly diverse pool of small RNAs. The latter can be further classified depending on their biogenesis pathway and include microRNAs (miRNAs), encoded by MIR genes, as well as several classes of short interfering RNAs (siRNAs), generated from double-stranded RNA (Borges and Martienssen, 2015). Non-coding RNAs exert different regulatory functions in plants, playing important roles in developmental programs but also integrating environmental cues and participating in stress responses (Chekanova, 2015; D'ario et al., 2017; Li et al., 2017; Sun et al., 2018). In this context, Sánchez-Retuerta et al. summarize the current knowledge on light regulation of miRNAs and lncRNAs and their functions in photomorphogenesis, circadian clock regulation, and photoperiod-dependent flowering. On the other hand, the review by Cho highlights the potential role of transposon-derived lncRNAs and small RNAs, which suppress harmful effects of transposition, but can also act on non-transposon transcripts affecting development and stress responses. Furthermore, Calixto et al. analyze the effect of cold stress on the expression and processing of regulatory non-coding RNAs, performing an ultra-deep sequencing of a diel time-series of Arabidopsis plants. The authors identified specific lncRNAs and primary miRNAs differentially expressed and/or alternatively spliced in response to cold stress that potentially contribute to acclimation and freezing tolerance. Bazin et al. explore the role of Nuclear Speckle RNA-binding proteins (NSRs), which are known alternative splicing modulators, through a regulatory loop involving the lncRNA *Alternative Splicing Competitor RNA (ASCO)* (Bardou et al., 2014). Using RNA immunoprecipitation coupled with RNA-seq, the authors showed that NSRs directly interact with lncRNAs, many of which are upregulated in the *nsra nsrb* mutant. This suggests that NSRs regulate the steady-state abundance of lncRNAs, which in turn may contribute to splicing regulation.

In the plant cell nucleus, miRNAs are processed from primary miRNA transcripts by the Microprocessor complex, which is composed of three core proteins in Arabidopsis: Dicer-Like 1 (DCL1), Hyponastic Leaves 1 (HYL1), and Serrate (SE) (Song et al., 2019). Dolata et al. review how the Microprocessor complex is regulated, emphasizing the posttranslational regulation of its components, but also covering posttranscriptional regulatory mechanisms. Chitarra et al. reveal miRNA changes associated to development, photosynthesis, jasmonate signaling, and disease

resistance in *Vitis vinifera* infected with the phytoplasma *Flavescence dorée*. Importantly, the authors provide a valuable tool for the research community working on *Vitis* spp. (miRVIT, available at <http://mirvit.ipsp.cnr.it/>); they assemble a comprehensive database of novel putative grapevine miRNAs, uniformly reannotating and aligning all described accessions to the latest version of the grape genome and listing their validated and predicted target transcripts. In common bean, Martin-Rodriguez et al. explore the role of the nodule-expressed miR319d in the regulation of rhizobial infection and symbiotic nodule development. The authors identified the transcription factor *TEOSINTE BRANCHED/CYCLOIDEA/PCF 10 (TCP10)* as a target of miR319d and propose that the effect of miR319d-TCP10 on nodulation involves the modulation of jasmonate signaling through the transcriptional regulation of the jasmonate biosynthetic gene *LIPOXYGENASE 2 (LOX2)*. The 5' fragments of ARGONAUTE 1 (AGO1)-mediated cleavage guided by miRNAs and siRNAs are uridylated at the 3' end to facilitate further degradation. In Arabidopsis, uridylation of these fragments is carried out by two terminal uridylyltransferases (TUTases) HEN1 SUPPRESSOR 1 (HESO1) and URIDYLTRANSFERASE 1 (URT1), whose activity is explored in the study of Zuber et al. The authors optimize a 3'RACE-seq method to analyze quantitatively and qualitatively the uridylation at the 3' ends of AGO1 5' cleavage products. Showcasing the applicability of 3'RACE-seq, the 5' cleavage fragments of the miR159 target *MYB33* and the miR156/157 target *SQUAMOSA-PROMOTER BINDING PROTEIN-LIKE 13 (SPL13)* are compared in the wild type and mutants impaired in TUTase activity. The authors uncover different contributions of HESO1 and URT1 to the uridylation of the cleavage fragments, with HESO1 being the main TUTase.

The diverse collection of articles in this Research Topic demonstrates the current vibrant and rapid progression of the plant RNA biology field. Current major challenges include, but are not limited to, identifying functional protein-RNA interactions (Foley et al., 2017), ascribing biological functions to different RNAs (Zhao et al., 2019), and dissecting tissue complexity using single cell transcriptomics (Ryu et al., 2019). Technological advances coupled to the increase in publicly-available data are expected to allow the community to address these and other issues at a fast pace in upcoming years, bringing key novel insights into our understanding of the mechanisms through which RNA is regulated and regulates biological processes in plant cells.

## AUTHOR CONTRIBUTIONS

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

## ACKNOWLEDGMENTS

We thank all authors who submitted their work for this Research Topic as well as the invaluable help of reviewers in manuscript evaluation and the support of professional editorial staff at

Frontiers. DS, AC, CV, and PD are funded by Fundação para a Ciência e a Tecnologia (grants: UID/Multi/04555/2013, PTDC/BIA-PLA/1084/2014, PTDC/BIA-FBT/31018/2017, and

PTDC/BIA-BID/32347/2017). DSt is funded by the German research foundation through STA653/9, STA653/13, and STA653/15.

## REFERENCES

- Bardou, F., Ariel, F., Simpson, C. G., Romero-Barrios, N., Laporte, P., Balzergue, S., et al. (2014). Long noncoding RNA modulates alternative splicing regulators in Arabidopsis. *Dev. Cell* 30, 166–176. doi: 10.1016/j.devcel.2014.06.017
- Borges, F., and Martienssen, R. A. (2015). The expanding world of small RNAs in plants. *Nat. Rev. Mol. Cell. Biol.* 16, 727–741. doi: 10.1038/nrm4085
- 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
- D'ario, M., Griffiths-Jones, S., and Kim, M. (2017). Small RNAs: big impact on plant development. *Trends Plant Sci.* 22, 1056–1068. doi: 10.1016/j.tplants.2017.09.009
- Foley, S. W., Gosai, S. J., Wang, D., Selamoglu, N., Sollitti, A. C., Köster, T., et al. (2017). A global view of RNA-protein interactions identifies post-transcriptional regulators of root hair cell fate. *Dev. Cell* 41, 204–220.e5. doi: 10.1016/j.devcel.2017.03.018
- Li, S., Castillo-González, C., Yu, B., and Zhang, X. (2017). The functions of plant small RNAs in development and in stress responses. *Plant J.* 90, 654–670. doi: 10.1111/tpj.13444
- Ryu, K. H., Huang, L., Kang, H. M., and Schiefelbein, J. (2019). Single-cell RNA sequencing resolves molecular relationships among individual plant cells. *Plant Physiol.* 179, 1444–1456. doi: 10.1104/pp.18.01482
- Song, X., Li, Y., Cao, X., and Qi, Y. (2019). MicroRNAs and their regulatory roles in plant-environment interactions. *Ann. Rev. Plant Biol.* 70:489–525. doi: 10.1146/annurev-arplant-050718-100334
- Sun, X., Zheng, H., and Sui, N. (2018). Regulation mechanism of long non-coding RNA in plant response to stress. *Biochem. Biophys. Res. Commun.* 503, 402–407. doi: 10.1016/j.bbrc.2018.07.072
- Zhao, W., Chu, S., and Jiao, Y. (2019). Present scenario of circular RNAs (circRNAs) in plants. *Front. Plant Sci.* 10:379. doi: 10.3389/fpls.2019.00379

**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 Szakonyi, Confraria, Valerio, Duque and Staiger. 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 of Studying RNA Secondary Structure and Its Influence on Gene Regulation in Plants

Xiaofei Yang<sup>1†</sup>, Minglei Yang<sup>1†</sup>, Hongjing Deng<sup>1,2,3†</sup> and Yiliang Ding<sup>1\*</sup>

<sup>1</sup> Department of Cell and Developmental Biology, John Innes Centre, Norwich, United Kingdom, <sup>2</sup> State Key Laboratory of Plant Genomics and National Center for Plant Gene Research, Institute of Genetics and Developmental Biology, Chinese Academy of Sciences, Beijing, China, <sup>3</sup> College of Life Sciences, University of Chinese Academy of Sciences, Beijing, China

## OPEN ACCESS

### Edited by:

Dora Szakonyi,  
Instituto Gulbenkian de Ciência (IGC),  
Portugal

### Reviewed by:

Cuncong Zhong,  
University of Kansas, United States  
Carmen Hernandez,  
Instituto de Biología Molecular y  
Celular de Plantas (IBMCP), Spain  
Aleksandar Spasic,  
University of Rochester, United States

### \*Correspondence:

Yiliang Ding  
yiliang.ding@jic.ac.uk

<sup>†</sup> These authors have contributed  
equally to this work.

### Specialty section:

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

**Received:** 28 February 2018

**Accepted:** 02 May 2018

**Published:** 22 May 2018

### Citation:

Yang X, Yang M, Deng H and Ding Y  
(2018) New Era of Studying RNA  
Secondary Structure and Its Influence  
on Gene Regulation in Plants.  
Front. Plant Sci. 9:671.  
doi: 10.3389/fpls.2018.00671

The dynamic structure of RNA plays a central role in post-transcriptional regulation of gene expression such as RNA maturation, degradation, and translation. With the rise of next-generation sequencing, the study of RNA structure has been transformed from *in vitro* low-throughput RNA structure probing methods to *in vivo* high-throughput RNA structure profiling. The development of these methods enables incremental studies on the function of RNA structure to be performed, revealing new insights of novel regulatory mechanisms of RNA structure in plants. Genome-wide scale RNA structure profiling allows us to investigate general RNA structural features over 10s of 1000s of mRNAs and to compare RNA structuromes between plant species. Here, we provide a comprehensive and up-to-date overview of: (i) RNA structure probing methods; (ii) the biological functions of RNA structure; (iii) genome-wide RNA structural features corresponding to their regulatory mechanisms; and (iv) RNA structurome evolution in plants.

**Keywords:** RNA structurome, gene regulation, regulatory RNAs, RNA structure and function, plant RNA biology

RNA secondary structure plays many essential roles in RNA synthesis, metabolism, and regulatory pathways (Bevilacqua et al., 2016; Vandivier et al., 2016). Previous efforts to determine RNA structure depended on classical and time-consuming techniques, such as nuclear magnetic resonance spectroscopy (NMR), X-ray crystallography, and cryo-electron microscopy (Table 1) (Lengyel et al., 2014). However, these methods yielded data limited to a few key RNAs with comparatively short length (less than 200 nt) and high abundance (~1 μmol).

More recently, enzymatic and chemical structure probing methods have been developed to routinely and efficiently obtain structural information of individual RNAs. Ribonucleases (RNase) cleave either single-stranded (ss) RNA regions or double-stranded (ds) RNA regions to indicate RNA base-pairing status. The most commonly used enzymatic probing reagents include RNase V1 (for dsRNA), RNase S1 (for ssRNA), RNase A (for C/U in ssRNA), and RNase T1 (for G in ssRNA) (Knapp, 1989). The RNase-based RNA structure probing method has been used extensively in studying RNA structure with less toxicity, but with the limitation of cell permeability (Kwok, 2016). For chemical probing, two main types of chemical reagent can be used. One modifies the Watson-Crick base-pairing face on the nucleobase, as a direct measure of single-strandedness. Dimethyl sulfate (DMS) is one of the most commonly used

nucleobase probing reagents as it easily penetrates the cell, a pre-requisite for *in vivo* chemical probing (Zaug and Cech, 1995). Another example is 1-cyclohexyl-(2-morpholinoethyl) carbodiimide metho-*p*-toluene sulfonate (CMCT), which targets the unpaired N3 position of uracil and the unpaired N1 position of guanine (Incarnato et al., 2014); while 3-ethoxy-1,1-dihydroxy-2-butanone (kethoxal) attacks the unpaired N1 and unpaired exocyclic amine positions of guanine (Noller and

Chaires, 1972). Among these reagents, DMS is predominantly used to probe RNA structures in different organisms (Ding et al., 2014; Rouskin et al., 2014; Talkish et al., 2014; Deng et al., 2018). The other type of chemical reagent modifies the ribose, by selective 2'-hydroxyl acylation and which can be analyzed by primer extension (SHAPE) (Mortimer and Weeks, 2007; Spitale et al., 2013). A particular advantage of SHAPE is that it generates structural information for all four nucleotides at the same time.

**TABLE 1 |** RNA structure probing methods for both individual RNAs and at the genome-wide scale.

	Experiments	Capabilities	<i>In silico, in vitro, or in vivo</i>	Application	Reaction conditions	Reference
Targeted individual RNA structure probing	NMR; X-ray crystallography; Cryo-electron microscopy	Determine the three-dimensional structure at high resolution	<i>In vitro</i>	Ribosome, HIV	Specific buffer conditions	Lengyel et al., 2014
	Gel-based enzymatic and chemical probing	Determine the RNA secondary structure of high abundant RNAs up to 200 nt length	<i>In vitro or in vivo</i>	<i>E. coli</i>	Specific buffer conditions/cellular conditions	Noller and Chaires, 1972; Zaug and Cech, 1995
	SHAPE-CE	Determine the RNA secondary structure of high abundant RNAs up to 400 nt length	<i>In vitro</i>	<i>Arabidopsis</i>	Specific buffer conditions	Hawkes et al., 2016
	DMS/SHAPE-LMPCR	Determine the RNA secondary structure of low abundant RNAs	<i>In vivo</i>	<i>Arabidopsis</i> , human	Cellular conditions	Kwok et al., 2013
	SHAPE-Seq	Determine the RNA secondary structure of long RNAs	<i>In vitro</i>	RNase P pT181 sense RNA	Specific buffer conditions	Lucks et al., 2011
	SHAPE-MaP; DMS-MaP	Determine the RNA secondary structure of low abundant and long RNAs	<i>In vitro or in vivo</i>	TPP, HIV Yeast, human	Specific buffer conditions/cellular conditions	Siegfried et al., 2014; Smola et al., 2016; Zubradt et al., 2017
Genome-wide RNA structure profiling	FragSeq; PARS; PARTE; PIP-seq	Determine genome-wide <i>in vitro</i> RNA secondary structure with enzymatic probing	<i>In vitro</i>	Mouse Yeast, human <i>Arabidopsis</i>	Specific buffer conditions	Kertesz et al., 2010; Underwood et al., 2010; Li et al., 2012b; Wan et al., 2012, 2014; Gosai et al., 2015; Foley and Gregory, 2016
	CIRS-seq	Determine genome-wide <i>in vitro</i> RNA secondary structure with chemical probing	<i>In vitro</i>	Mouse	Specific buffer conditions	Incarnato et al., 2014
	DMS-seq; Structure-seq; Mod-seq	Determine genome-wide <i>in vivo</i> RNA secondary structure with chemical probing	<i>In vivo</i>	<i>Arabidopsis</i> Yeast <i>Oryza sativa</i>	Cellular conditions	Ding et al., 2014; Rouskin et al., 2014; Talkish et al., 2014; Deng et al., 2018
	icSHAPE	Determine genome-wide <i>in vivo</i> RNA secondary structure with chemical probing	<i>In vivo</i>	Mouse	Cellular conditions	Spitale et al., 2015

Polyacrylamide gel electrophoresis (PAGE) assays were traditionally used to measure the modified pattern of both enzymatic and chemical reactions (Noller and Chaires, 1972; Knapp, 1989; Zaug and Cech, 1995). However, these gel-based assays were limited to highly abundant and short (less than 200 nt) RNAs. The application of capillary electrophoresis (CE) improved the detection limits of both the length (up to 400 nt) and the abundance of RNA (Table 1) (Watts et al., 2009). A recent application of CE on *Arabidopsis thaliana* long non-coding RNA (lncRNA), *COOLAIR*, revealed the remarkable complexity of RNA structure up to 750 nt (Hawkes et al., 2016). A further improved method on probing sensitivity, DMS/SHAPE-LMPCR, was developed in *Arabidopsis thaliana* (Table 1). This method achieved “attomole” sensitivity allowing RNA structure probing of low abundance RNAs in living cells (Kwok et al., 2013). By subsequently combining the action of DMS with next-generation sequencing high-depth RNA structural information of very long RNAs was achieved (Lucks et al., 2011; Smola et al., 2016). For instance, the structural information of over 18 kb lncRNA, *Xist*, was fulfilled in a single experiment (Smola et al., 2016). The development of these approaches has significantly improved the sensitivity and resolution for probing individual RNA structure both *in vitro* and *in vivo*. The capability for single nucleotide-resolution quantitative measurements on any RNA down to 1 attomole and up to 18 kb enables efficient functional investigation of RNA structure in biological processes.

Genome-wide RNA structure profiling was initially achieved by coupling enzymatic probing with next-generation sequencing, PARS (parallel analysis of RNA structure) (Table 1). It was developed in yeast by measuring the catalytic activity of two enzymes, RNase V1 (for dsRNA) and S1 (for ssRNA) (Kertesz et al., 2010). This method was extended in *Arabidopsis thaliana*, *Caenorhabditis elegans*, *Drosophila melanogaster*, and *Homo sapiens* (Li et al., 2012a,b; Wan et al., 2014). An enhanced method, PIP-seq (protein interaction profiling sequencing), complements RNA–protein interaction information with *in vitro* RNA structure profiling (Table 1) (Foley et al., 2015; Gosai et al., 2015; Foley and Gregory, 2016). A further improvement on genome-wide scale RNA structure profiling extended to living cells and addressed native RNA folding status. By harnessing the cell permeability of DMS, the first genome-wide *in vivo* RNA structure profiling method, Structure-seq, was developed in *Arabidopsis* (Ding et al., 2014, 2015) in parallel with DMS-seq and Mod-seq in yeast (Rouskin et al., 2014; Talkish et al., 2014) (Table 1). Both methods reveal *in vivo* RNA structures are more single-stranded than *in vitro* and *in silico* computational predicted RNA structures. Use of the Structure-seq method was recently extended to rice (Deng et al., 2018). A follow-up genome-wide *in vivo* RNA structure profiling method, icSHAPE (*in vivo* click SHAPE), was developed in mouse by using the SHAPE chemical reagent with the power of four-nucleotide probing (Table 1) (Spitale et al., 2015). In addition to measuring reverse transcription stopping, chemical modification can also be determined by mutational profiling (Table 1) (Siegfried et al., 2014; Smola et al., 2016; Zubradt et al., 2017).

These powerful genome-wide methods can provide an accurate and quantitative *in vivo* RNA structure map over tens

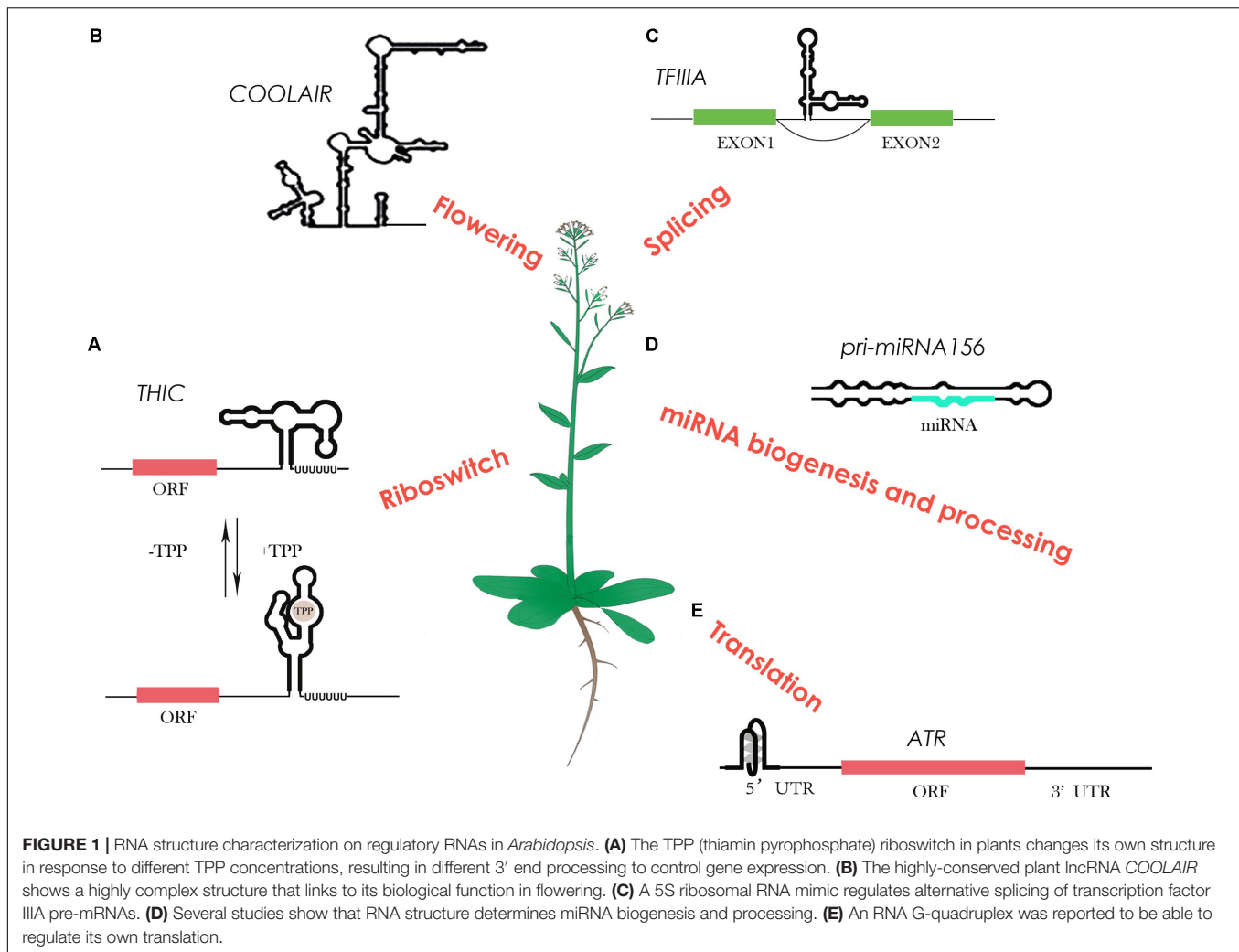
of thousands of RNA with single nucleotide-resolution. These technological advances create an unprecedented scale for the in-depth study of the global impact of RNA structure in gene regulation. For example, regulatory RNAs are able to act as a master regulator in gene expression. In general, these regulatory RNAs directly turn on or off gene expression by altering RNA secondary structure. A recent study of RNA structure characterization on a range of regulatory RNAs in *Arabidopsis* is illustrated below (Figure 1).

A riboswitch is a type of regulatory RNA that contains specific RNA structure segments, which can change conformation depending on specific ligand binding, e.g., metabolites. A well-studied example of a riboswitch is the vitamin B1 derivative thiamin pyrophosphate (TPP), which resides in the 3′ UTR region of the thiamin biosynthetic gene *THIC* (Wachter et al., 2007) (Figure 1A). With a low TPP concentration, the 3′ end processing of *THIC* mRNA results in a short 3′ UTR that permits high expression of the *THIC* gene. Conversely, with a high TPP concentration, TPP binds directly with the 3′ end of the RNA and induces a structural change that prevents splicing. This results in a long 3′ UTR inducing RNA degradation, subsequently reducing *THIC* gene expression (Wachter et al., 2007). Unlike riboswitches in bacteria that control translation through a structural change in the 5′ UTR, plants may have evolved a diverse and more complicated alternative 3′ end processing mechanism in order to cope with a large number of metabolites (Wachter et al., 2007).

Not only are some metabolites able to bind to specific RNA structures to regulate their synthesis pathways, but some are also able to regulate their own expression levels. A plant conserved pre-mRNA of *transcription factor IIIA* (*TFIIIA*) contains a 5S rRNA mimic structural element in one of its exons (Hammond et al., 2009) (Figure 1C). When ribosomal protein L5 binds to this 5S rRNA mimic, it triggers exon skipping in *TFIIIA* mRNA to control *TFIIIA* levels (Hammond et al., 2009). This ribosomal protein–mRNA interaction provides a new-found class of RNAs regulating alternative splicing to control the protein level.

Furthermore, specific RNA structural motifs such as G-quadruplexes (GQS) also play an important role in gene expression regulation. RNA GQSs are typically more stable in the presence of potassium or sodium. Tens of thousands of putative GQSs were identified in *Arabidopsis* and other plant species (Mullen et al., 2010). A recent study reported the first highly-conserved plant RNA GQS located in the 5′ UTR of *ATAXIA TELANGIECTASIA-MUTATED AND RAD3-RELATED* (*ATR*), inhibiting its translation when forming stable GQS structures (Kwok et al., 2015) (Figure 1E). Interestingly, potassium concentrations in plant cells can dramatically increase under drought stress (Mullen et al., 2010). Thus, GQS structural motifs in plants may specifically act as a regulator in response to abiotic stress, such as drought and salinity.

Long non-coding RNAs have also been shown as important regulatory RNAs involved in various biological processes. The study of lncRNA structures has been limited in the past due to their long length and low abundance. Advances in probing methods has enabled the highly-conserved plant lncRNA *COOLAIR* structure to be determined by chemical profiling with CE (Hawkes et al., 2016). *COOLAIR* is a key regulator of a



major plant developmental gene *FLC* (*FLOWERING LOCUS C*), in response to vernalisation. The distal *COOLAIR* isoform in *Arabidopsis* (**Figure 1**) is highly-structured with numerous secondary structural motifs, an intricate multi-way junction, and two unusual asymmetric 5' internal loops (Hawkes et al., 2016) (**Figure 1B**). Interestingly, a single nucleotide polymorphism (SNP) in the natural variation accession, Var2-6, is able to change the structure to affect the RNA stability, resulting in a late-flowering phenotype in Var2-6 (Hawkes et al., 2016). RNA secondary structure determination has progressed our understanding of the structure–function relationship of lncRNAs for the first time in plants.

The other well-known regulatory RNAs, miRNAs, also heavily rely on RNA structure for their regulatory functions (Herr and Baulcombe, 2004). The double-stranded region of miRNA precursors (pri-miRNAs) are recognized and processed by Dicer protein, an RNase III-like enzyme (Herr and Baulcombe, 2004). Previous studies in plants on both individual miRNA precursors and genome-wide assessment of pri-miRNA processing products confirmed that different structure determinants within pri-miRNAs compete for the processing machinery (Song et al., 2010;

Bologna et al., 2013) (**Figure 1D**). A recent RNA structure characterization study by NMR shows the upper stem of a double-stranded region of pri-miR156 is important for Dicer processing at different temperatures, that substantiates the structure-determined Dicer processing feature (Kim et al., 2017). After Dicer processing, an Argonaute (AGO) protein will recognize the processed duplex miRNA to target mRNA containing complementary sequence for either RNA cleavage or translational inhibition (Valencia-Sanchez et al., 2006). Genome-wide *in vitro* RNA structure profiling in *Arabidopsis* revealed a less structured pattern in miRNA target sites that indicates the relationship between miRNA target efficiency and the single-stranded structural feature (Li et al., 2012b).

Apart from these studies on regulatory RNAs, recent genome-wide research also reveals the general role of mRNA structure in a variety of post-transcriptional regulations such as RNA maturation, RNA stability, RNA location and translation.

Alternative splicing is an important process in RNA maturation. More than 40% of *Arabidopsis* genes possess alternative spliced isoforms (Filichkin et al., 2010). The first *in vivo* RNA structure profiling in *Arabidopsis* revealed a



significantly less structural pattern in the 40 nt region upstream of the 5' splice site for unspliced events (including exon skipping and intron retention) (Ding et al., 2014). PIP-seq further revealed that this kind of structural pattern results in more RNA-protein interactions in *Arabidopsis* nuclei (Gosai et al., 2015; Foley and Gregory, 2016). Interestingly, PIP-seq also found the robust structure at the 3' splice site is responsible for more protein interactions (Gosai et al., 2015; Foley and Gregory, 2016). Thus, these RNA structural features indicate an important role of RNA structure in regulating alternative splicing.

Another RNA maturation process is alternative polyadenylation (APA) that is found in over 60% of mRNAs in *Arabidopsis* (Loke et al., 2005). *In vivo* RNA structure profiling shows a strong structural pattern in the U- and A-rich upstream region of the cleavage site as well as a single-stranded region at the cleavage site (Ding et al., 2014). These patterns may correlate with the recognition of endonucleases for regulating APA. Further study using PIP-seq shows more protein bound up- and downstream of the APA cleavage site as compared to constitutive polyadenylation events (Gosai et al., 2015). However, APA sites do not exhibit altered *in vitro* RNA secondary structure compared to constitutive sites (Gosai et al., 2015). This suggests there may be different effects of RNA structure on both protein binding and cleavage activity, that warrant closer investigation.

In addition to RNA maturation, the relationship between RNA structure and RNA degradation has also been uncovered by *in vitro* RNA structurome analysis in *Arabidopsis*. Unlike yeast, highly-structured mRNAs are more likely to be degraded in *Arabidopsis*, probably via specific siRNA processing (Li et al., 2012b).

An interesting study on RNA mobility in plants shows that a stem-bulge-stem-loop tRNA-derived structural motif is sufficient to mediate mRNA transport. A large number of mRNAs containing this motif can be moved through graft junction (Zhang et al., 2016). Thus, RNA structure might also affect intercellular communication across plants.

Another major impact of RNA structure is its regulatory role in translation. Both *in vitro* and *in vivo* RNA structure profiling show a single-stranded region upstream of the start codon that might facilitate ribosome initiation (Li et al., 2012b; Ding et al., 2014). Moreover, a triplet periodic trend is observed in the CDS region but not in UTRs. These structure patterns are obvious in mRNAs with high translation efficiency and are absent in those with low translation efficiency (Ding et al., 2014). This implies that ribosomes may recognize RNA secondary structure as an additional layer of information alongside sequence content.

Additionally, RNA structure is also strongly associated with RNA methylation sites and RNA binding protein (RBP) sites. For example, N<sup>6</sup>-methylation of adenosine alters the stability of the A·U pair (Harcourt et al., 2017). Cellular RNAs show a decrease in base pairing around sites of m<sup>6</sup>A when they were methylated (Harcourt et al., 2017). Recent genome-wide studies indicate that the N<sup>6</sup>-methyladenosine (m<sup>6</sup>A) prefers single-stranded conformations rather than double-stranded structures (Zhao et al., 2017). A genome-wide study in *Arabidopsis* shows an enrichment of m<sup>6</sup>A around the start codon, stop codon and 3' UTR region (Luo et al., 2014).

Interestingly, this enrichment region of m<sup>6</sup>A is well-correlated with the single-stranded region identified in RNA structure profiling (Luo et al., 2014; Zhao et al., 2017). A study of the RNA structurome in rice also confirmed that higher m<sup>6</sup>A modification sites tend to have less RNA structure (Deng et al., 2018). This indicates that m<sup>6</sup>A association may alter RNA structure to more single-strandedness to facilitate gene regulation.

Another key player in post-transcriptional regulations is RBP. Unlike DNA binding protein, RBP associates not only with the primary sequence motifs, but also RNA structural patterns. A recent study combining genome-wide RBP profiling and RNA secondary structure profiling shows that RBP binding sites tend to be more single-stranded (Gosai et al., 2015). Interestingly, a nuclear PIP-seq study confirms that both RNA secondary structure and RBP binding sites show quite different patterns between hair and non-hair cells in plants (Foley et al., 2017). This suggests that cell-type-specific RNA structure and RBP binding may be a new regulatory mechanism during plant development.

From an evolutionary perspective, the conservation and diversity of RNA structurome between species remains poorly understood. A recent study compared, for the first time, the conservation and divergence of *in vivo* RNA structurome between plant species, to assess the evolutionary adaptation of RNA structure (Deng et al., 2018). This study found that *in vivo* RNA secondary structure conservation does not correlate with sequence conservation between rice and *Arabidopsis*. The conservation and divergence in both sequence and RNA secondary structure are highly relevant with specific biological processes (Deng et al., 2018). This indicates evolutionary selection not only modifies sequence, but also alters RNA structure to regulate gene expression. This in turn suggests that RNA secondary structure may serve a different layer of selection to sequence in plants.

Recent methodology advances have overcome previous limitations of both low-throughput and *in vitro* conditions for studying RNA secondary structure. These new methods, with single-nucleotide resolution, genome-wide scale and high sensitivity, significantly accelerate the study of *in vivo* RNA structure and associated biological functions. Plants are more sensitive than animals to varying environmental conditions, such as changes in temperature, salinity, acidity, and heavy metal concentrations (Schleuning et al., 2016). These factors are able to affect RNA folding (Tan and Chen, 2011; Wan et al., 2012; Sun et al., 2017). By applying these new RNA structure analysis methods under a range of environmental conditions, we will be able to determine how RNA structure alters in response to these changes. By integrating RNA structure profiling with mutagenesis assays and phenotypic analysis, the relationship between RNA structure and biological function can be investigated in greater details. Extending analyses to other plant species provides scope for exploring evolutionary selection at the RNA structure level. It is notable that this new era of studying RNA secondary structure provides unprecedented opportunity for discovering novel regulatory mechanisms of gene expression in plants.

## AUTHOR CONTRIBUTIONS

XY, MY, HD, and YD wrote the manuscript. MY produced the figure. XY, HD, and YD made the corrections. All authors listed have made a substantial, direct and intellectual contribution to the work, and approved it for publication.

## FUNDING

This study was supported by the Biotechnology and Biological Sciences Research Council (Grant No. BB/L025000/1 to

XY, HD, and YD) and the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation program (Grant No. 680324 to MY and YD).

## ACKNOWLEDGMENTS

We apologize to authors whose work has not been cited here, because of the limitations associated with the review format.

## REFERENCES

- Bevilacqua, P. C., Ritchey, L. E., Su, Z., and Assmann, S. M. (2016). Genome-wide analysis of RNA secondary structure. *Annu. Rev. Genet.* 50, 235–266. doi: 10.1146/annurev-genet-120215-035034
- Bologna, N. G., Schapire, A. L., Zhai, J., Chorostecki, U., Boisbouvier, J., Meyers, B. C., et al. (2013). Multiple RNA recognition patterns during microRNA biogenesis in plants. *Genome Res.* 23, 1675–1689. doi: 10.1101/gr.153387.112
- Deng, H., Cheema, J., Zhang, H., Woolfenden, H., Norris, M., Liu, Z., et al. (2018). Rice in vivo RNA structurome reveals RNA secondary structure conservation and divergence in plants. *Mol. Plant* 11, 607–622. doi: 10.1016/j.molp.2018.01.008
- Ding, Y., Kwok, C. K., Tang, Y., Bevilacqua, P. C., and Assmann, S. M. (2015). Genome-wide profiling of in vivo RNA structure at single-nucleotide resolution using structure-seq. *Nat. Protoc.* 10, 1050–1066. doi: 10.1038/nprot.2015.064
- Ding, Y., Tang, Y., Kwok, C. K., Zhang, Y., Bevilacqua, P. C., and Assmann, S. M. (2014). In vivo genome-wide profiling of RNA secondary structure reveals novel regulatory features. *Nature* 505, 696–700. doi: 10.1038/nature12756
- 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
- Foley, S. W., Gosai, S. J., Wang, D., Selamoglu, N., Sollitti, A. C., Koster, T., et al. (2017). A global view of RNA-protein interactions identifies post-transcriptional regulators of root hair cell fate. *Dev. Cell* 41, 204.e5–220.e5. doi: 10.1016/j.devcel.2017.03.018
- Foley, S. W., and Gregory, B. D. (2016). Protein interaction profile sequencing (PIP-seq). *Curr. Protoc. Mol. Biol.* 116, 27.5.1–27.5.15. doi: 10.1002/cpmb.21
- Foley, S. W., Vandivier, L. E., Kuksa, P. P., and Gregory, B. D. (2015). Transcriptome-wide measurement of plant RNA secondary structure. *Curr. Opin. Plant Biol.* 27, 36–43. doi: 10.1016/j.pbi.2015.05.021
- Gosai, S. J., Foley, S. W., Wang, D., Silverman, I. M., Selamoglu, N., Nelson, A. D., et al. (2015). Global analysis of the RNA-protein interaction and RNA secondary structure landscapes of the *Arabidopsis* nucleus. *Mol. Cell* 57, 376–388. doi: 10.1016/j.molcel.2014.12.004
- Hammond, M. C., Wachter, A., and Breaker, R. R. (2009). A plant 5S ribosomal RNA mimic regulates alternative splicing of transcription factor IIIA pre-mRNAs. *Nat. Struct. Mol. Biol.* 16, 541–549. doi: 10.1038/nsmb.1588
- Harcourt, E. M., Kietrys, A. M., and Kool, E. T. (2017). Chemical and structural effects of base modifications in messenger RNA. *Nature* 541, 339–346. doi: 10.1038/nature21351
- Hawkes, E. J., Hennelly, S. P., Novikova, I. V., Irwin, J. A., Dean, C., and Sanbonmatsu, K. Y. (2016). COOLAIR antisense RNAs form evolutionarily conserved elaborate secondary structures. *Cell Rep.* 16, 3087–3096. doi: 10.1016/j.celrep.2016.08.045
- Herr, A. J., and Baulcombe, D. C. (2004). RNA silencing pathways in plants. *Cold Spring Harb. Symp. Quant. Biol.* 69, 363–370. doi: 10.1101/sqb.2004.69.363
- Incarnato, D., Neri, F., Anselmi, F., and Oliviero, S. (2014). Genome-wide profiling of mouse RNA secondary structures reveals key features of the mammalian transcriptome. *Genome Biol.* 15:491. doi: 10.1186/s13059-014-0491-2
- Kertesz, M., Wan, Y., Mazar, E., Rinn, J. L., Nutter, R. C., Chang, H. Y., et al. (2010). Genome-wide measurement of RNA secondary structure in yeast. *Nature* 467, 103–107. doi: 10.1038/nature09322
- Kim, W., Kim, H. E., Lee, A. R., Jun, A. R., Jung, M. G., Ahn, J. H., et al. (2017). Base-pair opening dynamics of primary miR156a using NMR elucidates structural determinants important for its processing level and leaf number phenotype in *Arabidopsis*. *Nucleic Acids Res.* 45, 875–885. doi: 10.1093/nar/gkw747
- Knapp, G. (1989). Enzymatic approaches to probing of RNA secondary and tertiary structure. *Methods Enzymol.* 180, 192–212. doi: 10.1016/0076-6879(89)80102-8
- Kwok, C. K. (2016). Dawn of the in vivo RNA structurome and interactome. *Biochem. Soc. Trans.* 44, 1395–1410. doi: 10.1042/BST20160075
- Kwok, C. K., Ding, Y., Shahid, S., Assmann, S. M., and Bevilacqua, P. C. (2015). A stable RNA G-quadruplex within the 5'-UTR of *Arabidopsis thaliana* ATR mRNA inhibits translation. *Biochem. J.* 467, 91–102. doi: 10.1042/BJ20141063
- Kwok, C. K., Ding, Y., Tang, Y., Assmann, S. M., and Bevilacqua, P. C. (2013). Determination of in vivo RNA structure in low-abundance transcripts. *Nat. Commun.* 4:2971. doi: 10.1038/ncomms3971
- Lengyel, J., Hnath, E., Storms, M., and Wohlfarth, T. (2014). Towards an integrative structural biology approach: combining Cryo-TEM, X-ray crystallography, and NMR. *J. Struct. Funct. Genomics* 15, 117–124. doi: 10.1007/s10969-014-9179-9
- Li, F., Zheng, Q., Ryykin, P., Dragomir, I., Desai, Y., Aiyer, S., et al. (2012a). Global analysis of RNA secondary structure in two metazoans. *Cell Rep.* 1, 69–82. doi: 10.1016/j.celrep.2011.10.002
- Li, F., Zheng, Q., Vandivier, L. E., Willmann, M. R., Chen, Y., and Gregory, B. D. (2012b). Regulatory impact of RNA secondary structure across the *Arabidopsis* transcriptome. *Plant Cell* 24, 4346–4359. doi: 10.1105/tpc.112.104232
- Loke, J. C., Stahlberg, E. A., Strenski, D. G., Haas, B. J., Wood, P. C., and Li, Q. Q. (2005). Compilation of mRNA polyadenylation signals in *Arabidopsis* revealed a new signal element and potential secondary structures. *Plant Physiol.* 138, 1457–1468. doi: 10.1104/pp.105.060541
- Lucks, J. B., Mortimer, S. A., Trapnell, C., Luo, S., Aviran, S., Schroth, G. P., et al. (2011). Multiplexed RNA structure characterization with selective 2'-hydroxyl acylation analyzed by primer extension sequencing (SHAPE-Seq). *Proc. Natl. Acad. Sci. U.S.A.* 108, 11063–11068. doi: 10.1073/pnas.1106501108
- Luo, G. Z., MacQueen, A., Zheng, G., Duan, H., Dore, L. C., Lu, Z., et al. (2014). Unique features of the m6A methylome in *Arabidopsis thaliana*. *Nat. Commun.* 5:5630. doi: 10.1038/ncomms5630
- Mortimer, S. A., and Weeks, K. M. (2007). A fast-acting reagent for accurate analysis of RNA secondary and tertiary structure by SHAPE chemistry. *J. Am. Chem. Soc.* 129, 4144–4145. doi: 10.1021/ja0704028
- Mullen, M. A., Olson, K. J., Dallaire, P., Major, F., Assmann, S. M., and Bevilacqua, P. C. (2010). RNA G-Quadruplexes in the model plant species *Arabidopsis thaliana*: prevalence and possible functional roles. *Nucleic Acids Res.* 38, 8149–8163. doi: 10.1093/nar/gkq804
- Noller, H. F., and Chaires, J. B. (1972). Functional modification of 16S ribosomal RNA by kethoxal. *Proc. Natl. Acad. Sci. U.S.A.* 69, 3115–3118. doi: 10.1073/pnas.69.11.3115
- Rouskin, S., Zubradt, M., Washietl, S., Kellis, M., and Weissman, J. S. (2014). Genome-wide probing of RNA structure reveals active unfolding of mRNA structures in vivo. *Nature* 505, 701–705. doi: 10.1038/nature12894
- Schleuning, M., Frund, J., Schweiger, O., Welk, E., Albrecht, J., Albrecht, M., et al. (2016). Ecological networks are more sensitive to plant than to

- animal extinction under climate change. *Nat. Commun.* 7:13965. doi: 10.1038/ncomms13965
- Siegfried, N. A., Busan, S., Rice, G. M., Nelson, J. A., and Weeks, K. M. (2014). RNA motif discovery by SHAPE and mutational profiling (SHAPE-MaP). *Nat. Methods* 11, 959–965. doi: 10.1038/nmeth.3029
- Smola, M. J., Christy, T. W., Inoue, K., Nicholson, C. O., Friedersdorf, M., Keene, J. D., et al. (2016). SHAPE reveals transcript-wide interactions, complex structural domains, and protein interactions across the Xist lncRNA in living cells. *Proc. Natl. Acad. Sci. U.S.A.* 113, 10322–10327. doi: 10.1073/pnas.1600008113
- Song, L., Axtell, M. J., and Fedoroff, N. V. (2010). RNA secondary structural determinants of miRNA precursor processing in *Arabidopsis*. *Curr. Biol.* 20, 37–41. doi: 10.1016/j.cub.2009.10.076
- Spitale, R. C., Crisalli, P., Flynn, R. A., Torre, E. A., Kool, E. T., and Chang, H. Y. (2013). RNA SHAPE analysis in living cells. *Nat. Chem. Biol.* 9, 18–20. doi: 10.1038/nchembio.1131
- Spitale, R. C., Flynn, R. A., Zhang, Q. C., Crisalli, P., Lee, B., Jung, J. W., et al. (2015). Structural imprints in vivo decode RNA regulatory mechanisms. *Nature* 519, 486–490. doi: 10.1038/nature14263
- Sun, L. Z., Zhang, D., and Chen, S. J. (2017). Theory and modeling of RNA structure and interactions with metal ions and small molecules. *Annu. Rev. Biophys.* 46, 227–246. doi: 10.1146/annurev-biophys-070816-033920
- Talkish, J., May, G., Lin, Y., Woolford, J. L. Jr., and McManus, C. J. (2014). Mod-seq: high-throughput sequencing for chemical probing of RNA structure. *RNA* 20, 713–720. doi: 10.1261/rna.042218.113
- Tan, Z. J., and Chen, S. J. (2011). Salt contribution to RNA tertiary structure folding stability. *Biophys. J.* 101, 176–187. doi: 10.1016/j.bpj.2011.05.050
- Underwood, J. G., Uzilov, A. V., Katzman, S., Onodera, C. S., Mainzer, J. E., Mathews, D. H., et al. (2010). FragSeq: transcriptome-wide RNA structure probing using high-throughput sequencing. *Nat. Methods* 7, 995–1001. doi: 10.1038/nmeth.1529
- Valencia-Sanchez, M. A., Liu, J., Hannon, G. J., and Parker, R. (2006). Control of translation and mRNA degradation by miRNAs and siRNAs. *Genes Dev.* 20, 515–524. doi: 10.1101/gad.1399806
- Vandivier, L. E., Anderson, S. J., Foley, S. W., and Gregory, B. D. (2016). The conservation and function of RNA secondary structure in plants. *Annu. Rev. Plant Biol.* 67, 463–488. doi: 10.1146/annurev-arplant-043015-111754
- Wachter, A., Tunc-Ozdemir, M., Grove, B. C., Green, P. J., Shintani, D. K., and Breaker, R. R. (2007). Riboswitch control of gene expression in plants by splicing and alternative 3' end processing of mRNAs. *Plant Cell* 19, 3437–3450. doi: 10.1105/tpc.107.053645
- Wan, Y., Qu, K., Ouyang, Z., Kertesz, M., Li, J., Tibshirani, R., et al. (2012). Genome-wide measurement of RNA folding energies. *Mol. Cell* 48, 169–181. doi: 10.1016/j.molcel.2012.08.008
- Wan, Y., Qu, K., Zhang, Q. C., Flynn, R. A., Manor, O., Ouyang, Z., et al. (2014). Landscape and variation of RNA secondary structure across the human transcriptome. *Nature* 505, 706–709. doi: 10.1038/nature12946
- Watts, J. M., Dang, K. K., Gorelick, R. J., Leonard, C. W., Bess, J. W. Jr., Swanson, R., et al. (2009). Architecture and secondary structure of an entire HIV-1 RNA genome. *Nature* 460, 711–716. doi: 10.1038/nature08237
- Zaug, A. J., and Cech, T. R. (1995). Analysis of the structure of *Tetrahymena* nuclear RNAs in vivo: telomerase RNA, the self-splicing rRNA intron, and U2 snRNA. *RNA* 1, 363–374.
- Zhang, W., Thieme, C. J., Kollwig, G., Apelt, F., Yang, L., Winter, N., et al. (2016). tRNA-related sequences trigger systemic mRNA transport in plants. *Plant Cell* 28, 1237–1249. doi: 10.1105/tpc.15.01056
- Zhao, B. S., Roundtree, I. A., and He, C. (2017). Post-transcriptional gene regulation by mRNA modifications. *Nat. Rev. Mol. Cell Biol.* 18, 31–42. doi: 10.1038/nrm.2016.132
- Zubradt, M., Gupta, P., Persad, S., Lambowitz, A. M., Weissman, J. S., and Rouskin, S. (2017). DMS-MaPseq for genome-wide or targeted RNA structure probing in vivo. *Nat. Methods* 14, 75–82. doi: 10.1038/nmeth.4057

**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 © 2018 Yang, Yang, Deng and Ding. 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 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 of U12 Intron in Proper Pre-mRNA Splicing of Plant Cap Binding Protein 20 Genes

Marcin Pieczynski<sup>1†</sup>, Katarzyna Kruska<sup>1†</sup>, Dawid Bielewicz<sup>1</sup>, Jakub Dolata<sup>1</sup>, Michał Szczesniak<sup>2</sup>, Wojciech Karłowski<sup>3</sup>, Artur Jarmolowski<sup>1\*</sup> and Zofia Szweykowska-Kulinska<sup>1\*</sup>

<sup>1</sup> Department of Gene Expression, Institute of Molecular Biology and Biotechnology, Faculty of Biology, Adam Mickiewicz University in Poznań, Poznań, Poland, <sup>2</sup> Department of Integrative Genomics, Institute of Anthropology, Faculty of Biology, Adam Mickiewicz University in Poznań, Poznań, Poland, <sup>3</sup> Department of Computational Biology, Institute of Molecular Biology and Biotechnology, Faculty of Biology, Adam Mickiewicz University in Poznań, Poznań, Poland

## OPEN ACCESS

### Edited by:

Dora Szakonyi,  
Instituto Gulbenkian de Ciência (IGC),  
Portugal

### Reviewed by:

Song Li,  
Duke University, United States  
Hunseung Kang,  
Chonnam National University,  
South Korea

### \*Correspondence:

Artur Jarmolowski  
artjarmo@amu.edu.pl  
Zofia Szweykowska-Kulinska  
zofszywey@amu.edu.pl

<sup>†</sup> These authors have contributed  
equally to this work.

### Specialty section:

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

**Received:** 12 January 2018

**Accepted:** 27 March 2018

**Published:** 16 April 2018

### Citation:

Pieczynski M, Kruska K,  
Bielewicz D, Dolata J, Szczesniak M,  
Karłowski W, Jarmolowski A and  
Szyweykowska-Kulinska Z (2018) A  
Role of U12 Intron in Proper  
Pre-mRNA Splicing of Plant Cap  
Binding Protein 20 Genes.  
Front. Plant Sci. 9:475.  
doi: 10.3389/fpls.2018.00475

The nuclear cap-binding complex (CBC) is composed of two cap-binding proteins: CBP20 and CBP80. The *CBP20* gene structure is highly conserved across land plant species. All studied *CBP20* genes contain eight exons and seven introns, with the fourth intron belonging to the U12 class. This highly conserved U12 intron always divides the plant *CBP20* gene into two parts: one part encodes the core domain containing the RNA binding domain (RBD), and the second part encodes the tail domain with a nuclear localization signal (NLS). In this study, we investigate the importance of the U12 intron in the *Arabidopsis thaliana* *CBP20* gene by moving it to different intron locations of the gene. Relocation of the U12 intron resulted in a significant decrease in the U12 intron splicing efficiency and the accumulation of wrongly processed transcripts. These results suggest that moving the U12 intron to any other position of the *A. thaliana* *CBP20* gene disturbs splicing, leading to substantial downregulation of the level of properly spliced mRNA and CBP20 protein. Moreover, the replacement of the U12 intron with a U2 intron leads to undesired alternative splicing events, indicating that the proper localization of the U12 intron in the *CBP20* gene secures correct *CBP20* pre-mRNA maturation and CBP20 protein levels in a plant. Surprisingly, our results also show that the efficiency of U12 splicing depends on intron length. In conclusion, our study emphasizes the importance of proper U12 intron localization in plant *CBP20* genes for correct pre-mRNA processing.

**Keywords:** U12 introns, U2 introns, mRNA splicing, CBP20, *Arabidopsis thaliana*

## INTRODUCTION

The cap-binding complex (CBC) is a nuclear heterodimer composed of two cap-binding proteins: cap-binding protein 20 (CBP20) and cap-binding protein 80 (CBP80, also known as ABA Hypersensitive 1, ABH1), which binds to the 5' cap of all RNA polymerase II transcripts. In plants, CBC is important for pre-mRNA and pri-miRNA first intron splicing and regulation of pre-mRNA alternative splicing (Szarzynska et al., 2009; Raczynska et al., 2010). The CBC is also involved in miRNA biogenesis (Kim et al., 2008; Laubinger et al., 2008; Ren and Yu, 2012). Amino acid sequence comparison of *Arabidopsis thaliana* and *Oryza sativa* CBP20 and CBP80 proteins



revealed high sequence conservation of the small subunit of CBC (CBP20) and a considerably lower level of conservation of its large subunit (CBP80) (Kmieciak et al., 2002). Interestingly, in comparison to animal CBP20 proteins, plant CBP20s contain an additional carboxy-terminal fragment. The structure of the CBP20 protein in higher plants can be divided into two parts. The core part, built mostly of 138 amino acids, contains the conservative RNA binding domain (RBD), which plays a crucial role in cap structure recognition and binding (Izaurre et al., 1994), and the tail part, built of approximately 120 amino acids, contains a nuclear localization signal (NLS). In animals, this is the CBP80 protein that carries a functional NLS and is responsible for the import of CBC into the nucleus (Dias et al., 2009). In *A. thaliana*, two functionally independent NLSs are located in the tail part of the CBP20 protein (Kmieciak et al., 2002). In contrast, plants CBP80s do not contain any NLS; thus, the CBP20 protein targets the whole plant CBC to the nucleus. Moreover, in *A. thaliana*, the CBP20 protein is stabilized by CBP80 (Kierzkowski et al., 2009). The down-regulation of nuclear cap-binding proteins in *Arabidopsis* leads to mild developmental abnormalities, such as serrated rosette leaves and delayed development. Interestingly, loss of CBC functions in *Arabidopsis* plants confers hypersensitivity to abscisic acid (ABA) during seeds germination, significant reduction of stomatal conductance, and in consequence greatly enhances tolerance of the *cbp20* and *cbp80* mutants to drought (Hugouvieux et al., 2001; Papp et al., 2004; Jäger et al., 2011; Pieczynski et al., 2013) and salinity (Kong et al., 2014).

*Arabidopsis* and rice *CBP20* gene structures are conserved and contain eight exons. The core part of the CBP20 protein is encoded by the first four exons of the gene, while the tail part is encoded by the last four exons. Six out of seven introns of the *CBP20* gene belong to the classical and abundant U2 introns, while intron no. 4 represents the much rarer U12 introns. Computational analyses allowed the identification in the *A. thaliana* genome 246 U12 introns constituting 0.17% of all predicted introns in this species (Alioto, 2007). Transcriptome analyses, however, identified about eight times more U12 introns (2069 vs. 246) than previously found using the computational approach (Marquez et al., 2012). The majority of U12 introns (89.4%) contain GT-AG terminal dinucleotides; however, a small portion (10.6%) is characterized by the presence of other terminal dinucleotides, of which AT-AC comprises almost half of such non-GT-AG introns (4.8%) (Marquez et al., 2012). U12 introns contain a very characteristic and highly conserved branch point region (TTCCTTRAY), and unlike U2 introns, they do not have any polypyrimidine tract (Lewandowska et al., 2004; Simpson and Brown, 2008; Marquez et al., 2012; Turunen et al., 2013). It has been shown that several proteins of U11/U12 di-snRNP are indispensable for the correct splicing of the U12 intron-containing genes, which is crucial for the normal development of *A. thaliana* (Kim et al., 2010; Kwak et al., 2012; Jung and Kang, 2014; Xu et al., 2016). Moreover, the *Arabidopsis* *quatre-quart1* (*QQT1*) gene is an indispensable U12 intron-containing gene whose correct splicing is necessary for the wild type phenotype and development of *Arabidopsis* plants (Kwak et al., 2017).

The phylogenetic distribution of U12 introns shows that the minor (U12) splicing pathway appeared very early in eukaryotic evolution, but during the course of evolution, most U12 introns systematically changed to U2 introns (Bartschat and Samuelsson, 2010; Lin et al., 2010). Despite this process, a few U12 introns were retained in selected genes and remain very stable in some taxa (Basu et al., 2008). *CBP20* genes from both *Arabidopsis* and rice were identified as belonging to this small number of genes containing U12 introns with AT-AC terminal dinucleotides. Moreover, the U12 intron in *CBP20* gene in both plant species is located between exons no. 4 and no. 5, splitting the *CBP20* coding sequence into the core and tail CBP20 protein parts in both species studied. Surprisingly, the exons encoding the core part of the protein have identical lengths (18t, 224, 139, and 34 nt), but the exons encoding the tail part of CBP20s differ considerably in length.

In this paper, we show that the *CBP20* gene structure is conserved across land plant species from liverworts to higher plants. All studied *CBP20* genes contain eight exons and seven introns, with the fourth intron belonging to the U12 class. In addition, the length of the first four exons is conserved, while the length of the exons encoding the CBP20 tail part varies considerably in all plant species studied. Furthermore, we show that the *CBP20* U12 introns in plants may be as short as 76 nt and as long as 2733 nt. Our experiments demonstrate that the efficiency of U12 splicing depends on intron length. The experiments carried out to explain the conserved localization of the *CBP20* U12 intron show that the exchange of the *CBP20* U12 intron with the U2 intron leads to undesired alternative splicing events and that the proper localization of the U12 intron in the *CBP20* gene secures correct *CBP20* pre-mRNA maturation and CBP20 protein levels in a plant.

## MATERIALS AND METHODS

### Plant Material and Growth Conditions

The experiments were performed using *A. thaliana* (L.) Columbia-0 wt plants (Lehle Seeds, Round Rock, TX, United States) and a homozygous T-DNA insertion line *cbp20* (Papp et al., 2004). *Solanum tuberosum* ssp. *tuberosum* var. Sante, *Nicotiana tabacum* var. Xanthi, *Hordeum vulgare* var. Sebastian and liverwort *Pellia endiviifolia* sp B were used for the *CBP20* gene sequencing.

*Arabidopsis* plants were grown in 'Jiffy-7 42mm' soil (Jiffy Products International BV, Moerdijk, Nederland) in an MLR-350H Versatile Environmental Test Chamber (Sanyo, Loughborough, Leicestershire, United Kingdom) with a 16 h light/8 h dark photoperiod (approx. 80  $\mu\text{mol m}^{-2}\text{s}^{-1}$ ), constant temperature of 22°C and humidity of 70%. Potato plants were grown in sterile conditions in a greenhouse (22°C with constant light, approximately 80  $\mu\text{mol m}^{-2}\text{s}^{-1}$ ) on  $1/2$  Murashige-Skoog medium pH 5.5–5.6. *N. tabacum* and *H. vulgare* plants were grown in a greenhouse (22°C with 12 h light/12 h dark photoperiod, approx. 120  $\mu\text{mol m}^{-2}\text{s}^{-1}$ ) on soil irrigated with mineral nutrients. *P. endiviifolia* sp B was grown as described by Sierocka et al. (2011).

*Arabidopsis thaliana* transformation was performed using the floral-dip method according to Clough and Bent (1998).

## DNA and RNA Isolation

Total genomic DNA, RNA and plasmid DNA were isolated using a DNeasy Plant Mini Kit (Qiagen), RNeasy Plant Mini Kit (Qiagen), and QIAprep Spin Miniprep Kit, respectively, according to protocols supplied by the manufacturers. Purity and amounts of DNA and RNA were determined using NanoDrop Spectrophotometer (Thermo Scientific).

## DNA Sequencing

DNA sequencing was performed with a BigDye v3.1 sequencing kit (Applied Biosystems, Foster City, CA, United States) on a ABI Prism 3130XL Analyzer (Applied Biosystems) in the Molecular Biology Techniques Laboratory, Faculty of Biology, Adam Mickiewicz University in Poznan, Poland.

## cDNA Synthesis, PCR and DNA Cloning

Four micrograms of total RNA was reverse-transcribed using Superscript III Reverse Transcriptase (Invitrogen) and oligo(dT)<sub>15</sub> primer (Novazym). cDNA of *CBP20* genes from potato, tobacco, barley and liverwort was amplified by PCR using primers designed according to ESTs from NCBI database (Benson et al., 2012). In the case of potato, we designed primers according to the EST 706129 sequence. Tobacco *CBP20* gene cDNA was amplified using primers designed according to the contig sequence assembled on the basis of nine different EST sequences (accession numbers: AM815125, CV017334, AM818189, EB679251, EB440185, AM827799, AM808999, EB444475, and EB678046). Barley *CBP20* cDNA was assembled on the basis of the following EST sequences: TC111197 and TC78828 from TIGR database and BU991417, BJ480285, AJ463125, and AJ475973 from NCBI. Primers for cDNA *CBP20* amplification from *P. endiviifolia* were designed according to the results of whole transcriptome sequencing.

PCR was carried out as described in Szarzynska et al. (2009) and Sierocka et al. (2011). Genomic sequence of the *CBP20* gene from potato and tobacco was amplified using the Expand Long Template PCR System (Roche). PCR products were separated on a 1% agarose gel, purified using a QIAquick PCR Purification Kit (Qiagen), cloned into the pGEM-T Easy vector (Promega) and sequenced. Primers used for amplification of genomic and cDNA sequences of the *CBP20* gene from different plant species are shown in Supplementary Table S1. Amplification of *CBP20* cDNAs as well as genes from potato, tobacco, barley and liverwort is shown in Supplementary Figure S1.

## Expression Construct Preparation

Mini- and midi-gene constructs numbers 1, 8, and 9 were prepared by PCR amplification of appropriate *A. thaliana* and *Physcomitrella patens* *CBP20* gene fragments. Mini- and midi-gene constructs numbers 2–7 were prepared as presented in Supplementary Figure S2. Each construct contains exons no. 4 and no. 5 from the *A. thaliana* *CBP20* gene. Depending on the construct, these two exons are separated by a U12 or U2 intron

originating from different plant species. Mini-gene constructs were prepared in 3- or 4-step PCR. In PCR-1, the U2 or U12 intron was amplified using a forward primer containing on its 5' end 20 nucleotides complementary to exon no. 4. The reverse primer used in PCR-1 contained at its 5' end 20 nucleotides complementary to exon no. 5. In PCR-2, the sequence of exon no. 5 was amplified using a forward primer complementary to the last 20 nucleotides of the U2 or U12 intron, depending on the individual mini-gene construct. The final mini-gene sequence was amplified in PCR-3, in which PCR-1 and PCR-2 products were used as templates. In mini-gene constructs containing the U12 intron derived from the *S. tuberosum* or *Vitis vinifera* *CBP20* gene, an additional PCR-0 was performed in which the U12 intron sequence was amplified.

Each mini-gene construct was cloned into the pDH515 vector at a unique BamHI restriction site within an intronless zein gene encoding a maize seed storage protein. Additionally, the vector contains the 35S promoter of *Cauliflower mosaic virus* (CaMV 35S) and terminator sequence (Lewandowska et al., 2004).

The maxi-gene wt transgene construct was prepared by PCR amplification of the *Arabidopsis* *CBP20* gene sequence encompassing its native promoter (1228 bp) (Kmieciak et al., 2002). The PCR product was cloned into the NotI restriction site in the pENTR vector. The other five maxi-gene constructs were obtained performing a series of PCRs using the wt transgene construct or *A. thaliana* genomic DNA as a template. Specific sub-fragments of each maxi-gene construct were obtained in the PCR, followed by the consecutive cohesive ends annealing and final PCR amplification (Supplementary Figure S3). The individual maxi-gene constructs were transferred from the pENTR cloning vector to the pEarleyGate302 expression vector using Gateway LR Clonase II Enzyme Mix (Thermo Fisher Scientific). All constructs were confirmed by DNA sequencing. Primers used for construct preparation are listed in Supplementary Table S1.

## Protoplast Isolation, Transfection and Splicing Analysis

Splicing of mini- and midi-gene constructs was analyzed in protoplasts isolated from leaves of *N. tabacum* var. Xanthi (Lewandowska et al., 2004). Protoplast suspension was transfected with each mini- or midi-gene construct as described by Lewandowska et al. (2004). After overnight incubation, total RNA from protoplasts was isolated. cDNA template was synthesized using the zein3'-R reverse primer complementary to the zein sequence flanking mini- and midi-gene constructs in the pDH515 vector (Simpson and Filipowicz, 1996). Splicing analyses were carried out by RT-PCR using following primers: zeinF-FAM labeled with fluorescent phosphoramidite 6-FAM at 5' end and zeinR (see Supplementary Table S1).

6-FAM-labeled RT-PCR products were quantitatively analyzed using capillary electrophoresis on an ABI 3130xl Genetic Analyzer. The length of the labeled products was calculated using Peak Scanner Software v1.0 (Applied Biosystems) by comparison with the GeneScan-350 TAMRA

size standards (Applied Biosystems). Quantification of RT-PCR products was carried out after 23rd amplification cycle by measurement of the fluorescent peak areas of the detected fragments. RT-PCR was carried out in three biological replicates. The same RT-PCR products were separated in parallel on a 1% agarose gel. 6-FAM-labeled fragments were eluted from the gel, cloned into the pGEM T-Easy (Promega) vector and sequenced to confirm specific splicing products.

Splicing analysis of maxi-gene transcripts that were introduced into the *A. thaliana cbp20* mutant was also carried out using capillary electrophoresis, as described above. Primers used for RT-PCR amplification hybridized to the sequence of exons no. 3 and no. 6 of the *CBP20* gene transcript (Supplementary Table S1). The forward primer Splice-FAM-F was labeled at its 5' end with the fluorescent phosphoramidite 6-FAM.

## RT-qPCR

Real-time RT-qPCR was performed with Power SYBR® Green PCR Master MIX (Applied Biosystems, Warrington, United Kingdom) on a 7900HT Fast Real-Time PCR System (Applied Biosystems) in 10-μl reaction volumes in 384-well plates. To assess the splicing efficiency of mini-genes, an absolute quantification approach was applied. Standard curves were prepared using plasmids containing individual mini-genes. Splicing efficiency was calculated by comparison of spliced and unspliced transcript levels to the total transcript amount. To estimate the total level of the *CBP20* transcript in maxi-gene transgenic lines, the fold change was calculated using the 2- $\Delta\Delta C_t$  method. The mRNA fragment of elongation factor 1- $\alpha$  (EF1- $\alpha$ , TAIR locus: At1g07930) was amplified as a reference gene in *A. thaliana*.

## Western Blot

Protein extracts were separated by 10% SDS-PAGE, transferred to a polyvinylidene fluoride membrane (PVDF; Immobilon®-P, Millipore), and analyzed by western blot using antibodies at the indicated dilutions: anti-Actin (691001; MP Biomedicals) at 1:5000 and anti-CBP20 (AS09 530; Agrisera) at 1:1000.

## Bioinformatics Tools

Comparison between DNA and protein sequences was carried out using bioinformatics tools: ClustalW<sup>1</sup> and BoxShade<sup>2</sup>. Sequencing results were assembled together using the ContigExpress program from Vector NTI (Invitrogen).

## Sequence Accession Numbers

*Arabidopsis thaliana*, Gene ID: 834443; *Populus trichocarpa*, Gene ID: 7469563; *V. vinifera*, Gene ID: 100265980; *S. tuberosum*, GU046516.1; *N. tabacum*, GU058037.1; *O. sativa*, Gene ID: 4329963; *H. vulgare*, FJ548567.1; *S. moellendorffii* – *Selaginella*

*moellendorffii* v1.0, Scaffold 3333196:1<sup>3</sup>; *P. patens* – *P. patens* V1.2\_genome scaffold\_105<sup>4</sup>.

## RESULTS

### Plant *CBP20* Gene Structure Is Evolutionarily Conserved and Reveals Preserved Localization of a U12 Intron

The analysis of *CBP20* genes from *A. thaliana* and *O. sativa* revealed high sequence and structural conservation. To gain deeper insight into the evolutionary conservation of the *CBP20* gene, we decided to compare *CBP20* sequences from various plant species across the plant kingdom. The *CBP20* gene and cDNA sequences from *P. trichocarpa*, *V. vinifera*, *Selaginella moellendorffii*, and *P. patens* were found in publicly available databases (Figure 1). In addition, full-length *CBP20* cDNA sequences from *S. tuberosum* var. Sante, *N. tabacum* var. Xanthi, *H. vulgare* var. Sebastian and the liverwort *P. endiviifolia* subspecies B were established in our laboratory. Based on the obtained *CBP20* cDNA sequences, we determined the full-length *CBP20* gene sequences in potato, tobacco, barley and liverwort genomes.

We compared the *CBP20* gene structure from ten plant species, including representatives of Bryophyta (liverworts, mosses), Leucophyta (lycophyte), and Spermatophyta (monocots and dicots) (Figure 1A). In all species analyzed, the *CBP20* genes contain 8 exons and 7 introns. Interestingly, their first four exons, coding for the core part of the protein, are of the same length with one exception, *S. moellendorffii*, whose second exon is three nucleotides shorter than that in the other *CBP20* genes studied. In all *CBP20* genes, a U12 intron was found separating the evolutionarily conserved core part that encodes a canonical RBD from the carboxy-terminal fragment (tail) that encodes the NLS and generally exhibits a much lower degree of evolutionary conservation. All these U12 introns contain the canonical AT-AC terminal dinucleotides and a branch point sequence typical of U12 introns (Figure 1B). The *CBP20* U12 intron length varies from 76 nt in *S. moellendorffii* to 2733 nt in *V. vinifera*. In general, plant species can be divided into two classes: those carrying short U12 introns (from 76 to 284 nt) and those carrying long U12 introns (from 1619 to 2733 nt) (Figure 1A).

There are considerable differences between different plant *CBP20* gene lengths. The *S. moellendorffii* *CBP20* gene (1245 bp) is 10 times shorter than that from *S. tuberosum* (11430 bp). In the potato *CBP20* gene, the fifth intron is extremely long, approximately 6050 bp, which largely accounts for the unusual length of this gene. Computational analysis of this intron sequence has revealed that it contains characteristic signatures of the LINE1 retrotransposon. Since we could not identify any full open reading frame for reverse transcriptase, we assumed that this retroelement is defective.

Comparison of all plant *CBP20* amino acid sequences revealed the presence of conserved residues within the core domain that

<sup>1</sup> www.ebi.ac.uk

<sup>2</sup> www.ch.embnet.org

<sup>3</sup> https://genome.jgi.doe.gov

<sup>4</sup> www.cosmos.org



A

Plant species	E1	i1	E2	i2	E3	i3	E4	i4	E5	i5	E6	i6	E7	i7	E8
<i>A. thaliana</i>	18	206	224	84	139	88	34	134	111	335	66	88	107	105	75
<i>P. trichocarpa</i>	18	95	224	635	139	87	34	197	114	1006	54	170	110	278	75
<i>V. vinifera</i>	18	836	224	2200	139	88	34	2733	111	3696	54	138	107	892	75
<i>S. tuberosum</i>	18	290	224	2430	139	73	34	1619	114	~6050	57	115	107	75	75
<i>N. tabacum</i>	18	164	224	2813	139	78	34	1764	112	968	57	116	107	293	93
<i>O. sativa</i>	18	106	224	1017	139	88	34	239	99	507	54	77	86	95	78
<i>H. vulgare</i>	18	82	224	2007	139	87	34	224	99	809	51	80	83	82	78
<i>S. moellendorffii</i>	18	55	221	70	139	58	34	76	114	82	23	62	74	80	139
<i>P. patens</i>	18	870	224	472	139	115	34	153	138	173	42	262	77	187	74
<i>P. endiviifolia</i>	18	566	224	503	139	431	34	284	126	163	66	261	80	194	74

CORE

U12

TAIL

B

<i>P.patens</i>	ATATCCTT-----TT	CCTTAATTTAAACAATTGCAC	
<i>P.endiviifolia</i>	ATATCCTC-----TT	CCTTGACCAAATGC....CAC	
<i>P.trichocarpa</i>	ATATCCTT-----TT	CCTTAACCTGTGC.....TAC	
<i>S.tuberosum</i>	ATATCCTT-----TA	CCTTAACCTAGCAA.....TAC	
<i>N.tabacum</i>	ATATCCTT-----TA	CCTTAACCTAGTTA.....TAC	
<i>V.vinifera</i>	ATATCCTT-----TT	CCTTAATGTTTA.....TAC	
<i>O.sativa</i>	ATATCCTT-----TT	CCTTAATGAATCT.....TAC	
<i>H.vulgare</i>	ATACCCTT-----TT	CCTTAATGAATCT.....AAC	
<i>A.thaliana</i>	ATATCCTT-----TT	CCTTAACCTCATTGAT...TAC	
consensus sequence	RTATCCTY	TT.CCTTRAY	YAC
	5' splice site	branchpoint 10-20nt	3' splice site

**FIGURE 1 |** *CBP20* genes derived from different plant species present very similar gene structures and conserved localization of U12 introns. **(A)** Comparison of *CBP20* gene structures. Green color depicts exons, yellow color marks U2 introns, red color indicates U12 introns, and orange color depicts the second exon of the *CBP20* gene from *S. moellendorffii*, which is shorter than other exons in the same position in other plants studied by 3 nucleotides. Note that with the exception of *S. moellendorffii*, the exon lengths are the same in all plant species studied until exon no. 5 is reached, which is separated from the protein core-encoding part of the gene by the U12 intron. E – exon, i – intron. **(B)** A comparison of consensus elements of *CBP20* U12 introns derived from the plant species studied (Simpson and Brown, 2008). Dashed lines represent individual nucleotide sequences between the 5' splice site and branch point. Dots mark alignment gaps. Nucleotides that differ from the consensus sequence in the branch point region are marked in black. The conserved adenine nucleotides in the branch point are marked in red.

were reported to be responsible for binding to the cap structure and for recognition of RNA (Supplementary Figure S4) (Mazza et al., 2002). In the tail fragment of all *CBP20* proteins, we identified potential NLS sequences. In the majority of plant species, the *CBP20* gene contains two NLS motives, as was shown before for *A. thaliana*, while in *P. trichocarpa* and *P. patens*, we identified only one motif: the proximal NLS. Generally, sequence homology of full-length *CBP20* proteins in plants ranges from 61 to 79% identity, and that of the highly conserved core part ranges from 80 to 89% identity. The homology within the tail part is rather low (39–69% identity), but the distal part of this tail region shows again a relatively high degree of conservation, suggesting that this part of the protein may play a structural or functional role.

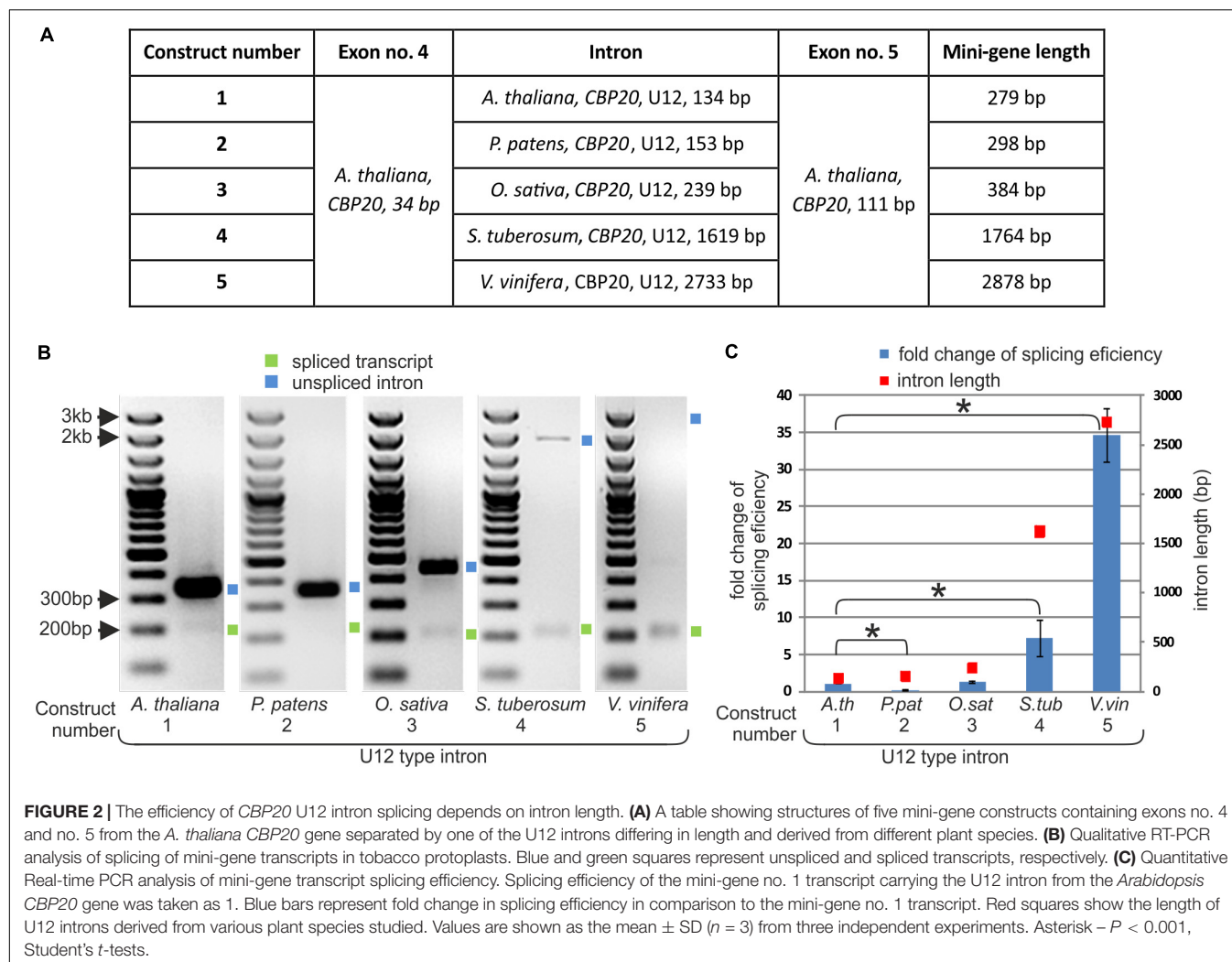
It was shown that U12 introns over the course of evolution were usually exchanged with U2 introns (Burge et al., 1998;

Bartschat and Samuelsson, 2010). The evolutionarily conserved localization of the plant *CBP20* U12 intron separating the conserved core domain from the tail part of the protein in all *CBP20* plant genes characterized encouraged us to investigate the role of this intron in *CBP20* pre-mRNA maturation.

## Efficiency of U12 Splicing Depends on the Intron Length

As we have shown, the length of *CBP20* U12 introns in the studied plant species varies considerably. To test whether U12 intron length may influence U12 intron splicing efficiency, we prepared five mini-gene constructs. These mini-genes consist of *A. thaliana* *CBP20* exon no. 4 (34 bp) and exon no. 5 (111 bp) separated by the following: construct no. 1 – the original *A. thaliana* *CBP20* U12 intron (134 bp), construct no. 2 – the *P. patens* *CBP20* U12 intron (153 bp), construct





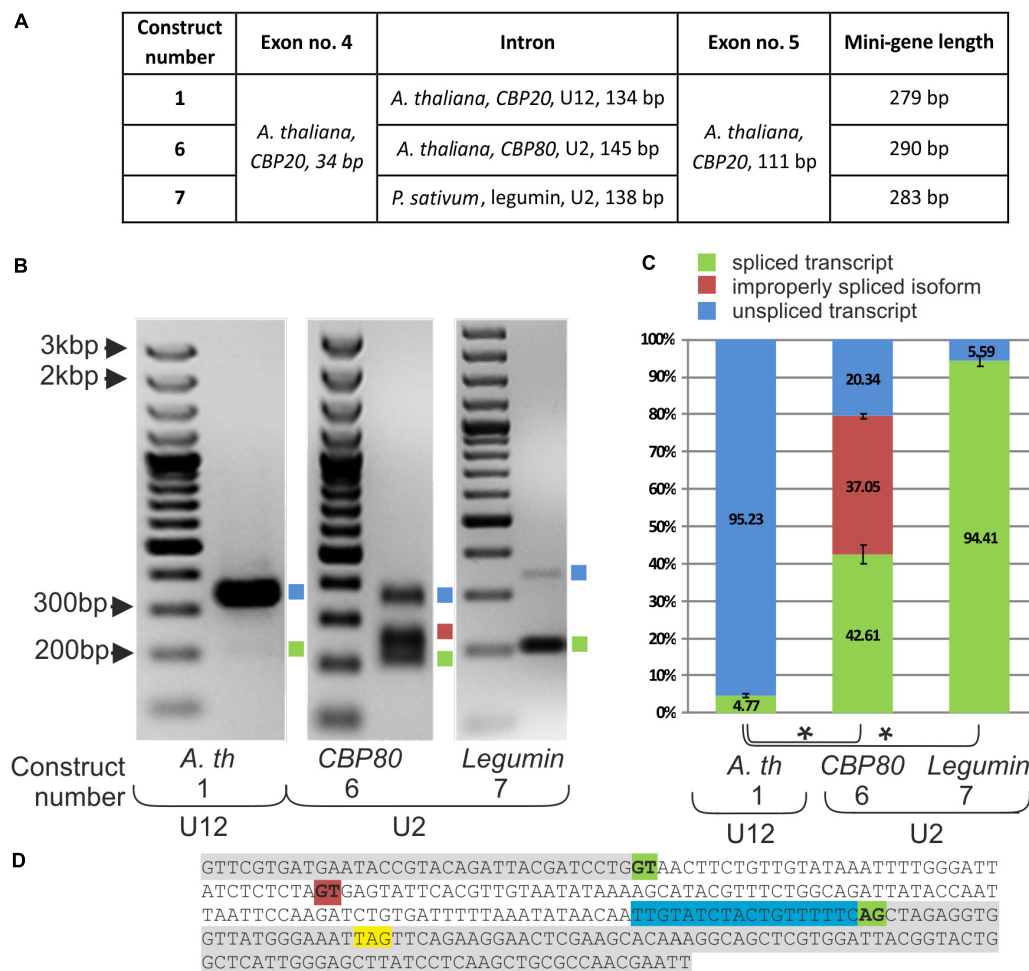
no. 3 – the *O. sativa* *CBP20* U12 intron (239 bp), construct no. 4 – the *S. tuberosum* *CBP20* U12 intron (1619 bp) and construct no. 5 – the *V. vinifera* *CBP20* U12 intron (2733 bp) (Figure 2A). The mini-genes were cloned into the pDH515 expression vector (Lewandowska et al., 2004) and sequenced. These recombinant plasmids were then used for tobacco mesophyll protoplasts transfection. Total RNA was isolated from the transfected protoplasts incubated overnight, and cDNA derived from the mini-gene transcripts was prepared. Splicing analyses were carried out applying RT-PCR with fluorescent-labeled primers. The splicing efficiency of each construct was determined as the mean of three independent experiments. All U12 introns were spliced correctly; no alternative splicing events were observed (Figure 2B). However, the splicing efficiency of different mini-genes varied considerably. We found a correlation between the U12 intron length and splicing efficiency as revealed by RT-qPCR using the absolute quantification method (Figure 2C and Supplementary Table S2). The splicing efficiency of the 134 nt long U12 intron from *A. thaliana* was assumed in this comparison to be 1. The U12 introns derived from *O. sativa*, *S. tuberosum* and

*V. vinifera* *CBP20* genes were spliced 1.28, 7.17, and 34.65 times more efficiently than that from *A. thaliana*, respectively (Figure 2C).

An exception to this rule was observed in the splicing efficiency of the U12 intron derived from *P. patens*. Despite its almost identical length with the U12 intron from *A. thaliana*, the splicing efficiency of this U12 intron was five times lower.

## The Replacement of a U12 Intron With a U2 Intron in the *CBP20* Mini-Gene Improves Splicing Efficiency but Leads to Deleterious and Undesired Improper Splicing Events

To learn more about the role of the U12 intron in *CBP20* plant genes, we constructed a series of mini-genes containing *A. thaliana* exons no. 4 and no. 5 separated by the original *CBP20* U12 intron (134 bp – construct no. 1), the U2 intron derived from the *A. thaliana* *CBP80* gene (intron no. 11, 146 bp – construct no. 6) and the U2 intron derived from the *P. sativum* legumin gene (intron no. 1, 138 bp – construct no. 7) (Figure 3A). Both



**FIGURE 3 |** Replacement of the original *CBP20* U12 intron by a U2 intron may lead to deleterious splicing events. **(A)** A table showing structures of three mini-gene constructs containing exons no. 4 and no. 5 from the *A. thaliana* *CBP20* gene separated by the original U12 intron (construct no. 1), by the U2 intron from the *Arabidopsis* *CBP80* gene (construct no. 6), and by the U2 intron from the *P. sativum* legumin gene (construct no. 7). **(B)** Qualitative RT-PCR analysis of splicing of mini-gene transcripts in tobacco protoplasts. Blue and green squares represent unspliced and spliced transcripts, respectively. The red square represents the alternatively spliced *CBP80* U2 intron. **(C)** Quantitative analysis of mini-gene transcripts RT-PCR products using capillary electrophoresis. Splicing efficiency is calculated as a percentage of the sum of all detectable products. Values are shown as the mean  $\pm$  SD ( $n = 3$ ) from three independent experiments. **(D)** Nucleotide sequence of the mini-gene no. 6 containing the *Arabidopsis* *CBP80* gene U2 intron. Green color depicts constitutive U2 5' and 3' splice site dinucleotides, red color represents an alternative 5' splice site, the STOP codon is marked in yellow, and blue color stands for the polypyrimidine tract. Asterisk –  $P < 0.005$ , Student's  $t$ -tests.

U2 introns were chosen because of their similar length to that of the original *A. thaliana* *CBP20* U12 intron. The selected U2 introns have canonical GT-AG splice site dinucleotides and a polypyrimidine tract near the 3' splice site and show a low GC content, 26% and 29% for the pea legumin and *Arabidopsis* *CBP80* genes, respectively. U12 intron from the *A. thaliana* *CBP20* gene has also a low GG content – 36%.

Splicing of the mini-gene transcripts was studied in the tobacco protoplast system, as described in the case of mini-genes containing various U12 introns. The original U12 intron was spliced with only 4.77% efficiency, whereas both studied U2 introns, from the *Arabidopsis* *CBP80* gene and pea legumin gene, were spliced, reaching 79.66 and 94.4% efficiency, respectively (Figure 3B). These results are supported by

data that have been previously obtained in our laboratory (Lewandowska et al., 2004), showing that *A. thaliana* U12 introns are spliced less efficiently than are U2 introns in tobacco protoplasts. However, in the mini-gene containing the U2 intron derived from the *A. thaliana* *CBP80* gene, an alternative splicing event was observed: two alternative 5' splice sites were recognized by the tobacco U2 splicing machinery, leading to the proper splicing of the *CBP20* mini-gene transcript (42.61%) or to the inclusion of an additional 38 nt from the 5' intron end into the spliced mRNA (37.05%) (Figures 3C,D). The 38 nt long insertion introduces a premature stop codon resulting in premature translation termination. To test whether this alternative splicing event also occurs naturally in the *CBP80* gene, we analyzed the splicing of this intron in the *CBP80* transcript.

We were able to detect only constitutively, properly spliced *CBP80* mRNA, without any additional alternative splicing events (Supplementary Figure S5).

## ***Arabidopsis* CBP20 U12 Intron Surrounded by Additional Exons and U2 Introns Is Not Efficiently Recognized by the U12 Splicing Machinery**

It has been shown that the extension of mini-genes composed of two exons from *GSH2* interrupted by a U12 intron with neighboring exons and U2 introns improves U12 splicing efficiency (Lewandowska et al., 2004). To test whether this is also true for other genes containing U12 introns, we prepared two midi-genes that were based on *A. thaliana* and *P. patens* *CBP20* genes. *A. thaliana* (construct no. 8) and *P. patens* (construct no. 9) midi-genes consist of E3-i3(U2)-E4-i4(U12)-E5-i5(U2)-E6 from the *A. thaliana* and *P. patens* *CBP20* genes, respectively (Figure 4A). The midi-genes were studied in tobacco protoplasts as described before in the case of mini-genes.

The splicing patterns of *A. thaliana* and *P. patens* midi-transcripts were similar (Figure 4B). We obtained one main and many additional minor products. Sequencing of these products has shown that the main band (265 bp in *A. thaliana* and 241 bp in *P. patens*) represents a spliced product consisting of exons no. 3 and no. 6. Thus, splicing in that case leads to the exon skipping event by the removal of all three introns and exons no. 4 and no. 5. The remaining minor bands represent various spliced products listed in Figure 4C, among which a fully and properly spliced transcript containing all four exons (E3-E4-E5-E6) was also detected. Quantitative analysis shows that the main improperly spliced product was present in approximately 50% of all other differentially spliced RNA molecules, in both *Arabidopsis* and *Physcomitrella* midi-genes (Figures 4C,D). The fully spliced transcript containing all four exons (E3-E4-E5-E6) was present only in approximately 12.47% of all splicing products in *Arabidopsis* and 3.75% in *Physcomitrella*. Thus, in contrast to the *GSH2* midi-gene, the neighboring U2 introns and exons do not enhance the recognition of the U12 intron in the case of *Arabidopsis* and moss *CBP20* midi-genes. Instead, the U12 intron signals are mainly not recognized properly in these constructs.

## **The Proper Localization of the U12 Intron in the *CBP20* Gene Secures Correct Pre-mRNA Maturation**

To study the influence of the U12 intron on the splicing efficiency of the full *CBP20* pre-mRNA in plants, we prepared constructs composed of the natural *A. thaliana* *CBP20* gene promoter and the whole wild-type (wt) gene sequence or its variants (the maxi-genes series) (Figure 5B). In addition to the wt maxi-gene (wt transgene), we constructed *CBP20* gene variants in which (i) the U12 intron was replaced with the U2 intron derived from the *A. thaliana* *CBP80* gene (intron no. 11; U12→U2), (ii) the U12 intron was removed ( $\Delta$ U12), and (iii) exons no. 4 and

no. 5 were swapped (exon swap). Additionally, we prepared two constructs based on the *CBP20* gene variant U12→U2. In the first construct, the U12 intron was moved upstream in the *CBP20* gene and replaced the U2 intron no. 3 (U12 in core); by this, we moved the U12 intron into the core-encoding part of the *CBP20* gene. In the second construct, the U12 intron replaced the original U2 intron localized between exons no. 5 and no. 6 (U12 in tail); in this construct, the U12 intron was moved into the tail-encoding part of the *CBP20* gene. *A. thaliana* *cbp20* mutant plants were transformed with these constructs, and three independent homozygous lines of each gene variant were further tested.

Qualitative RT-PCR-based analyses of the *CBP20* mRNA levels revealed the presence of one dominant product representing fully spliced transcripts in the case of wt plants as well as transgenic plants carrying the wt *CBP20* gene construct (wt transgene) and the U12→U2 construct. A small amount of partially unspliced products was also detected (Figure 5A, upper panel). In transgenic plants carrying mutated *CBP20* genes in which the U12 intron was deleted ( $\Delta$ U12), we also observed the presence of fully and properly spliced *CBP20* mRNA. Moreover, the transgenic plants exhibited a wt phenotype, indicating that the *CBP20* protein is produced in the transgenic lines analyzed (Figure 5C, upper panel).

Only transgenic plants carrying the *CBP20* gene with swapped exons (exon swap) gave mRNAs in which the *CBP20* coding sequence was disrupted. This impaired mRNA accumulated at a lower level than in wt and wt transgene plants containing a wt copy of the *CBP20* gene (Figure 5A, upper panel). The exon swapping mRNA contains a premature stop codon (in exon no. 5) that might cause the synthesis of putative shorter proteins with a disturbed core fragment. As expected, this mutant plants' phenotype is very similar to that of the null *cbp20* mutant, exhibiting serrated rosette leaves and growth retardation (Figure 5C, lower panel), confirming that *CBP20* is indeed not produced in these plants. In the case of mutants in which the U12 intron was moved into the core- or tail-encoding parts of the gene, additional splice products were observed (Figure 5A, upper panel; the lower panel represents a zoomed in part of the upper one).

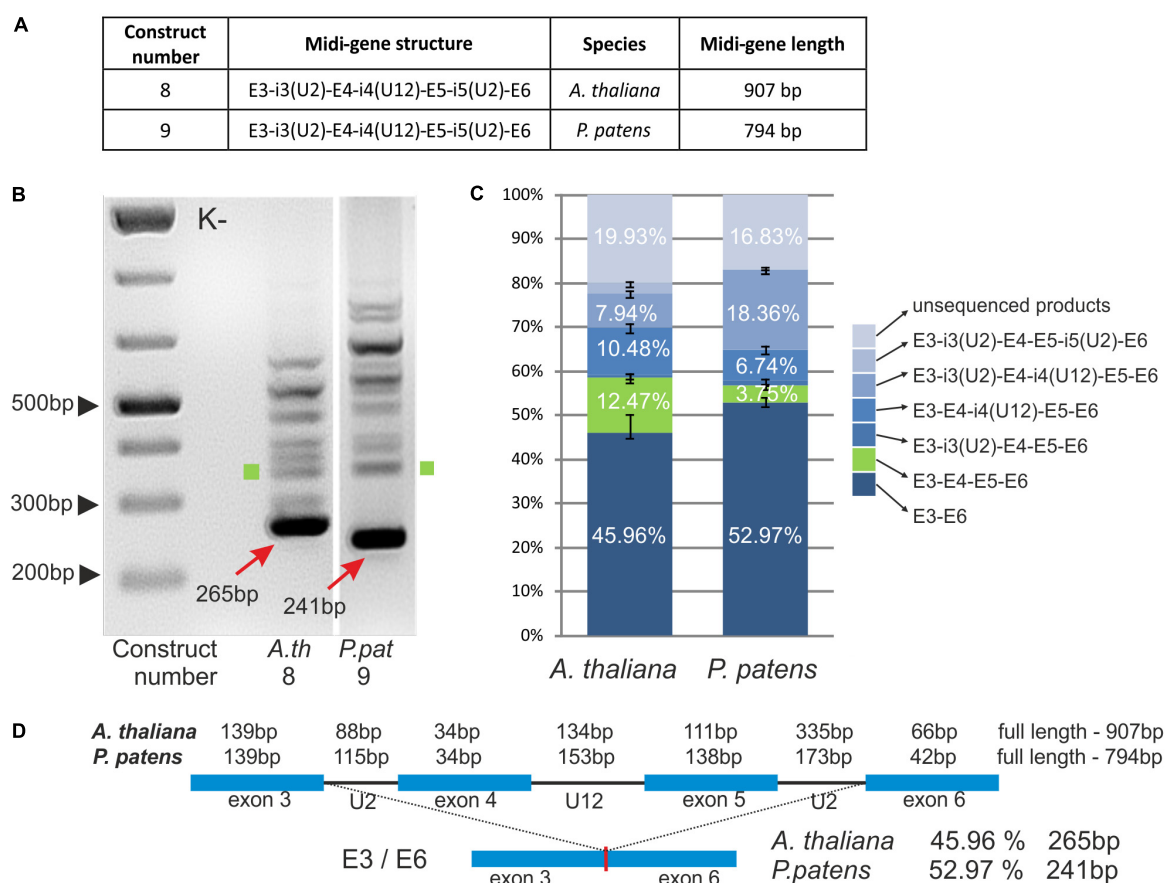
Quantitative RT-qPCR analyses using fluorescent-labeled primers and capillary electrophoresis were carried out to measure the levels of individual spliced products. Cloning and sequencing of these products were performed to identify particular splicing isoforms. In the case of  $\Delta$ U12, apart from the fully spliced product, we did not detect any additional mRNA isoforms. In the wt, wt transgene and U12→U2 plants, the fully spliced mRNA represented 96–97% of all spliced products, while U12- or U2-containing mRNA isoforms were present in 3–4% of all spliced products (Figure 5B). The exchange of exons no. 4 and no. 5 within the *CBP20* gene resulted in the presence of properly spliced mRNA (approximately 95% of all isoforms), the isoform still containing the U12 intron (approximately 3% of all splicing isoforms) and an additional mRNA isoform in which the 3' alternative U12 splice site was selected within exon no. 4 (approximately 2.5% of all splicing isoforms).

Two *A. thaliana* lines in which the U12 intron was relocated into the core- or tail-encoding part of the *CBP20* gene (U12

in core and U12 in tail) produced complex patterns of mRNA isoforms. In the case of the U12 in core variant, we identified and quantitatively measured the levels of five mRNA isoforms. The fully and correctly spliced *CBP20* mRNA isoform represented only 26% of all splicing products (**Figure 5B**). Surprisingly, the most abundant mRNA isoform showed an improperly spliced U2 intron replacing the original U12 intron. The alternative 5' splice site was selected, leading to the inclusion of an additional 38 nt from the 5' intron end into the processed mRNA (58.1%) (**Figure 5B**). However, these transgenic plants exhibit the wt phenotype, suggesting that the lower level of properly spliced *CBP20* mRNA is sufficient to fulfill plant requirements for the *CBP20* protein (**Figure 5C**). Interestingly, an identical improper splicing event was observed in the mini-gene containing the same U2 intron instead of the original U12 intron (**Figure 3B**). Three additional mRNA isoforms were observed in small amounts: (1) the isoform in which the U12 intron as well as the U2 intron that replaced the original U12 intron were retained (8.16%), (2) the

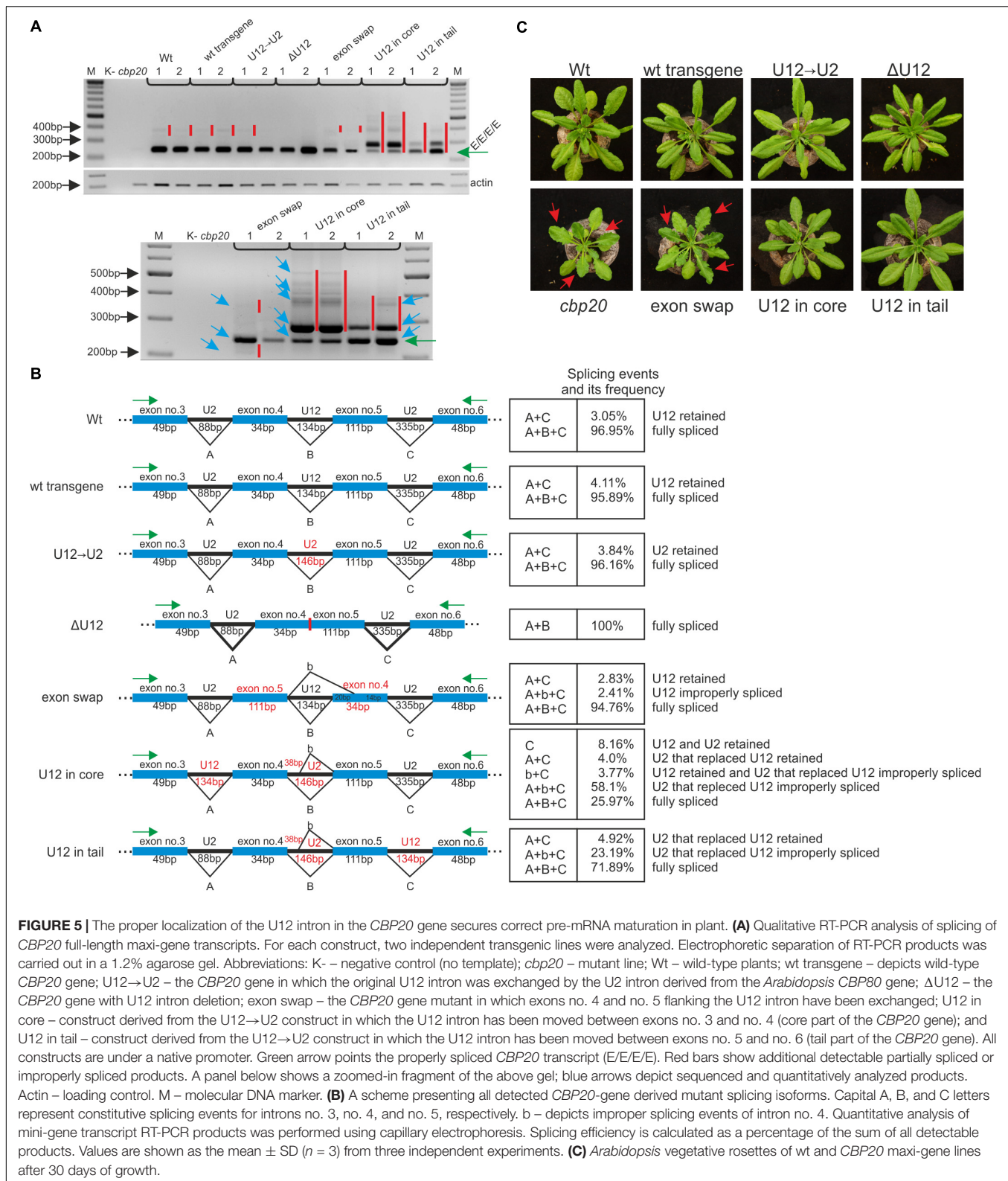
isoform in which the U2 intron that replaced the original U12 intron was retained (4.0%), and (3) the isoform in which the U12 intron was retained and the U2 intron located between exons no. 3 and no. 4 was incorrectly spliced (3.8%).

In the U12 in tail maxi-gene variant we identified and measured the levels of three mRNA isoforms. The fully and correctly spliced *CBP20* mRNA isoform represented almost 72% of all mRNA isoforms (**Figure 5B**). A minor portion of spliced products were identified as an isoform still containing the U2 intron that replaced the original U12 intron (4.9%). We also detected an mRNA isoform identical to that in the U12 in core maxi-gene, which contained the improperly spliced U2 intron that replaced the original U12 intron (23.2%) (**Figure 5B**). As expected, all these plants exhibit a wt phenotype (**Figure 5C**). These results suggest that shifting the U12 intron into other intron positions in the *Arabidopsis CBP20* gene disturbs splicing events, leading to substantial downregulation of the properly spliced mRNA level.



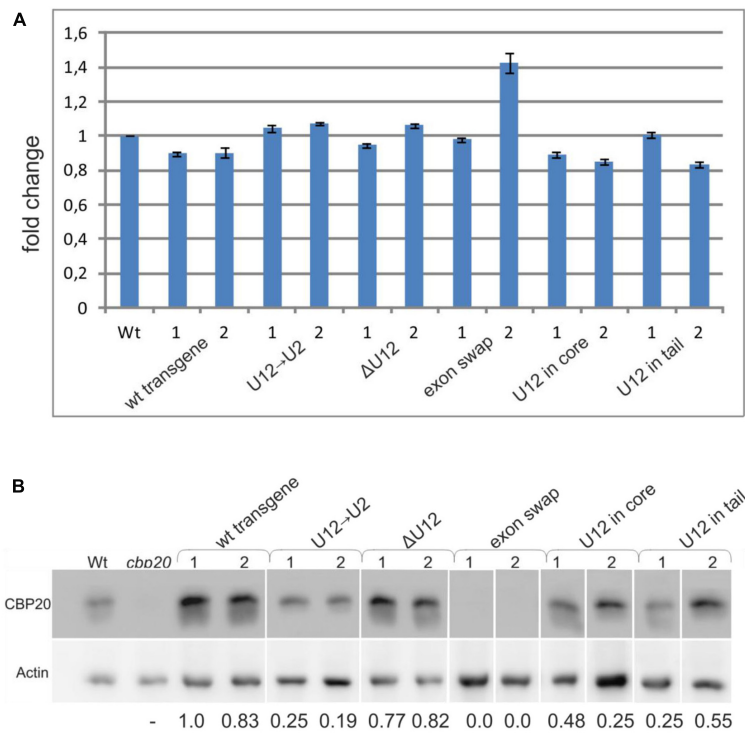
**FIGURE 4 |** Plant *CBP20* U12 intron surrounded by four exons and two U2 introns is not efficiently recognized by the U12 splicing machinery. **(A)** A table showing structures of two midi-gene constructs containing *A. thaliana* or *P. patens* *CBP20* gene fragment of the following composition: exon 3 – U2 intron 3 – exon 4 – U12 intron 4 – exon 5 – U2 intron 5 – exon 6. **(B)** Qualitative RT-PCR analysis of splicing of midi-gene transcripts in tobacco protoplasts. Arrows point to the improperly spliced products in which exons no. 4 and no. 5 are skipped. Green squares represent properly spliced midi-gene *CBP20* transcripts. The rest of the bands above the main amplification products represent a variety of partially spliced transcripts. **(C)** Quantitative analysis of midi-gene transcript RT-PCR products using capillary electrophoresis. Splicing efficiency is calculated as a percentage of the sum of all detectable products. Values are shown as the mean  $\pm$  SD ( $n = 3$ ) from three independent experiments. Green color depicts properly spliced *CBP20* mRNA isoform. **(D)** Structure of midi-genes and fully spliced transcript representing the most abundant splicing event.





To quantify the total level of *CBP20* transcript in the wt and maxi-gene transgenic lines analyzed, we performed real-time RT-PCR using primers designed to recognize the 3' ends of the

transcripts. No significant differences in the *CBP20* mRNA levels between wt and transgenic plants were observed (Figure 6A). However, we did detect changes between maxi-gene transgenic



**FIGURE 6 |** The localization, lack or replacement of the U12 intron in the *CBP20* gene does not influence the total level of the *CBP20* transcript but impacts the level of CBP20 protein in plants. **(A)** Real-time qPCR analysis of the total level of the *CBP20* transcript in maxi-gene transgenic lines. For each construct, two independent transgenic lines were analyzed. Wt – wild-type plants; *cbp20* – *cbp20* mutant line; wt transgene – wild-type *CBP20* gene structure; U12→U2 – *CBP20* gene in which the original U12 intron was replaced with the U2 intron derived from the *Arabidopsis CBP80* gene; ΔU12 – *CBP20* gene with U12 intron deletion; exon swap – the *CBP20* gene in which exons no. 4 and no. 5 flanking the U12 intron have been exchanged; U12 in core – a derivative of the U12→U2 construct in which the U12 intron has been introduced between exons no. 3 and no. 4; and U12 in tail – derivative of the U12→U2 construct in which the U12 intron has been introduced between exons no. 5 and no. 6. The *CBP20* mRNA level in Wt was taken as 1. Values are shown as the mean ± SD ( $n = 3$ ) from three independent experiments. **(B)** Western blot analysis of CBP20 protein levels in transgenic *Arabidopsis* plants. For each construct, two independent transgenic lines were analyzed. Upper panel – immunoblot using antibodies against CBP20 protein; lower panel – immunoblot with antibodies against actin used as a loading control. Numbers below the western blot image are relative intensities of CBP20 bands calculated using the wt transgene (line 1) CBP20 level as 1. Lines are described as previously. The western blot representing recognition of CBP20 in wild type and various mutants of this gene in *A. thaliana* is fused because the original blots contained data from additional transgenic lines, not included in the paper. The data are from two independent western blots. All calculations concerning the amount of the CBP20 protein were carried out separately for each western.

lines when the CBP20 protein level was analyzed by western blotting (**Figure 6B**). A similar protein level was observed for the wt transgene and ΔU12 lines. As expected, no CBP20 protein was detected in the exon swapping maxi-gene line due to the premature STOP codon existing in exon no. 5 of this construct. Interestingly, the decreased level of CBP20 was observed in the U12→U2 transgenic line as well as in the U12 in core and U12 in tail transgenic lines.

## DISCUSSION

The evolutionary conservation from liverworts to higher plants of the U12 intron localization within the plant *CBP20* genes encouraged us to study the role of this intron in the *CBP20* pre-mRNA maturation. It has been reported that U12 intron positions are more strongly conserved between animals and plants than are the positions of U2 introns (Basu et al., 2008).

Similar to plant *CBP20* genes, human, chicken, and starlet sea anemone *CBP20* genes also contain U12 introns. However, the animal *CBP20* gene structures and positions of U12 introns differ significantly from those of plant *CBP20* genes (Supplementary Figure S6). In contrast to their animal orthologs, plant *CBP20* genes contain an additional coding sequence representing the so-called tail part of CBP20. In each plant *CBP20* gene, a U12 intron separates the core part of the protein from the tail fragment. The tail part of plant CBP20s usually contains two NLSs, which is characteristic of plant small CBC subunits. Our *in silico* analyses revealed, however, that the *Chlamydomonas reinhardtii CBP20* gene does not have the tail-encoding part (GenBank: EDP01640.1). Interestingly, the *Chlamydomonas CBP20* gene also does not contain a U12 intron. Moreover, we have not found U12 introns in fungal *CBP20* genes that also lack the plant-specific CBP20 tail. Therefore, it can be speculated that the U12 intron in *CBP20* genes appeared during land plant evolution and was introduced into the

gene coding sequence together with the tail-encoding part of *CBP20*.

Comparison of the *CBP20* U12 introns of various plants revealed their broad spectrum of length: they can be as short as 76 bp (*S. moellendorffii*) and as long as 2733 bp (*V. vinifera*). Browsing the U12 database (U12DB) for U12 introns present in the *A. thaliana* genome revealed the presence of 246 U12 introns (Alioto, 2007). The length of these introns varies between 68 and 3731 bp. However, the number of U12 introns exceeding 1000 bp in length is rather low (only 6 such introns were described). The number of *Arabidopsis* U12 introns presented in the U12DB is underestimated, because more recent studies based on transcriptomic data revealed 2069 U12 introns in *Arabidopsis* (Marquez et al., 2012). Our data show that plant *CBP20* U12 introns derived from different plant species, when inserted into the mini-genes containing two flanking *A. thaliana* exons, exhibit differential impact on their splicing efficiency. Surprisingly, the longer the U12 intron was, the higher splicing efficiency was observed. We have previously reported that the splicing efficiency of plant U12 introns depends on a combination of factors, including TA content, exon splicing enhancer sequences (ESEs), and the presence of an adenosine at the upstream purine position in the plant U12 branch point consensus TCCTTATY (Lewandowska et al., 2004). All *CBP20* U12 introns studied in this paper exhibit a similar high TA content (approximately 65%), and almost all (except *P. endiviifolia*) contain an adenosine at the upstream position in their branch point sequences (Figure 1). Since the flanking exons are identical in all constructs, one can suggest that putative ESEs should influence all studied U12 intron splicing in the same way. Thus, the only evident difference between short and long *CBP20* U12 introns is that the longer U12 introns contain more TA-rich sequences, which may help in recognition of these introns by the U12 splicing machinery. Therefore, it is possible that one of the yet-unidentified hnRNP proteins recognizing TA-rich sequences within U12 introns plays an important role in plant minor spliceosome recruitment.

We noticed that the *CBP20* U12 intron from *P. patens* was extremely inefficiently spliced in tobacco protoplasts (Figure 2). Previously we also observed very low splicing efficiency of mini-genes containing U12 introns originated from different plant genes (Lewandowska et al., 2004). We showed that U12 intron splicing was enhanced by increased TA-richness of the intron. In this study we found that the branch point of the *P. patens* *CBP20* U12 intron has an additional C nucleotide in the 5' part of the consensus sequence compared to the *CBP20* U12 introns from other plant species (Figure 1). We analyzed all U12 introns from the *P. patens* genome and compared their branch points to the consensus sequence. Only four *P. patens* U12 introns containing the additional C nucleotide in the 5' part of the branch point were found (124 *P. patens* U12 introns were analyzed). Therefore, the low efficiency of *P. patens* *CBP20* U12 intron splicing could be due to the specific sequence of its branch point. Searching the U12DB for this additional C in the *A. thaliana* 5' part of all U12 branch point sequences revealed the presence of only six such examples (Alioto, 2007). This shows that the presence of an additional cytidine in the branch point sequence is acceptable for both *Physcomitrella* and

*Arabidopsis* U12 splicing machineries but may be a reason for the inefficient splicing of introns containing such branch point sequence. Further investigations are needed to uncover the role of this additional C in plant U12 intron splicing. On the other hand, the low splicing efficiency of the *P. patens* *CBP20* U12 intron may also be associated with the compatibility of specific intron/exon sequences as well as ESEs that are present or absent in different exonic sequences. Comparison of *P. patens* and other plant species exon sequences that surround *CBP20* U12 introns revealed a high similarity between exon no. 4, but in the case of exon no. 5, four insertions (3, 6, 14, and 5 bp long) and a single nucleotide deletion were found in *P. patens* (Supplementary Figure S7). Therefore, we cannot exclude the possibility that these exon differences between *Arabidopsis* and *Physcomitrella* may cause the incompatibility in the proper recognition of *P. patens* U12 intron-containing mini-gene transcripts by the tobacco U12 minor spliceosome.

To explain a role of the U12 intron in *CBP20* pre-mRNA maturation, several mini-, midi- and maxi-gene constructs were prepared. This approach allowed us to dissect the effect of given *CBP20* transcript fragments on U12 intron splicing. The shortest constructs (the mini-genes series) contained two *CBP20* exons that originally flank the U12 intron. The U12-containing mini-transcripts were correctly but poorly spliced (4.77%). In a similar assay, the pea U2 legumin intron was spliced correctly and very efficiently (94.41%). The *CBP80* U2 intron was spliced more efficiently (42.61%) than the original U12 intron but less efficiently than the legumin gene-derived U2 intron. In addition, alternatively spliced products were frequently observed (37.05%) (Figure 3C). These results have shown that the replacement of the U12 intron with a U2 intron increases splicing efficiency but may lead to undesired improper splicing events. Our observation agrees with the published data showing that U12 splicing slows down pre-mRNA processing (Lewandowska et al., 2004; Simpson and Brown, 2008). We demonstrated, however, that U2 introns when inserted between exons normally flanking a U12 intron are spliced improperly, which may lead to the downregulation of *CBP20* protein expression (Figure 4). It has been shown in animal cells that U12 intron splicing is a limiting step in pre-mRNA processing (Patel et al., 2002; Niemelä and Frilander, 2014).

Since the *CBP20* U12 intron splicing efficiency of the *Arabidopsis* mini-gene transcript (construct no. 1, Figure 2) was very low (0.34%), we decided to extend the construct, and we modified it by the introduction of two U2 introns and two flanking exons of the *Arabidopsis* *CBP20* gene (the midi-gene series). Indeed, we improved the proper splicing efficiency almost three times (from 4.77 to 12.47%, Figure 4). Unexpectedly, this correct splicing was predominated by an extensive skipping event resulting in the production of fused exons no. 3 and no. 6 (approximately 50% of all spliced products). In addition, a plethora of minor alternatively spliced products were observed. A similar effect was observed in the case of midi-gene transcripts containing the *Physcomitrella* *CBP20* intron, suggesting the general character of the results obtained (Figure 4). In our earlier studies, we used a similar approach to study analogous mini- and midi-gene constructs of the *Arabidopsis* *GSH2* gene containing a U12 intron (Lewandowska et al., 2004). Unlike the

results of the experiment presented in this paper, in the case of *GSH2*, we detected only constitutive splicing events, and as a result, we obtained the fully spliced products consisting of two or four exons when the mini- and midi-genes were analyzed, respectively. In addition, the transcript derived from the midi-gene (containing additional exon and intron sequences that flanked the U12 intron) was spliced more efficiently. This suggested that the U2 introns located upstream and downstream of U12 participated in the mechanism of definition of the U12 intron (Lewandowska et al., 2004). Interestingly, overexpression of the RBP45 protein, an hnRNP-like RNA binding protein with affinity for U-rich sequences (Lorković et al., 2000), resulted in a complete abolishment of the normal splicing pattern of the *GSH2* midi-gene transcript. RBP45 increased the overall splicing efficiency of the transcript, but caused an extensive skipping event resulting in the accumulation of RNA molecules that were composed of fused two terminal exons (Lewandowska et al., 2004). In the majority of *CBP20* midi-gene constructs analyzed in this paper, the presence of the U12 intron was ignored, and the most proximal 5' as well as distal 3' U2 splice sites were predominantly selected, without overexpression of RBP45. This can suggest that within the *CBP20* midi-gene transcripts are sequences with high affinity to RBP45 or other yet-unidentified tobacco RBP45-like proteins that can stimulate selection of the most external 5' and 3' splice sites.

Interestingly, the skipping effects that were observed during splicing of the midi-gene constructs were not detected when full-length *CBP20* gene variants (the maxi-genes series) were tested in transgenic plants. The wt transgene transcripts were correctly spliced, and removing the U12 intron (the  $\Delta$ U12 *CBP20* gene version) did not change the correct splicing pattern. The splicing pattern was changed, however, when the position of the U12 was altered (moved to the core or tail part of the *CBP20* gene), and a U2 intron replaced the original *Arabidopsis* U12 intron. In transgenic lines expressing these two constructs, incorrect splicing of the U2 intron, which served in the tested gene variants as intron no. 4, was detected. This is interesting since the same intron surrounded only by two U2 introns (a midi-gene) was spliced correctly, without any alternative events. Thus, in context of the whole *CBP20* gene, the U12 intron present upstream or downstream of its natural location induced selection of alternative splice sites. Moreover, the exon-swapping version of the *CBP20* gene (exons no. 4 and no. 5 were swapped) caused alterations in splice site selection of the U12 intron. These results show that sequences of both exons no. 4 and no. 5, as well as location of the U12 intron, are necessary for the proper

maturation of *A. thaliana* *CBP20* pre-mRNA. However, in all maxi-gene constructs analyzed, the levels of their transcripts were comparable to those of the wt (**Figure 6A**), suggesting that the alternative splicing events detected in some *CBP20* gene variants do not dramatically influence the levels of primary transcripts produced from the transgenes used in this study. Surprisingly, the protein level of CBP20 originating from the U12→U2 construct was lower than that expressed from the wt transgene. This low level of CBP20 was observed in two independent transgenic lines. Less CBP20 was also observed in the U12 in core and U12 in tail transgenes, and as expected, no CBP20 was detected in the exon swapping. These differences in the levels of CBP20, which are not supported by alterations in *CBP20* gene expression at the mRNA level, suggest that mature mRNAs originating from some of our constructs are partially retained in the nucleus and are not exported to the cytoplasm to be used as templates in translation. Further studies must be carried out to fully uncover the connection between U12 intron splicing, mRNA export and translation in the cytoplasm.

## AUTHOR CONTRIBUTIONS

MP and KK designed and performed the experiments, analyzed the data, and wrote the manuscript. DB and JD performed the experiments and contributed to data analysis. MS and WK conducted the bioinformatic analyses. AJ advised on experiments and data analysis and assisted in drafting the manuscript. ZS-K conceived the study contributed to data interpretation and participated in the manuscript writing. All authors read and approved the final manuscript.

## FUNDING

This work was supported by the National Science Centre project UMO-2013/11/B/NZ1/02099 and the Faculty of Biology at Adam Mickiewicz University in Poznan, Poland (Badania Statutowe). Funding for open access charge: Polish Ministry of Science and Higher Education [01/KNOW2/2014].

## SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fpls.2018.00475/full#supplementary-material>

## REFERENCES

- Alioto, T. S. (2007). U12DB: a database of orthologous U12-type spliceosomal introns. *Nucleic Acids Res.* 35, D110–D115. doi: 10.1093/nar/gkl796
- Bartschat, S., and Samuelsson, T. (2010). U12 type introns were lost at multiple occasions during evolution. *BMC Genomics* 11:106. doi: 10.1186/1471-2164-11-106
- Basu, M. K., Makalowski, W., Rogozin, I. B., and Koonin, E. V. (2008). U12 intron positions are more strongly conserved between animals and plants than U2 intron positions. *Biol. Direct.* 3:19. doi: 10.1186/1745-6150-3-19
- Benson, D. A., Karsch-Mizrachi, I., Clark, K., Lipman, D. J., Ostell, J., and Sayers, E. W. (2012). GenBank. *Nucleic Acids Res.* 40, D48–D53. doi: 10.1093/nar/gkr1202



- Burge, C. B., Padgett, R. A., and Sharp, P. A. (1998). Evolutionary fates and origins of U12-type introns. *Mol. Cell.* 2, 773–785. doi: 10.1016/S1097-2765(00)80292-0
- Clough, S. J., and Bent, A. F. (1998). Floral dip: a simplified method for *Agrobacterium*-mediated transformation of *Arabidopsis thaliana*. *Plant J.* 16, 735–743. doi: 10.1046/j.1365-313x.1998.00343.x
- Dias, S. M., Wilson, K. F., Rojas, K. S., Ambrosio, A. L., and Cerione, R. A. (2009). The molecular basis for the regulation of the cap-binding complex by the importins. *Nat. Struct. Mol. Biol.* 16, 930–937. doi: 10.1038/nsmb.1649
- Hugouvieux, V., Kwak, J. M., and Schroeder, J. I. (2001). An mRNA cap binding protein, ABH1, modulates early abscisic acid signal transduction in *Arabidopsis*. *Cell* 106, 477–487. doi: 10.1016/S0092-8674(01)00460-3
- Izaurrealde, E., Lewis, J., McGuigan, C., Jankowska, M., Darzynkiewicz, E., and Mattaj, I. W. (1994). A nuclear cap binding protein complex involved in pre-mRNA splicing. *Cell* 78, 657–668. doi: 10.1016/0092-8674(94)90530-4
- Jäger, K., Fábrián, A., Tompa, G., Deák, C., Höhn, M., Olmedilla, A., et al. (2011). New phenotypes of the drought-tolerant *cbp20 Arabidopsis thaliana* mutant have changed epidermal morphology. *Plant Biol.* 13, 78–84. doi: 10.1111/j.1438-8677.2010.00343.x
- Jung, H. J., and Kang, H. (2014). The *Arabidopsis* U11/U12-65K is an indispensable component of minor spliceosome and plays a crucial role in U12 intron splicing and plant development. *Plant J.* 78, 799–810. doi: 10.1111/tjp.12498
- Kierzkowski, D., Kmiecik, M., Piontek, P., Wojtaszek, P., Szweykowska-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, 814–825. doi: 10.1111/j.1365-313X.2009.03915.x
- Kim, S., Yang, J. Y., Xu, J., Jang, I. C., Prigge, M. J., and Chua, N. H. (2008). Two cap-binding proteins CBP20 and CBP80 are involved in processing primary MicroRNAs. *Plant Cell Physiol.* 49, 1634–1644. doi: 10.1093/pcp/pcn146
- Kim, W. Y., Jung, H. J., Kwak, K. J., Kim, M. K., Oh, S. H., Han, Y. S., et al. (2010). The *Arabidopsis* U12-type spliceosomal protein U11/U12-31K is involved in U12 intron splicing via RNA chaperone activity and affects plant development. *Plant Cell* 22, 3951–3962. doi: 10.1105/tpc.110.079103
- Kmiecik, M., Simpson, C. G., Lewandowska, D., Brown, J. W., and Jarmolowski, A. (2002). Cloning and characterization of two subunits of *Arabidopsis thaliana* nuclear cap-binding complex. *Gene* 283, 171–183. doi: 10.1016/S0378-1119(01)00859-9
- Kong, X., Ma, L., Yang, L., Chen, Q., Xiang, N., Yang, Y., et al. (2014). Quantitative proteomics analysis reveals that the nuclear cap-binding complex proteins *arabidopsis* CBP20 and CBP80 modulate the salt stress response. *J. Proteome Res.* 1, 2495–2510. doi: 10.1021/pr4012624
- Kwak, K. J., Jung, H. J., Lee, K. H., Kim, Y. S., Kim, W. Y., and Ahn, S. J. (2012). The minor spliceosomal protein U11/U12-31K is an RNA chaperone crucial for U12 intron splicing and the development of dicot and monocot plants. *PLoS One* 7:e43707. doi: 10.1371/journal.pone.0043707
- Kwak, K. J., Kim, B. M., Lee, K., and Kang, H. (2017). *quatre-quart1* is an indispensable U12 intron-containing gene that plays a crucial role in *Arabidopsis* development. *J. Exp. Bot.* 68, 2731–2739. doi: 10.1093/jxb/erx138
- Laubinger, S., Sachsenberg, T., Zeller, G., Busch, W., Lohmann, J. U., Rätsch, 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, 8795–8800. doi: 10.1073/pnas.0802493105
- 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
- Lin, C. F., Mount, S. M., Jarmolowski, A., and Makalowski, W. (2010). Evolutionary dynamics of U12-type spliceosomal introns. *BMC Evol. Biol.* 10:47. doi: 10.1186/1471-2148-10-47
- Lorković, Z. J., Wieczorek Kirk, D. A., Klahre, U., Hemmings-Mieszczyk, M., and Filipowicz, W. (2000). RBP45 and RBP47, two oligouridylylate-specific hnRNP-like proteins interacting with poly(A)+ RNA in nuclei of plant cells. *RNA* 6, 1610–1624. doi: 10.1017/S1355838200001163
- 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
- Mazza, C., Segref, A., Mattaj, I. W., and Cusack, S. (2002). Large-scale induced fit recognition of an m(7)GpppG cap analogue by the human nuclear cap-binding complex. *EMBO J.* 21, 5548–5557. doi: 10.1093/emboj/cdf538
- Niemelä, E. H., and Frilander, M. J. (2014). Regulation of gene expression through inefficient splicing of U12-type introns. *RNA Biol.* 11, 1325–1329. doi: 10.1080/15476286.2014.996454
- Papp, I., Mur, L. A., Dalmadi, A., Dulai, S., and Koncz, C. (2004). A mutation in the cap binding protein 20 gene confers drought tolerance to *Arabidopsis*. *Plant Mol. Biol.* 55, 679–686. doi: 10.1007/s11103-004-1680-2
- Patel, A. A., McCarthy, M., and Steitz, J. A. (2002). The splicing of U12-type introns can be a rate-limiting step in gene expression. *EMBO J.* 21, 3804–3815. doi: 10.1093/emboj/cdf297
- Pieczynski, M., Marczewski, W., Hennig, J., Dolata, J., Bielewicz, D., Piontek, P., et al. (2013). Down-regulation of CBP80 gene expression as a strategy to engineer a drought-tolerant potato. *Plant Biotechnol. J.* 11, 459–469. doi: 10.1111/pbi.12032
- 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, 265–278. doi: 10.1093/nar/gkp869
- Ren, G., and Yu, B. (2012). Critical roles of RNA-binding proteins in miRNA biogenesis in *Arabidopsis*. *RNA Biol.* 9, 1424–1428. doi: 10.4161/rna.22740
- Sierocka, I., Rojek, A., Bielewicz, D., Karłowski, W., Jarmolowski, A., and Szweykowska-Kulinska, Z. (2011). Novel genes specifically expressed during the development of the male thalli and antheridia in the dioecious liverwort *Pellia endiviifolia*. *Gene* 485, 53–62. doi: 10.1016/j.gene.2011.06.012
- Simpson, C. G., and Brown, J. W. (2008). U12-dependent intron splicing in plants. *Curr. Top. Microbiol. Immunol.* 326, 61–82. doi: 10.1007/978-3-540-76776-3\_4
- Simpson, G. G., and Filipowicz, W. (1996). Splicing of precursors to mRNA in higher plants: mechanism, regulation and sub-nuclear organisation of the spliceosomal machinery. *Plant Mol. Biol.* 32, 1–41. doi: 10.1007/BF00039375
- Szarzynska, B., Sobkowiak, L., Pant, B. D., Balazadeh, S., Scheible, W. R., Mueller-Roeber, B., et al. (2009). Gene structures and processing of *Arabidopsis thaliana* HYL1-dependent pri-miRNAs. *Nucleic Acids Res.* 37, 3083–3093. doi: 10.1093/nar/gkp189
- Turunen, J. J., Niemelä, E. H., Verma, B., and Frilander, M. J. (2013). The significant other: splicing by the minor spliceosome. *Wiley Interdiscip. Rev. RNA* 4, 61–76. doi: 10.1002/wrna.1141
- Xu, T., Kim, B. M., Kwak, K. J., Jung, H. J., and Kang, H. (2016). The *Arabidopsis* homolog of human minor spliceosomal protein U11-48K plays a crucial role in U12 intron splicing and plant development. *J. Exp. Bot.* 67, 3397–3406. doi: 10.1093/jxb/erw158

**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 © 2018 Pieczynski, Kruska, Bielewicz, Dolata, Szczesniak, Karłowski, Jarmolowski and Szweykowska-Kulinska. 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 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.



# Ectopic Transplastomic Expression of a Synthetic MatK Gene Leads to Cotyledon-Specific Leaf Variegation

Yujiao Qu<sup>1</sup>, Julia Legen<sup>1</sup>, Jürgen Arndt<sup>1</sup>, Stephanie Henkel<sup>1</sup>, Galina Hoppe<sup>1</sup>, Christopher Thieme<sup>1</sup>, Giovanna Ranzini<sup>1</sup>, Jose M. Muino<sup>1</sup>, Andreas Weihe<sup>1</sup>, Uwe Ohler<sup>2</sup>, Gert Weber<sup>1,3</sup>, Oren Ostersehter<sup>4</sup> and Christian Schmitz-Linneweber<sup>1\*</sup>

<sup>1</sup> Institut für Biologie, Humboldt-Universität zu Berlin, Berlin, Germany, <sup>2</sup> Computational Regulatory Genomics, Berlin Institute for Medical Systems Biology, Max Delbrück Center for Molecular Medicine, Berlin, Germany, <sup>3</sup> Helmholtz-Zentrum Berlin für Materialien und Energie, Joint Research Group Macromolecular Crystallography, Berlin, Germany, <sup>4</sup> Department of Plant and Environmental Sciences, The Alexander Silberman Institute of Life Sciences, The Hebrew University of Jerusalem, Jerusalem, Israel

## OPEN ACCESS

### Edited by:

Paula Duque,  
Instituto Gulbenkian de Ciência (IGC),  
Portugal

### Reviewed by:

Gorou Horiguchi,  
Rikkyo University, Japan  
Xiao-Ning Zhang,  
St. Bonaventure University,  
United States

### \*Correspondence:

Christian Schmitz-Linneweber  
smitzlic@rz.hu-berlin.de

### Specialty section:

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

**Received:** 29 March 2018

**Accepted:** 12 September 2018

**Published:** 04 October 2018

### Citation:

Qu Y, Legen J, Arndt J, Henkel S,  
Hoppe G, Thieme C, Ranzini G,  
Muino JM, Weihe A, Ohler U,  
Weber G, Ostersehter O and  
Schmitz-Linneweber C (2018) Ectopic  
Transplastomic Expression of a  
Synthetic MatK Gene Leads  
to Cotyledon-Specific Leaf  
Variegation. *Front. Plant Sci.* 9:1453.  
doi: 10.3389/fpls.2018.01453

Chloroplasts (and other plastids) harbor their own genetic material, with a bacterial-like gene-expression systems. Chloroplast RNA metabolism is complex and is predominantly mediated by nuclear-encoded RNA-binding proteins. In addition to these nuclear factors, the chloroplast-encoded intron maturase MatK has been suggested to perform as a splicing factor for a subset of chloroplast introns. MatK is essential for plant cell survival in tobacco, and thus null mutants have not yet been isolated. We therefore attempted to over-express MatK from a neutral site in the chloroplast, placing it under the control of a theophylline-inducible riboswitch. This ectopic insertion of MatK lead to a variegated cotyledons phenotype. The addition of the inducer theophylline exacerbated the phenotype in a concentration-dependent manner. The extent of variegation was further modulated by light, sucrose and spectinomycin, suggesting that the function of MatK is intertwined with photosynthesis and plastid translation. Inhibiting translation in the transplastomic lines has a profound effect on the accumulation of several chloroplast mRNAs, including the accumulation of an RNA antisense to *rp133*, a gene coding for an essential chloroplast ribosomal protein. Our study further supports the idea that MatK expression needs to be tightly regulated to prevent detrimental effects and establishes another link between leaf variegation and chloroplast translation.

**Keywords:** chloroplast, variegation, MatK, splicing, transplastomic, *Nicotiana tabacum*

## INTRODUCTION

Land plant chloroplast RNAs are heavily processed post-transcriptionally. Next to trimming, and RNA editing, a prominent post-transcriptional processing step is RNA splicing of group II introns. Group II intron splicing is mediated predominantly by nuclear-encoded RNA binding proteins, of which about a dozen have been described in maize or Arabidopsis (de Longevialle et al., 2010; Stern et al., 2010). In addition to this nuclear complement of splicing factors, there is a single chloroplast-encoded splicing factor named intron maturase K, short MatK. The MatK protein is related to prokaryotic intron maturases (Neuhaus and Link, 1987). Bacterial intron maturases are

encoded within group II introns, i.e., their “home introns.” The home intron of MatK is located in the *trnK* gene.

Group II introns are characterized by six secondary structure elements named domain (D) I – VI that fold into a globular tertiary structure (Michel and Ferat, 1995). The maturase reading frame is always found inside DIV, and the maturases are typically required for splicing of their host intron RNAs (Schmitz-Linneweber et al., 2015). Maturases are generally identified by several conserved motifs. These include a region with sequence similarity to retroviral-type reverse transcriptases (i.e., the RT domain), and a conserved sequence motif similar to the thumb domain of retroviral RTs, denoted domain X (Mohr et al., 1993). Biochemical studies have indicated that maturases make direct contacts with selected intron sequence elements and these contacts are essential for the splicing reaction. For example, the bacterial Maturase LtrA interacts with sequence stretches within DI, DII and DIV. These contacts help to attain a splicing-competent intron conformation (Matsuura et al., 2001; Rambo and Doudna, 2004). A major advance in the understanding of the roles of maturases in splicing has been recently accomplished by structural analyses of bacterial MATs bound to their cognate group II intron RNA targets (Piccirilli and Staley, 2016; Qu et al., 2016; Zhao and Pyle, 2016). These including the crystal structures of the RT domains of MATs from *Roseburia intestinalis* and *Eubacterium rectale* (Zhao and Pyle, 2016), and a cryo-EM analysis of the ribonucleoprotein complex of the *Lactobacillus lactis* intron-encoded LtrA maturase bound to its host *ltrB* intron RNA (Qu et al., 2016). The structures of the spliced *ltrB* intron (at 4.5 Å resolution) and of the *ltrB* intron in its ribonucleoprotein complex with LtrA (at 3.8 Å resolution) are further revealing functional coordination between the intron RNA with its cognate maturase protein. Remarkably, these structures reveal close relationships between the RT catalytic domain and telomerases, whereas the ‘active splicing centers’ resemble that of the Prp8 protein (Dlakic and Mushegian, 2011; Galej et al., 2013; Yan et al., 2015), which also resides at the core of the spliceosome.

Aside of their essential role in splicing, intron maturases are also required for genetic mobility of the intron (Lambowitz and Zimmerly, 2004). Bacterial group II introns can spread to other genomic positions in a process called retrohoming that depends on the maturase protein. Chloroplast group II introns of embryophytes are no longer mobile and MatK has lost the protein domains required for intron mobility (Mohr et al., 1993; Barthet and Hilu, 2007). By contrast, it has retained the so-called domain X, which was shown to be required for RNA splicing activity of bacterial introns (Lambowitz and Zimmerly, 2004).

A role in splicing for MatK is strongly supported by studies on chloroplasts devoid of a translational apparatus. In these mutants, the *trnK* precursor-RNA is not spliced (Vogel et al., 1997). In addition, an entire subgroup of chloroplast introns termed group IIA introns fails to splice as well (Hess et al., 1994; Vogel et al., 1999). The only conceivable factor that would require functioning chloroplast translation for splicing is MatK, which led to the proposition that MatK serves splicing of all group IIA introns. Indeed, direct association of MatK with intron RNA was demonstrated *in vitro* (Liere and Link, 1995) and with seven group IIA introns also *in vivo* (Zoschke et al., 2010).

The *matK* reading frame is found in all known autotrophic land-plant chloroplast genomes that contain group II introns, and is also present in basal streptophyte algae (Turmel et al., 2006). In the streptophyte algae *Zygnema*, in the fern *Adiantum capillus-veneris* and also in the parasitic land plants *Epifagus virginiana*, *Cuscuta exaltata*, and *Cuscuta reflexa*, *matK* is present as a stand-alone reading frame, while the *trnK* gene has been lost (Wolfe et al., 1992; Turmel et al., 2005; Funk et al., 2007; McNeal et al., 2007). This suggests a function of *matK* “*in trans*,” most likely for the splicing of pre-RNAs other than its cognate *trnK* intron. Among all embryophytes analyzed, only parasitic species have lost *matK*, among them species in the genera *Cuscuta*, *Cytinus*, *Rafflesia*, *Pilostyles*, as well as the orchid *Rhizantella* and other orchid genera (Funk et al., 2007; McNeal et al., 2009; Delannoy et al., 2011; Molina et al., 2014; Bellot and Renner, 2015; Roquet et al., 2016). These non-autotrophic plants have also lost their group IIA intron sequences, with the exception of the structurally derived and ‘evolutionarily younger’ *clpP-2* intron (Turmel et al., 2006; Funk et al., 2007; McNeal et al., 2009). Together, these phylogenetic analyses further support the notion that MatK is required for more than just splicing of its home intron. Also, its continuous presence in chloroplast genomes starting from early streptophytes suggests that its retention inside the chloroplast is not a chance event. Unlike *matK*, many other chloroplast and also mitochondrial genes have been transferred multiple times in different land plant lineages to the nucleus (Kleine et al., 2009). This includes several genes coding for maturases that have been found in the Arabidopsis nuclear genome (Mohr and Lambowitz, 2003). These are all targeted to the mitochondria (one is dually targeted to the mitochondria and chloroplasts), and at least three of them serve splicing of multiple mitochondrial group II introns (Nakagawa and Sakurai, 2006; Keren et al., 2009, 2012; Cohen et al., 2014; Zmudjak et al., 2017). These data demonstrate that maturases can be transferred to the nucleus and still function in splicing *in trans*. An important question therefore is: ‘why is *matK* evolutionary maintained on the chloroplast chromosome?’. To address this, functional genetic analyses are required. Unfortunately, the *matK* gene has been recalcitrant to reverse genetic tampering. Mutagenesis of the chloroplast *matK* reading frame by transplastomic mutagenesis failed so far, which has been taken as evidence that *matK* is an essential gene (Drescher, 2003; Zoschke et al., 2010). We therefore decided to study *matK* using a gain of function approach, i.e., by ectopic over-expression from the chloroplast genome in addition to the endogenous copy.

## RESULTS

### Introducing an Additional Copy of Chloroplast MatK Into a Neutral Integration Site Leads to Homoplastomic Tobacco Lines With Variegated Cotyledons

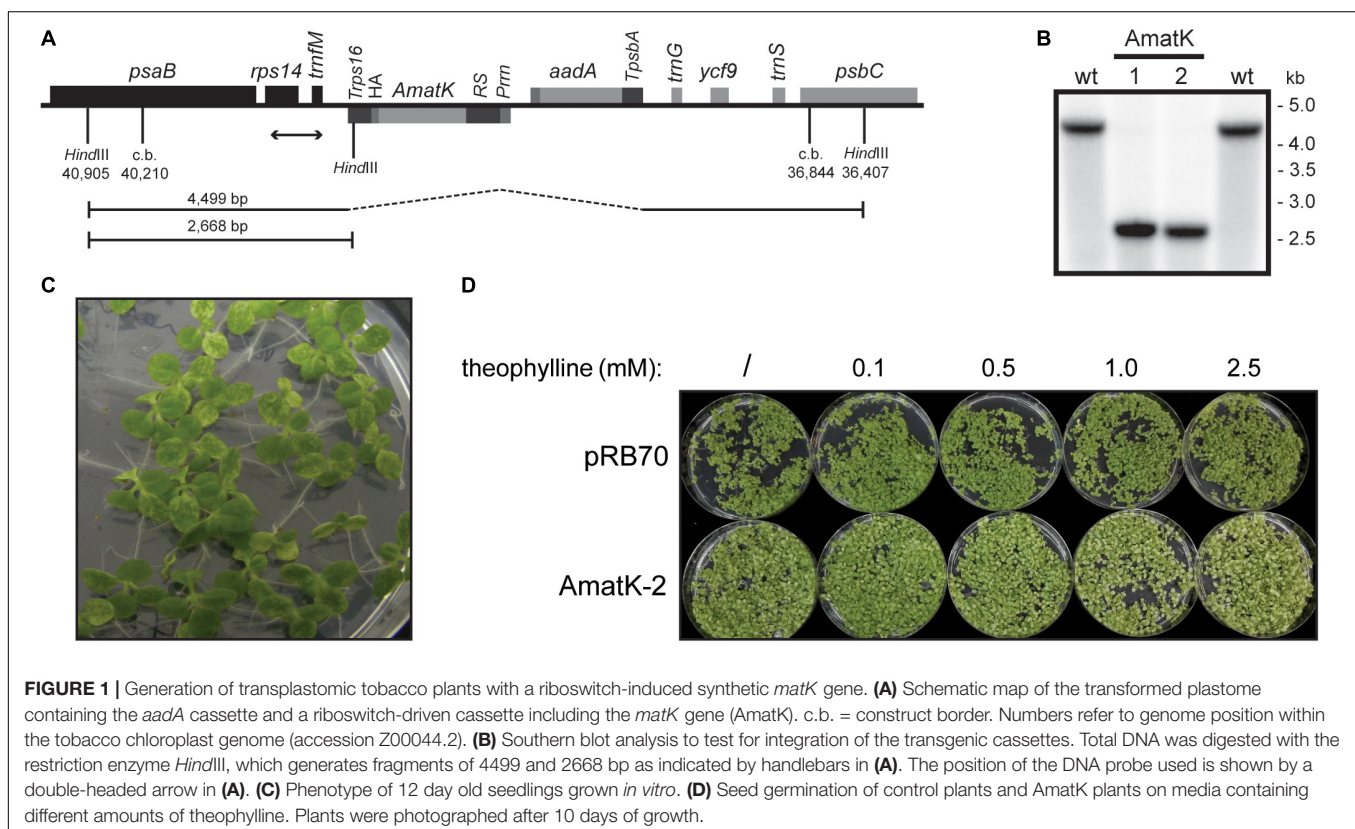
Given that *matK* is considered to be an essential gene, we opted for engineering an inducible over-expressor of MatK to



avoid lethal affects after constitutive over-expression. Therefore, we used a transformation vector that allows the expression of transgenes based on the theophylline-responsive riboswitch (Verhounig et al., 2010). In brief, the riboswitch utilized in this system is a translational on-switch, i.e., modulates translation initiation through the formation of an alternative structure that sequesters the Shine-Dalgarno and initiation (AUG) codon sequences in the absence of the inducer. The addition of theophylline leads to conformational changes in the 5'-UTR, which serves as an entry point for chloroplast ribosomes, allowing the translation of the transgene (Winkler and Breaker, 2005). This system was successfully used for the induced expression of GFP, using the chloroplast transformation vector pAV (Verhounig et al., 2010). The pAV vector also contains a selectable marker, i.e., a chimeric spectinomycin resistance gene, *aadA*, that facilitates the selection of plants with transgenic chloroplast genomes. The vector includes flanking sequences that allow to target the transgenes to a neutral intergenic spacer region in the chloroplast genome of tobacco by homologous recombination. Here, we replaced the GFP construct within the pAV vector with a *matK* reading frame, that has only 72% nucleotide sequence similarity with the endogenous tobacco *matK* sequence, while maintaining the encoded amino acid sequence (resulting vector: pRSAmatK, see also **Figure 1A** and **Supplementary Figure S1**). This prevents undesirable homologous recombination events within the endogenous *matK* sequence. In addition, we added a C-terminal triple HA tag to allow detection of the transplastomic MatK protein. Previously, we showed that C-terminal tagging of

the endogenous MatK does not interfere with intron splicing and does not entail any detectable macroscopic phenotype (Zoschke et al., 2010).

The pRSAmatK construct was introduced into tobacco plants by biolistic transformation of the chloroplast genome, followed by selection of spectinomycin-resistant cell lines. We isolated a total of 11 putative transplastomic lines. All lines were subjected to additional rounds of regeneration under antibiotic selection. This eliminated residual wild-type copies of the plastid genome, i.e., produced homoplastomic lines. We further tested for homoplastomy and correct integration by restriction fragment length polymorphism analyses (**Figure 1B**). We concluded that the lines were homoplastomic. We named these lines AmatK to reflect the use of a synthetic *matK* reading frame. When growing F1 plants from various AmatK lines, we noticed the mottled, variegated appearance of cotyledons in AmatK seedlings (**Figure 1C**). By contrast, the primary leaves were indistinguishable from those of the wild-type plants (**Figure 1C** and **Supplementary Figure S2**; in some cases, pale-green tissue was also observed in early primary leaves). The effect on the plants in the absence of the riboswitch inducer, theophylline, can be attributed to the known leakiness of the construct (Emadpour et al., 2015). We next asked, whether addition of theophylline to AmatK lines would modulate the observed variegated phenotype. Indeed, increasing concentrations of theophylline in the growth medium exacerbated cotyledon bleaching, while the control plants did not respond visibly to the addition of theophylline (**Figure 1D**). Together, these phenotypic analyses indicate that





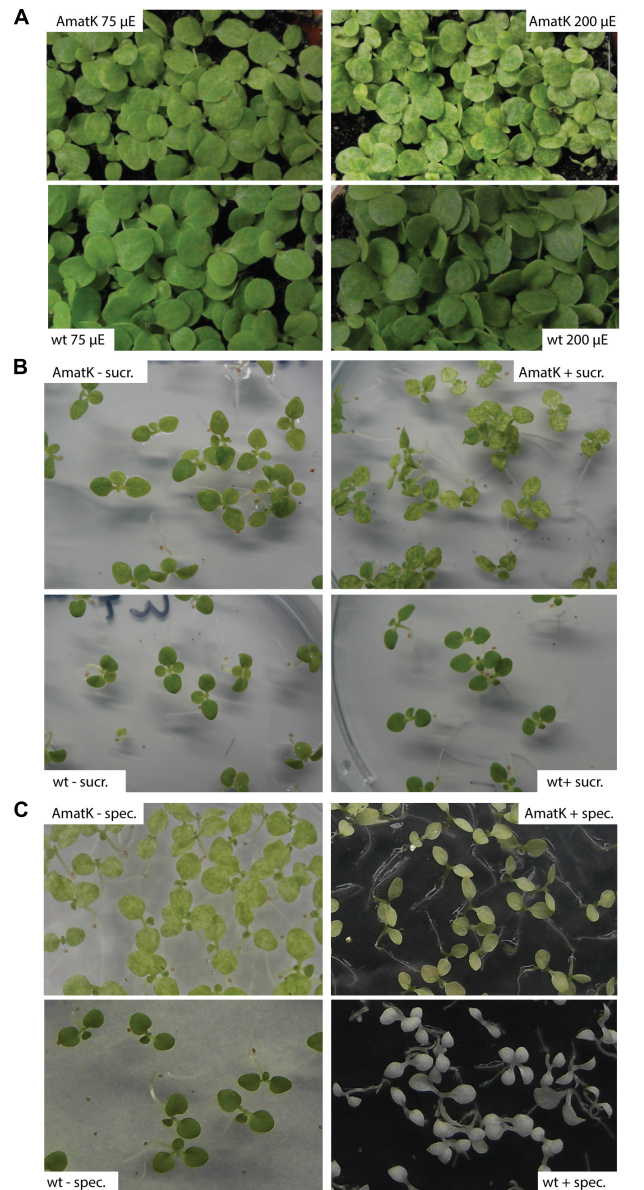
the *AmatK* insertion (i) interferes with chloroplast biogenesis in a cell-autonomous fashion, (ii) is restricted to cotyledon tissue, (iii) and responds to the riboswitch activation.

## ***AmatK* Leaf Variegation Is Modulated by Light, Sugar and the Inhibition of Chloroplast Translation**

Variegation of genetically homogenous tissue has been observed before, and often was shown to be modulated by external signals. For example, variegation of the *immutans* mutant is strongly affected by light and temperature signals (Rodermeier, 2002). We noticed that variegation was variable depending on in which growth cabinet and under which conditions we grew the plants. Both, the extent of pale areas and different degrees of paleness were observed. We therefore systematically tested the effects of different growth conditions on the extent of variegation. Specifically, we applied conditions that are known to affect photosynthesis – the core function of chloroplast genetic information. First, we tested different light conditions. At higher light fluences of  $200 \mu\text{E m}^{-2}\cdot\text{s}^{-1}$ , *AmatK* cotyledons are paler than at  $75 \mu\text{E m}^{-2}\cdot\text{s}^{-1}$ , while wt plants tolerate both light conditions with no visible alteration in leaf coloration (**Figure 2A**). This indicates that the ectopic insertion of *AmatK* leads to sensitivity to higher light intensities, i.e., potentially to a compromised photosynthetic apparatus and thus eventually to photoinhibition. Noteworthy, plants grown on soil show less severe variegation and paleness than plants grown *in vitro* on MS medium. This indicates that different growth conditions, like for example the provision with nutrients, gas exchange and/or light differences between the petri-dish grown plants and the soil-grown plants affect variegation as well.

Sugar is known to be a key regulator of plant metabolism, affecting the expression of many different plant genes (Pego et al., 2000). Accordingly, photosynthesis is known to be affected by the availability of exogenous sugars, which are thought to limit photosynthesis and prevent the proper development of the photosynthetic apparatus (Hdider and Desjardins, 1994; Koch, 1996; Van Huylenbroeck and Debergh, 1996; Rybczyński et al., 2007). *AmatK* plants grown on sugar show paler cotyledons, with larger patches of reduced chlorophyll, while wt plants were unaffected, suggesting that the effects of sugar and the *AmatK* transgene are additive (**Figure 2B**).

MatK has been suggested to be important for splicing of RNAs with essential roles for the translational apparatus, including several tRNAs and mRNAs for ribosomal proteins. Therefore, the effect of *AmatK* might well run via partially compromised translational activity in plastids. This should in turn effect the expression of the selectable marker introduced into *AmatK* plants, the *aadA* cassette. *aadA* confers resistance against spectinomycin. We tested spectinomycin resistance by germinating *AmatK* seeds on spectinomycin containing medium. As expected, wt seedlings are albino, devoid of chloroplast development. *AmatK* lines show yellowish cotyledons with almost no green sectors remaining (**Figure 2C**). We conclude that an inhibition of translation exacerbates variegation of *AmatK* cotyledons.



**FIGURE 2 |** *AmatK* variegation is modified by light, sugar and translation inhibition. **(A)** *AmatK* and wt plants germinated under two light regimes and photographed after 10 days. **(B)** *AmatK* and wt plants germinated on MS media with or without a sucrose supplement (3% w/v). **(C)** *AmatK* and wt plants germinated on MS media with or without a spectinomycin supplement (500  $\mu\text{g/L}$ ).

## **Transgene-Derived mRNA, but No Full-Length MatK:HA Protein Is Detected in *AmatK* Lines**

By design, the *AmatK* transgene was supposed to be translationally silent in the absence of the inducer theophylline. Yet, even in the absence of theophylline, *AmatK* plants showed a variegated cotyledon phenotype. These effects may be due to basal expression in the absence of inducer (Emadpour et al., 2015).

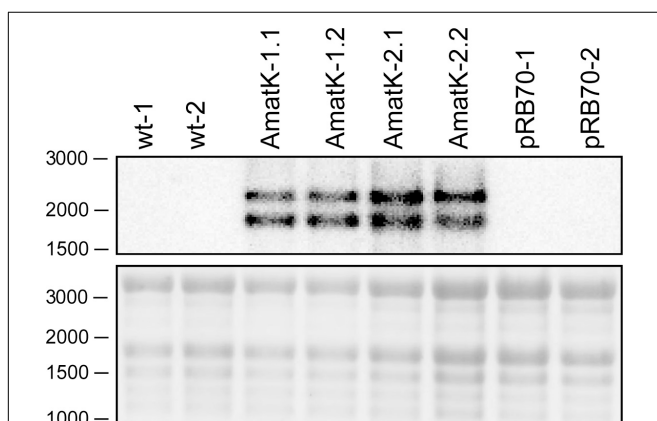
Possibly, low expression of *AmatK* already impacts chloroplast development at least in cotyledons. We therefore tested expression of the *AmatK* gene on the RNA and protein level.

RNA gel blot analyses were carried out with total RNA from *AmatK* lines grown on MS-medium with  $75 \mu\text{E m}^{-2}\cdot\text{s}^{-1}$  of light at  $25^\circ\text{C}$  (16 h light, 8 h darkness). Plants were harvested 12 days after imbibition. The oligonucleotide used for probe preparation is complementary to the junction region of the *AmatK* reading frame and the triple HA-tag, ensuring that the signal obtained from radioactive probe hybridization is specific to transcripts of the *AmatK* sequence. Two biological replicates were analyzed for each of two *AmatK* lines. In addition, we extracted RNA from two independent wt samples and two plants from the previously described pRB70 line, which carries an *aadA* cassette without the additional *AmatK* gene within the intergenic region between *trnfM* and *trnG* (Ruf et al., 2001). This controls for unwanted signals from the strongly expressed selectable marker gene. The *AmatK* lines display two signals at 1.8 and 2.2 kb, while no signal was observed for wt or pRB70 control samples (Figure 3). The expected size for a complete transcript of the synthetic *matK* gene spanning the region from the Prn promoter to the *rps16* terminator sequence is approximately 1.8 kb. Termination of transcription in chloroplasts is poorly understood, and transcriptional read through across so-called terminators of transcription of transplastomes has been seen before (Schmitz-Linneweber et al., 2001). Therefore, the longer, 2.2 kb transcript possibly represents read-through transcription across the *rps16* terminator and ends at the downstream *trnfM* gene. The absence of both signals from wt RNA samples demonstrates that the signals are not due to endogenous RNA species but represent mRNAs derived specifically from the introduced transgene.

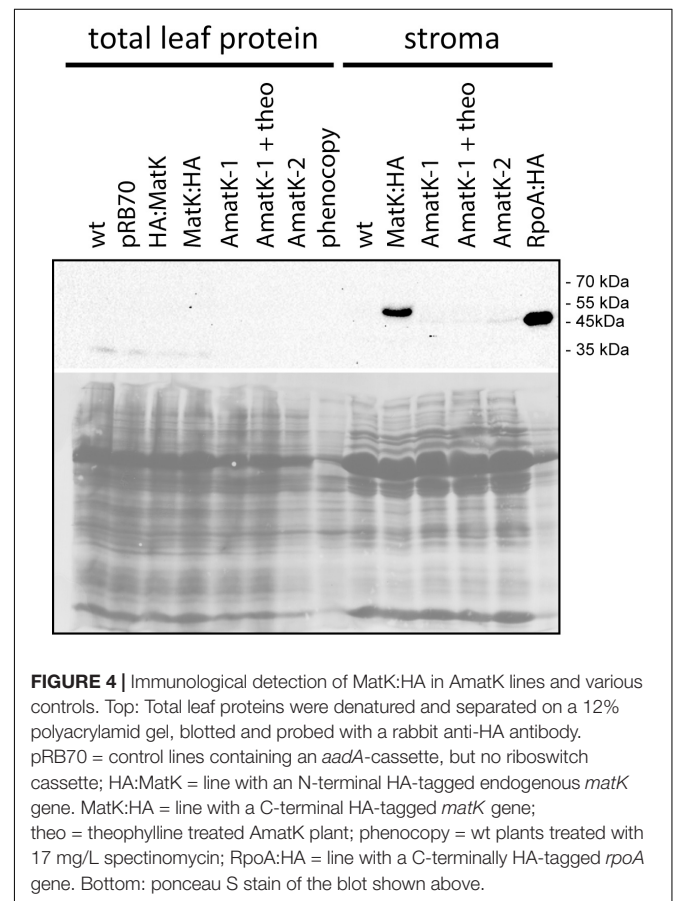
We next tested accumulation of the transgene-derived MatK protein using an antibody against the C-terminal HA epitope. We prepared total plant protein and/or stroma from *AmatK* and

wt lines. Some *AmatK* plants were grown in the presence of the riboswitch inducer theophylline. We used the following controls: (1). As controls for potential signals from the *aadA* cassette, we used the pRB70 lines mentioned above. (2). To control for effects caused by the pale cotyledon tissue, we used a phenocopy generated by treating wt plants with low doses of spectinomycin. This treatment leads to bleached cotyledons, but still allows greening of primary leaves, i.e., mimics the phenotype of *AmatK* seedlings (Supplementary Figure S3). (3). As a positive control, we added a line that carries an HA-tagged plastome-encoded subunit of the plastid RNA polymerase named RpoA:HA (Finster et al., 2013). (4). Total protein preparations from plants that carry an N- or C- terminally HA-tagged MatK, designated HA:MatK or MatK:HA, (Zoschke et al., 2010), were used as positive controls for the detection on MatK in transformed tobacco plants.

Our immunoblots show that MatK can be detected in stroma preparations from the tagged endogenously locus, confirming previous analyses (Zoschke et al., 2010) (Figure 4). The protein runs at approximately 55 kDa, well below the calculated molecular weight of 64 kDa, but in line with previous analyses of MatK (Zoschke et al., 2010). We could not detect MatK in any sample from total leaf protein preparations. Also, no signal corresponding to the size of the MatK:HA protein was observed in *AmatK* lines, including lines treated with theophylline, and independent of sample preparation (stroma or total protein).



**FIGURE 3 |** Detection of synthetic *matK* transcripts from *AmatK*, wt and *aadA* leaf tissue using RNA gel blot hybridization. 5  $\mu\text{g}$  of total plant RNA was extracted and analyzed by RNA gel blot hybridization using a probe directed against the junction of the *AmatK* and HA sequence, which is specific to the transgene. The wt lines -1 and -2 as well as the *AmatK* lines -1 and -2 represent biological replicates.



**FIGURE 4 |** Immunological detection of MatK:HA in *AmatK* lines and various controls. Top: Total leaf proteins were denatured and separated on a 12% polyacrylamide gel, blotted and probed with a rabbit anti-HA antibody. pRB70 = control lines containing an *aadA*-cassette, but no riboswitch cassette; HA:MatK = line with an N-terminal HA-tagged endogenous *matK* gene; MatK:HA = line with a C-terminal HA-tagged *matK* gene; theo = theophylline treated *AmatK* plant; phenocopy = wt plants treated with 17 mg/L spectinomycin; RpoA:HA = line with a C-terminally HA-tagged *rpoA* gene. Bottom: ponceau S stain of the blot shown above.

However, all stroma preparations from AmatK lines had a faint signal of about 47 kDa, possibly a degradation product (**Figure 4**). This signal is specific to the AmatK lines and is found consistently in different probings (**Supplementary Figure S4**). The signal must represent a C-terminal fragment of MatK, since it is detected by a C-terminal tag. In sum, our analysis indicates that protein accumulation from the synthetic transgene is aberrant and thus does not represent an over-expressor. While the transgene is expressed on the RNA level, no full-length MatK was detected on the protein level. Still, AmatK leads to the accumulation of a well-defined N-terminally truncated MatK fragment and a tissue specific chloroplast defect. This variegated tissue was used for all further experiments, without further transgene induction by theophylline.

## Chloroplast RNA Splicing Is Unaffected in AmatK Lines

There is as yet no formal proof, but strong biochemical and phylogenetic evidence for MatK being a splicing factor. Splicing defects could explain the altered chloroplast development in AmatK, in particular since MatK associates with four tRNA genes (Zoschke et al., 2010), which are essential for translation and thus chloroplast development (Alkatib et al., 2012a,b). We therefore analyzed total RNA from young seedlings at the cotyledon stage for the accumulation of spliced and unspliced RNAs that are considered MatK targets based on previous RNA-co-immunoprecipitation assays (Zoschke et al., 2010). We include phenocopies generated by mild spectinomycin treatment to control for secondary effects caused by failed chloroplast development. Both, probes complementary to intron as well as exon sequences were used to detect spliced as well as unspliced RNAs. Accumulation of unspliced versus spliced RNAs would be indicative of a splicing effect. However, no such differential accumulation was observed in AmatK lines versus controls (**Figure 5**). The stronger signals for unspliced precursors seen for *trnK* and *trnV* in AmatK samples versus wt samples are also seen for the phenocopy control and appear to be caused by differences in loading (see methylene blue stains). In sum, RNA splicing of the analyzed target RNAs is not impaired, arguing against an RNA splicing effect leading to the observed variegation.

## AmatK Plants Accumulate RNA Antisense to *rpl33* When Treated With Spectinomycin

In the absence of specific splicing defects, we wanted to examine other parts of chloroplast RNA metabolism in AmatK lines. Therefore, we analyzed global chloroplast RNA accumulation by microarray analysis. To this end, we extracted RNA from AmatK seedlings as well as from spectinomycin-treated phenocopy controls. The AmatK and phenocopy control RNAs were labeled with Cy3 and Cy5 fluorescent dyes, respectively, and hybridized competitively on a microarray that represents the chloroplast genome of tobacco in a tiling fashion (Finster et al., 2013). However, none of the probes showed a significant change (**Figure 6A**, **Supplementary Table S1**, and

**Supplementary Figure S5**). Mild reductions were seen for several tRNAs and a mild, but non-significant increased steady-state level was observed for two mRNAs (marked in **Figure 6A**). Of these, we randomly picked the *rpl33-clpP* region for validation by RNA gel blot hybridization. We probed four genes located on opposite strands, using strand-specific probes (**Figures 6B,C**). The steady state levels of these RNAs were found to be similar between wild-type and AmatK plants, further supporting the microarray data.

Given that spectinomycin exacerbates the bleaching phenotype (**Figure 2C**), we tested RNA accumulation also on spectinomycin-containing medium. Indeed, transcript levels are dramatically increasing for *rpl33*, *rps18* and *clpP* under these conditions (**Figure 6B**). An increase can also be observed for *rps12*, in particular when considering that less total RNA is loaded in this lane. Intended as a negative control, we also probed for RNA accumulating antisense to *rpl33*. To our surprise, we obtained strong signals for AmatK lines treated with spectinomycin, whereas for all other samples, signals were barely above background.

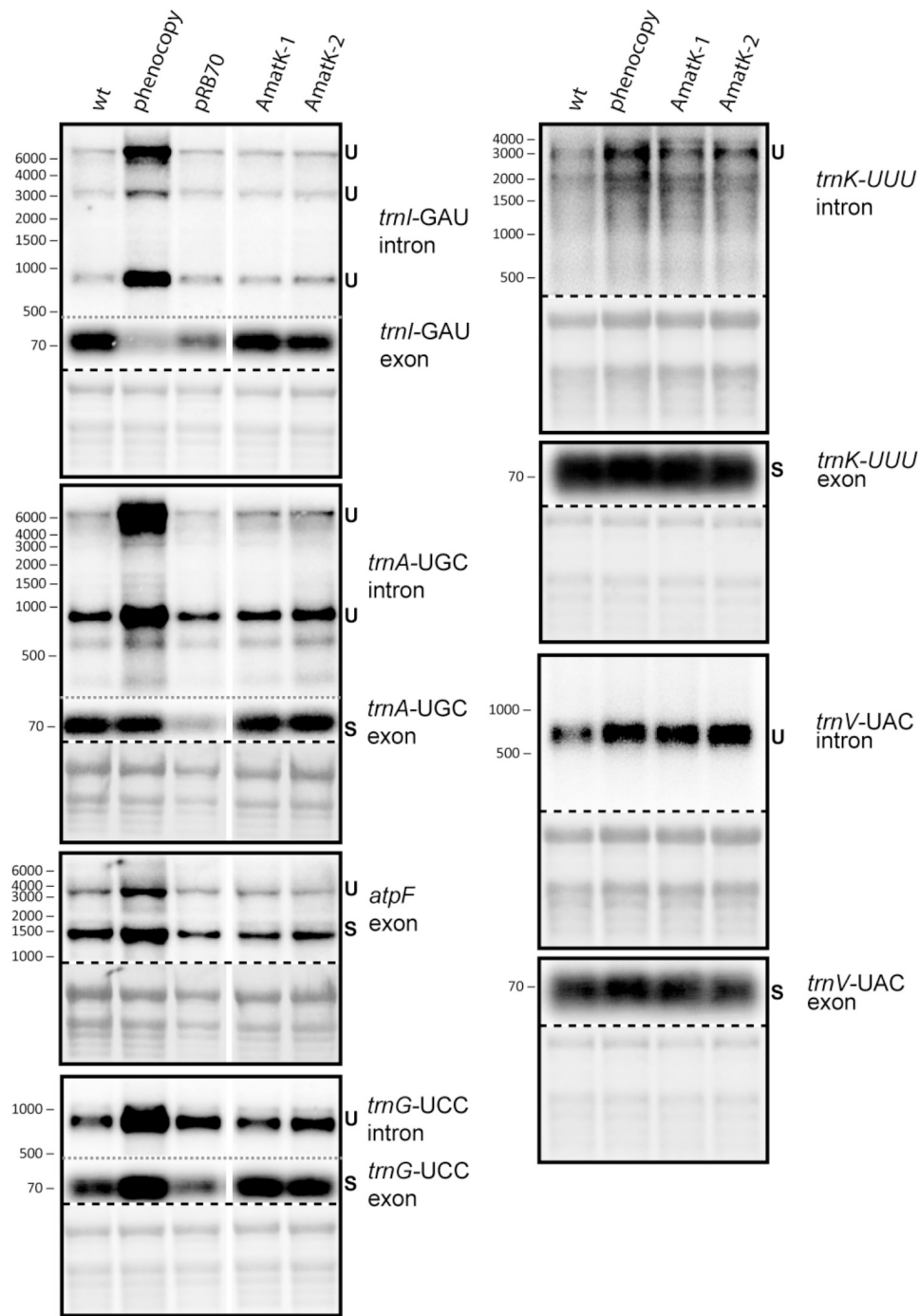
In sum, plastid RNA levels in AmatK plants are very similar to RNA levels in control plants, thus suggesting that the phenotype is not caused by altered RNA accumulation in the mutants. The situation changes drastically, when plants are treated with spectinomycin, which is also reflected in the more pronounced cotyledon bleaching: Spectinomycin induces over-accumulation of several mRNAs in AmatK lines and also leads to strongly increased antisense transcripts to the *rpl33* gene. In conclusion, inhibition of plastid translation in AmatK insertion lines leads to reprogramming of chloroplast gene expression at least for the *rpl33-clpP* genomic region.

## DISCUSSION

### AmatK-Induced Plastome-Based Leaf Variegation Is Modulated by External and Internal Cues

The differentiation of chloroplasts in land plants depends foremost on light and developmental-dependent signals (Lopez-Juez and Pyke, 2005). A multitude of mutants perturbed in chloroplast development have been isolated by forward and reverse genetic approaches (Leister, 2003). Such mutants usually display uniform bleaching phenotypes, and are often embryonic lethal or seedling lethal. Some mutants have variegated leaves. Leaf variegation is based on the occurrence of tissue sectors that either contain fully developed chloroplasts or pale, aberrant plastids (Kirk and Tilney-Bassett, 1978; Sakamoto, 2003; Aluru et al., 2006). Recessive mutants that cause variegation have been isolated in monocot as well as in dicot plants (e.g., Hagemann and Scholz, 1962; Walbot and Coe, 1979; Carol et al., 1999; Wang et al., 2000, 2004). These are typically caused by nuclear-loci. By contrast, AmatK plants represent a chloroplast genetic alteration that is maternally inherited. Plastome-based variegation can be caused by segregation of two genetically distinct plastids, one of



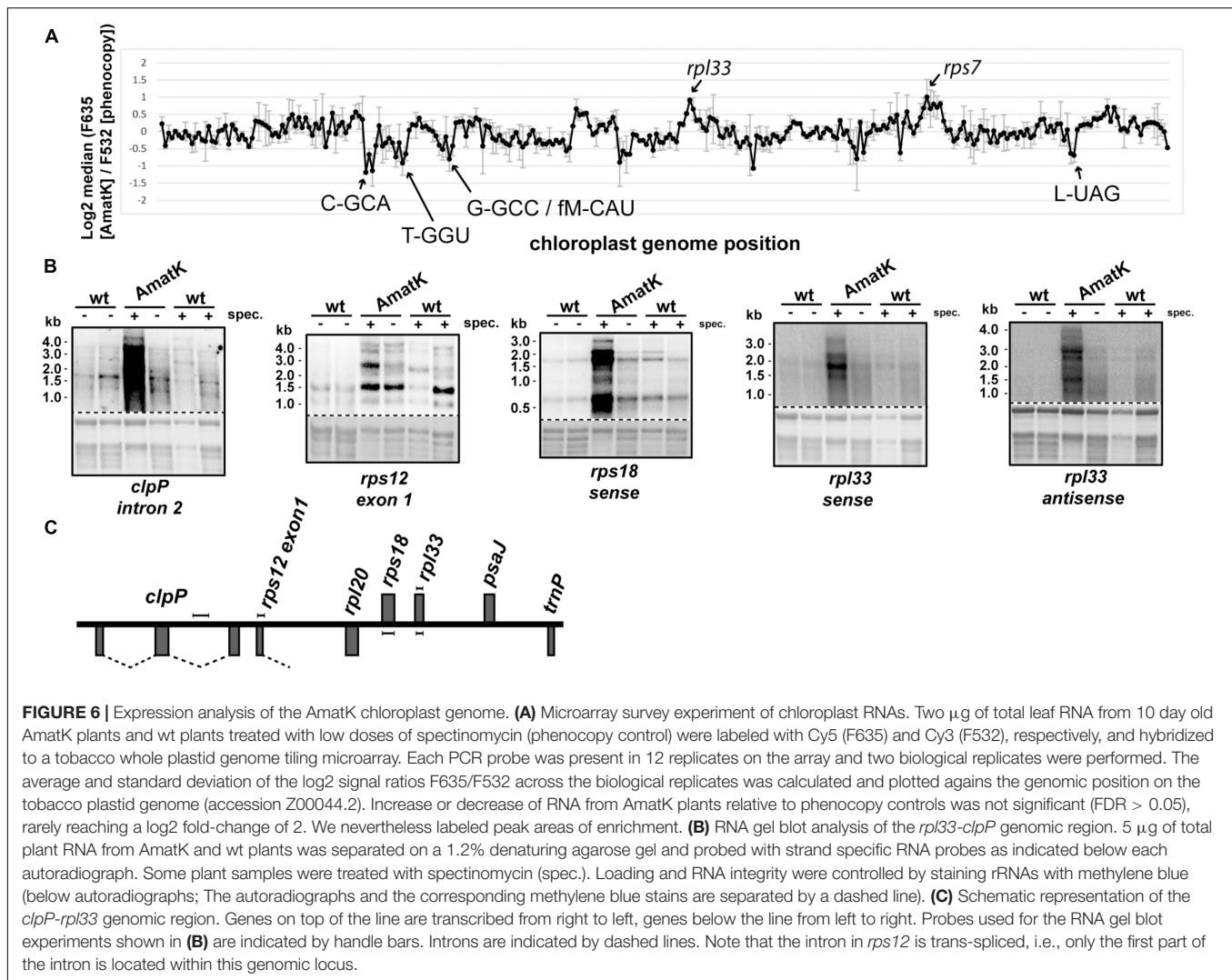


**FIGURE 5 |** Analysis of RNA splicing in AmatK and control plants. 5  $\mu$ g of total leaf RNA was separated on 1.2% agarose gels and analyzed by hybridization to probes specifying target RNAs of MatK (see **Supplementary Table S1** online for primer sequences). The following RNA samples were analyzed: RNA from AmatK, *aadA*, wt plants, and wt plants treated with spectinomycin (phenocopy control). Methylene blue staining identifies rRNAs and was carried out prior to hybridization. The autoradiographs and the corresponding methylene blue stains (below) are separated by a dashed line. Note that exon probeings for *trnG*, *trnA*, and *trnI* were done after the respective intron probeings on the same blots (separated by gray, dotted lines). White vertical lines indicate, where irrelevant lanes have been removed.

which does not develop into fully green chloroplasts (Baur, 1909; Correns, 1909; Börner and Sears, 1986; Greiner and Bock, 2013). It has also been described for the segregation of two genetically distinct plastid chromosomes, for example segregation of

chromosomes lacking a *trnN*-GUU gene versus wt chromosomes (Legen et al., 2007). Depletion of *trnN*-GUU pools during segregation of homoplastomic tissue was considered to cause leaf variegation. This cannot be the case with AmatK plants, since





these lines are homoplasmic according to Southern analyses. Thus, since *AmatK* plants are genetically uniform, variegation must be linked to a different, variable parameter. In *AmatK* cells, this parameter drops below a hypothetical threshold, which prevents chloroplast development. Three external factors, light, sugar and the antibiotic spectinomycin were shown to modulate the extent of variegation and are thus informative in terms of identifying the putative threshold. Light has been noted before as a booster of variegation (Rosso et al., 2009). Light intensity during growth correlates with pale leaf area in the variegation mutants *immutans*, *spotty*, *var1*, *var3*, and *thf1* (Rosso et al., 2009). This effect has been attributed to increased excitation pressure and the resulting photo-oxidative damage under higher light fluences (Rosso et al., 2009; Huang et al., 2013). In analogy, we hypothesize that the *AmatK* insertion makes developing chloroplasts more vulnerable to increased excitation pressure. Given the role of *MatK* as a splicing factor, altered gene expression in *AmatK* plants might interfere with the correct expression of the photosynthetic apparatus and thus lead to detrimental changes in electron transport that are exacerbated

under high light conditions. The nature of the primary defect remains to be determined.

Sugar as the second modifier of the *AmatK* phenotype has been shown to lead to the repression of nuclear genes for photosynthetic functions, e.g., the light harvesting proteins (Krapp et al., 1991; Price et al., 2004; Li et al., 2006). In addition, sugar represses plastid gene expression in liquid culture (Price et al., 2004; Gonzali et al., 2006; Osuna et al., 2007) and also in seedlings grown on sugar-containing MS media (Van Dingenen et al., 2016). Interestingly, the effect on chloroplast gene expression was observed as early as 3 h following sucrose treatment – roughly half of all repressed genes are from the chloroplast genome (Van Dingenen et al., 2016). Notably, *matK* and *rps18* are the most strongly reacting non-photosynthetic chloroplast genes after sugar treatment (Table 1 in Van Dingenen et al., 2016). We hypothesize that the detrimental effect of *AmatK* on chloroplast gene expression adds up with the repressive effects of sugars on chloroplast gene expression (including on the endogenous *matK*), which ultimately results in increased variegation. This implies a threshold of plastid gene expression

during chloroplast development, below which no differentiation of chloroplasts is possible. This is further supported by our results from spectinomycin treatments, which showed an even more pronounced effect on variegation than light and sugar application. Spectinomycin blocks chloroplast translation, but should be detoxified by the expression of the *aadA* cassette in *AmatK* plants. This function is compromised given the almost complete bleaching of cotyledons on spectinomycin-containing medium. In conclusion, the expression of the *aadA* cassette is hampered in the *AmatK* background. Expression of the *aadA* cassette is driven by a PEP promoter and requires an active and intact plastid translational machinery. In summary, our data suggest that the *AmatK* insertion negatively affects plastid gene expression, which in some cotyledon cells is lowered beyond a point necessary for chloroplast development.

### ***AmatK* Is Involved in Cotyledon-Specific Chloroplast Development**

The main difference between cotyledons and primary leaves is their developmental origin. While all primary leaves are derived from the shoot apical meristem (SAM) post-germination, cotyledons emerge already at the late globular stage of embryo development. This is a time in development, when seedlings are submerged in the soil and live heterotrophically. The plastids in cotyledons develop into etioplasts and only when they emerge from the soil and are irradiated, etioplasts transform into chloroplasts (Mansfield and Briarty, 1996). This is a very different trajectory than plastids take in the SAM, where they differentiate from proplastids directly into chloroplasts during leaf primordia development. The genetic basis of chloroplast development in cotyledons is mostly unknown, but a smaller number of mutants have been described that display pale or even albino cotyledons, while the primary leaves are able to undergo normal greening. This includes *white cotyledons* (*wco*; Yamamoto et al., 2000), and *snowy cotyledon 1, 2, and 3* (*sc1,2,3*; Albrecht et al., 2006, 2008, 2010; Ruppel and Hangarter, 2007; Shimada et al., 2007; Zagari et al., 2017). Other mutants, including *sig6* and *dg1* mutants, exert a dominant effect on cotyledons, but also lead to measurable alterations in chlorophyll content in primary leaves (Ishizaki et al., 2005; Chi et al., 2008). Only one mutant has been described to date that leads to cotyledon-specific variegation, caused by a disruption of the *spd1* gene (Ruppel et al., 2011). All of the listed mutants are recessive and affect nuclear genes. *AmatK* is distinct in eliciting cotyledon variegation via a plastome modification. It is interesting that – like MatK – several of the mutated nuclear genes underlying variegation phenotypes are involved in plastid gene expression: *WCO* and *SCO1* have been implicated in rRNA maturation and translation elongation, respectively. *SIG6* and *DG1* interact with each other and play a role for plastid transcription. This demonstrates that defects in chloroplast gene expression can selectively affect chloroplast development in cotyledons – a possible scenario also for the *AmatK* lines. *SCO2* encodes a DNAJ-like protein and is required for the accumulation of photosystem II-LHCII complexes (Zagari et al., 2017). This mechanism for variegation

is likely distinct from *AmatK*-induced variegation in that it does not involve plastid gene expression and since recent data indicate functions for *SCO2* beyond cotyledons (Zagari et al., 2017). The functions of *SPD1* and *SCO3* are unknown. In sum, while the molecular mechanisms behind cotyledon-specific variegation remain unknown, it is intriguing that of the few factors known, most play a role for plastid gene expression. Thus, chloroplasts in cotyledons depend on a yet to be determined threshold of chloroplast gene expression, that is no longer crossed in all plastids during development in variegation mutants.

### **A Gain of Function in *AmatK* Lines Impairs Chloroplast Development**

The initial aim of our study was to generate a strong overexpressor of MatK. However, while *AmatK* mRNA is readily detected, we did not find any full-length MatK protein derived from the transgene, but instead only a weakly accumulating degradation product. Since the tag used for the detection of the degradation product is located at MatK's C-terminus, we conclude that translation proceeds to the stop codon, and that the protein is degraded co- or post-translationally. This is surprising, since the amino acid sequence of *AmatK* is identical to the stable, endogenously HA-tagged version of MatK (Zoschke et al., 2010). We speculate that translation dynamics of the *AmatK* mRNA are different from those of the endogenous *matK* mRNA since the two genes have massively different codon usage (**Supplementary Figure S1**). Changes in translation speed can lead to different folding kinetics of the nascent protein and thus to alternative degradation routes (Kimchi-Sarfaty et al., 2007; Gloge et al., 2014). We speculate that such a truncated protein could act as a dominant negative factor in the chloroplast. However, the truncated *AmatK* protein levels are not rising after theophylline treatment, while at the same time variegation increases. Thus, the N-terminally truncated *AmatK* cannot be the cause of increased defective chloroplast development after theophylline addition. We, however, cannot rule out that an *AmatK* protein fragment representing the N-terminal part of *AmatKs* exerts the effect. Such a fragment would, however, not be detectable, because it would lack the C-terminal HA tag and thus remains a speculation at present. Alternative explanations, i.e., that the *AmatK* gene or the *AmatK* mRNA cause the defect, cannot be ruled out, but are also not in line with the theophylline-dependence of the phenotype. While this needs to be tested in the future, our data suggest that the downstream effects represent a gain of function rather than a loss of function of MatK. First, splicing of MatK target introns appears unaffected in *AmatK* plants: If we agree to take MatK as a splicing factor (Zoschke et al., 2010), these data suggest that *AmatK* does not compete or interfere with the endogenous MatK. Second, we discovered a number of effects on the accumulation of RNAs that are dependent on inhibiting translation in *AmatK* plants. These RNAs are not associated with MatK in RIP-chip assays. For instance, the *rp133* operon accumulates more RNA in *AmatK* plants than in phenocopy controls, including the accumulation of antisense RNA. In the green alga *Chlamydomonas reinhardtii*, chloroplast

antisense RNAs are stabilized in the absence of translation by forming double-stranded RNA (Cavauiuolo et al., 2017). In general, over-accumulation of RNAs has been described for mutants with global defects in chloroplast translation and is caused at least in part by an increased activity of selected chloroplast RNA polymerases in albino tissue (Emanuel et al., 2004). We tried to control for such secondary effects by analyzing phenocopy controls, but in fact cannot be sure, whether the extent of translational inhibition in the control is truly mimicking the effect seen in AmatK plants. Therefore, whether the RNA defects seen in AmatK tissue are primary or secondary effects cannot be decided at present. Still, a defect in the expression of *rpl33* and its neighboring genes, *rps18* and *rpl20*, could very well explain the bleaching phenotype. *rps18* and *rpl20* encode essential ribosomal proteins (Rogalski et al., 2006; Rogalski et al., 2008). The antisense RNA detected here is long enough to also make hybrids with *rps18* and *rpl20* mRNAs. This could interfere with sense translation of both mRNAs and thus eventually to reduced ribosome biogenesis and translation capacity. That reduced translational capacity can lead to bleached tissue is exemplified by spectinomycin treatment and several mutants in genes for translational components (e.g., Bellaoui and Gruissem, 2004; Legen et al., 2007; Meurer et al., 2017). Future transplastomic expression of antisense RNA to selected mRNAs could test this hypothesis. Importantly, the option to modulate the level of variegation in cotyledons of AmatK lines makes these plants a valuable tool to determine the time-window and unravel the threshold at which chloroplast gene expression exerts its effect on plastid differentiation.

## MATERIALS AND METHODS

### Plant Material and Growth Conditions

Tobacco plants (*N. tabacum* cv. *Petit Havana*) were grown on agar medium containing 30 g/L sucrose for plastid transformation assays. Transplastomic AmatK or pRB70 plants were grown on the same medium or on medium supplemented with 500 mg/L spectinomycin or supplemented with various concentrations of theophylline. wt plants grown on medium with 17 mg/L spectinomycin served as phenocopy controls. For light treatments, plants were grown under standard conditions in a walk-in chamber (25°C, humidity 55%, 16 h light/8 dark). Homoplastomy of the regenerates and seedlings of T1 generation were tested by Southern Blot Analysis and/or germination assay on a medium containing 500 mg/L spectinomycin.

### Construction of the Synthetic *matK* (AmatK) Overexpression Vector

The tobacco plastid rRNA operon promoter Prn and riboswitch RS were amplified from plasmid pAV6 (Verhounig et al., 2010) with primer pair SacI 5'-GAACAAAAGCTGGAGCTCG-3' (SacI site underlined) and AmatKRSr 5'-GTAACGCTGGATTCTTCCATATGATCCTCTCCACGAGA G-3' (sequence overlap with AmatK underlined). The PCR product was ligated to the AmatK fragment via

overlap PCR with primer pair SacI and HA5AmatK 5'-CATAATCAGGAACATCATAAGGATACTGGTAGTTTGCCA GGTC-3' (sequence overlap with HA underlined). The resulting PCR product was ligated to the 3XHA tag (Zoschke et al., 2010) via overlap PCR with primer pair SacI and pAVHAR 5'-CCTTAATTGAATTCTCTAGAGCCTCATTAAGCATAATCA GGTACATC-3' (XbaI site underlined). The resulting PCR product and pAV6 vector were digested with SacI and XbaI and ligated, resulting in the pRSAmatK vector.

### Plastid Transformation

Stable transformation of plastids was performed by particle bombardment followed by spectinomycin selection. The transformation protocol was carried out according to the protocols of Svab (Svab and Maliga, 1993) and Okuzaki (Okuzaki and Tabei, 2012). Two rounds of transformation were carried out; in the first round, two transplastomic lines with variegated cotyledons were retrieved, in the second round more than a dozen independent transplastomic lines with cotyledon variegation were isolated. Briefly, sterile tobacco leaves from wild type plants cultured on MS were cut into 0.5 cm × 0.5 cm squares. The leaf pieces were placed onto MS plates with the lower leaf side up and cultured overnight. The particle bombardment was carried out using the Biolistic PDS-1000/He Particle Delivery System (1,100 psi, L2 = 6 cm, 10 shots per construct, Bio-Rad) with gold particles (0.6 μM, Bio-Rad) loaded with the transformation plasmids. Spectinomycin resistant calli were regenerated on RMOP medium. Several putative transplastomic lines were obtained and further rounds of selection were carried out to generate homoplastomic lines. The correct integration of AmatK was verified by genotyping PCR and Southern blot. The PCR was performed with the primer pair AmatKseqf 5'-GTTCTCGCATTTGGATCTC-3' and trnMf 5'-TTCAAATCCTGTCTCCGCAA-3'.

### Isolation of DNA and Southern Blot

Genomic DNA was isolated from leaves according to standard protocols (Murray and Thompson, 1980). For Southern blot analysis, *Hind*III-digested total plant DNA was separated on a 0.8% agarose gel, blotted to a Nylon membrane and probed with a PCR product that was body-labeled using α-<sup>32</sup>P-CTP. The Decaprime DNA labeling Kit (ThermoFisher, Berlin, Germany) was used to label PCR-generated probe DNA fragments (see **Supplementary Table S1** for primers). Hybridization was carried out at 55°C overnight using standard protocols (Sambrook et al., 1989).

### Isolation of RNA and RNA Gel Blot Hybridization

Total cellular RNA was extracted using TriZol (ThermoFisher, Berlin, Germany). 5 μg of total cellular RNA was separated on formaldehyde-containing 1.2% agarose gels and blotted to Hybond N membrane (GE healthcare, Munich, Germany). Strand-specific RNA probes were synthesized using T7 Polymerase (ThermoFisher, Berlin, Germany) with radioactive γ-<sup>32</sup>P-UTP



(Hartmann Analytic) using PCR products as templates (see **Supplementary Table S1** for primer sequences). Strand-specific DNA probes were generated with the DecaLabel kit and PCR products according to the manufacturer's instructions (ThermoFisher, Berlin, Germany; see **Supplementary Table S1** for primer sequences). Alternatively, oligonucleotides were end-labeled using T4 polynucleotide kinase (ThermoFisher, Berlin, Germany) and  $\gamma$ - $^{32}$ P-UTP (Hartmann Analytic) and used as probes (see **Supplementary Table S1** for oligonucleotide sequences). Hybridizations were done according to standard procedures (Sambrook et al., 1989). Signals were detected using a PMI Imaging system (Bio-Rad, Munich, Germany).

## Immunoblot Analyses

Protein extraction was performed according to standard protocols (Schagger and von Jagow, 1987). Protein fractions were separated on polyacrylamide gels and subsequently transferred in Towbin buffer (Towbin et al., 1979) to a PVDF membrane (0.2  $\mu$ m pore size; GE Healthcare, Munich, Germany). Integrity and equal loading of proteins was detected by Ponceau S staining of the membrane. MatK:HA detection was performed with a rat HA-specific antibody (Roche, Mannheim, Germany). Chemoluminescent signals were detected using a ChemiDoc system (BioRad, Munich, Germany).

## Microarray Analysis

2  $\mu$ g of total cellular RNA from 10 day old AmatK and phenocopy control plants with red- (Cy5) and green-fluorescing (Cy3) dyes, respectively, using the ULS RNA Fluorescent Labeling Kit according to the manufacturer's instructions (Kreatech, Amsterdam, Netherlands). The samples were hybridized to a whole-genome tobacco chloroplast tiling microarrays (Finster et al., 2013). Hybridization, array washings and signal detection were carried out as described previously (Kupsch et al., 2012).

Microarray analysis was done in R using the packages *marray* (version 1.52) and *multtest* (version 2.28). Namely, expression values were loaded in R using *marray*. After this, loess normalization at the probe level was applied using *maNorm* function from the *marray* package for the two biological replicates performed. The array contains 12 copies of each probe located at different spots to reduce any bias regarding its location on the array; the median log2 fold-change expression for the 12 copies was used to calculate the log2 fold-change

of the region represented by the probe for each biological replicate independently. A *t*-test was applied to calculate which regions show an average log2 fold-change different of zero. Multiple hypotheses correction was applied using the Benjamini & Hochberg method (BH) implemented in the *multitest* package.

## AUTHOR CONTRIBUTIONS

YQ designed the transformation vectors, performed chloroplast transformation and selection for homoplasmic lines, and performed genetic crosses. JA and AW analyzed transgene expression and splicing efficiency by RNA gel blot hybridization. JL performed genotyping PCRs and Southern analyses. SH and GH performed RNA gel blot analyses and microarray analyses. CT performed immunological analyses. GR participated in the screening for transplastomic plants. JM and UO evaluated the microarrays and contributed to manuscript preparation. GW supported the design of the transformation vector and immunological analyses. OO contributed to project design and manuscript preparation. CS-L conceived the study, prepared the figures, and wrote the manuscript.

## FUNDING

Generous funding by the DFG to CS-L (TR175; project A02) and by GIF to CS-L and OO (I-1213-319.13/2012) is gratefully acknowledged. YQ was supported by a fellowship from the China Scholarship Council.

## ACKNOWLEDGMENTS

Technical support by T. Börner (HU Berlin) is gratefully acknowledged. We are grateful to Ralph Bock and Stephanie Ruf for making available the pAV6 vector and pRB70 plants.

## SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fpls.2018.01453/full#supplementary-material>

## REFERENCES

- Albrecht, V., Ingenfeld, A., and Apel, K. (2006). Characterization of the snowy cotyledon 1 mutant of *Arabidopsis thaliana*: the impact of chloroplast elongation factor G on chloroplast development and plant vitality. *Plant Mol. Biol.* 60, 507–518. doi: 10.1007/s11103-005-4921-0
- Albrecht, V., Ingenfeld, A., and Apel, K. (2008). Snowy cotyledon 2: the identification of a zinc finger domain protein essential for chloroplast development in cotyledons but not in true leaves. *Plant Mol. Biol.* 66, 599–608. doi: 10.1007/s11103-008-9291-y
- Albrecht, V., Simkova, K., Carrie, C., Delannoy, E., Giraud, E., Whelan, J., et al. (2010). The cytoskeleton and the peroxisomal-targeted snowy cotyledon3 protein are required for chloroplast development in *Arabidopsis*. *Plant Cell* 22, 3423–3438. doi: 10.1105/tpc.110.074781
- Alkatib, S., Fleischmann, T. T., Scharff, L. B., and Bock, R. (2012a). Evolutionary constraints on the plastid tRNA set decoding methionine and isoleucine. *Nucleic Acids Res.* 40, 6713–6724. doi: 10.1093/nar/gks350
- Alkatib, S., Scharff, L. B., Rogalski, M., Fleischmann, T. T., Matthes, A., Seeger, S., et al. (2012b). The contributions of wobbling and superwobbling to the reading of the genetic code. *PLoS Genet.* 8:e1003076. doi: 10.1371/journal.pgen.1003076
- Aluru, M. R., Yu, F., Fu, A., and Rodermel, S. (2006). *Arabidopsis* variegation mutants: new insights into chloroplast biogenesis. *J. Exp. Bot.* 57, 1871–1881. doi: 10.1093/jxb/erj008
- Barthet, M. M., and Hilu, K. W. (2007). Expression of matK: functional and evolutionary implications. *Am. J. Bot.* 94, 1402–1412. doi: 10.3732/ajb.94.8.1402
- Baur, E. (1909). Das Wesen und die Erblichkeitsverhältnisse der "Varietates albomarginatae hort." von Pelargonium zonale. *Z. Indukt. Abstamm. Vererbungs.* 1, 330–351.



- Bellaoui, M., and Gruissem, W. (2004). Altered expression of the Arabidopsis ortholog of DCL affects normal plant development. *Planta* 219, 819–826. doi: 10.1007/s00425-004-1295-5
- Bellet, S., and Renner, S. S. (2015). The plastomes of two species in the endoparasite genus *pilostyles* (Apodanthaceae) each retain just five or six possibly functional genes. *Genome Biol. Evol.* 8, 189–201. doi: 10.1093/gbe/evv251
- Börner, T., and Sears, B. B. (1986). Plastome mutants. *Plant Mol. Biol. Rep.* 4, 69–92. doi: 10.1007/BF02732104
- Carol, P., Stevenson, D., Bisanz, C., Breitenbach, J., Sandmann, G., Mache, R., et al. (1999). Mutations in the Arabidopsis gene IMMUTANS cause a variegated phenotype by inactivating a chloroplast terminal oxidase associated with phytoene desaturation. *Plant Cell* 11, 57–68. doi: 10.1105/tpc.11.1.57
- Cavauiolo, M., Kuras, R., Wollman, F. A., Choquet, Y., and Vallon, O. (2017). Small RNA profiling in *Chlamydomonas*: insights into chloroplast RNA metabolism. *Nucleic Acids Res.* 45, 10783–10799. doi: 10.1093/nar/gkx668
- Chi, W., Ma, J., Zhang, D., Guo, J., Chen, F., Lu, C., et al. (2008). The pentatricopeptide repeat protein delayed greening1 is involved in the regulation of early chloroplast development and chloroplast gene expression in Arabidopsis. *Plant Physiol.* 147, 573–584. doi: 10.1104/pp.108.116194
- Cohen, S., Zmudjak, M., Colas des Francs-Small, C., Malik, S., Shaya, F., Keren, I., et al. (2014). nMAT4, a maturase factor required for nad1 pre-mRNA processing and maturation, is essential for holocomplex I biogenesis in Arabidopsis mitochondria. *Plant J.* 78, 253–268. doi: 10.1111/tjp.12466
- Correns, C. (1909). Vererbungsversuche mit blaß(gelb)grünen und buntblättrigen Sippen bei *Mirabilis jalapa*, *Urtica pilulifera* und *Lunaria annua*. *Z. Indukt. Abstamm. Vererbungslehre* 1, 291–329.
- de Longevialle, A. F., Small, I. D., and Lurin, C. (2010). Nuclear encoded splicing factors implicated in RNA splicing in higher plant organelles. *Mol. Plant* 3, 691–705. doi: 10.1093/mp/ssq025
- Delannoy, E., Fujii, S., Colas des Francs-Small, C., Brundrett, M., and Small, I. (2011). Rampant gene loss in the underground orchid *Rhizanthella gardneri* highlights evolutionary constraints on plastid genomes. *Mol. Biol. Evol.* 28, 2077–2086. doi: 10.1093/molbev/msr028
- Đlakic, M., and Mushegian, A. (2011). Prp8, the pivotal protein of the spliceosomal catalytic center, evolved from a retroelement-encoded reverse transcriptase. *RNA* 17, 799–808. doi: 10.1261/rna.2396011
- Drescher, A. (2003). *ycf1, ycf14 und RNA-Edierung: Untersuchungen an im Lauf der Plastidenevolution neu Hinzugewonnenen Genen und Eigenschaften*. Ph.D. thesis, LMU München, München.
- Emadpour, M., Karcher, D., and Bock, R. (2015). Boosting riboswitch efficiency by RNA amplification. *Nucleic Acids Res.* 43:e66. doi: 10.1093/nar/gkxv165
- Emanuel, C., Weihe, A., Graner, A., Hess, W. R., and Börner, T. (2004). Chloroplast development affects expression of phage-type RNA polymerases in barley leaves. *Plant J.* 38, 460–472. doi: 10.1111/j.0960-7412.2004.02060.x
- Finster, S., Eggert, E., Zoschke, R., Weihe, A., and Schmitz-Linneweber, C. (2013). Light-dependent, plastome-wide association of the plastid-encoded RNA polymerase with chloroplast DNA. *Plant J.* 76, 849–860. doi: 10.1111/tjp.12339
- Funk, H. T., Berg, S., Krupinska, K., Maier, U. G., and Krause, K. (2007). Complete DNA sequences of the plastid genomes of two parasitic flowering plant species, *Cuscuta reflexa* and *Cuscuta gronovii*. *BMC Plant Biol.* 7:45. doi: 10.1186/1471-2229-7-45
- Galej, W. P., Oubridge, C., Newman, A. J., and Nagai, K. (2013). Crystal structure of Prp8 reveals active site cavity of the spliceosome. *Nature* 493, 638–643. doi: 10.1038/nature11843
- Gloge, F., Becker, A. H., Kramer, G., and Bukau, B. (2014). Co-translational mechanisms of protein maturation. *Curr. Opin. Struct. Biol.* 24, 24–33. doi: 10.1016/j.sbi.2013.11.004
- Gonzali, S., Loreti, E., Solfanelli, C., Novi, G., Alpi, A., and Perata, P. (2006). Identification of sugar-modulated genes and evidence for in vivo sugar sensing in Arabidopsis. *J. Plant Res.* 119, 115–123. doi: 10.1007/s10265-005-0251-1
- Greiner, S., and Bock, R. (2013). Tuning a menage a trois: co-evolution and co-adaptation of nuclear and organellar genomes in plants. *Bioessays* 35, 354–365. doi: 10.1002/bies.201200137
- Hagemann, R., and Scholz, F. (1962). Ein fall gen-induzierter mutationen des plasmotyps bei gerste. *Der. Züchter* 32, 50–59.
- Hdider, C., and Desjardins, Y. (1994). Effects of sucrose on photosynthesis and phosphoenolpyruvate carboxylase activity of in vitro cultured strawberry plantlets. *Plant Cell Tissue Organ Cult.* 36, 27–33. doi: 10.1007/BF00048312
- Hess, W. R., Hoch, B., Zeltz, P., Hübschmann, T., Kössel, H., and Börner, T. (1994). Inefficient rpl2 splicing in barley mutants with ribosome-deficient plastids. *Plant Cell* 6, 1455–1465. doi: 10.1105/tpc.6.10.1455
- Huang, W., Chen, Q., Zhu, Y., Hu, F., Zhang, L., Ma, Z., et al. (2013). Arabidopsis thylakoid formation 1 is a critical regulator for dynamics of PSII-LHCII complexes in leaf senescence and excess light. *Mol. Plant* 6, 1673–1691. doi: 10.1093/mp/ss069
- Ishizaki, Y., Tsunoyama, Y., Hatano, K., Ando, K., Kato, K., Shinmyo, A., et al. (2005). A nuclear-encoded sigma factor, Arabidopsis SIG6, recognizes sigma-70 type chloroplast promoters and regulates early chloroplast development in cotyledons. *Plant J.* 42, 133–144. doi: 10.1111/j.1365-313X.2005.02362.x
- Keren, I., Bezawork-Geleta, A., Kolton, M., Maayan, I., Belausov, E., Levy, M., et al. (2009). AtnMat2, a nuclear-encoded maturase required for splicing of group-II introns in Arabidopsis mitochondria. *RNA* 15, 2299–2311. doi: 10.1261/rna.1776409
- Keren, I., Tal, L., des Francs-Small, C. C., Araujo, W. L., Shevtsov, S., Shaya, F., et al. (2012). nMAT1, a nuclear-encoded maturase involved in the trans-splicing of nad1 intron 1, is essential for mitochondrial complex I assembly and function. *Plant J.* 71, 413–426. doi: 10.1111/j.1365-313X.2012.04998.x
- Kimchi-Sarfaty, C., Oh, J. M., Kim, I. W., Sauna, Z. E., Calcagno, A. M., Ambudkar, S. V., et al. (2007). A “silent” polymorphism in the MDR1 gene changes substrate specificity. *Science* 315, 525–528. doi: 10.1126/science.1135308
- Kirk, J., and Tilney-Bassett, R. (1978). *The Plastids: Their Chemistry, Structure, Growth, and Inheritance*. New York, NY: Elsevier/North Holland Biomedical Press.
- Kleine, T., Maier, U. G., and Leister, D. (2009). DNA transfer from organelles to the nucleus: the idiosyncratic genetics of endosymbiosis. *Annu. Rev. Plant Biol.* 60, 115–138. doi: 10.1146/annurev.arplant.043008.092119
- Koch, K. E. (1996). Carbohydrate-modulated gene expression in plants. *Annu. Rev. Plant Physiol. Plant Mol. Biol.* 47, 509–540. doi: 10.1146/annurev.arplant.47.1.509
- Krapp, A., Quick, W. P., and Stitt, M. (1991). Ribulose-1,5-bisphosphate carboxylase-oxygenase, other Calvin-cycle enzymes, and chlorophyll decrease when glucose is supplied to mature spinach leaves via the transpiration stream. *Planta* 186, 58–69. doi: 10.1007/BF00201498
- Kupsch, C., Ruwe, H., Gusewski, S., Tillich, M., Small, I., and Schmitz-Linneweber, C. (2012). Arabidopsis chloroplast RNA binding proteins CP31A and CP29A associate with large transcript pools and confer cold stress tolerance by influencing multiple chloroplast RNA processing steps. *Plant Cell* 10, 4266–4280. doi: 10.1105/tpc.112.103002
- Lambowitz, A. M., and Zimmerly, S. (2004). Mobile group II introns. *Annu. Rev. Genet.* 38, 1–35. doi: 10.1146/annurev.genet.38.072902.091600
- Legen, J., Wanner, G., Herrmann, R. G., Small, I., and Schmitz-Linneweber, C. (2007). Plastid tRNA genes trnC-GCA and trnN-GUU are essential for plant cell development. *Plant J.* 51, 751–762. doi: 10.1111/j.1365-313X.2007.03177.x
- Leister, D. (2003). Chloroplast research in the genomic age. *Trends Genet.* 19, 47–56. doi: 10.1016/S0168-9525(02)00003-3
- Li, Y., Lee, K. K., Walsh, S., Smith, C., Hadingham, S., Sorefan, K., et al. (2006). Establishing glucose- and ABA-regulated transcription networks in Arabidopsis by microarray analysis and promoter classification using a relevance vector machine. *Genome Res.* 16, 414–427. doi: 10.1101/gr.4237406
- Liere, K., and Link, G. (1995). RNA-binding activity of the matK protein encoded by the chloroplast trnK intron from mustard (*Sinapis alba* L.). *Nucleic Acids Res.* 23, 917–921. doi: 10.1093/nar/23.6.917
- Lopez-Juez, E., and Pyke, K. A. (2005). Plastids unleashed: their development and their integration in plant development. *Int. J. Dev. Biol.* 49, 557–577. doi: 10.1387/ijdb.051997el
- Mansfield, S. G., and Briarty, L. G. (1996). The dynamics of seedling and cotyledon cell development in *Arabidopsis thaliana* during reserve mobilization. *Int. J. Plant Sci.* 157, 280–295. doi: 10.1086/297347
- Matsuura, M., Noah, J., and Lambowitz, A. (2001). Mechanism of maturase-promoted group II intron splicing. *EMBO J.* 20, 7259–7270. doi: 10.1093/emboj/20.24.7259
- McNeal, J. R., Arumugunathan, K., Kuehl, J. V., Boore, J. L., and Depamphilis, C. W. (2007). Systematics and plastid genome evolution of the cryptically

- photosynthetic parasitic plant genus *Cuscuta* (Convolvulaceae). *BMC Biol.* 5:55. doi: 10.1186/1741-7007-5-55
- McNeal, J. R., Kuehl, J. V., Boore, J. L., Leebens-Mack, J., and dePamphilis, C. W. (2009). Parallel loss of plastid introns and their maturation in the genus *Cuscuta*. *PLoS One* 4:e5982. doi: 10.1371/journal.pone.0005982
- Meurer, J., Schmid, L. M., Stoppel, R., Leister, D., Brachmann, A., and Manavski, N. (2017). PALE CRESS binds to plastid RNAs and facilitates the biogenesis of the 50S ribosomal subunit. *Plant J.* 92, 400–413. doi: 10.1111/tpj.13662
- Michel, F., and Ferat, J. L. (1995). Structure and activities of group II introns. *Annu. Rev. Biochem.* 64, 435–461. doi: 10.1146/annurev.bi.64.070195.002251
- Mohr, G., and Lambowitz, A. M. (2003). Putative proteins related to group II intron reverse transcriptase/maturases are encoded by nuclear genes in higher plants. *Nucleic Acids Res.* 31, 647–652. doi: 10.1093/nar/gkg153
- Mohr, G., Perlman, P. S., and Lambowitz, A. M. (1993). Evolutionary relationships among group II intron-encoded proteins and identification of a conserved domain that may be related to maturase function. *Nucleic Acids Res.* 21, 4991–4997. doi: 10.1093/nar/21.22.4991
- Molina, J., Hazzouri, K. M., Nickrent, D., Geisler, M., Meyer, R. S., Pentony, M. M., et al. (2014). Possible loss of the chloroplast genome in the parasitic flowering plant *Rafflesia lagascae* (Rafflesiaceae). *Mol. Biol. Evol.* 31, 793–803. doi: 10.1093/molbev/msu051
- Murray, M. G., and Thompson, W. F. (1980). Rapid isolation of high molecular weight plant DNA. *Nucleic Acids Res.* 8, 4321–4325. doi: 10.1093/nar/8.19.4321
- Nakagawa, N., and Sakurai, N. (2006). A mutation in At-nMat1a, which encodes a nuclear gene having high similarity to group II intron maturase, causes impaired splicing of mitochondrial NAD4 transcript and altered carbon metabolism in *Arabidopsis thaliana*. *Plant Cell Physiol.* 47, 772–783. doi: 10.1093/pcp/pcj051
- Neuhaus, H., and Link, G. (1987). The chloroplast tRNA(Lys)(UUU) gene from mustard (*Sinapis alba*) contains a class II intron potentially encoding for a maturase-related polypeptide. *Curr. Genet.* 11, 251–257. doi: 10.1007/BF00355398
- Okuzaki, A., and Tabei, Y. (2012). Improvement of the plastid transformation protocol by modifying tissue treatment at pre- and post-bombardment in tobacco. *Plant Biotechnol.* 29, 307–310. doi: 10.5511/plantbiotechnology.12.0319a
- Osuna, D., Usadel, B., Morcuende, R., Gibon, Y., Blasing, O. E., Hohne, M., et al. (2007). Temporal responses of transcripts, enzyme activities and metabolites after adding sucrose to carbon-deprived *Arabidopsis* seedlings. *Plant J.* 49, 463–491. doi: 10.1111/j.1365-3113X.2006.02979.x
- Pego, J. V., Kortstee, A. J., Huijser, C., and Smeekens, S. C. M. (2000). Photosynthesis, sugars and the regulation of gene expression. *J. Exp. Bot.* 51(Suppl. 1), 407–416. doi: 10.1093/jexbot/51.suppl\_1.407
- Piccirilli, J. A., and Staley, J. P. (2016). Reverse transcriptases lend a hand in splicing catalysis. *Nat. Struct. Mol. Biol.* 23, 507–509. doi: 10.1038/nsmb.3242
- Price, J., Laxmi, A., St Martin, S. K., and Jang, J. C. (2004). Global transcription profiling reveals multiple sugar signal transduction mechanisms in *Arabidopsis*. *Plant Cell* 16, 2128–2150. doi: 10.1105/tpc.104.022616
- Qu, G., Kaushal, P. S., Wang, J., Shigematsu, H., Piazza, C. L., Agrawal, R. K., et al. (2016). Structure of a group II intron in complex with its reverse transcriptase. *Nat. Struct. Mol. Biol.* 23, 549–557. doi: 10.1038/nsmb.3220
- Rambo, R. P., and Doudna, J. A. (2004). Assembly of an active group II intron-maturase complex by protein dimerization. *Biochemistry* 43, 6486–6497. doi: 10.1021/bi049912u
- Rodermel, S. (2002). *Arabidopsis* variegation mutants. *Arabidopsis Book* 1:e0079. doi: 10.1199/tab.0079
- Rogalski, M., Ruf, S., and Bock, R. (2006). Tobacco plastid ribosomal protein S18 is essential for cell survival. *Nucleic Acids Res.* 34, 4537–4545. doi: 10.1093/nar/gkl634
- Rogalski, M., Schottler, M. A., Thiele, W., Schulze, W. X., and Bock, R. (2008). Rpl33, a nonessential plastid-encoded ribosomal protein in tobacco, is required under cold stress conditions. *Plant Cell* 20, 2221–2237. doi: 10.1105/tpc.108.060392
- Roquet, C., Coissac, E., Cruaud, C., Boleda, M., Boyer, F., Alberti, A., et al. (2016). Understanding the evolution of holoparasitic plants: the complete plastid genome of the holoparasite *Cytinus hypocistis* (Cytinaceae). *Ann. Bot.* 118, 885–896. doi: 10.1093/aob/mcw135
- Rosso, D., Bode, R., Li, W., Krol, M., Saccon, D., Wang, S., et al. (2009). Photosynthetic redox imbalance governs leaf sectoring in the *Arabidopsis thaliana* variegation mutants immutans, spotty, var1, and var2. *Plant Cell* 21, 3473–3492. doi: 10.1105/tpc.108.062752
- Ruf, S., Hermann, M., Berger, I. J., Carrer, H., and Bock, R. (2001). Stable genetic transformation of tomato plastids and expression of a foreign protein in fruit. *Nat. Biotechnol.* 19, 870–875. doi: 10.1038/nbt0901-870
- Ruppel, N. J., and Hangarter, R. P. (2007). Mutations in a plastid-localized elongation factor G alter early stages of plastid development in *Arabidopsis thaliana*. *BMC Plant Biol.* 7:37. doi: 10.1186/1471-2229-7-37
- Ruppel, N. J., Logsdon, C. A., Whippe, C. W., Inoue, K., and Hangarter, R. P. (2011). A mutation in *Arabidopsis* seedling plastid development1 affects plastid differentiation in embryo-derived tissues during seedling growth. *Plant Physiol.* 155, 342–353. doi: 10.1104/pp.110.161414
- Rybczyński, J., Borkowska, B., Fiuk, A., Gawrońska, H., Śliwińska, E., and Mikula, A. (2007). Effect of sucrose concentration on photosynthetic activity of in vitro cultures *Gentiana kurroo* (Royle) germplings. *Acta Physiol. Plant.* 29, 445–453. doi: 10.1007/s11738-007-0054-1
- Sakamoto, W. (2003). Leaf-variegated mutations and their responsible genes in *Arabidopsis thaliana*. *Genes Genet. Syst.* 78, 1–9. doi: 10.1266/ggs.78.1
- Sambrook, J., Fritsch, E. F., and Maniatis, T. (1989). *Molecular Cloning - A Laboratory Manual*. Cold Spring Harbor, New York, NY: Cold Spring Harbor Laboratory Press.
- Schagger, H., and von Jagow, G. (1987). Tricine-sodium dodecyl sulfate-polyacrylamide gel electrophoresis for the separation of proteins in the range from 1 to 100 kDa. *Anal. Biochem.* 166, 368–379. doi: 10.1016/0003-2697(87)90587-2
- Schmitz-Linneweber, C., Lampe, M. K., Sultan, L. D., and Ostersetzer-Biran, O. (2015). Organellar maturases: a window into the evolution of the spliceosome. *Biochim. Biophys. Acta* 1847, 798–808. doi: 10.1016/j.bbabi.2015.01.009
- Schmitz-Linneweber, C., Tillich, M., Herrmann, R. G., and Maier, R. M. (2001). Heterologous, splicing-dependent RNA editing in chloroplasts: allotetraploidy provides trans-factors. *EMBO J.* 20, 4874–4883. doi: 10.1093/emboj/20.17.4874
- Shimada, H., Mochizuki, M., Ogura, K., Froehlich, J. E., Osteryoung, K. W., Shirano, Y., et al. (2007). *Arabidopsis* cotyledon-specific chloroplast biogenesis factor CYO1 is a protein disulfide isomerase. *Plant Cell* 19, 3157–3169. doi: 10.1105/tpc.107.051714
- Stern, D. B., Goldschmidt-Clermont, M., and Hanson, M. R. (2010). Chloroplast RNA metabolism. *Annu. Rev. Plant Biol.* 61, 125–155. doi: 10.1146/annurev-arplant-042809-112242
- Svab, Z., and Maliga, P. (1993). High-frequency plastid transformation in tobacco by selection for a chimeric aadA gene. *Proc. Natl. Acad. Sci. U.S.A.* 90, 913–917. doi: 10.1073/pnas.90.3.913
- Towbin, H., Staehelin, T., and Gordon, J. (1979). Electrophoretic transfer of proteins from polyacrylamide gels to nitrocellulose sheets: procedure and some applications. *Proc. Natl. Acad. Sci. U.S.A.* 76, 4350–4354. doi: 10.1073/pnas.76.9.4350
- Turnell, M., Otis, C., and Lemieux, C. (2005). The complete chloroplast DNA sequences of the charophycean green algae *Staurostrum* and *Zygnema* reveal that the chloroplast genome underwent extensive changes during the evolution of the *Zygnematales*. *BMC Biol.* 3:22. doi: 10.1186/1741-7007-3-22
- Turnell, M., Otis, C., and Lemieux, C. (2006). The chloroplast genome sequence of *Chara vulgaris* sheds new light into the closest green algal relatives of land plants. *Mol. Biol. Evol.* 23, 1324–1338. doi: 10.1093/molbev/msk018
- Van Dingenen, J., De Milde, L., Vermeersch, M., Maleux, K., De Rycke, R., De Bruyne, M., et al. (2016). Chloroplasts are central players in sugar-induced leaf growth. *Plant Physiol.* 171, 590–605. doi: 10.1104/pp.15.01669
- Van Huylbroeck, J. M., and Debergh, P. C. (1996). Impact of sugar concentration in vitro on photosynthesis and carbon metabolism during ex vitro acclimatization of *Spathiphyllum* plantlets. *Physiol. Plant.* 96, 298–304. doi: 10.1111/j.1365-3054.1996.tb00217.x
- Verhounig, A., Karcher, D., and Bock, R. (2010). Inducible gene expression from the plastid genome by a synthetic riboswitch. *Proc. Natl. Acad. Sci. U.S.A.* 107, 6204–6209. doi: 10.1073/pnas.0914423107
- Vogel, J., Borner, T., and Hess, W. R. (1999). Comparative analysis of splicing of the complete set of chloroplast group II introns in three higher plant mutants. *Nucleic Acids Res.* 27, 3866–3874. doi: 10.1093/nar/27.19.3866

- Vogel, J., Hübschmann, T., Börner, T., and Hess, W. R. (1997). Splicing and intron-internal RNA editing of trnK-matK transcripts in barley plastids: support for MatK as an essential splice factor. *J. Mol. Biol.* 270, 179–187. doi: 10.1006/jmbi.1997.1115
- Walbot, V., and Coe, E. H. (1979). Nuclear gene iojap conditions a programmed change to ribosome-less plastids in *Zea mays*. *Proc. Natl. Acad. Sci. U.S.A.* 76, 2760–2764. doi: 10.1073/pnas.76.6.2760
- Wang, Q., Sullivan, R. W., Kight, A., Henry, R. L., Huang, J., Jones, A. M., et al. (2004). Deletion of the chloroplast-localized Thylakoid formation1 gene product in *Arabidopsis* leads to deficient thylakoid formation and variegated leaves. *Plant Physiol.* 136, 3594–3604. doi: 10.1104/pp.104.04.9841
- Wang, Y., Duby, G., Purnelle, B., and Boutry, M. (2000). Tobacco VDL gene encodes a plastid DEAD box RNA helicase and is involved in chloroplast differentiation and plant morphogenesis. *Plant Cell* 12, 2129–2142. doi: 10.1105/tpc.12.11.2129
- Winkler, W. C., and Breaker, R. R. (2005). Regulation of bacterial gene expression by riboswitches. *Annu. Rev. Microbiol.* 59, 487–517. doi: 10.1146/annurev.micro.59.030804.121336
- Wolfe, K. H., Morden, C. W., and Palmer, J. D. (1992). Function and evolution of a minimal plastid genome from a nonphotosynthetic parasitic plant. *Proc. Natl. Acad. Sci. U.S.A.* 89, 10648–10652. doi: 10.1073/pnas.89.22.10648
- Yamamoto, Y. Y., Puente, P., and Deng, X. W. (2000). An *Arabidopsis* cotyledon-specific albino locus: a possible role in 16S rRNA maturation. *Plant Cell Physiol.* 41, 68–76. doi: 10.1093/pcp/41.1.68
- 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
- Zagari, N., Sandoval-Ibanez, O., Sandal, N., Su, J., Rodriguez-Concepcion, M., Stougaard, J., et al. (2017). Snowy cotyledon 2 promotes chloroplast development and has a role in leaf variegation in both *Lotus japonicus* and *Arabidopsis thaliana*. *Mol. Plant* 10, 721–734. doi: 10.1016/j.molp.2017.02.009
- Zhao, C., and Pyle, A. M. (2016). Crystal structures of a group II intron maturase reveal a missing link in spliceosome evolution. *Nat. Struct. Mol. Biol.* 23, 558–565. doi: 10.1038/nsmb.3224
- Zmudjak, M., Shevtsov, S., Sultan, L. D., Keren, I., and Ostersetter-Biran, O. (2017). Analysis of the roles of the *Arabidopsis* nMAT2 and PMH2 proteins provided with new insights into the regulation of group II intron splicing in land-plant mitochondria. *Int. J. Mol. Sci.* 18:E2428. doi: 10.3390/ijms18112428
- Zoschke, R., Nakamura, M., Liere, K., Sugiura, M., Börner, T., and Schmitz-Linneweber, C. (2010). An organellar maturase associates with multiple group II introns. *Proc. Natl. Acad. Sci. U.S.A.* 107, 3245–3250. doi: 10.1073/pnas.0909400107

**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 © 2018 Qu, Legen, Arndt, Henkel, Hoppe, Thieme, Ranzini, Muino, Weihe, Ohler, Weber, Ostersetter and Schmitz-Linneweber. 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 as a Regulator of Early Plant Development

Dóra Szakonyi\* and Paula Duque\*

*Instituto Gulbenkian de Ciência, Oeiras, Portugal*

## OPEN ACCESS

### Edited by:

Simon Gilroy,  
University of Wisconsin–Madison,  
United States

### Reviewed by:

Ligeng Ma,  
Capital Normal University, China  
Lydia Gramzow,  
Friedrich-Schiller-Universität Jena,  
Germany

### \*Correspondence:

Dóra Szakonyi  
dszakonyi@igc.gulbenkian.pt;  
Dora.Szakonyi@gmail.com  
Paula Duque  
duquep@igc.gulbenkian.pt

### Specialty section:

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

**Received:** 03 April 2018

**Accepted:** 23 July 2018

**Published:** 15 August 2018

### Citation:

Szakonyi D and Duque P (2018)  
Alternative Splicing as a Regulator  
of Early Plant Development.  
Front. Plant Sci. 9:1174.  
doi: 10.3389/fpls.2018.01174

Most plant genes are interrupted by introns and the corresponding transcripts need to undergo pre-mRNA splicing to remove these intervening sequences. Alternative splicing (AS) is an important posttranscriptional process that creates multiple mRNA variants from a single pre-mRNA molecule, thereby enhancing the coding and regulatory potential of genomes. In plants, this mechanism has been implicated in the response to environmental cues, including abiotic and biotic stresses, in the regulation of key developmental processes such as flowering, and in circadian timekeeping. The early plant development steps – from embryo formation and seed germination to skoto- and photomorphogenesis – are critical to both execute the correct body plan and initiate a new reproductive cycle. We review here the available evidence for the involvement of AS and various splicing factors in the initial stages of plant development, while highlighting recent findings as well as potential future challenges.

**Keywords:** alternative splicing, early seedling development, embryogenesis, photomorphogenesis, seed dormancy, seed maturation, seed germination, splicing factors

## mRNA PROCESSING AND ALTERNATIVE SPLICING

Accurate processing of precursor mRNAs (pre-mRNAs) is a major step in gene expression crucial for performing everyday housekeeping functions, executing developmental programs, and responding to intrinsic and environmental cues. It involves modification steps to remove non-coding sequences as well as add the cap and the poly(A) tail to the 5' and 3' ends of the mRNA, respectively (reviewed in Proudfoot, 2011; Shi and Manley, 2015; Ramanathan et al., 2016). Pre-mRNA splicing, the excision of introns followed by joining of exons, is catalyzed by the spliceosome, a large ribonucleoprotein complex. The spliceosomal subunits assemble at conserved nucleotides at the exon-intron boundaries also known as the 5' (or donor) and 3' (or acceptor) splice sites (SS), the branch point and the polypyrimidine tract. In addition to the core spliceosomal components, many RNA-binding proteins play key roles in mRNA processing, SS selection and splicing (reviewed in Meyer et al., 2015). In higher eukaryotes, intron-containing genes frequently give rise to multiple mRNAs through alternative splicing (AS) (Figure 1A), during which differential recognition of SS can lead to intron retention, exon skipping and/or alternative 5'/3'SS selection. AS can significantly enhance a genome's coding capacity by producing protein variants with altered function. It also often affects mRNA stability by introducing premature stop codons in the coding sequence, thus targeting these transcripts to degradation by nonsense-mediated decay (NMD). Furthermore, AS can modify gene expression by modulating transcription elongation and/or translation efficiency (reviewed in Reddy et al., 2013; Laloum et al., 2017). It is hence not surprising that AS fulfills important biological functions. In plants, it has been found to control key processes like the circadian clock or flowering time as well as the response to environmental cues, including abiotic stress or pathogen attack (reviewed in Staiger and Brown, 2013; Yang et al., 2014; Laloum et al., 2017; Shang et al., 2017).



## EARLY PLANT DEVELOPMENT

The first stages of a plant's life are essential to establish the basic body pattern, develop different tissue types and initiate a new reproductive cycle (**Figure 1B**). Sexual reproduction of land plants involves the alternation of haploid and diploid stages. Angiosperms have a dominant diploid sporophyte and a relatively short haploid phase consisting of a few microscopic cells. Seeds are produced by double fertilization. One sperm cell fuses with the egg cell to form the diploid embryo, while a second sperm cell fertilizes the diploid central cell to give rise to the endosperm (reviewed in Raghavan, 2003; Berger et al., 2008). During embryogenesis, the one-cell zygote undergoes a tightly regulated developmental program to form a mature embryo. In dicots such as *Arabidopsis thaliana* (arabidopsis), this process includes distinct morphological stages, called globular, heart, torpedo, and bent cotyledon, leading to the establishment of the basic body plan and main tissue/organ initials including the shoot and root apical meristems (reviewed in Palovaara et al., 2016). Embryo morphogenesis is followed by seed maturation, which involves the accumulation of reserves, acquisition of desiccation tolerance, reduction of metabolic activities and induction of dormancy to enable survival of the embryo until favorable environmental conditions allow germination (reviewed in Graeber et al., 2012). Fresh seeds usually show high dormancy that gradually decreases over time in a process called after-ripening. The release from dormancy depends on environmental factors (e.g., light quality, day length, temperature, water availability, exposure to cold) and internal regulators (e.g., hormones, regulatory proteins, chromatin status) (reviewed in Kucera et al., 2005; Nee et al., 2017). Germination starts with water uptake (imbibition) and rapid expansion of the embryo, leading to rupture of the seed coat and emergence of the radicle. Seedlings growing in the dark display skotomorphogenic development (etiolated growth), characterized by elongated hypocotyls, apical hook, pale cotyledons and short roots. When exposed to light, the seedling undergoes photomorphogenesis to activate vegetative growth, displaying shorter and thicker hypocotyls as well as green and expanded cotyledons (reviewed in Wu, 2014). Hormones are important regulators of early plant development. Embryo formation is governed by auxins and cytokinins, while abscisic acid (ABA) is important for the completion of seed maturation and building up dormancy. ABA is also the major inhibitor of seed germination, with its effect being counteracted by gibberellic acid, ethylene, and brassinosteroids (reviewed in Palovaara et al., 2016).

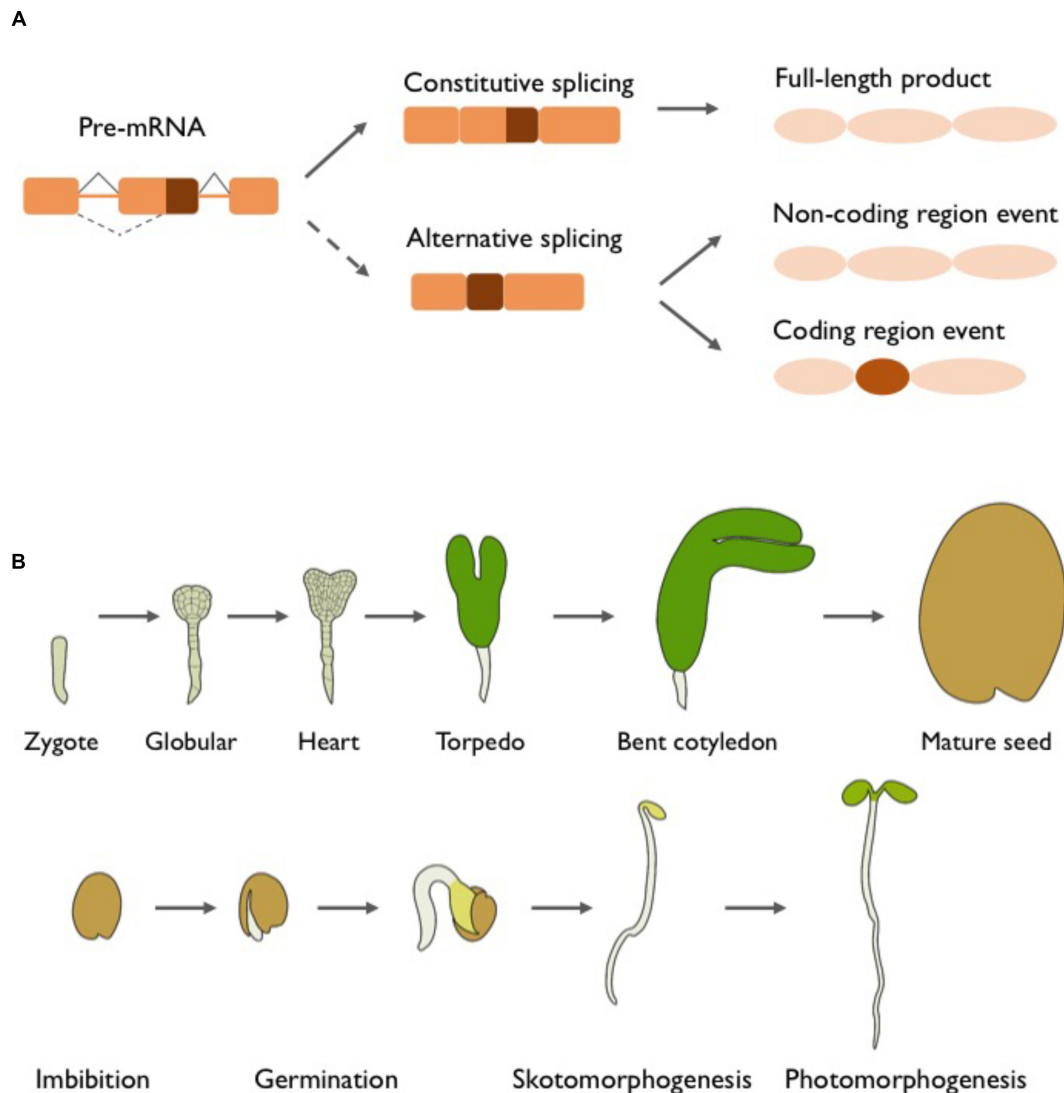
## GLOBAL ALTERNATIVE SPLICING CHANGES DURING EARLY PLANT DEVELOPMENT

Next-generation sequencing has revolutionized transcriptomic studies. The latest RNA-seq data gathered in higher plants showed that traditional approaches largely underestimated the proportion of genes undergoing AS. Current assessments indicate that up to 70% of plant multiexon genes generate more than one

transcript via this mechanism, with intron retention representing the predominant mode of AS (Lu et al., 2010; Zhang et al., 2010; Marquez et al., 2012; Shen et al., 2014; Thatcher et al., 2014; Chamala et al., 2015; Sun and Xiao, 2015; Iniguez et al., 2017; Zhang et al., 2017). In fact, increased sequencing coverage revealed a large number of non-annotated AS events and splice variants (Marquez et al., 2012; Zhang et al., 2017). Most of the plant AS events map to coding regions, thereby altering protein sequence and potentially function or compromising mRNA stability. Indeed, a significant proportion of intron-containing genes are potentially regulated by NMD (Zhang et al., 2010; Kalyna et al., 2012; Drechsel et al., 2013). Although thousands of alternatively spliced mRNAs are detected in genome-wide analyses, detailed genetic and molecular studies will be required to identify functionally relevant AS events.

Numerous plant large-scale studies have focused on gene expression and AS patterns in different tissues and during development, identifying many novel organ- or stage-specific mRNAs with dynamic expression changes and a stage-dependent switch in isoform dominance for many genes (Zhang et al., 2010; Thatcher et al., 2014; Klepikova et al., 2016; Vaneechoutte et al., 2017). Notably, genes encoding alternatively spliced transcripts are not necessarily differentially expressed during developmental transitions, suggesting that AS shapes the transcriptome independently from transcriptional regulation (Srinivasan et al., 2016). These findings are confirmed by deep-sequencing studies tracking expression and AS changes during the first stages of plant development (Aghamirzaie et al., 2013; Lu et al., 2013; Sun and Xiao, 2015; Qu et al., 2016; Thatcher et al., 2016; Narsai et al., 2017). The detection of prominent AS switches and of development-specific splice variants corroborates an important regulatory layer of early plant development at the splicing level. Interestingly, RNA-processing factors themselves undergo AS resulting in a potential autoregulatory feedback loop.

During embryogenesis in soybean (Aghamirzaie et al., 2013), AS of 47,331 genes produced 217,371 different transcripts, most of which had not been previously identified. Nearly one third of the genes showed variations in transcript levels during embryo development, including those encoding enzymes involved in carbon or nitrogen metabolism and hormone-mediated signaling pathways. Most AS events were detected during the later stages of embryogenesis, i.e., embryo maturation, dehydration, establishment of dormancy, and at the quiescent state. This induction of AS may be explained by the striking clustering of both splicing-related and ABA-associated factors observed at the late phases of seed development. Seed maturation and desiccation, which involve very specific developmental, hormonal, and biochemical processes, were also examined in arabidopsis (Srinivasan et al., 2016), where RNA-seq profiling was performed on developing and mature seeds. Interestingly, transcription and AS showed opposite trends, with transcription declining during seed maturation, while AS increased. Over a quarter of the loci undergoing AS expressed stage-specific splice variants or showed a marked isoform switch, with a striking 88% of the detected AS events being absent from the TAIR10 genome annotation. Again, there were no significant changes in total transcript levels of many alternatively spliced genes,



**FIGURE 1 |** Alternative splicing and early plant development. **(A)** Constitutive and alternative splicing. Nascent multi exonic mRNAs need to undergo pre-mRNA splicing. Constitutive splicing removes the non-coding introns, producing a mature mRNA that encodes the full-length protein or transcript with biological functions. The same pre-mRNA molecule can undergo alternative splicing (AS) and produce different transcript variants. AS events occurring in non-coding sequences often impact gene expression, but will result in a protein identical to the full-length isoform. In numerous cases, coding regions are affected by AS, thus originating markedly different mRNAs and potentially distinct proteins that can vary in virtually all functional aspects. **(B)** Embryogenesis, seed maturation and germination, and early seedling development. After fertilization, the zygote undergoes a rapid succession of highly coordinated cell divisions to form globular stage embryos, which show establishment of the apical–basal axis and a first distinction between outer and inner cells. The embryonic cells further differentiate during the heart stage, when many of the basic cell types (provasculature, endodermis, cortex, and protoderm) and organ primordia (cotyledons, hypocotyl, and primary root) are formed and a bilateral body pattern appears. Expanding cotyledons give the embryo a torpedo shape, and the formation of the shoot and root apical meristems is completed. The next steps involve further cell growth and divisions until the embryo reaches its final shape and size. Seed maturation is completed with the accumulation of reserves and the establishment of desiccation tolerance and seed dormancy. Dry seeds are released from dormancy in response to a combination of environmental cues and internal signals. After water uptake, key biochemical and molecular processes are restored, followed by the rupture of the seed coat and emergence of the radicle, marking the completion of germination. Under darkness, the buried seedling undergoes skotomorphogenesis characterized by a short root, elongated hypocotyl, apical hook and absence of photosynthetic pigments. Upon light exposure, photomorphogenesis is activated leading to inhibition of hypocotyl elongation, opening and expansion of the cotyledons, and initiation of photosynthesis after chloroplast maturation.

pointing to AS as an important regulatory mechanism operating independently from transcription. Most of the genes exhibiting differential splicing were involved in RNA processing, potentially amplifying the AS regulatory effect in preparation for seed germination.

Two recent studies addressed the AS contribution during seed germination. In barley embryos, 14–20% of multiexon genes expressed multiple mRNA isoforms, some of which displayed clear changes during early germination (Zhang et al., 2016). Surprisingly, the most prominent AS event was alternative 3'SS

selection, and there were no substantial alterations in total transcript levels for most genes. Assessment of the biological functions of the genes undergoing AS during germination indicated involvement in protein synthesis, energy and carbon metabolism as well as RNA transport and splicing. Overall, seed germination appears to require expression of a specific set of genes, with AS playing a widespread role. The regulatory potential of AS during germination is underscored by a subsequent report in *arabidopsis* (Narsai et al., 2017) confirming the expression of time- and tissue-specific mRNA variants, the occurrence of dynamic changes in isoform abundance, and that splicing regulators are major AS targets during this developmental process.

AS regulation during early plant growth is also relevant in the context of environmental responses. Light, which is perceived by various photoreceptors, strongly impacts the life cycle of plants, regulating among others early developmental steps such as seed germination and the transition to autotrophic growth. Genome-wide effects of light on plant AS were recently analyzed by RNA-seq (Wu et al., 2014; Mancini et al., 2016), including in very young seedlings (Shikata et al., 2014; Hartmann et al., 2016). Shikata et al. (2014) reported that, during the initial response of etiolated seedlings to red light, the number of genes showing phytochrome-mediated differential gene expression or changed AS pattern is comparable, while later transcription becomes the dominant regulatory mechanism. In the phytochrome-dependent AS dataset, splicing-related genes were overrepresented, including SR proteins and the U1 and U2 spliceosomal subunits, while transcription factors comprised the major group of differentially expressed genes. AS seemed to play a significant role in light-induced chloroplast differentiation, as photosynthesis- and plastid-related genes were also enriched in the differential AS sets. When Hartmann et al. (2016) analyzed the response of etiolated *arabidopsis* seedlings exposed blue, red, or white light treatments, ~20% of genes were found to be differentially expressed, with ~700 AS events being detected, most of which mapped to coding sequences. Again, gene ontology analysis revealed overrepresentation of the RNA-binding category, including many splicing factors. A link between light-induced AS and mRNA stability was also uncovered, with 77.2% of the detected mRNA isoforms more abundant in the dark samples being potential NMD targets. Remarkably, in most of AS events, an isoform switch from a putative instable mRNA variant to a protein-coding alternative occurred upon light exposure. Moreover, mutants lacking the major red or blue light receptors showed impaired AS mainly when subjected to monochromatic red or blue light, indicating that additional signaling pathways influence AS under white light. The authors suggested that metabolic signals, sugars in particular, are implicated in light-mediated AS regulation.

## SPLICING FACTORS REGULATING EARLY PLANT DEVELOPMENT

Compelling evidence from large-scale analyses pointing to an important role for AS during early plant development is being

substantiated by accumulating *in vivo* genetic studies (Table 1). Overexpression or complete abrogation of splicing function often causes embryo lethality, indicating that the corresponding genes are essential for viability and development of a functional plant (Kalyna et al., 2003; Schmitz-Linneweber et al., 2006; Liu et al., 2009; Kim et al., 2010; Fouquet et al., 2011; Swaraz et al., 2011; Perea-Resa et al., 2012; Shikata et al., 2012; Sasaki et al., 2015; Tsugeki et al., 2015). Some studies have established a hormonal basis for the embryo and early seedling development defects caused by altered expression of splicing factors (Kalyna et al., 2003; Casson et al., 2009; Tsugeki et al., 2015), with abnormal spatial distribution of auxin arising from erroneous splicing and expression of auxin biosynthesis, transport, and signaling genes. A link between mRNA splicing and auxin signaling was also uncovered in flowers, where subcellular compartmentation of an auxin biosynthetic gene is regulated by AS (Kriechbaumer et al., 2012).

Seed dormancy and germination are also strongly affected in mRNA processing mutants. These effects were mostly reported to relate to splicing (Dolata et al., 2015) and polyadenylation (Cyrek et al., 2016) of the *DOG1* gene, a key seed dormancy regulator and known AS target, and to changes in ABA signaling (Xiong et al., 2001; Sugliani et al., 2010; Jiang et al., 2012). Early seedling development can be affected as a manifestation of wider pleiotropic defects (Liu et al., 2010; Swaraz et al., 2011; Perea-Resa et al., 2012; Shikata et al., 2012; Hsieh et al., 2015; Yap et al., 2015) or in weak alleles of embryo lethal mutants (Kalyna et al., 2003; Gutierrez-Marcos et al., 2007; Fouquet et al., 2011; Tsugeki et al., 2015). Observed phenotypes include disturbed cotyledons, hypocotyls, vasculature patterning, roots and/or seedling viability and growth. Notably, mRNA splicing in plastids and mitochondria appears to be crucial for seed development and plant growth in both *arabidopsis* and maize (Schmitz-Linneweber et al., 2006; Gutierrez-Marcos et al., 2007; Liu et al., 2010; Hernando et al., 2015; Hsieh et al., 2015; Yap et al., 2015; Chen et al., 2017).

Genetic and molecular analyses have confirmed a role for splicing factors in photomorphogenesis, particularly in red-light responses. Phytochrome-dependent light signaling influences AS through specific splicing components, with additional splicing factors such as SR proteins being differentially processed in loss-of-function mutants of these effectors under various light conditions (Shikata et al., 2012; Hernando et al., 2015; Xin et al., 2017). Interestingly, Xin et al. (2017) demonstrated red light-dependent direct interaction and colocalization of a splicing factor and phytochrome B.

## ALTERNATIVE SPLICING TARGETS AFFECTING EARLY PLANT DEVELOPMENT

Despite massive transcriptome changes imposed by AS during early plant development, only a handful of alternatively spliced transcripts have had their functional significance analyzed in detail (Table 1). While one group, including the *arabidopsis*

**TABLE 1 |** Splicing factors and targets functioning in early plant development.

Gene	Organism	Molecular function	Biological process	Reference
<b>RNA splicing factors involved in early development</b>				
<i>ABO5</i>	Arabidopsis	Mitochondrion-targeted PPR protein	Early seedling development	Liu et al., 2010
<i>AEF1</i>	Arabidopsis	Plastid-targeted PPR protein	Early seedling development	Yap et al., 2015
<i>CUV</i>	Arabidopsis	DEAH-box RNA-dependent ATPase homolog	Embryogenesis, early seedling development	Tsugeki et al., 2015
<i>Dek35</i>	Maize	Mitochondrion-targeted PPR protein	Seed development	Chen et al., 2017
<i>EMP4</i>	Maize	Mitochondrion-targeted PPR protein	Seed development, seedling growth	Gutierrez-Marcos et al., 2007
<i>ESP1</i>	Arabidopsis	RNA 3' end processing factor	Seed dormancy	Cyrek et al., 2016
<i>FY</i>	Arabidopsis	RNA 3' end processing factor	Seed dormancy	Jiang et al., 2012; Cyrek et al., 2016
<i>GFA1</i>	Arabidopsis	U5 snRNP component (spliceosomal protein)	Embryogenesis	Liu et al., 2009
<i>LSM8</i>	Arabidopsis	U6 snRNP component (Sm-like protein)	Seed and early seedling development	Perea-Resa et al., 2012
<i>MDF</i>	Arabidopsis	RS domain protein	Early seedling development	Casson et al., 2009
<i>NTR1</i>	Arabidopsis	Spliceosome disassembly factor	Seed dormancy	Dolata et al., 2015
<i>PCSF4</i>	Arabidopsis	RNA 3' end processing factor	Seed dormancy	Cyrek et al., 2016
<i>PPR4</i>	Arabidopsis	Chloroplast-targeted PPR protein	Embryogenesis	Schmitz-Linneweber et al., 2006
<i>PPR4</i>	Maize	Chloroplast-targeted PPR protein	Seedling growth	Schmitz-Linneweber et al., 2006
<i>PRMT4</i>	Arabidopsis	Protein arginine methyltransferase	Hypocotyl elongation in response to light	Hernando et al., 2015
<i>PRMT5</i>	Arabidopsis	Protein arginine methyltransferase	Hypocotyl elongation in response to light	Hernando et al., 2015
<i>PRP8</i>	Arabidopsis	Core spliceosomal protein	Embryogenesis	Sasaki et al., 2015
<i>RGH3</i>	Maize	U2AF35-related (spliceosomal protein)	Embryo, endosperm and seedling development	Fouquet et al., 2011
<i>RRC1</i>	Arabidopsis	RS domain (SR-like) protein	Embryogenesis, early seedling development, hypocotyl elongation in response to light	Shikata et al., 2012
<i>RSZ33</i>	Arabidopsis	SR-protein splicing factor	Embryogenesis, early seedling development	Kalya et al., 2003
<i>RTF2</i>	Arabidopsis	Rtf2-domain splicing-related protein	Embryogenesis	Sasaki et al., 2015
<i>SAD1/LSM5</i>	Arabidopsis	U6 snRNP component (Sm-like protein)	Seed dormancy	Xiong et al., 2001
<i>SFPS</i>	Arabidopsis	U2-associated splicing factor	Hypocotyl elongation in response to light	Xin et al., 2017
<i>SLO3</i>	Arabidopsis	Mitochondrion-targeted PPR protein	Seed development, germination, early seedling development	Hsieh et al., 2015
<i>SmD3-a SmD3-b</i>	Arabidopsis	snRNP core subunits (spliceosomal proteins)	Embryogenesis, early seedling development	Swaraz et al., 2011
<i>SUA</i>	Arabidopsis	RNA-binding protein	Seed dormancy	Sugliani et al., 2010
<i>U11/U12-31K</i>	Arabidopsis	U12-type spliceosomal protein	Embryogenesis	Kim et al., 2010
<b>Functional AS targets in early development</b>				
<i>ABI3</i>	Arabidopsis	B3 domain-containing transcription factor	Seed maturation and dormancy	Sugliani et al., 2010
<i>ABI3</i>	Pea	B3 domain-containing transcription factor	Seed maturation and dormancy	Gagete et al., 2009
<i>ABI3</i>	Linseed flax	B3 domain-containing transcription factor	Seed maturation and dormancy	Wang et al., 2018
<i>ABI3</i>	Tomato	B3 domain-containing transcription factor	Seed maturation and dormancy	Gao et al., 2013

(Continued)



TABLE 1 | Continued

<i>ABI5</i>	Rice	bZIP transcription factor	Seed maturation and dormancy	Zou et al., 2007
<i>COP1</i>	Arabidopsis	E3 ubiquitin protein ligase	Skotomorphogenesis	Zhou et al., 1998
<i>DOG1</i>	Arabidopsis	Unknown	Seed dormancy	Bentsink et al., 2006; Nakabayashi et al., 2015
<i>HPR</i>	Pumpkin	Hydroxypyruvate reductase	Early seedling development	Mano et al., 1999
<i>HYH</i>	Arabidopsis	bZIP transcription factor	Hypocotyl elongation in response to light	Sibout et al., 2006; Li et al., 2017
<i>PIF6</i>	Arabidopsis	bHLH transcription factor	Seed dormancy, seed germination, hypocotyl elongation in response to light	Penfield et al., 2010
<i>SPA3</i>	Arabidopsis	WD40 protein with kinase domain	Hypocotyl elongation in response to light	Shikata et al., 2014
<i>SR30</i>	Arabidopsis	SR protein splicing factor	Light-regulated AS during photomorphogenesis	Hartmann et al., 2016, 2018
<i>VP1</i>	Rice	B3 domain-containing transcription factor	Seed maturation and dormancy	Fan et al., 2007
<i>Vp1</i>	Wheat	B3 domain-containing transcription factor	Seed maturation and dormancy	McKibbin et al., 2002; Wilkinson et al., 2005

*DOG1* gene as well as the *OsABI5* and *ABI3/VP1* transcription factors, plays roles in seed maturation, dormancy, and ABA responses (McKibbin et al., 2002; Wilkinson et al., 2005; Bentsink et al., 2006; Fan et al., 2007; Zou et al., 2007; Gagate et al., 2009; Sugliani et al., 2010; Gao et al., 2013; Nakabayashi et al., 2015; Wang et al., 2018), another is important for light signaling and includes *COP1*, *HYH*, and *SPA3* (Zhou et al., 1998; Sibout et al., 2006; Shikata et al., 2014; Li et al., 2017). Moreover, *PIF6* regulates both seed dormancy and light responses (Penfield et al., 2010). Despite the few individual events studied, AS is known to act via diverse mechanisms, as illustrated below.

In agreement with results from large-scale studies, the expression of numerous individual mRNA variants was found to be development- or tissue-specific (Zhou et al., 1998; Fan et al., 2007; Gagate et al., 2009; Sugliani et al., 2010; Gao et al., 2013; Wang et al., 2018), with some turning out to be non-functional, either because they did not produce an active protein or no phenotypic consequence was observed as a result of ectopic expression (Wang et al., 2018). In another study, genetic complementation tests indicated that the different splice variants perform functions equivalent to the constitutive form, even when lacking crucial amino acid sequences or domains (Li et al., 2017). Similarly, AS did not fundamentally influence DNA-binding or protein-protein interaction ability of the *ABI3* and *ABI5* transcription factors from different plant species, though the binding strength appeared to differ among the various isoforms (Zou et al., 2007; Gagate et al., 2009; Gao et al., 2013).

Alternative splice variants can also fulfill similar or distinct functions depending on developmental stage. The constitutive and alternative *PIF6* mRNA variants similarly influenced light responses in seedlings, while only the short isoform displayed evident functions during seed germination (Penfield et al., 2010).

In the case of *COP1* and *SPA3*, ectopic overexpression of alternative splice forms phenocopied knock-out mutant phenotypes, indicating that some alternative forms can interfere with the function of the full-length protein (Zhou et al., 1998; Shikata et al., 2014). Strikingly, co-expression and direct protein interactions were found to be necessary for full *DOG1* function (Nakabayashi et al., 2015). In genetic complementation assays, independent expression of individual *DOG1* isoforms driven by the native promoter did not restore seed dormancy, whereas transgenic lines carrying two or more *DOG1* variants showed improved dormancy. Detailed analysis of these results supported the hypothesis that, although single isoforms are active, the presence of multiple isoforms is required for adequate *DOG1* function. On the other hand, AS-induced changes in protein sequence may lead not only to diminished biological function but, as demonstrated for an *HYH* isoform lacking a protein interaction domain for proteasomal degradation, also to a more stable and hence more active protein isoform (Sibout et al., 2006).

Subcellular targeting provides specialized locations for intracellular processes and can interfere with the regulatory and biochemical potential of proteins. AS of a pumpkin hydroxypyruvate reductase (*HPR*) acting in photorespiration affected the C-terminal targeting sequence, with one splice form localizing in the peroxisome and another in the cytosol (Mano et al., 1999). The two mRNAs were expressed at similar levels in darkness, while light promoted the production of the shorter, cytosol-localized variant. Most recently, retention of an mRNA variant of the arabidopsis *SR30* splicing regulator in the nucleus was shown to influence mRNA stability by preventing the degradation of a potential NMD target in the cytoplasm and its association to the translation machinery (Hartmann et al., 2018).

## CONCLUSION AND PERSPECTIVES

Recent transcriptome-wide, genetic and molecular studies have demonstrated that regulation of the complex developmental steps from embryogenesis to establishment of a functional plant includes posttranscriptional control via AS. Seed maturation, establishment and maintenance of seed dormancy, and young seedling responses to light stand out as significant AS-regulated processes. The detection of time- and tissue-specific mRNA variants and of notable switches in splicing patterns substantiate crucial roles for AS in other early development processes. Further large-scale analyses in different tissue types using the latest sequencing technologies and single-cell approaches will be key to understand the full extent of AS events occurring during the initial stages of plant development. Improved standardization of data processing and analysis along with more meticulous experimental set-ups should also allow for more reliable comparative studies. Comprehensive publicly available databases, providing a detailed and up-to-date view of AS in plants are still lacking. These will be pivotal in pinpointing promising novel splice forms and assist in functional studies to distinguish biologically relevant AS contributing to proteomic diversity or gene expression regulation from non-functional AS events and splicing noise. Importantly, state-of-the-art methodology such as iCLIP is proving successful in

plant systems and should allow identification of the mRNAs targeted directly by splicing factors to control early plant development.

## AUTHOR CONTRIBUTIONS

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

## FUNDING

DS was supported by a Postdoctoral Fellowship (SFRH/BPD/94796/2013) from Fundação para a Ciência e a Tecnologia (FCT), which also finances research in our lab through Grant PTDC/BIA-PLA/1084/2014. Funding from the GREEN-it research unit (UID/Multi/04551/2013) is also acknowledged.

## ACKNOWLEDGMENTS

Our apologies to the authors whose work was not cited due to space restrictions.

## REFERENCES

- Aghamirzaie, D., Nabiyouni, M., Fang, Y., Klumas, C., Heath, L. S., Grene, R., et al. (2013). Changes in rna splicing in developing soybean (glycine max) embryos. *Biology* 2, 1311–1337. doi: 10.3390/biology2041311
- Bentsink, L., Jowett, J., Hanhart, C. J., and Koornneef, M. (2006). Cloning of DOG1, a quantitative trait locus controlling seed dormancy in *Arabidopsis*. *Proc. Natl. Acad. Sci. U.S.A.* 103, 17042–17047. doi: 10.1073/pnas.0607877103
- Berger, F., Hamamura, Y., Ingouff, M., and Higashiyama, T. (2008). Double fertilization - caught in the act. *Trends Plant Sci.* 13, 437–443. doi: 10.1016/j.tplants.2008.05.011
- Casson, S. A., Topping, J. F., and Lindsey, K. (2009). MERISTEM-DEFECTIVE, an RS domain protein, is required for the correct meristem patterning and function in *Arabidopsis*. *Plant J.* 57, 857–869. doi: 10.1111/j.1365-3113.2008.03738.x
- 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, X., Feng, F., Qi, W., Xu, L., Yao, D., Wang, Q., et al. (2017). Dek35 Encodes a PPR Protein that Affects cis-Splicing of Mitochondrial nad4 Intron 1 and Seed Development in Maize. *Mol. Plant* 10, 427–441. doi: 10.1016/j.molp.2016.08.008
- Cyrek, M., Fedak, H., Ciesielski, A., Guo, Y., Sliwa, A., Brzezniak, L., et al. (2016). Seed Dormancy in *Arabidopsis* Is Controlled by Alternative Polyadenylation of DOG1. *Plant Physiol.* 170, 947–955. doi: 10.1104/pp.15.01483
- 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
- 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
- Fan, J., Niu, X., Wang, Y., Ren, G., Zhuo, T., Yang, Y., et al. (2007). Short, direct repeats (SDRs)-mediated post-transcriptional processing of a transcription factor gene OsVP1 in rice (*Oryza sativa*). *J. Exp. Bot.* 58, 3811–3817. doi: 10.1093/jxb/erm231
- 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
- Gagate, A. P., Riera, M., Franco, L., and Rodrigo, M. I. (2009). Functional analysis of the isoforms of an ABI3-like factor of *Pisum sativum* generated by alternative splicing. *J. Exp. Bot.* 60, 1703–1714. doi: 10.1093/jxb/erp038
- Gao, Y., Liu, J., Zhang, Z., Sun, X., Zhang, N., Fan, J., et al. (2013). Functional characterization of two alternatively spliced transcripts of tomato ABCISIC ACID INSENSITIVE3 (ABI3) gene. *Plant Mol. Biol.* 82, 131–145. doi: 10.1007/s11103-013-0044-1
- Graeber, K., Nakabayashi, K., Miatton, E., Leubner-Metzger, G., and Soppe, W. J. (2012). Molecular mechanisms of seed dormancy. *Plant Cell Environ.* 35, 1769–1786. doi: 10.1111/j.1365-3040.2012.02542.x
- Gutierrez-Marcos, J. F., Dal Pra, M., Giulini, A., Costa, L. M., Gavazzi, G., Cordelier, S., et al. (2007). empty pericarp4 encodes a mitochondrion-targeted pentatricopeptide repeat protein necessary for seed development and plant growth in maize. *Plant Cell* 19, 196–210. doi: 10.1105/tpc.105.039594
- 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., Wiefner, 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
- Hernando, C. E., Sanchez, S. E., Mancini, E., and Yanovsky, M. J. (2015). Genome wide comparative analysis of the effects of PRMT5 and PRMT4/CARM1 arginine methyltransferases on the *Arabidopsis thaliana* transcriptome. *BMC Genomics* 16:192. doi: 10.1186/s12864-015-1399-2
- Hsieh, W. Y., Liao, J. C., Chang, C. Y., Harrison, T., Boucher, C., and Hsieh, M. H. (2015). The SLOW GROWTH3 pentatricopeptide repeat protein is required for the splicing of Mitochondrial NADH dehydrogenase subunit7 intron 2 in *Arabidopsis*. *Plant Physiol.* 168, 490–501. doi: 10.1104/pp.15.00354

- Iniguez, L. P., Ramirez, M., Barbazuk, W. B., and Hernandez, G. (2017). Identification and analysis of alternative splicing events in *Phaseolus vulgaris* and glycine max. *BMC Genomics* 18:650. doi: 10.1186/s12864-017-4054-2
- Jiang, S., Kumar, S., Eu, Y. J., Jami, S. K., Stasolla, C., and Hill, R. D. (2012). The *Arabidopsis* mutant, *fy-1*, has an ABA-insensitive germination phenotype. *J. Exp. Bot.* 63, 2693–2703. doi: 10.1093/jxb/err452
- Kalyna, M., Lopato, S., and Barta, A. (2003). Ectopic expression of *atRSZ33* reveals its function in splicing and causes pleiotropic changes in development. *Mol. Biol. Cell* 14, 3565–3577. doi: 10.1091/mbc.E03-02-0109
- 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
- Kim, W. Y., Jung, H. J., Kwak, K. J., Kim, M. K., Oh, S. H., Han, Y. S., et al. (2010). The *Arabidopsis* U12-type spliceosomal protein U11/U12-31K is involved in U12 intron splicing via RNA chaperone activity and affects plant development. *Plant Cell* 22, 3951–3962. doi: 10.1105/tpc.110.079103
- 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/tbj.13312
- Kriebchaumer, 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
- Kucera, B., Cohn, M. A., and Leubner-Metzger, G. (2005). Plant hormone interactions during seed dormancy release and germination. *Seed Sci. Res.* 15, 281–307. doi: 10.1079/ssr2005218
- Laloum, T., Martin, G., and Duque, P. (2017). Alternative Splicing Control of Abiotic Stress Responses. *Trends Plant Sci* 23, 140–150. doi: 10.1016/j.tplants.2017.09.019
- Li, C., Zheng, L., Zhang, J., Lv, Y., Liu, J., Wang, X., et al. (2017). Characterization and functional analysis of four HYH splicing variants in *Arabidopsis* hypocotyl elongation. *Gene* 619, 44–49. doi: 10.1016/j.gene.2017.04.001
- 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
- Liu, Y., He, J., Chen, Z., Ren, X., Hong, X., and Gong, Z. (2010). ABA over-sensitive 5 (ABO5), encoding a pentatricopeptide repeat protein required for cis-splicing of mitochondrial *nad2* intron 3, is involved in the abscisic acid response in *Arabidopsis*. *Plant J.* 63, 749–765. doi: 10.1111/j.1365-313X.2010.04280.x
- 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
- Lu, X., Chen, D., Shu, D., Zhang, Z., Wang, W., Klukas, C., et al. (2013). The differential transcription network between embryo and endosperm in the early developing maize seed. *Plant Physiol.* 162, 440–455. doi: 10.1104/pp.113.214874
- Mancini, E., Sanchez, S. E., Romanowski, A., Schlaen, R. G., Sanchez-Lamas, M., Cerdan, 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
- 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
- McKibbin, R. S., Wilkinson, M. D., Bailey, P. C., Flintham, J. E., Andrew, L. M., Lazzeri, P. A., et al. (2002). Transcripts of Vp-1 homeologues are misspliced in modern wheat and ancestral species. *Proc. Natl. Acad. Sci. U.S.A.* 99, 10203–10208. doi: 10.1073/pnas.152318599
- 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
- Nakabayashi, K., Bartsch, M., Ding, J., and Soppe, W. 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
- 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
- Nee, G., Xiang, Y., and Soppe, W. J. (2017). The release of dormancy, a wake-up call for seeds to germinate. *Curr. Opin. Plant Biol.* 35, 8–14. doi: 10.1016/j.pbi.2016.09.002
- Palovaara, J., De Zeeuw, T., and Weijers, D. (2016). Tissue and organ initiation in the plant embryo: a first time for everything. *Annu. Rev. Cell Dev. Biol.* 32, 47–75. doi: 10.1146/annurev-cellbio-111315-124929
- 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
- Perea-Resa, C., Hernandez-Verdeja, T., Lopez-Cobollo, R., Del Mar Castellano, M., and Salinas, J. (2012). LSM proteins provide accurate splicing and decay of selected transcripts to ensure normal *Arabidopsis* development. *Plant Cell* 24, 4930–4947. doi: 10.1105/tpc.112.103697
- Proudfoot, N. J. (2011). Ending the message: poly(A) signals then and now. *Genes Dev.* 25, 1770–1782. doi: 10.1101/gad.172684.11
- Qu, J., Ma, C., Feng, J., Xu, S., Wang, L., Li, F., et al. (2016). Transcriptome dynamics during maize endosperm development. *PLoS One* 11:e0163814. doi: 10.1371/journal.pone.0163814
- Raghavan, V. (2003). Some reflections on double fertilization, from its discovery to the present. *New Phytol.* 159, 565–583. doi: 10.1046/j.1469-8137.2003.00846.x
- Ramanathan, A., Robb, G. B., and Chan, S. H. (2016). mRNA capping: biological functions and applications. *Nucleic Acids Res.* 44, 7511–7526. doi: 10.1093/nar/gkw551
- 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
- Sasaki, T., Kanno, T., Liang, S. C., Chen, P. Y., Liao, W. W., Lin, W. D., et al. (2015). An Rtf2 domain-containing protein influences pre-mRNA splicing and is essential for embryonic development in *Arabidopsis thaliana*. *Genetics* 200, 523–535. doi: 10.1534/genetics.115.176438
- Schmitz-Linneweber, C., Williams-Carrier, R. E., Williams-Voelker, P. M., Kroeger, T. S., Vichas, A., and Barkan, A. (2006). A pentatricopeptide repeat protein facilitates the trans-splicing of the maize chloroplast *rps12* pre-mRNA. *Plant Cell* 18, 2650–2663. doi: 10.1105/tpc.106.046110
- 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
- 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
- Shi, Y., and Manley, J. L. (2015). The end of the message: multiple protein-RNA interactions define the mRNA polyadenylation site. *Genes Dev.* 29, 889–897. doi: 10.1101/gad.261974.115
- 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.1365-313X.2012.04937.x
- Sibout, R., Sukumar, P., Hettiarachchi, C., Holm, M., Muday, G. K., and Hardtke, C. S. (2006). Opposite root growth phenotypes of *hy5* versus *hy5* *hyh* mutants correlate with increased constitutive auxin signaling. *PLoS Genet.* 2:e202. doi: 10.1371/journal.pgen.0020202
- Srinivasan, A., Jimenez-Gomez, J. M., Fornara, F., Soppe, W. J., and Brambilla, V. (2016). Alternative splicing enhances transcriptome complexity in desiccating seeds. *J. Integr. Plant Biol.* 58, 947–958. doi: 10.1111/jipb.12482
- 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

- Sugliani, M., Brambilla, V., Clercx, E. J., Koornneef, M., and Soppe, W. 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
- 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
- Swaraz, A. M., Park, Y. D., and Hur, Y. (2011). Knock-out mutations of *Arabidopsis* SmD3-b induce pleiotropic phenotypes through altered transcript splicing. *Plant Sci.* 180, 661–671. doi: 10.1016/j.plantsci.2011.01.011
- 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
- Tsugeki, R., Tanaka-Sato, N., Maruyama, N., Terada, S., Kojima, M., Sakakibara, H., et al. (2015). CLUMSY VEIN, the *Arabidopsis* DEAH-box Prp16 ortholog, is required for auxin-mediated development. *Plant J.* 81, 183–197. doi: 10.1111/tbj.12721
- 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/tbj.13746
- Wang, Y., Zhang, T., Song, X., Zhang, J., Dang, Z., Pei, X., et al. (2018). 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
- Wilkinson, M., Lenton, J., and Holdsworth, M. (2005). Transcripts of Vp-1 homoeologues are alternatively spliced within the Triticeae tribe. *Euphytica* 143, 243–246. doi: 10.1007/s10681-005-7856-2
- 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. H. (2014). Gene expression regulation in photomorphogenesis from the perspective of the central dogma. *Annu. Rev. Plant Biol.* 65, 311–333. doi: 10.1146/annurev-arplant-050213-040337
- Xin, R., Zhu, L., Salome, P. A., Mancini, E., Marshall, C. M., Harmon, F. G., et al. (2017). SPF45-related splicing factor for phytochrome signaling promotes photomorphogenesis by regulating pre-mRNA splicing in *Arabidopsis*. *Proc. Natl. Acad. Sci. U.S.A.* 114, E7018–E7027. doi: 10.1073/pnas.1706379114
- Xiong, L., Gong, Z., Rock, C. D., Subramanian, S., Guo, Y., Xu, W., et al. (2001). Modulation of abscisic acid signal transduction and biosynthesis by an Sm-like protein in *Arabidopsis*. *Dev. Cell* 1, 771–781. doi: 10.1016/S1534-5807(01)00087-9
- Yang, S., Tang, F., and Zhu, H. (2014). Alternative splicing in plant immunity. *Int. J. Mol. Sci.* 15, 10424–10445. doi: 10.3390/ijms150610424
- Yap, A., Kindgren, P., Colas Des Francs-Small, C., Kazama, T., Tanz, S. K., Toriyama, K., et al. (2015). AEF1/MPR25 is implicated in RNA editing of plastid atpF and mitochondrial nad5, and also promotes atpF splicing in *Arabidopsis* and rice. *Plant J.* 81, 661–669. doi: 10.1111/tbj.12756
- Zhang, G., Guo, G., Hu, X., Zhang, Y., Li, Q., Li, R., 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, 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, 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
- Zhou, D. X., Kim, Y. J., Li, Y. F., Carol, P., and Mache, R. (1998). COP1b, an isoform of COP1 generated by alternative splicing, has a negative effect on COP1 function in regulating light-dependent seedling development in *Arabidopsis*. *Mol. Gen. Genet.* 257, 387–391. doi: 10.1007/s004380050662
- Zou, M., Guan, Y., Ren, H., Zhang, F., and Chen, F. (2007). Characterization of alternative splicing products of bZIP transcription factors OsABI5. *Biochem. Biophys. Res. Commun.* 360, 307–313. doi: 10.1016/j.bbrc.2007.05.226

**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 © 2018 Szakonyi 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.





# Functional Characterization of SMG7 Paralogs in *Arabidopsis thaliana*

Claudio Capitaio<sup>1†</sup>, Neha Shukla<sup>2†</sup>, Aneta Wandrolova<sup>2</sup>, Ortrun Mittelsten Scheid<sup>1</sup> and Karel Riha<sup>2\*</sup>

<sup>1</sup> Gregor Mendel Institute of Molecular Plant Biology, Austrian Academy of Sciences, Vienna Biocenter, Vienna, Austria,

<sup>2</sup> Central European Institute of Technology, Masaryk University, Brno, Czechia

## OPEN ACCESS

### Edited by:

Dora Szakonyi,  
Instituto Gulbenkian de Ciência (IGC),  
Portugal

### Reviewed by:

Daniel Silhavy,  
Agricultural Biotechnology Institute,  
Hungary  
Uener Kolukisaoglu,  
Eberhard Karls Universität Tübingen,  
Germany

### \*Correspondence:

Karel Riha  
karel.riha@ceitec.muni.cz

<sup>†</sup> Shared first co-authorship

### Specialty section:

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

**Received:** 31 May 2018

**Accepted:** 17 October 2018

**Published:** 06 November 2018

### Citation:

Capitaio C, Shukla N,  
Wandrolova A, Mittelsten Scheid O  
and Riha K (2018) Functional  
Characterization of SMG7 Paralogs  
in *Arabidopsis thaliana*.  
Front. Plant Sci. 9:1602.  
doi: 10.3389/fpls.2018.01602

SMG7 proteins are evolutionary conserved across eukaryotes and primarily known for their function in nonsense mediated RNA decay (NMD). In contrast to other NMD factors, SMG7 proteins underwent independent expansions during evolution indicating their propensity to adopt novel functions. Here we characterized SMG7 and SMG7-like (SMG7L) paralogs in *Arabidopsis thaliana*. SMG7 retained its role in NMD and additionally appears to have acquired another function in meiosis. We inactivated SMG7 by CRISPR/Cas9 mutagenesis and showed that, in contrast to our previous report, SMG7 is not an essential gene in *Arabidopsis*. Furthermore, our data indicate that the N-terminal phosphoserine-binding domain is required for both NMD and meiosis. Phenotypic analysis of SMG7 and SMG7L double mutants did not indicate any functional redundancy between the two genes, suggesting neofunctionalization of SMG7L. Finally, protein sequence comparison together with a phenotyping of T-DNA insertion mutants identified several conserved regions specific for SMG7 that may underlie its role in NMD and meiosis. This information provides a framework for deciphering the non-canonical functions of SMG7-family proteins.

**Keywords:** nonsense mediated RNA decay, meiosis, SMG7, gene duplication, *Arabidopsis*

## INTRODUCTION

In eukaryotic cells, gene expression is controlled by several surveillance mechanisms that assure accurate and robust production of functional proteins. One such control mechanism is nonsense mediated RNA decay (NMD), which degrades aberrant RNAs (He and Jacobson, 2015; Lykke-Andersen and Jensen, 2015; Hug et al., 2016; Raxwal and Riha, 2016). NMD typically targets transcripts containing premature termination codons (PTC) that arise as a consequence of a missense mutation or alternative splicing, leading to the translation of a truncated protein. NMD is an evolutionary highly conserved pathway that may have co-evolved with the acquisition of introns (Farlow et al., 2010). Although NMD is one of the most well-studied RNA surveillance pathways, its exact molecular mechanism is not yet fully resolved. The consensus model, based on studies across a range of organisms, suggests that NMD is induced by the premature translation termination at stop codons in the absence of a full complement of canonical translation termination signals (He and Jacobson, 2015; Lykke-Andersen and Jensen, 2015).

The most detailed insights into the NMD mechanism were obtained in mammals (He and Jacobson, 2015; Lykke-Andersen and Jensen, 2015; Hug et al., 2016). The key step

is the phosphorylation and activation of the RNA helicase UPF1 by the SMG1 kinase in response to a ribosome stalled at a PTC. Mammalian NMD primarily targets transcripts harboring introns in their 3'UTR and UPF1 activation is facilitated by the presence of an exon-exon junction 50–55 nucleotides downstream of the PTC. This is sensed via interaction with the exon junction complex, and is mediated by UPF3B and UPF2. Once phosphorylated, UPF1 is recognized by the phosphoserine-binding proteins SMG5, SMG6, and SMG7 that induce degradation of the aberrant mRNA. SMG5, SMG6, and SMG7 belong to the same protein family, which is characterized by a phosphoserine-binding domain that structurally resembles 14-3-3 proteins (Fukuhara et al., 2005). Their association with UPF1 defines two major RNA degradation pathways. SMG6 itself acts as an endonuclease cleaving mRNA in the vicinity of the PTC. SMG5 and SMG7 bind phosphorylated UPF1 as a heterodimer (Jonas et al., 2013; Nicholson et al., 2018). SMG7 recruits the CCR4-NOT deadenylase complex and induces deadenylation-dependent decapping, which is followed by 5'-3'mRNA degradation (Unterholzner and Izaurralde, 2004; Loh et al., 2013). NMD in human cells mainly relies on SMG6 mediated degradation, while the SMG5-SMG7 dimer seems to act as a back-up pathway.

The NMD mechanism described in mammals appears to be conserved in plants, although some aspects differ (Kerenyi et al., 2008; Shaul, 2015). NMD in plants acts efficiently on transcripts containing 3'UTR introns, but also on transcripts that possess long 3'UTR without introns (Kertesz et al., 2006; Kurihara et al., 2009; Kalyna et al., 2012; Drechsel et al., 2013; Degtiar et al., 2015). This mechanism, known as the faux 3'UTR model, was described in budding yeast and postulates that the long distance between the PTC and poly(A)-tail fails to provide the proper context for translation termination and sensitizes transcripts to NMD (He and Jacobson, 2015; Hug et al., 2016). Functional studies in tobacco showed that both NMD pathways require UPF1, UPF2, and SMG7, while components of the exon junction complex are only involved in the intron-based mechanism (Kerenyi et al., 2008). As in mammals, phosphorylation of UPF1 is critical for late steps of plant NMD (Merai et al., 2012; Kerenyi et al., 2013). The SMG1 kinase is present in the majority of plants, but has been repeatedly lost in the *Brassicaceae* family, suggesting that its function can be substituted by an as yet unknown kinase (Lloyd and Davies, 2013; Causier et al., 2017). Phosphorylated UPF1 recruits SMG7 through its terminal 14-3-3 domain, which in turn mediates RNA degradation and UPF1 relocalization to P-bodies (Merai et al., 2012). Plants do not possess SMG5 and SMG6 paralogs, but the genomes of dicots encode an additional SMG7-like (SMG7L) protein (Riehs et al., 2008). Nevertheless, functional studies in grapevine indicate that SMG7L has lost its NMD activity (Benkovics et al., 2011).

Downregulation of NMD in Arabidopsis leads to a strong immune response that is caused by derepression of TIR-NB-LRR immune receptors (Jeong et al., 2011; Rayson et al., 2012; Riehs-Kearnan et al., 2012; Gloggnitzer et al., 2014; Raxwal and Riha, 2016). UPF1 and SMG7 are understood to act in subsequent steps of a linear pathway; therefore, their inactivation should have similar biological effects. However, the phenotypes of UPF1

and SMG7 in Arabidopsis differ substantially. A null *upf1-3* mutation causes seedling lethality due to massive activation of the immune response (Yoine et al., 2006; Riehs-Kearnan et al., 2012). This lethality can be rescued by the inactivation of PAD4, a key component of pathogen signaling, but the *upf1-3 pad4* double mutants still exhibit retarded growth and pleiotropic developmental defects. Analysis of plants carrying different T-DNA insertions in the SMG7 gene showed range of different phenotypes. The most N-terminal disruption in the 14-3-3 domain (*smg7-5*) was proposed to cause embryonic lethality, as no mutant plants could be recovered (Riehs et al., 2008). *Smg7-1* and *smg7-3* mutants carrying insertions in the central region of SMG7 are viable, but exhibit retarded growth caused by pathogen response activation and are infertile due to abortive meiosis (Riehs et al., 2008; Bulankova et al., 2010). While the vegetative growth defects are fully rescued in *smg7-1 pad4* double mutants, meiosis is still defective. Furthermore, the meiotic defects are not observed in *upf1-3 pad4* mutants, suggesting that the role of SMG7 in meiosis is not linked to NMD (Riehs-Kearnan et al., 2012).

Here, we aimed to clarify whether Arabidopsis SMG7 is an essential gene by characterizing SMG7-loss of function mutants generated by CRISPR-Cas9 mutagenesis. Furthermore, we investigated the requirements for the 14-3-3 domain in meiosis and the redundancy between SMG7 and SMG7L paralogs. Finally, we used phylogenetic analysis to search for motifs that may define SMG7 function in NMD and meiosis.

## MATERIALS AND METHODS

### Plant Material

Seeds of *Arabidopsis thaliana* were grown in soil in phytotrons at 20–22°C with a photoperiod of 16 h light/8 h dark. All wild type and mutant plants were derived from *A. thaliana* ecotype Col-0. The *smg7-1* and *smg7-6* mutants used in this study were characterized previously (Riehs et al., 2008; Riehs-Kearnan et al., 2012) and the *smg7l-1* mutation was obtained from the Arabidopsis Stock Centre (SAIL\_634H06).

### Vectors and Transgenic Plants

We used the CRISPR/Cas9 system developed for targeted mutagenesis in plants (Fauser et al., 2014). The CRISPR sequence guiding the Cas9 nuclease was designed to target the second exon of the SMG7 gene downstream of and near to the ATG codon. Briefly, the guide RNA was created by cloning the duplex oligonucleotide 5'-CTTGGCTCGCTCCCATGAAG-3' into the *BbsI* site of pEN-Chimera (Fauser et al., 2014). The region coding for the gRNA was transferred into pDe-CAS9 using the Gateway<sup>TM</sup> cloning system, creating the binary vector pCCC40. Transgenic plants were generated by Agrobacterium-mediated floral-dip transformation of *A. thaliana* ecotype Col-0. Transgenic plants were screened in the T2 generation by high resolution melting (HRM) point analysis for the presence of sequence polymorphisms at the target site. The CRISPR/Cas9 transgene was out segregated in the T3 generation and plants carrying the *smg7-7* allele in a heterozygous

constitution were identified by HRM analysis and sequencing. Plants were characterized in the T4 and T5 generations. The complementation constructs for functional significance of the 14-3-3 domain of SMG7 were as follows: 6.9 kb of SMG7 genomic sequence including 1.5 kb promoter and 3'UTR was PCR amplified using primers SMG7::SMG7-SbfI and SMG7::SMG7-SacI (**Supplementary Table S1**) and cloned into pJET1.2/blunt Cloning Vector (Thermo Fisher Scientific). For site directed mutagenesis, a smaller fragment of SMG7 gene was amplified with the primers F-SMG7-SpeI and R-SMG7-HpaI and the PCR product was subcloned in the pJET1.2/blunt Cloning Vector. Site directed mutagenesis of K77E and R185E was done in two rounds of PCR, first with primers F-SMG7-K77E and R-SMG7-K77E, and second with primers F-SMG7-R185E and R-SMG7-R185E. The mutated fragment was subcloned back into SpeI/HpaI of the SMG7 gene construct. Wild type and mutated versions of the SMG7 gene were subcloned into the PstI/SacI sites of the binary vector pCBK06 (Riha et al., 2002) and resulting constructs were transformed into *Agrobacterium tumefaciens* GV3101. Complementation lines were generated by transforming Arabidopsis heterozygous for the *smg7-1* allele by floral dip technique.

## Fertility Assays

Pollen viability was determined by Alexander staining (Alexander, 1969). Meiosis in pollen mother cells was analyzed by staining whole anthers with DAPI followed by confocal microscopy according to (Brownfield et al., 2015).

## Gene Expression Analysis

Total RNA was isolated from plant leaf tissue using RNA Blue (Top-Bio) following the manufacturer's protocol. Quality and quantity of isolated RNA were checked on a 1.2% denaturing agarose gel and with the NanoDrop 2000C spectrophotometer (Thermo Fisher Scientific), respectively. For each sample, 10 µg of total RNA was treated with DNase I (Roche) to remove genomic DNA and 2 µg DNA-free RNA was reverse transcribed using Superscript IV (Invitrogen) following the manufacturer's protocol. For qRT-PCR reactions, cDNA was diluted two times and all reactions were carried out using EvaGreen dye (Biotium), GoTaq polymerase (Promega, United States), and gene-specific primers on a LightCycler 96 (Roche). Expression was normalized to At2G28390 and relative fold change was calculated using the delta-delta Ct method (Livak and Schmittgen, 2001). All qRT-PCR reactions were performed on at least three biological replicates and 2–3 technical replicates. The sequence of all primers used in this study are listed in **Supplementary Table S1**.

## Protein Alignment

We downloaded the full-length protein sequences of SMG7 and SMG7L from 13 plant species (*Physcomitrella patens*, *Selaginella moellendorffii*, *Amborella trichopoda*, *Medicago truncatula*, *Arabidopsis thaliana*, *Brassica oleracea*, *Cucumis sativus*, *Nicotiana attenuata*, *Helianthus annuus*, *Oryza sativa*, *Zea mays*, *Sorghum bicolor*, *Hordeum vulgare*) from Ensembl Plants (release 38) and of human (*Homo sapiens*) from Ensembl (release 91). The following sequences

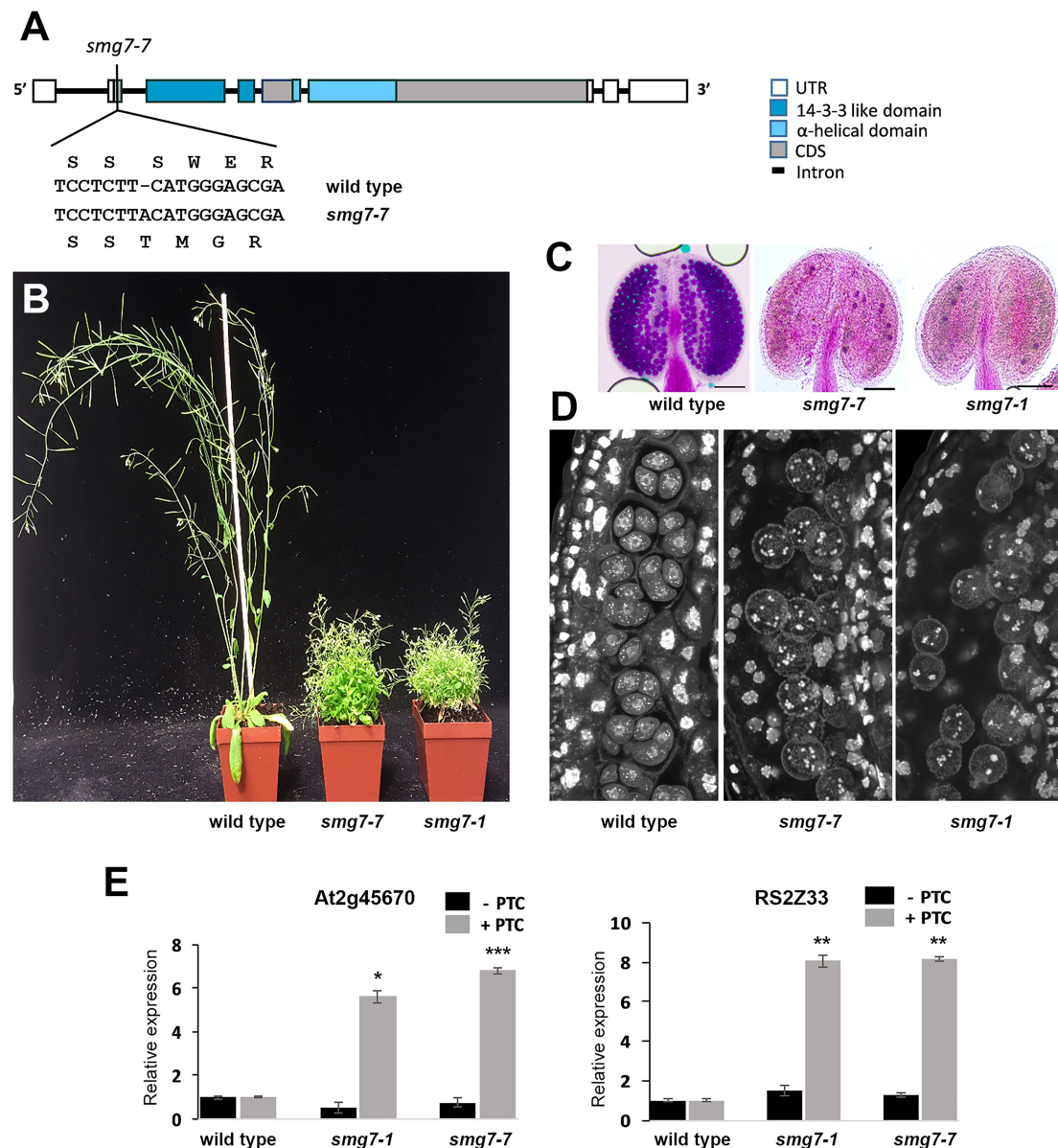
were used for alignment: *HsSMG7* (ENSP00000425133), *PpSMG7a* (PP1S80\_14V6.1), *PpSMG7b* (PP1S311\_73V6.1), *SmSMG7a* (EFJ27061), *SmSMG7b* (EFJ21470), *AmtSMG7* (ERN18017), *MtSMG7a* (KEH28378), *MtSMG7b* (KEH16467), *AtSMG7* (AT5G19400.1), *AtSMG7L* (AT1G28260.1), *BoSMG7a* (Bo9g153800.1), *BoSMG7b* (Bo2g018020.1), *BoSMG7La* (Bo5g054690.1), *BoSMG7Lb* (Bo3g143280.1), *CsSMG7* (KGN66550), *CsSMG7L* (KGN64688), *NaSMG7* (OIS97991), *NaSMG7L* (OIT28005), *HaSMG7a* (OTG27135), *HaSMG7b* (OTG30173), *HaSMG7L* (OTF97490), *OsSMG7* (Os08t0305300-01), *ZmSMG7a* (Zm00001d019920\_P002), *ZmSMG7b* (Zm00001d005502\_P002), *SbSMG7* (KXG35214), *HvSMG7a* (HORVU5Hr1G050800.5), *HvSMG7b* (HORVU0Hr1G029520.1). Protein alignment was performed with the MegAlign Pro function of DNASTAR Navigator (v12.2.0.80) using the MUSCLE (Multiple Sequence Comparison by Log-Expectation) method. The multiple sequence alignment (MSA) for selected sequence features of the SMG7 protein across the plant kingdom was used to generate a sequence logo using the online version of WebLogo3 (Crooks et al., 2004). SMG7 specific regions were identified by visually inspecting multiple sequence alignment of plant SMG7/SMG7L proteins for regions that are conserved in SMG7 paralogs of monocots and dicots, but diverged in SMG7L.

## RESULTS

### Full Inactivation of Arabidopsis SMG7 Does Not Result in Embryonic Lethality

In our attempts to genetically complement the *smg7-5* mutation with a functional SMG7 gene, we failed to recover plants homozygous for the *smg7-5* allele. This led us to ask whether the embryonic lethality observed in the *smg7-5* line (Riehs et al., 2008) is indeed linked to the T-DNA insertion in the SMG7 gene. To clarify this issue, we used CRISPR/Cas9 targeted mutagenesis to generate another mutation that disrupts the conserved 14-3-3 domain. We targeted the Cas9 nuclease to the second exon of the SMG7 gene and identified an allele that contains a frame shift mutation 13 amino acids downstream of the ATG codon and hence likely represents a full loss of function allele that we call *smg7-7* (**Figure 1A**). We were able to recover viable plants homozygous for the *smg7-7* allele that phenotypically resembled the growth-retarded *smg7-1* mutants (**Figure 1B**). These plants were sterile and did not produce any pollen (**Figure 1C**). As the infertility of *smg7-1* was caused by arrest of meiotic progression in anaphase of the second meiotic division (Riehs et al., 2008), we performed cytogenetic analysis of pollen mother cells in anthers of *smg7-7* mutants. This revealed meiocytes with irregularly distributed chromosomes, which is typical for this anaphase II arrest (**Figure 1D**). To test the effect of *smg7-7* on NMD, we used an assay that quantifies two alternatively spliced variants of the same transcript, one of which contains a PTC (Gloggnitzer et al., 2014; **Figure 1E**). We observed comparable upregulation of two different PTC containing transcripts in *smg7-1* and *smg7-7*, suggesting that NMD is impaired to the similar extent in both alleles. Thus, *smg7-1* and *smg7-7* lead to the same effect on NMD





**FIGURE 1 |** Comparison of Arabidopsis *smg7-7* and *smg7-1* mutants. **(A)** Schematic diagram of the SMG7 gene with the position of the *smg7-7* mutation indicated. **(B)** Approximately 6-week-old wild type and mutant plants. **(C)** Anthers with viable pollen visualized by Alexander staining. **(D)** Developing pollen mother cells within an anther stained by DAPI. Tetrad nuclei are apparent in wild type, while late *smg7* meiocytes contain randomly distributed chromatids. **(E)** Effect of *smg7* alleles on the relative abundance of two alternatively spliced variants of the same transcript as determined by real time RT-PCR. Error bars represent SEM of three biological replicates. Asterisks indicate statistical significance of difference from wild type (\* $P < 0.05$ , \*\* $P < 0.01$ , \*\*\* $P < 0.001$ , two-tailed  $t$ -test).

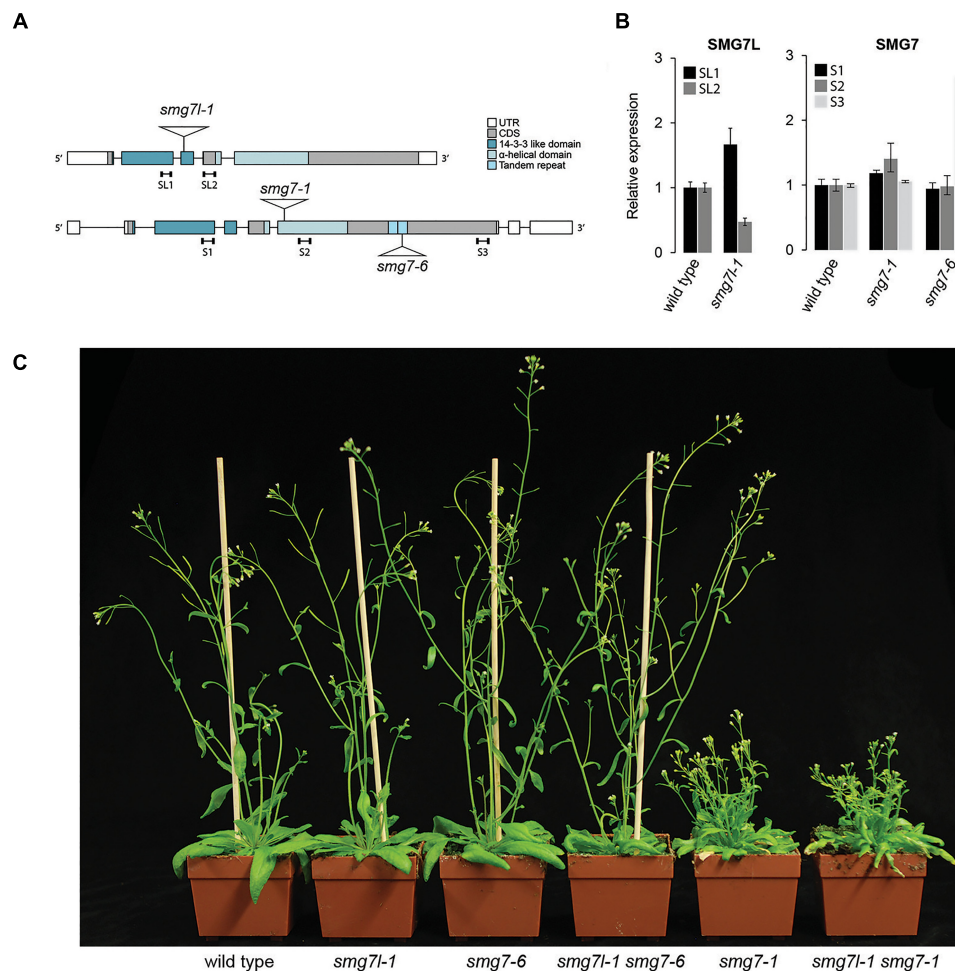
and meiosis, demonstrating that the *smg7-1* allele used in our previous studies likely represents a full loss of function mutation and that inactivation of SMG7 is not lethal, in contrast to the conclusion previously reported (Riehs et al., 2008).

## Arabidopsis SMG7L Does Not Compensate for the Loss of SMG7

Our observation that complete loss of SMG7 does not cause embryonic lethality implies that the consequence of SMG7

inactivation is milder than that of UPF1. This led us to ask whether SMG7L can partially compensate for the loss of SMG7 in NMD. To address this question, we generated plants with mutations in both genes. We combined *smg7-1*, which carries a T-DNA insertion in the N-terminal 14-3-3 like domain, with either the severe *smg7-1*, or the hypomorphic *smg7-6* (Figure 2A). *smg7-6* contains a T-DNA insertion in the C-terminus and was reported to be proficient for NMD but partially impaired in meiosis (Riehs-Kearnan et al., 2012). Quantitative RT-PCR analysis showed that mRNA encoding the





**FIGURE 2 |** Double mutants of *SMG7* and *SMG7L*. **(A)** Diagram of *SMG7* and *SMG7L* genes with positions of T-DNA insertions and functional domains indicated. **(B)** qRT-PCR analysis of *SMG7L* and *SMG7* transcripts. Positions of the regions amplified by PCR are indicated in **(A)**. Error bars represent SEM of three biological replicates. **(C)** Five week-old wild type and mutant plants.

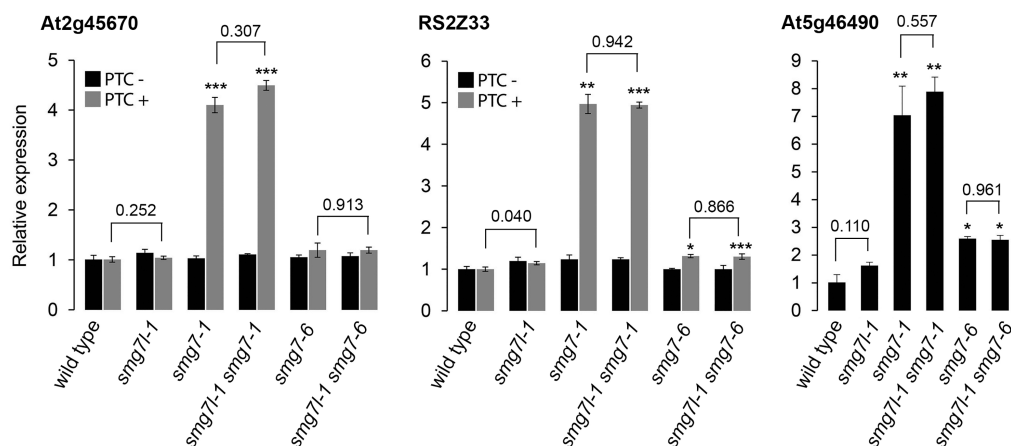
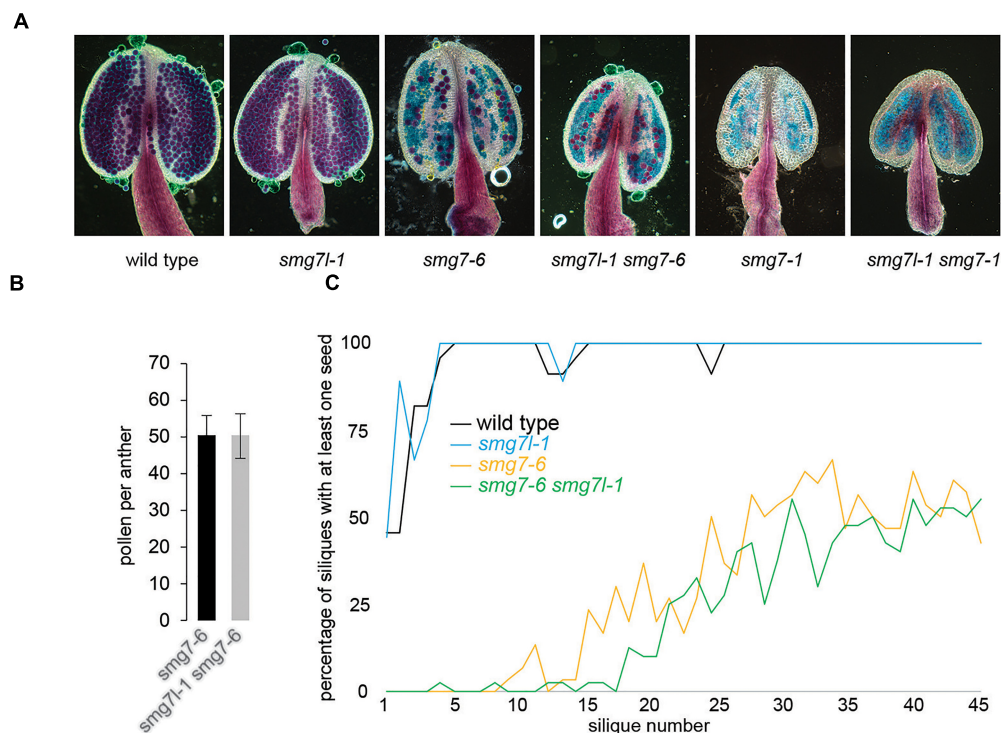
N-terminal portion of the *SMG7* protein is expressed at the same level in *smg7-6* as in wild type plants, suggesting the production of a truncated, partially functional protein (Figure 2B). Plants deficient for *SMG7L* did not show any obvious phenotype and were indistinguishable from wild type plants (Figure 2C).

We anticipated that if *SMG7L* acts redundantly with *SMG7*, it should exacerbate the phenotypes of both *smg7* alleles. However, neither *smg7l-1 smg7-1* nor *smg7l-1 smg7-6* double mutants exhibited any difference from the respective *smg7* single mutants with regards to growth defects and fertility (Figures 2C, 3). While *smg7-1* is infertile, *smg7-6* plants produce a reduced amount of viable pollen and are partially fertile (Figures 3A,B). Usually, the first 15–20 flowers on the main inflorescence bolt are infertile, while later flowers produce viable seeds (Figure 3C). Quantification of pollen and seed production did not reveal any difference between *smg7-6* and *smg7l-1 smg7-6*, arguing that *SMG7L* does not compensate for the meiotic function of *SMG7*. We assessed the efficiency of NMD by quantifying three different endogenous transcripts known to be degraded

by NMD (Figure 4; Gloggnitzer et al., 2014). The *smg7-1* null mutation displayed increased amounts of all transcripts targeted by NMD (Figure 4). Although we have previously reported that the hypomorphic *smg7-6* allele is NMD proficient (Riehs-Kearnan et al., 2012), we observed a slight but reproducible increase of the three NMD-targeted transcripts in this study (Figure 4) suggesting a very mild NMD defect. Importantly, the *smg7l-1 smg7-1* and *smg7l-1 smg7-6* double mutants did not show any additional increase in NMD-regulated transcripts. Together, these data allow the conclusion that *SMG7L* does not act redundantly with *SMG7* either in meiosis or in NMD.

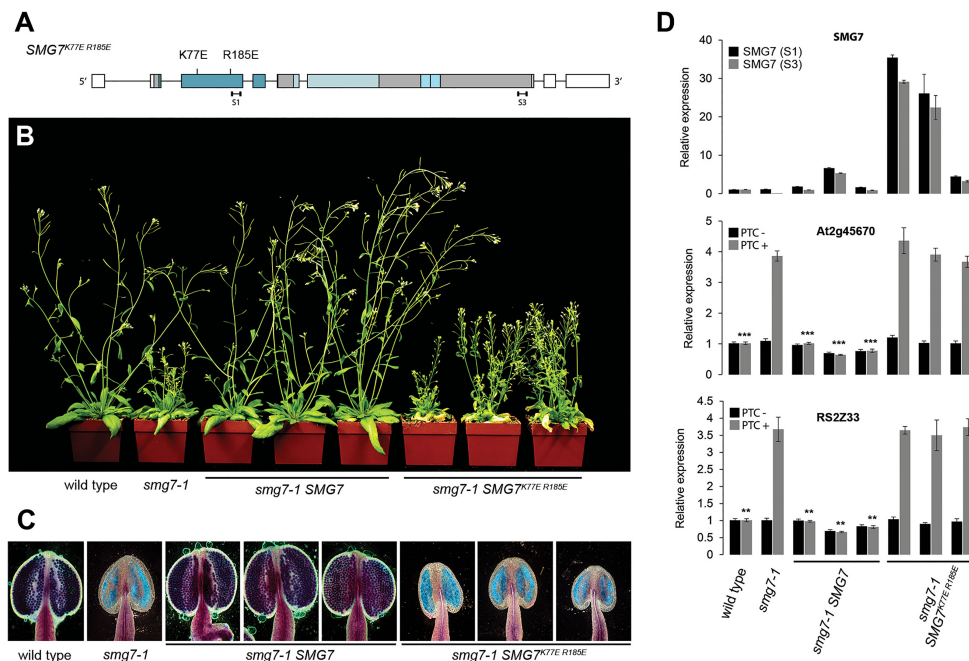
## The N-Terminal 14-3-3 Domain of *SMG7* Is Important for Meiosis

The genetic separation of NMD and meiotic functions in *smg7-6* mutants together with the fact that *UPF1* and *UPF3* deficient plants lack a meiotic phenotype (Riehs-Kearnan et al., 2012) indicate that *SMG7* acts in meiosis through a different



mechanism than in NMD. Thus, we wanted to know whether the conserved 14-3-3 domain that mediates interaction with phosphorylated UPF1 in NMD is also required for meiosis. We complemented the *smg7-1* with a copy of the Arabidopsis SMG7 gene carrying mutations in the conserved K77 and R185 residues that form the phosphoserine-binding pocket and are required for NMD in tobacco (Figure 5A; Fukuhara et al., 2005;

Merai et al., 2012). We transformed plants heterozygous for *smg7-1* with constructs carrying either wild type SMG7 or the mutant SMG7<sup>K77E R185E</sup> and identified transformants homozygous for *smg7-1* in segregating T2 populations. Three independent lines were analyzed for each construct. While the wild type construct readily complemented the growth phenotype, fertility, and pollen production, SMG7<sup>K77E R185E</sup> plants were



**FIGURE 5 |** Complementation of *smg7-1* with *SMG7<sup>K77E R185E</sup>*. **(A)** Schematic diagram of the SMG7 gene with the K77E and R185E mutations indicated. **(B)** Five week-old T2 *smg7-1* plants complemented with either the wild type or *SMG7<sup>K77E R185E</sup>* construct. **(C)** Anthers assayed by Alexander staining. **(D)** Relative abundance of SMG7 and two NMD reporter genes determined by real time RT-PCR. Two sets of primers were used for quantification of SMG7 mRNA (S1, S3) and their position is indicated in panel **A**. Error bars represent SEM from three biological replicas. Asterisks indicate statistical significance of difference from *smg7-1* mutants (\* $P < 0.5$ , \*\* $P < 0.01$ , \*\*\* $P < 0.001$ , two-tailed  $t$ -test). Plants derived from three independent transformants are shown in each category in **(B–D)**.

indistinguishable from *smg7-1* mutants despite higher level of SMG7 mRNA expression (Figures 5B,C). As expected, the *SMG7<sup>K77E R185E</sup>* plants were deficient in NMD, but they also failed to produce any pollen (Figures 5C,D). This indicates that the phosphoserine-binding activity mediated by the conserved N-terminal 14-3-3 motif is important not only for NMD, but also for the meiotic function of SMG7.

## The Meiotic Function of SMG7 May Depend on Conserved Motifs in Its C-Terminus

Studies in grapevine using virus induced gene silencing (Benkovics et al., 2011) together with our phenotypic characterization of Arabidopsis mutants demonstrated that SMG7 and SMG7L paralogs functionally diverged in dicots. While SMG7 retained its function in NMD, SMG7L apparently acquired a novel function(s) unrelated to NMD and meiosis. We hypothesized that this neofunctionalization would be accompanied by amino acid sequence divergence in regions exclusively important for NMD. According to this hypothesis, SMG7 should have retained motifs involved in NMD whereas some of these motifs may have been lost in SMG7L.

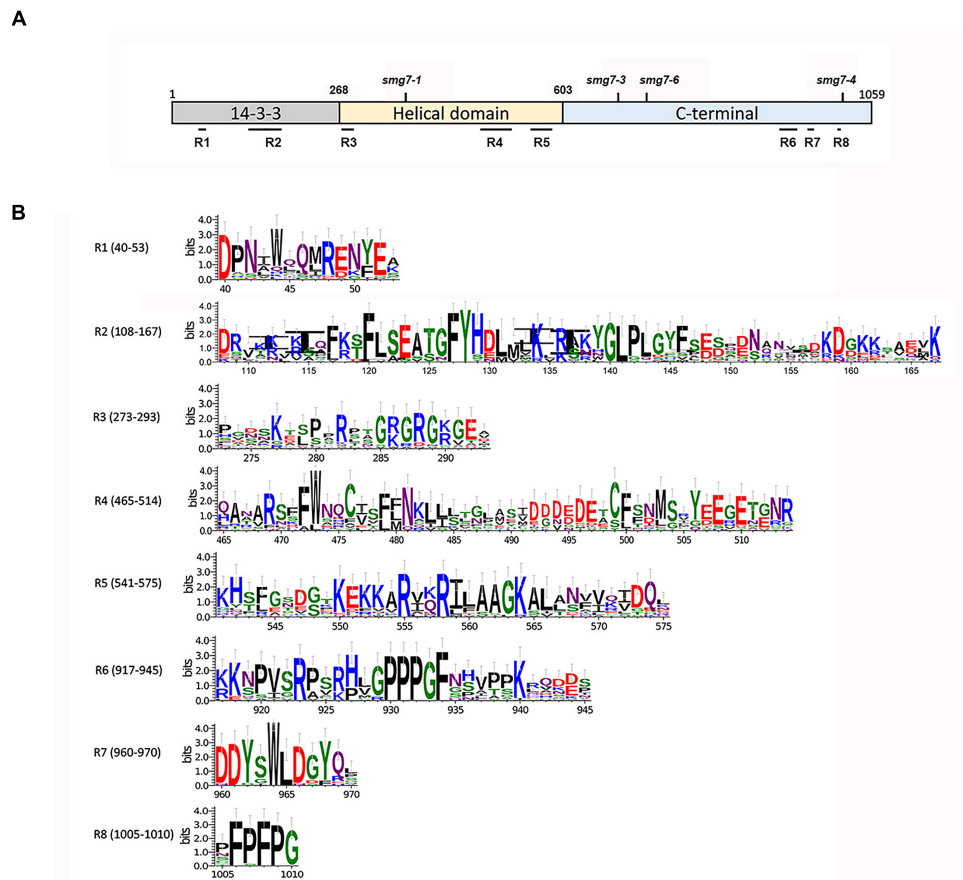
We performed a sequence comparison of multiple plant SMG7 and SMG7L proteins to identify regions that are conserved across the plant kingdom but which diverged during the evolution of SMG7L in dicots. Five such regions were identified in the evolutionary conserved N-terminal half of SMG7, and three

additional SMG7-specific motifs were found at the C-terminus (Figure 6). While all 6 amino acids residues that form the phosphoserine-binding pocket of the 14-3-3 domain are retained in both SMG7 and SMG7L, two regions in the 14-3-3 domain diverged between these paralogs. When mapped over the structure of human SMG7 (Fukuhara et al., 2005), the plant SMG7-specific regions 1 and 2 correspond to alpha helices  $\alpha 2$  and  $\alpha 4$  with the connecting loop to  $\alpha 5$ , respectively. The  $\alpha 2$  and  $\alpha 4$  helices are aligned on the convex surface of the 14-3-3 domain, and  $\alpha 4$  together with the extended loop was reported to form an interaction interface with SMG5 (Jonas et al., 2013). Thus, it is likely that these regions form a protein interaction surface and their divergence in SMG7L reflects altered protein-binding specificity. The SMG7-specific region 3 is located in the loop that connects the 14-3-3 domain with the helical domain and is larger in SMG7 than in SMG7L paralogs (Figure 6). The SMG7-specific regions 4 and 5 span helices  $\alpha 16$  and  $\alpha 18$  at the end of the helical domain. The C-terminal portion of SMG7 is generally less conserved than the N-terminus. Nevertheless, we found three motifs at the C-terminus of SMG7 proteins (motifs 6, 7, and 8) that are shared among monocots, dicots, and moss, but were lost in dicot SMG7L (Figure 6).

## DISCUSSION

SMG7 is an ancient phosphoserine-binding protein that is present in most eukaryotes. Its primary function is linked to





**FIGURE 6 |** Amino acid sequence motifs specific for the Arabidopsis SMG7 protein. **(A)** Schematic representation of the SMG7 protein and of its domains. Amino acid coordinates for domain boundaries, positions of SMG7-specific regions, and disruptions of SMG7 by T-DNA insertions are indicated. **(B)** Consensus sequences of individual SMG7-specific regions assembled from alignment of proteins from 13 plant species (**Supplementary Figure S1**). The x-axis represent the position in the *A. thaliana* SMG7 protein sequence. The y-axis labels represent the information content of the position in sequence (in bits).

NMD as its physical and functional interactions with UPF1 are conserved from plants to animals (Unterholzner and Izaurralde, 2004; Kerenyi et al., 2013). Nevertheless, proteins of the SMG7 family have acquired additional functions during evolution. The best known example is Est1p in budding yeast that recruits telomerase to chromosome ends via binding to phosphorylated Cdc13 (Li et al., 2009; Chen et al., 2018). Hence, SMG7 acts as an adaptor protein that can recruit different molecular machines to the sites of their action in a phosphorylation dependent manner. Arabidopsis SMG7 is involved in at least two independent molecular processes, NMD and meiotic progression. Here we show that its N-terminal 14-3-3 domain is required for both NMD and meiosis. While UPF1 is the substrate that defines the role of SMG7 in NMD (Kerenyi et al., 2013), the meiotic function is likely mediated by another binding protein as Arabidopsis deficient for UPF1 does not exhibit the meiotic defects described for *smg7* mutants (Riehs-Kearnan et al., 2012).

Proteins of the SMG7 gene family underwent independent multiplications during evolution that were accompanied by functional diversification. In vertebrates SMG5, SMG6, and SMG7 paralogs play distinct roles in NMD that define two

separate RNA degradation pathways. Multiple copies of SMG7 have also independently arisen in plants including maize, poplar, grapevine, or the moss *P. patens* (Benkovics et al., 2011; Li et al., 2015). In the majority of cases, these are relatively recent species-specific duplications that only occasionally span larger phylogenetic units. SMG7L represents the most ancient plant duplication, originating at the root of dicots (Riehs et al., 2008). Here we demonstrate that SMG7L does not act redundantly with SMG7 in either NMD or meiosis implying that it evolved a novel, yet unknown function. Domain swapping experiments indicated that the N-terminal portion of SMG7L retained its capability to bind UPF1 and complement NMD, whereas the C-terminus lost its capability to trigger RNA degradation in a tethering assay (Benkovics et al., 2011). Nevertheless, it is unlikely that UPF1 is a physiological substrate of SMG7L; sequence divergence in the loops and helices that form the surface interfaces of the conserved N-terminal domains rather suggest that SMG7L binds other proteins.

Phylogenetic comparison of SMG7 and SMG7L protein sequences revealed several regions that were conserved in SMG7, but diverged in SMG7L. Positions of the SMG7-specific motifs



in respect to mutations that have been functionally characterized (Figure 6A) may provide clues on their function. Arabidopsis *smg7-1* and *smg7-3* represent T-DNA insertions that truncate SMG7 at amino acid positions 355 and 675, and both alleles are deficient in NMD and meiosis (Riehs et al., 2008; Riehs-Kearnan et al., 2012). The *smg7-6* disruption is only 43 amino acids downstream from *smg7-3* at position 718 yet is only very mildly impaired in NMD, implying a functional significance of this region. Nevertheless, the region between amino acids 675 and 718 is highly divergent among plant SMG7 homologs (Supplementary Figure S1), which speaks against a specific role for this region in NMD. Since *smg7-3* is disrupted shortly after a motif conserved in both SMG7 and SMG7L paralogs, this mutation may affect the proper folding or stability of the truncated protein. Experiments using a transient expression in tobacco showed that the Arabidopsis SMG7 C-terminal domain is required for NMD and that tethering a C-terminal fragment (amino acids 517–1059) to RNA is sufficient to trigger its degradation (Benkovics et al., 2011). Thus, the region responsible for RNA degradation must be located between amino acids 517 and 718, which overlaps with the SMG7-specific region 5 (Figure 6). The impaired fertility of *smg7-6* indicates a requirement for the SMG7 C-terminus in meiosis. It is likely that the meiotic function of SMG7 is mediated additionally through conserved regions 6, 7, and 8. In support of this notion, Arabidopsis *smg7-4* which carry a T-DNA disruption at amino acid position 1013 immediately after the SMG7-specific region 8 are fully fertile (Riehs-Kearnan et al., 2012).

Whether the conserved function of these motifs is specifically linked to reproduction, or whether they underlie a general molecular function, for which deficiency in Arabidopsis is manifested as a meiotic defect, remains to be seen.

## AUTHOR CONTRIBUTIONS

CC, NS, KR, and OMS designed the experiments. CC performed the experiments shown in Figures 1–5. NS performed the

experiments shown in Figures 1, 6. AW generated the *smg7-7* mutant plants. KR and NS wrote the manuscript.

## FUNDING

The authors are grateful for financial support by the Austrian Science Fund (FWF W 1238, DK “Chromosome Dynamics” to KR and OMS), by the Ministry of Education, Youth and Sports of the Czech Republic, European Regional Development Fund-Project “REMAP” (No. CZ.02.1.01/0.0/0.0/15\_003/0000479 to KR), and by the Czech Science Foundation (16-18578S to KR).

## ACKNOWLEDGMENTS

We acknowledge the Plant Science core facility and the core facility CELLIM of CEITEC supported by the Czech-BioImaging large RI project (LM2015062 funded by MEYS CR) for their support with obtaining scientific data presented in this paper.

## SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fpls.2018.01602/full#supplementary-material>

**FIGURE S1** | Sequence alignment of the following SMG7 proteins: *HsSMG7* (ENSP00000425133), *PpSMG7a* (PP1S80\_14V6.1), *PpSMG7b* (PP1S311\_73V6.1), *SmSMG7a* (EFJ27061), *SmSMG7b* (EFJ21470), *AmtSMG7* (ERN18017), *MtSMG7a* (KEH28378), *MtSMG7b* (KEH16467), *AtSMG7* (AT5G19400.1), *AtSMG7L* (AT1G28260.1), *BoSMG7a* (Bo9g153800.1), *BoSMG7b* (Bo2g018020.1), *BoSMG7La* (Bo5g054690.1), *BoSMG7Lb* (Bo3g143280.1), *CsSMG7* (KGN66550), *CsSMG7L* (KGN64688), *NaSMG7* (OIS97991), *NaSMG7L* (OIT28005), *HaSMG7a* (OTG27135), *HaSMG7b* (OTG30173), *HaSMG7L* (OTF97490), *OsSMG7* (Os08t0305300-01), *ZmSMG7a* (Zm00001d019920\_P002), *ZmSMG7b* (Zm00001d005502\_P002), *SbSMG7* (KXG35214), *HvSMG7a* (HORVU5Hr1G050800.5), *HvSMG7b* (HORVU0Hr1G029520.1).

**TABLE S1** | Primers used in this study.

## REFERENCES

- Alexander, M. P. (1969). Differential staining of aborted and nonaborted pollen. *Stain Technol.* 44, 117–122. doi: 10.3109/10520296909063335
- Benkovics, A. H., Nyiko, T., Merai, Z., Silhavy, D., and Bisztray, G. D. (2011). Functional analysis of the grapevine paralogs of the SMG7 NMD factor using a heterolog VIGS-based gene depletion-complementation system. *Plant Mol. Biol.* 75, 277–290. doi: 10.1007/s11103-010-9726-0
- Brownfield, L., Yi, J., Jiang, H., Minina, E. A., Twell, D., and Kohler, C. (2015). Organelles maintain spindle position in plant meiosis. *Nat. Commun.* 6:6492. doi: 10.1038/ncomms7492
- Bulankova, P., Riehs-Kearnan, N., Nowack, M. K., Schnittger, A., and Riha, K. (2010). Meiotic progression in *Arabidopsis* is governed by complex regulatory interactions between SMG7, TDM1, and the meiosis I-specific cyclin TAM. *Plant Cell* 22, 3791–3803. doi: 10.1105/tpc.110.078378
- Causier, B., Li, Z., De Smet, R., Lloyd, J. P. B., Van de Peer, Y., and Davies, B. (2017). Conservation of nonsense-mediated mRNA decay complex components throughout eukaryotic evolution. *Sci. Rep.* 7:16692. doi: 10.1038/s41598-017-16942-w
- Chen, H., Xue, J., Churikov, D., Hass, E. P., Shi, S., Lemon, L. D., et al. (2018). Structural insights into yeast telomerase recruitment to telomeres. *Cell* 172, 331.e13–343.e13. doi: 10.1016/j.cell.2017.12.008
- 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
- Degtiar, E., Fridman, A., Gottlieb, D., Vexler, K., Berezin, I., Farhi, R., et al. (2015). The feedback control of UPF3 is crucial for RNA surveillance in plants. *Nucleic Acids Res.* 43, 4219–4235. doi: 10.1093/nar/gkv237
- 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
- Farlow, A., Meduri, E., Dolezal, M., Hua, L., and Schlötterer, C. (2010). Nonsense-mediated decay enables intron gain in *Drosophila*. *PLoS Genet.* 6:e1000819. doi: 10.1371/journal.pgen.1000819

- Fausser, F., Schiml, S., and Puchta, H. (2014). Both CRISPR/Cas-based nucleases and nickases can be used efficiently for genome engineering in *Arabidopsis thaliana*. *Plant J.* 79, 348–359. doi: 10.1111/tpj.12554
- Fukuhara, N., Ebert, J., Unterholzner, L., Lindner, D., Izaurralde, E., and Conti, E. (2005). SMG7 is a 14-3-3-like adaptor in the nonsense-mediated mRNA decay pathway. *Mol. Cell.* 17, 537–547. doi: 10.1016/j.molcel.2005.01.010
- Gloggnitzer, J., Akimcheva, S., Srinivasan, A., Kusenda, B., Riehs, N., Stampfl, H., et al. (2014). Nonsense-mediated mRNA decay modulates immune receptor levels to regulate plant antibacterial defense. *Cell Host Microbe* 16, 376–390. doi: 10.1016/j.chom.2014.08.010
- 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
- Hug, N., Longman, D., and Caceres, J. F. (2016). Mechanism and regulation of the nonsense-mediated decay pathway. *Nucleic Acids Res.* 44, 1483–1495. doi: 10.1093/nar/gkw010
- Jeong, H. J., Kim, Y. J., Kim, S. H., Kim, Y. H., Lee, I. J., Kim, Y. K., et al. (2011). Nonsense-mediated mRNA decay factors, UPF1 and UPF3, contribute to plant defense. *Plant Cell Physiol.* 52, 2147–2156. doi: 10.1093/pcp/pcr144
- Jonas, S., Weichenrieder, O., and Izaurralde, E. (2013). An unusual arrangement of two 14-3-3-like domains in the SMG5-SMG7 heterodimer is required for efficient nonsense-mediated mRNA decay. *Genes Dev.* 27, 211–225. doi: 10.1101/gad.206672.112
- Kalya, 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
- Kerenyi, F., Wawer, I., Sikorski, P. J., Kufel, J., and Silhavy, D. (2013). Phosphorylation of the N- and C-terminal UPF1 domains plays a critical role in plant nonsense-mediated mRNA decay. *Plant J.* 76, 836–848. doi: 10.1111/tpj.12346
- 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
- Kertesz, S., Kerenyi, Z., Merai, Z., Bartos, I., Palfy, T., Barta, E., et al. (2006). Both introns and long 3'-UTRs operate as cis-acting elements to trigger nonsense-mediated decay in plants. *Nucleic Acids Res.* 34, 6147–6157. doi: 10.1093/nar/gkl737
- Kurihara, Y., Matsui, A., Hanada, K., Kawashima, M., Ishida, J., Morosawa, T., et al. (2009). Genome-wide suppression of aberrant mRNA-like noncoding RNAs by NMD in *Arabidopsis*. *Proc. Natl. Acad. Sci. U.S.A.* 106, 2453–2458. doi: 10.1073/pnas.0808902106
- Li, S., Chen, L., Zhang, L., Li, X., Liu, Y., Wu, Z., et al. (2015). BnaC9.SMG7b functions as a positive regulator of the number of seeds per silique in *Brassica napus* by regulating the formation of functional female gametophytes. *Plant Physiol.* 169, 2744–2760. doi: 10.1104/pp.15.01040
- Li, S., Makovets, S., Matsuguchi, T., Blethrow, J. D., Shokat, K. M., and Blackburn, E. H. (2009). Cdk1-dependent phosphorylation of Cdc13 coordinates telomere elongation during cell-cycle progression. *Cell* 136, 50–61. doi: 10.1016/j.cell.2008.11.027
- Livak, K. J., and Schmittgen, T. D. (2001). Analysis of relative gene expression data using real-time quantitative PCR and the 2<sup>-ΔΔC<sub>T</sub></sup> Method. *Methods* 25, 402–408. doi: 10.1006/meth.2001.1262
- Lloyd, J. P., and Davies, B. (2013). SMG1 is an ancient nonsense-mediated mRNA decay effector. *Plant J.* 76, 800–810. doi: 10.1111/tpj.12329
- Loh, B., Jonas, S., and Izaurralde, E. (2013). The SMG5-SMG7 heterodimer directly recruits the CCR4-NOT deadenylase complex to mRNAs containing nonsense codons via interaction with POP2. *Genes Dev.* 27, 2125–2138. doi: 10.1101/gad.226951.113
- Lykke-Andersen, S., and Jensen, T. H. (2015). Nonsense-mediated mRNA decay: an intricate machinery that shapes transcriptomes. *Nat. Rev. Mol. Cell Biol.* 16, 665–677. doi: 10.1038/nrm4063
- Merai, Z., Benkovics, A. H., Nyiko, T., Debreczeny, M., Hiripi, L., Kerenyi, Z., et al. (2012). The late steps of plant nonsense-mediated mRNA decay. *Plant J.* 73, 50–62. doi: 10.1111/tpj.12015
- Nicholson, P., Gkratsou, A., Josi, C., Colombo, M., and Muhlemann, O. (2018). Dissecting the functions of SMG5, SMG7, and PNRC2 in nonsense-mediated mRNA decay of human cells. *RNA* 24, 557–573. doi: 10.1261/rna.063719.117
- Raxwal, V. K., and Riha, K. (2016). Nonsense mediated RNA decay and evolutionary capacitance. *Biochim. Biophys. Acta* 1859, 1538–1543. doi: 10.1016/j.bbagr.2016.09.001
- 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:e31917. doi: 10.1371/journal.pone.0031917
- Riehs, N., Akimcheva, S., Puizina, J., Bulankova, P., Idol, R. A., Siroky, J., et al. (2008). *Arabidopsis* SMG7 protein is required for exit from meiosis. *J. Cell Sci.* 121, 2208–2216. doi: 10.1242/jcs.027862
- Riehs-Kearnan, N., Gloggnitzer, J., Dekrout, B., Jonak, C., and Riha, K. (2012). Aberrant growth and lethality of *Arabidopsis* deficient in nonsense-mediated RNA decay factors is caused by autoimmune-like response. *Nucleic Acids Res.* 40, 5615–5624. doi: 10.1093/nar/gks195
- Riha, K., Watson, J. M., Parkey, J., and Shippen, D. E. (2002). Telomere length deregulation and enhanced sensitivity to genotoxic stress in *Arabidopsis* mutants deficient in Ku70. *EMBO J.* 21, 2819–2826. doi: 10.1093/emboj/21.11.2819
- 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
- Unterholzner, L., and Izaurralde, E. (2004). SMG7 acts as a molecular link between mRNA surveillance and mRNA decay. *Mol. Cell.* 16, 587–596. doi: 10.1016/j.molcel.2004.10.013
- Yoine, M., Nishii, T., and Nakamura, K. (2006). *Arabidopsis* UPF1 RNA helicase for nonsense-mediated mRNA decay is involved in seed size control and is essential for growth. *Plant Cell Physiol.* 47, 572–580. doi: 10.1093/pcp/pcj035

**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 © 2018 Capitao, Shukla, Wandrolova, Mittelsten Scheid and Riha. 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.



# Under a New Light: Regulation of Light-Dependent Pathways by Non-coding RNAs

Camila Sánchez-Retuerta<sup>1</sup>, Paula Suárez-López<sup>1</sup> and Rossana Henriques<sup>1,2,3\*</sup>

<sup>1</sup> Centre for Research in Agricultural Genomics, CSIC-IRTA-UAB-UB, Barcelona, Spain, <sup>2</sup> School of Biological, Earth and Environmental Sciences, University College Cork, Cork, Ireland, <sup>3</sup> Environmental Research Institute, University College Cork, Cork, Ireland

## OPEN ACCESS

### Edited by:

Dora Szakonyi,  
Instituto Gulbenkian de Ciência (IGC),  
Portugal

### Reviewed by:

Santiago Mora-García,  
Consejo Nacional de Investigaciones  
Científicas y Técnicas (CONICET),  
Argentina  
Umesh K. Reddy,  
West Virginia State University,  
United States

### \*Correspondence:

Rossana Henriques  
rossana.henriques@ucc.ie

### Specialty section:

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

**Received:** 23 March 2018

**Accepted:** 14 June 2018

**Published:** 26 July 2018

### Citation:

Sánchez-Retuerta C, Suárez-López P and Henriques R (2018) Under a New Light: Regulation of Light-Dependent Pathways by Non-coding RNAs. *Front. Plant Sci.* 9:962. doi: 10.3389/fpls.2018.00962

The biological relevance of non-protein coding RNAs in the regulation of critical plant processes has been firmly established in recent years. This has been mostly achieved with the discovery and functional characterization of small non-coding RNAs, such as small interfering RNAs and microRNAs (miRNAs). However, recent next-generation sequencing techniques have widened our view of the non-coding RNA world, which now includes long non-coding RNAs (lncRNAs). Small and lncRNAs seem to diverge in their biogenesis and mode of action, but growing evidence highlights their relevance in developmental processes and in responses to particular environmental conditions. Light can affect *MIRNA* gene transcription, miRNA biogenesis, and RNA-induced silencing complex (RISC) activity, thus controlling not only miRNA accumulation but also their biological function. In addition, miRNAs can mediate several light-regulated processes. In the lncRNA world, few reports are available, but they already indicate a role in the regulation of photomorphogenesis, cotyledon greening, and photoperiod-regulated flowering. In this review, we will discuss how light controls *MIRNA* gene expression and the accumulation of their mature forms, with a particular emphasis on those miRNAs that respond to different light qualities and are conserved among species. We will also address the role of small non-coding RNAs, particularly miRNAs, and lncRNAs in the regulation of light-dependent pathways. We will mainly focus on the recent progress done in understanding the interconnection between these non-coding RNAs and photomorphogenesis, circadian clock function, and photoperiod-dependent flowering.

**Keywords:** microRNAs, long non-coding RNAs, light, circadian clock, photoperiod

## INTRODUCTION

Light is fundamental for plant life. It not only provides the energy necessary for photosynthesis but also functions as an indicator of time and place. Light quality and duration inform a plant about its neighbors in a process described as shade avoidance, which ultimately shapes its architecture and development. Light cues also function as indicators of time of the day and season [e.g., short days (SDs) in autumn and winter versus long days (LDs) in spring and summer]. Last but not least, light, together with temperature, is an input to the circadian clock, providing circadian entrainment and

contributing to maintain rhythmicity and robustness of the clock. Proper clock resetting is critical for the adequate timing of biological processes, which will then reflect in plant fitness.

In parallel with seasonal and positional information, light also provides important developmental cues. For instance, during photomorphogenesis, when the young seedling emerges through the soil reaching for light, hypocotyl elongation is inhibited, while roots grow, cotyledons open and expand, and the first true leaves develop (Leivar and Monte, 2014). These morphological changes are accompanied by the development of chloroplasts and chlorophyll accumulation. Later in development, light will also regulate the time of flowering. Depending on each species requirements, specific photoperiods (the relative duration of the daily light and dark periods) will provide molecular cues, such as the accumulation of specific regulators, required for the transition of shoot apical meristems into inflorescence meristems (Andrés and Coupland, 2012).

Therefore, considering all light-dependent and/or light-regulated processes, it is not surprising that light signals promote massive transcriptional changes affecting both protein coding genes as well as non-protein coding transcripts (López-Juez et al., 2008; Wang H. et al., 2014). Although light regulation of (protein-coding) gene expression has been studied in extreme detail, the role of light in the regulation of non-protein coding RNAs is less known.

Non-coding RNAs have been divided into different functional groups based on the size of their active forms. Non-coding RNAs shorter than 200 base pairs (bp) fall within the category of small non-coding RNAs, whereas longer ones constitute long non-coding RNAs (lncRNAs). Historically, small non-coding RNAs were the first identified and characterized in great detail. Within this group, microRNAs (miRNAs) are transcribed from *MIRNA* genes as longer precursors (primary miRNAs, pri-miRNAs) by RNA polymerase II. Plant pri-miRNAs fold into stem-loop structures and are processed by a complex containing DICER-LIKE1 (DCL1), HYPONASTIC LEAVES1 (HYL1), and SERRATE (SE), which will first release the stem-loop (pre-miRNA) structure, and then cleave it giving rise to a mature miRNA:miRNA\* duplex, normally 20–22 nucleotides (nt) long, with the rarer cases of 23–25 nt, with 2-nt 3' overhangs (Voinnet, 2009; Borges and Martienssen, 2015; Yu et al., 2017). This duplex is methylated at its 3' ends by the methyltransferase HUA ENHANCER1 (HEN1) and this methylation is necessary to stabilize miRNAs (Yu et al., 2005). The mature miRNA strand is then selectively bound by ARGONAUTE1 (AGO1) and loaded into the RNA-induced silencing complex (RISC), which recognizes transcripts with full or partial complementarity to the miRNA and cleaves and/or represses translation of these miRNA target transcripts, therefore silencing them (Voinnet, 2009; Yu et al., 2017) (**Figure 1**). miRNAs act in particular developmental processes and/or in response to certain environmental cues, sometimes in specific tissues where they accumulate at lower levels than their targets, protein-coding transcripts.

lncRNAs, on the other hand, can exert their biological function without further processing and accumulate to similar levels as miRNAs (Liu et al., 2012). lncRNAs are classified according to their genomic location into intergenic (expressed

in between coding regions), intronic (expressed from introns), or natural antisense transcripts (NATs, expressed from the complementary strand to a protein-coding gene or another lncRNA; Liu et al., 2012; Wang H. et al., 2014). Comprehensive screening approaches have revealed thousands of lncRNAs that display spatial-, temporal-, and developmental-specific patterns of expression. Interestingly, among the few lncRNAs whose biological function is known, two are involved in light-regulated processes. *HIDDEN TREASURE 1 (HID1)* is involved in photomorphogenesis and seedling greening (Wang Y. et al., 2014; Wang et al., 2018), whereas *CDF5 LONG NON-CODING RNA (FLORE)* mediates the circadian regulation of flowering time (Henriques et al., 2017).

Here, we review the role of miRNAs and lncRNAs associated with, or regulated by light-dependent pathways. We will discuss light regulation of miRNA expression, biogenesis, processing, and function, as well as their role in light-related processes, such as photomorphogenesis and photoperiod-dependent flowering. We will also discuss the available reports characterizing lncRNA function in light-dependent pathways.

## MicroRNAs

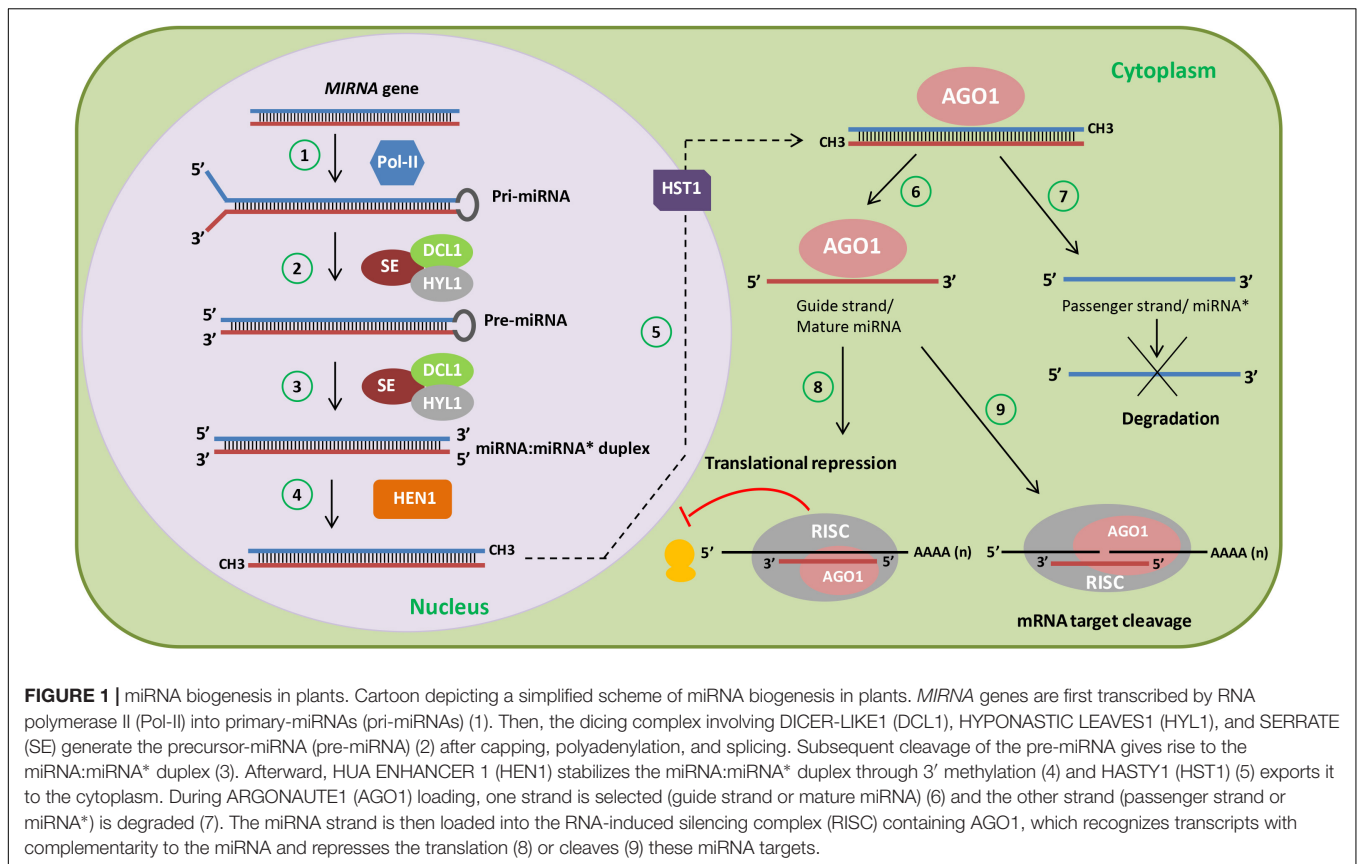
Light regulation of plant miRNAs can occur through at least three different mechanisms: regulation of *MIRNA* gene transcription and its levels, regulation of miRNA processing *via* effects on components of the miRNA biogenesis machinery, and regulation of miRNA function through control of RISC factors. In turn, miRNAs can target genes encoding components of the light signaling pathways, thereby affecting light responses such as photomorphogenesis and photoperiod-dependent flowering. Below we discuss the recent progresses made in these fields of research. Due to the interconnection between light and the circadian clock, we will also discuss the available reports on circadian-regulated ncRNAs.

## Light Regulation of miRNA Levels

Light, photoperiod, and the circadian clock modulate the expression levels of *MIRNA* genes. This regulation is achieved by the presence of light-responsive elements in their promoters, and can be triggered by specific light qualities (red, blue, far-red, UV-A, and UV-B). To identify the miRNA families under this control, several screening approaches have been developed, resulting in extensive listings of light-responsive miRNAs.

For instance, the comparison of miRNAs between wild-type rice (*Oryza sativa*) plants and a *phytochrome B (phyB)* mutant identified 135 differentially expressed miRNAs, of which 97 were upregulated and 38 were downregulated. These miRNAs include conserved and novel miRNAs, with sizes varying between 21-nt and 24-nt, such as miR156, miR166, miR171, and miR408 (Sun et al., 2015). Differently from a previous report showing that miR172 is regulated by PHYB in potato (Martin et al., 2009), Sun et al. (2015) did not find any miR172 family member that responded to PHYB in rice, suggesting that PHYB-mediated regulation of miRNAs may differ between plant species. Using degradome sequencing, it was found that 70 rice genes were





targeted by 32 differentially expressed miRNAs between the wild-type and the *phyB* mutant (Sun et al., 2015). A large proportion of them (42%) encode transcription factors, suggesting that regulation of gene expression by miRNA target genes may play an important role in PHYB-mediated light signaling. Members of the transcription factor families identified by Sun et al. (2015) are involved in light signaling in *Arabidopsis*. It is worth noting that one of the identified miRNA target genes is a homolog of the *Arabidopsis* circadian-regulated *TANDEM ZINC KNUCKLE PROTEIN (TZP)* gene, which is involved in blue light-associated growth (Loudet et al., 2008). However, further research is required to fully understand the role of these rice miRNAs and target genes in light-regulated processes (Sun et al., 2015).

Moreover, ELONGATED HYPOCOTYL 5 (HY5), a major regulator of photomorphogenesis that acts downstream of several photoreceptors, was shown to bind to the promoters and regulate the expression of several *MIRNA* genes, such as *MIR156D*, *MIR402*, *MIR408*, and *MIR858A* in *Arabidopsis* (Zhang et al., 2011). This, together with the effect of PHYB on miRNA levels (Sun et al., 2015), strongly suggests that light affects miRNA abundance. In support of this, small RNA sequencing indicates that the miRNA population of the seedling apical hook changes after treating dark-grown soybean (*Glycine max*) seedlings with far-red light. The levels of several miRNAs (miR166, miR167, miR168, miR394, miR396, miR530, miR1507, miR1508, miR1509, and miR2218) respond to far-red light in different parts of the soybean seedling, in most cases showing

upregulation by far-red light (Li et al., 2014). Therefore, light seems to upregulate miRNAs in the apical hook, leading to downregulation of target genes that presumably repress apical hook and cotyledon opening.

Light has strong effects on plant metabolism, especially in the regulation of fundamental plant processes such as photosynthesis (Kooke and Keurentjes, 2012). To understand the involvement of miRNAs in light regulation of potato metabolism, Qiao et al. (2017) used high-throughput sequencing to study the expression levels of miRNAs and mRNAs in dark-treated and red-light-treated leaves and detached tuber skin in potato. Among the 69 known and novel miRNAs that were differentially regulated, most were upregulated in leaves, whereas in tuber skin, about half of them were upregulated and half downregulated (Qiao et al., 2017). Notably, the miR399 family and the novel miRNA miRn55 showed opposite light responses in leaves (upregulation) and tuber skin (downregulation). General gene expression analyses revealed that most differentially expressed mRNAs (67%) were upregulated in tuber skin, whereas in leaves, about half were upregulated and half downregulated. Only 14% of differentially expressed miRNAs and 12% of differentially expressed genes were common to leaves and tuber skin, indicating that the molecular response to red light differs between these tissues. Nevertheless, there was an effect of light on genes involved in primary and secondary metabolism in both tissues, especially in carbohydrate metabolism, which is downregulated, and alkaloid metabolism, which is upregulated. Repression of carbohydrate

metabolism was supported by anti-correlation between miRNAs and mRNAs, which suggests downregulation of putative miRNA targets involved in this process (Qiao et al., 2017). Although confirmation of these findings using additional methods is required, this work points to a role of miRNAs in the regulation of metabolic pathways affected by red light in potato.

Additional support to the effect of red light on miRNA levels has been obtained in *Arabidopsis*. In this species, the most differentially expressed miRNAs in red light versus darkness are miR160, miR163, miR319, miR394, miR779, miR851, miR854, and miR2111 (Sun et al., 2018). At least part of this effect is mediated by PHYTOCHROME-INTERACTING FACTOR4 (PIF4), a transcription factor that is a negative regulator of PHYB-mediated red light signaling. In a *pif4* mutant, 22 mature miRNAs, representing 16 miRNA families, show altered levels relative to wild-type plants under red light. PIF4 promotes the expression of genes encoding miR156/157, miR160, miR165/166, miR167, miR170/171, and miR394 and represses the expression of genes encoding miR172 and miR319 by binding to the promoters of several of these genes. In addition to acting as a transcription factor for *MIRNA* genes, PIF4 regulates miRNA processing, as discussed below (Sun et al., 2018).

Besides light exposure, photoperiod can also control miRNA accumulation. Li et al. (2015) used high-throughput sequencing to identify miRNAs regulated by photoperiod in soybean. Comparing miRNA levels in seedlings grown under LDs and SDs revealed 37 miRNA families to be day-length-responsive. Five members of the miR156 family were induced under LD conditions and this correlated with downregulation of four miR156-target *SQUAMOSA PROMOTER-BINDING PROTEIN-LIKE* (*SPL*) genes. Given that miR156 overexpression delays flowering in soybean (Cao et al., 2015), lower miR156 levels under SD than LD may explain the fact that soybean flowering is induced by SD. Conversely, the expression of six members of the miR172 family was induced under SD conditions and this was coordinated with the repression of ten predicted targets belonging to the *APETALA2* (*AP2*)-like family (Li et al., 2015). In addition to miR156 and miR172, other conserved miRNAs, such as miR159/319, miR166, miR167, miR395, and miR408 were also shown to be photoperiod regulated in soybean.

The simplified model of the circadian clock places light as an input that confers clock entrainment. Similarly to light, the clock also controls the expression of a vast set of genes, especially at the dark/light transition (Michael et al., 2008), thus allowing plants to anticipate certain daily processes. Interestingly, the circadian clock can also regulate the expression of protein-coding as well as non-coding transcripts (Hazen et al., 2009). To uncover circadian-regulated miRNAs, Siré et al. (2009) compared mature and pri-miRNA transcript levels under light/dark cycles or under free-running (continuous light) conditions for two consecutive days. It was found that miR167, miR168, miR171, and miR398 levels oscillated under light/dark cycles, peaking around the light/dark transition. However, for certain miRNAs, the waving pattern was not accompanied by a clear antiphasic expression of their targets but rather by a phase delay. This was the case of miR171/*SCARECROW-LIKE 6* (*SCL6*, also named *SCL6-IV*, *HAM3*, and *LOM3*) and miR398/*Cu/Zn SUPEROXIDE*

*DISMUTASE 2* (*CSD2*) pairs, where target gene expression is reduced as the miRNA levels increase. However, in the case of the miR168/*AGO1* and miR167/*AUXIN RESPONSE FACTOR 6* (*ARF6*) pairs, an in-phase pattern of accumulation was found, suggesting that miRNAs could regulate their targets by incorporating feedback loops. However, when these modules were tested under free-running conditions, a clear oscillation was not found, an indication that these miRNAs are most likely light regulated.

In another report, Li et al. (2016) addressed the mechanism by which the circadian clock regulates carbon and nitrogen metabolism in rice plants grown in field conditions. Using a comprehensive approach, the authors investigated how light/dark cycles control certain biological processes. They investigated metabolic and enzymatic activities, as well as gene expression and alternative splicing events that could show rhythmic behavior. In their search for oscillatory miRNAs associated with carbon and/or nitrogen metabolism, Li et al. (2016) uncovered miR1440b, miR1877, miR2876-5p, and miR5799 as possible rice miRNA candidates. These miRNAs oscillate under light/dark cycles and negatively correlate with their targets at the majority of time points analyzed. However, it still remains to be seen whether these are *bona fide* circadian miRNAs oscillating under free-running conditions.

### UV-Responsive miRNAs

Since different light qualities can trigger diverse developmental responses, some of which can be detrimental, several screening approaches have been designed to understand how light, especially high light, UV-A, and UV-B radiation can modulate accumulation of miRNAs and their targets. Zhou et al. (2016), using seedlings from *Brassica rapa* exposed to blue light and UV-A, identified 15 conserved and 226 novel differentially expressed miRNAs that could target genes encoding regulators of plant growth, development, and photomorphogenesis (Zhou et al., 2016). Blue light and UV-A exposure slightly downregulated miR156/157 expression, which correlated with upregulation of their targets, *SPL9* (Bra004674) and *SPL15* (Bra003305), which encode transcription factors involved in the regulation of many processes, including the juvenile-to-adult transition, flowering, and secondary metabolism, especially anthocyanin biosynthesis (Wu et al., 2009; Gou et al., 2011; Wang and Wang, 2015). In *Arabidopsis*, miR156 and SPLs are involved in complex feedback loops (Wu et al., 2009; Yant et al., 2010). As plants age, miR156 levels decrease and SPLs (e.g., *SPL9*) accumulate, inhibiting anthocyanin biosynthesis (Gou et al., 2011). Zhou et al. (2016) hypothesize that blue and UV-A light activate genes involved in light signaling and some of them promote anthocyanin biosynthesis. When anthocyanin levels increase, then the regulatory feedback loops involving miR156/157 would act to balance its metabolism.

Similar to UV-A, UV-B radiation affects plant growth, physiology, and metabolism (Dotto and Casati, 2017). In one of the first reports on the effect of light on miRNAs, Zhou et al. (2007) used a computational approach to predict *MIRNA* genes induced by UV-B light in *Arabidopsis*. miRNAs belonging to 11

miRNA families were identified as putatively upregulated by UV-B light (Zhou et al., 2007). Although these results have not been experimentally confirmed in *Arabidopsis*, several reports have shown the effect of UV-B on miRNA levels in aspen (*Populus tremula*), maize (*Zea mays*), and wheat (*Triticum aestivum*). Jia et al. (2009) found that the level of eight miRNA families increased and that of seven families decreased in response to UV-B light stress in aspen. In maize, from the 16 UV-B-regulated miRNA families detected, 6 were upregulated and 10 downregulated (Casati, 2013). UV-B-treated wheat plants showed upregulation of three and downregulation of another three known miRNAs, in addition to a novel wheat miRNA, which was slightly upregulated shortly after UV-B exposure and then was downregulated (Wang et al., 2013). Thirteen UV-B-responsive miRNA families, i.e., miR156/157, miR159, miR160, miR164, miR165/166, miR167, miR169, miR170/171, miR172, miR393, miR395, miR398, and miR399, are common to at least two plant species, indicating substantial conservation in the miRNAs that respond to UV-B light. Although the direction of the effect (upregulation or downregulation) on several miRNAs differs between species, others are consistently promoted (miR165/166, miR167, and miR398) or repressed (miR395) by UV-B light in three species. This conservation probably reflects important roles of these miRNAs in plant responses to UV-B radiation.

The upstream regulatory regions of the UV-B-regulated *MIRNA* genes contain numerous light-related motifs, such as GT-1 sites, G boxes, I-boxes, CCAAT-boxes, and GATA-boxes (Zhou et al., 2007; Jia et al., 2009; Wang et al., 2013), consistent with the effect of light on these genes. Stress-responsive *cis*-elements were also found in several of those promoters (Jia et al., 2009; Wang et al., 2013). Given that some of these miRNAs are involved in responses to abiotic and biotic stresses (Casati, 2013; Li et al., 2017), we hypothesize that these miRNAs might coordinate adaptive responses to diverse threats.

Among the targets of the UV-B-regulated miRNAs are transcription factors, several factors involved in auxin signaling and factors involved in responses to other stresses (Zhou et al., 2007; Jia et al., 2009; Casati, 2013). In most cases, as would be expected, there is an inverse correlation between the effect of UV-B on miRNAs and on their corresponding target transcripts (Zhou et al., 2007; Jia et al., 2009; Casati, 2013; Wang et al., 2013). For instance, in maize, an increase in miR165/166 correlated with inhibition of their target gene, encoding a HD-ZIP III transcription factor, *ROLLED LEAF1* (*RLD1*; Casati, 2013). In agreement with the role of these miRNAs in delimitating *HD-ZIP III* transcripts to the adaxial side of leaf primordia (Juarez et al., 2004), UV-B-dependent repression of *RLD1* was associated with leaf arching, suggesting a role for miR165/166 in this UV-B response (Casati et al., 2006). In aspen and wheat, miR395 downregulation is associated with upregulation of its targets encoding ATP sulfurylases (APS), enzymes involved in inorganic sulfate assimilation (Jia et al., 2009; Wang et al., 2013). miR395 and APS are inversely regulated by sulfate starvation, and miR395 overexpression reduces APS transcript levels and affects the response of *Arabidopsis* to abiotic stress conditions (Jones-Rhoades and Bartel, 2004; Kim et al., 2010), linking again a UV-B-regulated miRNA with other stress responses.

This is also the case of miR164, whose downregulation in UV-B-treated maize leaves is associated with increased levels of two stress-responsive *NAC-DOMAIN PROTEIN* target transcripts, as well as *EXOSTOSIN PROTEIN-LIKE* and an *ASPARTYL PROTEASE* (Hegedus et al., 2003; Casati, 2013). Interestingly, other miRNAs (miR171, miR172, miR156, and miR529) that target genes involved in critical developmental transitions (juvenile-to-adult transition and/or flowering) were also downregulated by UV-B light (Lauter et al., 2005; Xie et al., 2006; Chuck et al., 2008; Cuperus et al., 2011). This suggests that UV-B exposure promotes miRNA-mediated responses that delay certain developmental transitions, allowing plants either to repair UV-B-induced damage or to adapt to these stressful conditions.

Overall, the analysis of UV-B-regulated miRNAs in several plant species indicates that there is a core set of conserved UV-B-responsive miRNAs. Nevertheless, the findings also suggest that different species may recruit distinct miRNAs in order to cope with UV-B stress. Understanding the role of these miRNAs in UV-B-regulated processes awaits further investigation.

### miRNAs Accumulated in High Light Conditions

Besides UV light, high light can also be detrimental for the development of different plant species. To address how light, especially high light, controls miRNA expression in ma bamboo (*Dendrocalamus latiflorus*) grown under LD conditions, Zhao et al. (2013) generated small RNA libraries and determined how miRNA families were regulated. Several conserved miRNAs were identified, miR168 being the most abundant family, followed by miR156/157, miR535, miR165/166, and miR167. Interestingly, miR172 was not found, probably due to the unique flowering cycle of ma bamboo. Using stem-loop RT-qPCR, Zhao et al. (2013) determined the level of novel miRNAs in plants grown under regular white light, high light, or dark. However, the role of these miRNAs in high-light-mediated processes still remains to be determined.

A global analysis of the light-regulated miRNAs reported so far reveals that numerous miRNA families are affected by diverse light conditions (Table 1). Some of them, such as miR156/157, miR159/319, miR164, miR165/miR166, miR167, miR170/171, miR172, miR395, miR398, miR399, and miR408, have been identified as responsive to several light conditions in at least five plant species, strongly suggesting that they may play roles in light responses (Table 1). However, light could also modulate miRNA processing and/or activity toward its targets. Below we present the most recent reports addressing this type of regulation.

### Regulation of miRNA Biogenesis, Processing, and Function by Light

Besides regulating *MIRNA* gene expression, light can also modulate the levels and the activity of mature miRNAs. This can be achieved by a direct connection between light and regulators of the miRNA biogenesis pathway. HYL1 is a RNA-binding protein involved in miRNA processing (Yu et al., 2017). In *Arabidopsis*, HYL1 protein levels are regulated by CONSTITUTIVE PHOTOMORPHOGENIC 1 (COP1), an E3 ubiquitin ligase that mediates the proteasomal degradation of light signaling factors, such as photoreceptors

**TABLE 1** | miRNA families differentially regulated by light.

miRNA <sup>1</sup> family	Light conditions <sup>2</sup>	Plant species <sup>3</sup>	Reference
miR156/157	W, FR, B, UV-A, UV-B, <i>phyB</i> mutant, <i>pif4</i> mutant, LD/SD, bagging treatment	Ath, Gma, Mdo, Osa, Ppy, Ptr, Stu, Tae, Zma	Zhou et al., 2007, 2016; Jia et al., 2009; Casati, 2013; Wang et al., 2013; Bhogale et al., 2014; Tsai et al., 2014; Li et al., 2015; Sun et al., 2015, 2018; Qu et al., 2016; Lin et al., 2017; Qian et al., 2017; Xie et al., 2017
miR167	W, FR, UV-B, <i>phyB</i> mutant, <i>pif4</i> mutant, L/D cycles, LD/SD, bagging treatment	Ath, Gma, Mdo, Osa, Ptr, Tae	Zhou et al., 2007; Jia et al., 2009; Siré et al., 2009; Wang et al., 2013; Li et al., 2014, 2016; Sun et al., 2015, 2018; Qu et al., 2016; Lin et al., 2017
miR170/171	W, UV-B, <i>phyB</i> mutant, <i>pif4</i> mutant, L/D cycles, LD/SD, bagging treatment	Ath, Gma, Mdo, Osa, Tae, Zma	Zhou et al., 2007; Siré et al., 2009; Casati, 2013; Wang et al., 2013; Fan et al., 2015; Li et al., 2015; Sun et al., 2015, 2018; Qu et al., 2016; Lin et al., 2017
miR165/166	W, FR, R, UV-B, <i>phyB</i> mutant, <i>pif4</i> mutant, LD/SD	Ath, Gma, Osa, Ptr, Stu, Zma	Zhou et al., 2007; Jia et al., 2009; Casati, 2013; Li et al., 2014, 2015; Sun et al., 2015, 2018; Lin et al., 2017; Qiao et al., 2017
miR398	W, R, UV-A, UV-B, L/D cycles, LD/SD, bagging treatment	Ath, Gma, Mdo, Ptr, Stu, Zma	Zhou et al., 2007, 2016; Jia et al., 2009; Siré et al., 2009; Casati, 2013; Li et al., 2015; Qu et al., 2016; Lin et al., 2017; Qiao et al., 2017
miR159/319	W, R, UV-B, <i>phyB</i> mutant, <i>pif4</i> mutant, LD/SD,	Ath, Gma, Osa, Ptr, Tae	Zhou et al., 2007; Jia et al., 2009; Wang et al., 2013; Tsai et al., 2014; Li et al., 2015; Sun et al., 2015; Lin et al., 2017; Sun et al., 2018
miR172	R, B, UV-B, <i>pif4</i> mutant, LD/SD, bagging treatment	Ath, Mdo, Gma, Stu, Zma	Jung et al., 2007; Zhou et al., 2007; Martin et al., 2009; Casati, 2013; Li et al., 2015; Qu et al., 2016; Sun et al., 2018
miR396	W, FR, UV-A, UV-B, <i>phyB</i> mutant, LD/SD	Ath, Gma, Osa, Zma	Casadevall et al., 2013; Casati, 2013; Li et al., 2014, 2015; Sun et al., 2015; Zhou et al., 2016; Lin et al., 2017
miR408	W, R, UV-B, <i>phyB</i> mutant, LD/SD, bagging treatment	Ath, Gma, Mdo, Osa, Ptr, Stu	Jia et al., 2009; Zhang et al., 2014; Li et al., 2015; Sun et al., 2015; Qu et al., 2016; Lin et al., 2017; Qiao et al., 2017
miR160	W, R, UV-B, <i>phyB</i> mutant, <i>pif4</i> mutant, bagging treatment	Ath, Mdo, Osa, Ptr	Zhou et al., 2007; Jia et al., 2009; Sun et al., 2015, 2018; Qu et al., 2016; Lin et al., 2017
miR169	W, UV-B, <i>phyB</i> mutant, <i>pif4</i> mutant, LD/SD	Ath, Gma, Osa, Ptr	Zhou et al., 2007; Jia et al., 2009; Li et al., 2015; Sun et al., 2015, 2018; Lin et al., 2017
miR164	UV-B, <i>phyB</i> mutant, bagging treatment	Mdo, Ptr, Tae, Osa, Zma	Jia et al., 2009; Casati, 2013; Wang et al., 2013; Sun et al., 2015; Qu et al., 2016
miR395	W, UV-B, bagging treatment	Ath, Mdo, Ptr, Tae, Zma	Jia et al., 2009; Casati, 2013; Wang et al., 2013; Qu et al., 2016; Lin et al., 2017
miR399	W, R, UV-B, bagging treatment	Ath, Mdo, Ptr, Stu, Zma	Jia et al., 2009; Casati, 2013; Qu et al., 2016; Lin et al., 2017; Qiao et al., 2017
miR168	FR, UV-B, L/D cycles, LD/SD	Ath, Gma, Ptr	Jia et al., 2009; Siré et al., 2009; Li et al., 2014, 2015
miR393	W, UV-B, bagging treatment	Ath, Mdo, Ptr	Zhou et al., 2007; Jia et al., 2009; Qu et al., 2016; Lin et al., 2017
miR858	W, bagging treatment	Ath, Mdo	Qu et al., 2016; Sharma et al., 2016; Wang et al., 2016; Lin et al., 2017
miR1511	B, <i>phyB</i> mutant, LD/SD, bagging treatment	Ath, Gma, Mdo, Osa	Li et al., 2015; Sun et al., 2015; Qu et al., 2016; Zhou et al., 2016
miR163	W, R, B	Ath	Chung et al., 2016; Lin et al., 2017; Sun et al., 2018
miR390	W, UV-B, LD/SD	Ath, Gma, Ptr	Jia et al., 2009; Li et al., 2015; Lin et al., 2017

(Continued)



TABLE 1 | Continued

miRNA <sup>1</sup> family	Light conditions <sup>2</sup>	Plant species <sup>3</sup>	Reference
miR482	R, LD/SD, bagging treatment	Gma, Mdo, Stu	Li et al., 2015; Qu et al., 2016; Qiao et al., 2017
miR391	B, UV-A, bagging treatment	Ath, Mdo	Qu et al., 2016; Zhou et al., 2016
miR394	FR, R, <i>pif4</i> mutant	Ath, Gma	Li et al., 2014; Sun et al., 2018
miR397	R, <i>pif4</i> mutant	Ath, Stu	Qiao et al., 2017; Sun et al., 2018
miR403	LD/SD, bagging treatment	Gma, Mdo	Li et al., 2015; Qu et al., 2016
miR444	UV-B, <i>phyB</i> mutant	Osa, Zma	Casati, 2013; Sun et al., 2015
miR903			
miR472	R, UV-B, <i>pif4</i> mutant	Ath, Ptr	Jia et al., 2009; Sun et al., 2018
miR477	R, bagging treatment	Mdo, Stu	Qu et al., 2016; Qiao et al., 2017
miR530	FR, <i>phyB</i> mutant	Gma, Osa	Li et al., 2014; Sun et al., 2015
miR535	<i>phyB</i> mutant, bagging treatment	Mdo, Osa	Sun et al., 2015; Qu et al., 2016
miR827	R, <i>phyB</i> mutant	Osa, Stu	Sun et al., 2015; Qiao et al., 2017
miR833	W, <i>pif4</i> mutant	Ath	Lin et al., 2017; Sun et al., 2018
miR1507	FR, LD/SD	Gma	Li et al., 2014, 2015
miR1508			
miR2111	R, B, UV-A	Ath	Zhou et al., 2016; Sun et al., 2018
miR2876	<i>phyB</i> mutant, L/D cycles	Osa	Sun et al., 2015; Li et al., 2016
miR5072	B, <i>phyB</i> mutant	Ath, Osa	Sun et al., 2015; Zhou et al., 2016
miR5139			
miR5368	LD/SD, <i>phyB</i> mutant	Gma, Osa	Li et al., 2015; Sun et al., 2015

In this table, we present a summary of the different miRNAs that respond to two or more light treatments. <sup>1</sup>miRNA families reported to be regulated by light in at least two papers have been included. miRNA families are ordered according to the number of papers reporting their light regulation. <sup>2</sup>W, white light; FR, far-red light; R, red light; B, blue light; L/D, light/dark; LD/SD, long days/short days. <sup>3</sup>Ath, *Arabidopsis thaliana*; Gma, *Glycine max* (soybean); Mdo, *Malus x domestica* (apple); Osa, *Oryza sativa* (rice); Ppy, *Pyrus pyrifolia* (Chinese sand pear); Ptr, *Populus tremula* (aspen); Stu, *Solanum tuberosum* (potato); Tae, *Triticum aestivum* (wheat); Zma, *Zea mays* (maize).

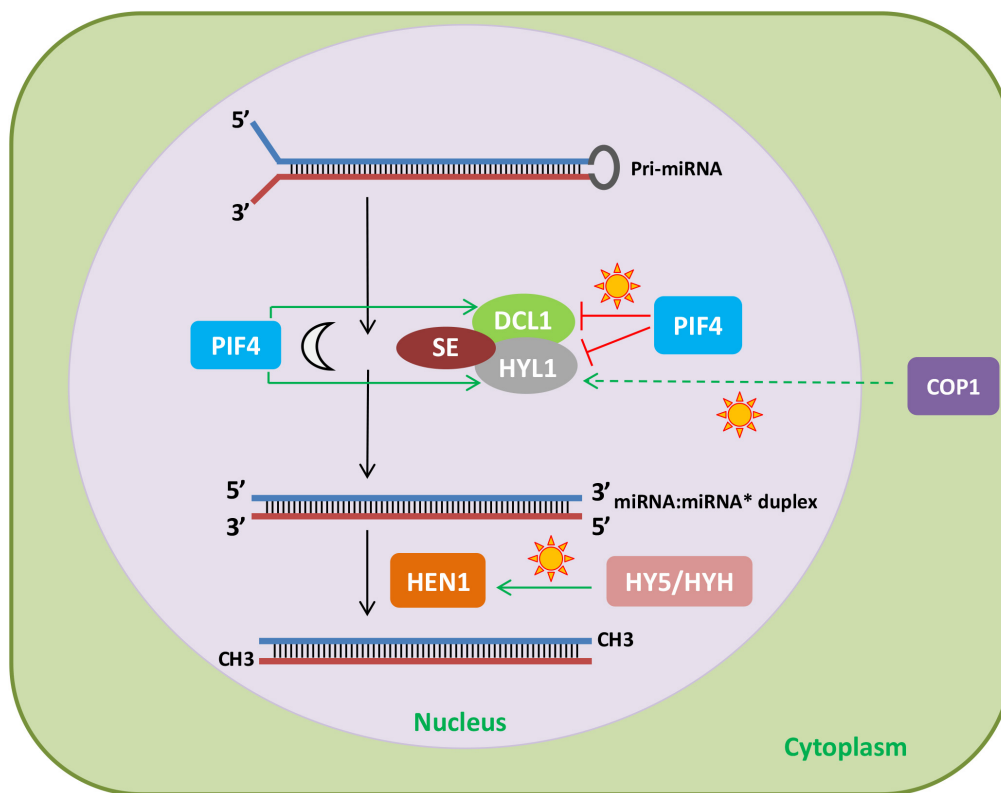
(Cho et al., 2014). It was found that the reduction of miRNA levels in *cop1* mutants relative to wild-type plants greatly overlapped with that observed in the *hyl1* mutant. Further analyses showed that COP1 regulates HYL1 protein levels but not through direct ubiquitination. Combining protein synthesis blockers with either autophagy or 26S proteasome inhibitors revealed that HYL1 was degraded by an unknown protease, which removed its N-terminal fragment, where two RNA-binding domains, essential for miRNA processing, are located (Cho et al., 2014). Light regulates COP1 nucleocytoplasmic shuttling (Jang et al., 2010) and, therefore, during the day, light stabilizes HYL1 due to cytoplasmic accumulation of COP1, which would target the protease acting on HYL1 degradation (Figure 2). This regulation could control miRNA processing and thus help understand the light/dark accumulation pattern of several mature miRNAs, such as miR167, miR168, miR171, and miR398, previously described by Siré et al. (2009).

HYPONASTIC LEAVES1 protein levels are also controlled by PIF4, another protein involved in photomorphogenesis (Sun et al., 2018). Both HYL1 and DCL1 interact with PIF4, which destabilizes them under red light and stabilizes them in the dark. The mechanism by which PIF4 controls the levels of HYL1 and DCL1 is unknown, but it does not involve transcriptional regulation or the ubiquitin-proteasome pathway. Given that PIF4 interacts with the photoreceptor PHYB and the miRNA processing factors DCL1 and HYL1, the question arises as to whether DCL1 and HYL1 affect light responses (Figure 2).

Indeed, *dcl1* and *hyl1* mutants show shorter hypocotyls than wild-type plants under red light, indicating that DCL1 and HYL1 play negative roles in photomorphogenesis (Sun et al., 2018).

Another miRNA processing factor regulated by light signaling pathways is HEN1. Light induction of *HEN1* expression results from the coordinated action of HY5 and HY5 homolog (HYH; Figure 2), which positively regulate *HEN1* downstream of several photoreceptors (Tsai et al., 2014). Consistent with this regulation, HEN1 and HYL1 are negative regulators of photomorphogenesis, in the case of HEN1, this is due to its role as a repressor of key transcription factors of this process (Tsai et al., 2014). In addition, alternative splicing of transcripts encoding several miRNA biogenesis and processing factors, such as DCL1, HYL1, and HEN1, is affected by different light qualities, adding a new layer to the regulation of miRNAs by light (reviewed in Hernando et al., 2017).

The de-etiolation process can also affect miRNA activity toward its targets. Lin et al. (2017) matched small RNA profiling with degradome sequencing in *Arabidopsis* and found that, although light exposure did not change dramatically the level of most miRNAs, it could modulate their target cleavage activity. In fact, with the exception of miR163, most miRNAs showed little fluctuations in their amount upon light transition. This did not correspond with the analysis of degradome signatures, implying that de-etiolation especially promoted the degradation of miR156/157 and miR396 targets, and thus suggesting higher cleavage activity for these families (Lin et al., 2017). However, although Lin et al. (2017) report that miR168 regulates *AGO1*



**FIGURE 2 |** Effect of light signaling components on the miRNA biogenesis and processing machinery. The light signaling factors COP1 and PIF4 regulate the stability of HYL1. In the light, COP1, probably through degradation of an unknown protease, increases HYL1 protein levels. In addition, PIF4 stabilizes HYL1 and DCL1 in the dark and destabilizes them under red light. Finally, HY5 and HYH promote *HEN1* expression in response to light. Therefore, light can affect miRNA levels by regulating light signaling components that modulate the abundance of miRNA biogenesis and processing factors.

levels under light and could modulate RISC activity and the efficiency of miRNA target cleavage, the mechanisms underlying this regulation by light still need to be demonstrated.

Interestingly, another report has shown AGO1 involvement in light regulation of adventitious rooting and hypocotyl elongation. *Arabidopsis ago1* mutants are hypersensitive to light, probably due to deregulation of the PHYA-dependent light signal transduction pathway (Sorin et al., 2005). Moreover, detiolation of an *Arabidopsis* hypomorphic *ago1-27* mutant in far-red light is impaired, indicating that normal photomorphogenesis requires AGO1 (Li et al., 2014). Given that AGO1 is an essential component of the RISC, these findings suggest that miRNA function is involved in far-red light responses.

Therefore, several components of the miRNA machinery are involved in certain light responses. On the other hand, light modulates *MIRNA* gene expression, accumulation of mature miRNAs by regulating components of the biogenesis pathway, as well as miRNA activity. How miRNAs participate in light as well as photoperiod signaling events will be discussed next.

## miRNAs Involved in Light and Photoperiod Signaling

Our analysis revealed several miRNA families that accumulate upon certain light treatments in different species. Most of these

miRNAs mediate critical biological processes necessary for plants to adjust to these changes. They can regulate metabolism, plant development, and certain responses to stress.

### miR396 and UV-B Regulation of Cell Proliferation

For instance, UV-B radiation affects cell proliferation in several species (Dotto and Casati, 2017). In *Arabidopsis*, UV-B light upregulates miR396 levels, leading to repression of the miR396 target genes *GROWTH REGULATING FACTOR1* (*GRF1*), *GRF2*, and *GRF3*, with consequent inhibition of cell proliferation in leaves (Casadevall et al., 2013). In fact, either reducing miR396 activity or expressing miR396-resistant forms of *GRF3* or *GRF2* results in reduced sensitivity of leaf growth to UV-B irradiation, confirming that *GRFs* mediate this phenotype. Nevertheless, downregulation of *GRF3* by UV-B light also seems to occur through miR396-independent mechanisms. In addition, miR396-mediated repression of leaf cell proliferation by UV-B radiation is dependent on the mitogen-activated protein kinase MPK3, known to be involved in the response to UV-B stress (Casadevall et al., 2013). Interestingly, in UV-B-treated maize leaves, miR396 inhibition had the opposite effect on *GRFs*, which were upregulated (Casati, 2013). This could be due to the different developmental stages analyzed in *Arabidopsis* and maize, although further studies are necessary to clarify the role of

miR396 in the regulation of UV-B-dependent responses in dicots and monocots.

### miRNAs Regulating Metabolism

As mentioned before, light controls many metabolic pathways in plants. Within this context, light-regulated miRNAs were shown to modulate metabolic events, such as methylation of specific signaling molecules (e.g., hormones), nutrient allocation, and pigment synthesis/accumulation.

miR163 as well as *pri-miR163* were recently shown to be highly induced by light in *Arabidopsis* (Chung et al., 2016; Lin et al., 2017). Moreover, miR163 may target *PXMT1* (At1g66700), a neighbor gene encoding a putative 1,7-paraxanthine methyltransferase that has been implicated in the methylation of natural chemicals such as hormones (Chung et al., 2016). Whereas *pri-miR163* is induced in red, blue, and white light, and *PXMT1* expression is inhibited upon light exposure, light does not affect *pri-miR163* maturation and processing. Both *pri-miR163* and its target *PXMT1* accumulate in roots, where miR163 accumulation increases root length, especially in the elongation/differentiation zone, most likely by inhibiting *PXMT1*. Interestingly, the miR163/*PXMT1* pair also seems to regulate germination, probably by controlling the early stages of plant development since radicle emergence to seedling de-etiolation. However, to fully understand the role of miR163 in this process, a more extensive functional characterization of *PXMT1* would be necessary.

Plant metabolism requires proper nutrient allocation, especially in the case of nutrients that constitute co-factors and are present in different proteins and biological processes. This is the case of copper, since it is required for electron transport in photosynthesis, defense against reactive-oxygen species as well as ethylene perception (Zhang et al., 2014). Maintenance of copper homeostasis is critical to photosynthesis efficiency and thus plant growth. This is achieved by the SPL7 transcription factor, which functions as a copper sensor able to adjust target gene expression upon changing copper levels. Surprisingly, light and copper seem to regulate gene expression of many shared biological processes, suggesting a direct connection between the transcriptional regulators SPL7 and HY5 (Zhang et al., 2014). Using chromatin immunoprecipitation and RNA sequencing, Zhang et al. (2014) characterized the SPL7 regulon and found it to overlap with HY5. This analysis revealed *MIR408* as a SPL7/HY5 common target, with SPL7 being the predominant transcription factor determining *MIR408* levels, thus confirming previous work showing that HY5 regulates *MIR408* expression (Zhang et al., 2011). Nevertheless, HY5/SPL7 coordinated regulation allowed miR408 accumulation upon copper deficiency and high light. Confirming this regulation, *MIR408* silencing phenotypes were similar to those of the *hy5 spl7* double mutant, whereas miR408 constitutive expression partially rescued them, suggesting that miR408 is a critical component of the SPL7/HY5 network (Zhang et al., 2014). Further analysis of the expression levels of the miR408 targets, *LACCASE12* (*LAC12*) and *LAC13*, showed that the SPL7/HY5/miR408 module allows differential regulation of very closely related genes. Therefore, this signaling network constitutes a molecular mechanism that integrates light and

copper signals to maintain copper homeostasis and consequently efficient photosynthetic rates that promote plant growth.

### miRNAs and pigment synthesis

Flavonoids are a group of phenolic compounds that include flavonols, flavones, isoflavones, and anthocyanins (Sharma et al., 2016). Due to their molecular diversity, flavonoids can confer protection against biotic and abiotic stress, as well as regulate plant growth and development. Anthocyanins, for instance, include a wide range of pigments (blue, red, and purple), which can not only protect against UV radiation but can also attract pollinators (Wang et al., 2016). Due to their relevance to plant life, the biosynthesis and regulation of all these compounds have been extensively studied. The transcription factors MYB-like 2 (MYBL2) and SPL9 are negative regulators of anthocyanin biosynthesis in *Arabidopsis* (Dubos et al., 2008; Matsui et al., 2008; Gou et al., 2011). On the other hand, overexpression of miR156 increases anthocyanin levels in *Arabidopsis*, since some miR156 target genes, including *SPL9*, repress anthocyanin biosynthesis (Gou et al., 2011; Cui et al., 2014). In Chinese sand pear (*Pyrus pyrifolia*) fruits, bagging and subsequent re-exposure to light causes anthocyanin accumulation in fruit peel, giving red color to the fruits (Qian et al., 2013). Qian et al. (2017) hypothesized that miR156 and SPLs would also be involved in anthocyanin accumulation in pears in response to bagging treatments. They detected miR156 in peels of bagging-treated pear fruits and found that the promoters of two pear *PpMIR156* genes contained light-responsive elements. After bag removal, miR156 levels increased, whereas *PpSPL* transcripts decreased. This was accompanied by upregulation of homologs of genes encoding transcription factors that control anthocyanin biosynthesis, and followed by upregulation of anthocyanin biosynthesis genes (Qian et al., 2017). Although functional analyses of miR156 and *PpSPL* genes in pear have not yet been reported, these results point to a possible role of miR156 and PpSPL proteins in light-induced anthocyanin biosynthesis in bagging-treated pear fruits.

In a related report, Qu et al. (2016) investigated the miRNA profiling of “Granny Smith” apple peels upon re-exposure to sunlight after fruit bagging, using high-throughput sequencing of small RNA libraries. They found that over 67% of differentially expressed miRNAs were downregulated in the bagged group as compared to the unbagged control. These findings suggest the existence of a light-regulated pool of miRNAs in apple peel. miR156, miR160, miR171, miR172, miR395, and miR398 were differentially expressed upon sunlight re-exposure. Interestingly, these miRNAs families had previously been identified as UV-B-regulated miRNAs in several plant species (see Section “UV-Responsive miRNAs”). In *Arabidopsis*, in addition to miR156, miR858 positively regulates anthocyanin biosynthesis by targeting specific repressors of this process (Gou et al., 2011; Qian et al., 2013; Cui et al., 2014; Wang et al., 2016). In tomato, miR858 also promotes anthocyanin biosynthesis (Jia et al., 2015). Interestingly, miR156, miR828, and miR858 expression levels varied upon apple fruit debagging, suggesting their involvement in anthocyanin biosynthesis in apple peels. In fact, miR156 correlated with anthocyanin accumulation only in “Granny Smith,” whereas miR5072 did it only in the “Starkrimson”

cultivar. These results highlight the high specificity of certain miRNA families to control particular responses, such as pigment accumulation to protect against light stress.

As previously mentioned, miR858 is a positive regulator of anthocyanin biosynthesis (Wang et al., 2016). In *Arabidopsis*, *MIR858A* expression is induced by light, especially high light, in a HY5-dependent manner. Upon transfer from dark to high light, *MIR858A* level oscillates, suggesting also circadian regulation. Interestingly, miR858a overexpression partially complements the long hypocotyl phenotype of the *hy5* mutant and rescues the anthocyanin levels of this mutant. miR858a and miR858b can promote anthocyanin biosynthesis in two ways: (1) by indirectly inducing the expression of genes encoding components of the anthocyanin biosynthetic pathway and (2) by inhibiting protein accumulation of their target gene encoding the repressor MYBL2 protein, most likely at the translational level. HY5 directly binds and represses the *MYBL2* promoter probably by a mechanism that involves the loss of H3K9Ac and H3K4me3-positive histone marks (Wang et al., 2016). Therefore, HY5 has a dual effect in promoting anthocyanin biosynthesis: it binds and represses *MYBL2*, and it promotes *MIR858A* expression in response to light (e.g., high light) signals.

It has been shown that *hen1* mutants, which have reduced levels of mature miRNAs, accumulate high anthocyanin levels (Tsai et al., 2014), a result somehow contradictory with miR156 and miR858 promoting anthocyanin biosynthesis. However, Sharma et al. (2016) propose a different function for miR858 in the regulation of flavonoid biosynthesis. They found that miR858 targets the positive regulators of flavonoid biosynthesis *MYB11*, *MYB12*, and *MYB111*. Moreover, transcriptional profiling of miR858-overexpressing lines (miR858-OX) revealed *MYB20*, *MYB42*, and *MYB83* as new targets of miR858. Since these transcription factors regulate several growth responses, the authors characterized growth phenotypes of miR858-OX and miR858a target mimic lines (*MIM858*), in which miR858a activity is silenced. While miR858-OX plants showed enhanced growth, early flowering, and an increase in seed size, *MIM858* plants displayed reduced growth, late flowering, and smaller seeds (Sharma et al., 2016). The metabolic profiling of these lines confirmed their opposite effect on flavonoid accumulation, since miR858 depletion resulted in the accumulation of major flavonoids, whereas they were significantly reduced in miR858-OX. This was accompanied by an opposite effect on lignin biosynthesis, since miR858-OX showed increased lignification and upregulation of lignin biosynthetic genes. These results indicate that miR858 could differently regulate anthocyanin, other flavonoids, and lignin biosynthesis. This could be achieved by targeting different MYB transcription factors in response to distinct light conditions (white light versus low or high light). This differential regulation could then help maintain a metabolic flux balance between the different biosynthetic pathways.

Chlorophylls are critical plant pigments involved in light absorption and energy transfer during photosynthesis. However, this fundamental process for plant life also generates reactive oxygen species, which are detrimental for plant growth and development. Considering that chlorophyll biosynthesis precursors are also sources for reactive oxygen species, it is

of paramount importance to tightly regulate this pathway. One of such mechanisms includes miR171 and miR171-targeted *SCL* genes, also known as *HAIRY MERISTEM (HAM)* or *LOST MERISTEMS (LOM)*. In *Arabidopsis*, miR171 positively regulates chlorophyll biosynthesis in the light through downregulation of miR171-targeted *SCLs*, leading to upregulation of protochlorophyllide oxidoreductase (POR), a key enzyme in chlorophyll biosynthesis (Ma et al., 2014). *MIR171C*-overexpressing plants and *scl6 scl22 scl27* mutants show higher chlorophyll and POR levels than wild-type plants, whereas plants expressing a miR171-resistant form of *SCL27* (rSCL27) show reduced chlorophyll and POR levels (Wang et al., 2010; Ma et al., 2014). On the other hand, downregulation of *POR* expression reduces the chlorophyll content of wild-type, *MIR171C*-OX, and *scl* triple mutants (Ma et al., 2014), indicating that the role of miR171c and *SCLs* on chlorophyll regulation is mediated by PORC. *SCL27* represses *PORC* expression by directly binding and inhibiting its promoter.

Moreover, the miR171-*SCL* module also mediates gibberellin-dependent effects on chlorophyll biosynthesis due to its regulation of DELLA proteins and *PORC* expression in the light, but not in the dark. In fact, *SCL27* interacts with the DELLA protein RGA and this interaction reduces the ability of *SCL27* to bind to the *PORC* promoter. Finally, *SCLs* induce the expression of *MIR171* genes, revealing the existence of a regulatory feedback loop (Ma et al., 2014). Ma et al. (2014) propose that this feedback loop may also contribute to maintain the diurnal oscillation of miR171.

Taken together, these findings emphasize the role of miRNAs in diverse light-regulated processes. In addition, miRNAs can also directly target major regulators of the light signaling pathways and/or the photoperiod pathway, thus modulating photomorphogenesis or the time to flower, respectively. The most recent findings on miRNAs involved in these processes are discussed below.

## Photomorphogenesis

Within the group of light-responsive miRNAs, the miR156/157 family occupies a prominent position (Table 1). In *Arabidopsis*, miR157d and miR319 promote the degradation of their target transcripts, encoding the positive and negative key photomorphogenesis regulators HY5 and TCP (TEOSINTE BRANCHED1, CYCLOIDEA, and PCF) transcription factors, respectively. Interestingly, both induction and stabilization of miR157d and miR319 were shown to depend on HEN1, with HEN1 accumulation increasing miR157d and miR319 levels in de-etiolating seedlings (Tsai et al., 2014). In turn, HY5 induces *HEN1* expression in a light-dependent manner. Therefore, HEN1 and HY5 constitute a negative feedback regulatory loop that is mediated by miR157, since this miRNA will ultimately target *HY5* transcripts for cleavage. In a *hen1* mutant, both *HY5* transcript and protein levels are increased, probably due to the decrease in miR157d, resulting in a light-hypersensitive phenotype. Oppositely, miR157 constitutive expression reduced *HY5* transcript and protein levels, resulting in seedlings displaying a light hyposensitivity phenotype. Nevertheless, the light-hypersensitive phenotype displayed by *hen1* mutants may



also be dependent on miR319, which promoted the cleavage of *TCP* mRNAs and repressed hypocotyl elongation (Tsai et al., 2014). The role of miR319 in photomorphogenesis has been confirmed using a loss of function mutant, which shows longer hypocotyls than the wild type under red light (Sun et al., 2018). Perhaps, the HEN1/HY5 feedback loop may help explain the inconsistencies found in the effect of HEN1, miR156, and miR858 on anthocyanin accumulation.

Three additional miRNAs, miR160, miR167, and miR848, affect *Arabidopsis* hypocotyl elongation under red light (Sun et al., 2018). It is worth mentioning that miR160 and miR167 respond to different light conditions in several species (Table 1). Further understanding of the role of these miRNAs in light responses will undoubtedly provide novel insights into how miRNAs contribute to adaptation to different light environments.

Besides the early stages of photomorphogenesis, when the seedling reaches for light and hypocotyl elongation is critical, miRNAs regulate other light responses during plant adult life, such as the perception of their surroundings and neighbors or the time to flower, for instance. When in isolation, plants receive light with a high red to far-red (R:FR) ratio. However, when they grow under a canopy or in close proximity to neighboring plants, the R:FR ratio is lower and this triggers diverse morphological and physiological responses that allow the plants to adapt to these light conditions (Ballaré and Pierik, 2017). When *Arabidopsis* plants grown under light/dark cycles are treated with a pulse of far-red light at the end of the light period (end-of-day far-red, EOD-FR), they show earlier flowering, increased petiole elongation, and a reduction in the number of rosette branches in comparison with plants grown under normal white-light/dark cycles (WL; Xie et al., 2017). EOD-FR treatment causes a decrease in mature miR156 levels by downregulating the expression of several *MIR156* genes. Consistent with this, there is upregulation of several miR156-target *SPL* genes in EOD-FR-treated plants (Xie et al., 2017). This suggests that the miR156-*SPL* module might be involved in EOD-FR responsiveness. Indeed, when plants with reduced miR156 activity (*MIM156* plants) are grown under WL, they show a constitutive EOD-FR response (Xie et al., 2017). These findings suggest that a reduction in miR156 levels is required for this response, and place miR156 as a repressor of this process.

PHYTOCHROME-INTERACTING FACTORS (PIFs) mediate the effect of phytochromes on numerous light responses (Leivar and Monte, 2014). PIF protein abundance increases under low R:FR light conditions (Lorrain et al., 2008; Leivar et al., 2012; Xie et al., 2017), and PIFs act as positive regulators of EOD-FR responses, showing therefore an opposite behavior to miR156. Further substantiating a previous report suggesting that *MIR156E* is a direct target gene of PIF5 (Hornitschek et al., 2012), Xie et al. (2017) showed that several PIFs repress transcription of five *MIR156* genes, including *MIR156E*, by directly binding to PIF-binding sites present in their promoters. This causes a reduction in mature miR156 levels and a concomitant increase in miR156-target *SPL* transcript abundance. Moreover, genetic analyses show that miR156 acts downstream of PIF5 (Xie et al., 2017). Altogether, these results demonstrate that miR156 responds to EOD-FR treatments and mediates, probably through

downregulation of *SPL* genes, the effect of PIFs on at least part of the low R:FR light responses. It will be interesting to determine whether these findings can be reproduced under low R:FR conditions more closely resembling canopy or proximity shade. Therefore, not only do miR156/157 respond to light but also affect several light-regulated processes.

### miRNAs and Photoperiod-Regulated Processes

Photoperiod affects many plant metabolic, physiological, and developmental processes. It also affects the levels of some miRNAs, as we have already discussed (Li et al., 2015). An example is miR156, which is regulated by photoperiod in potato and soybean, although not in *Arabidopsis* (Jung et al., 2012; Bhogale et al., 2014; Li et al., 2015). miR156, via its target genes, downregulates miR172 in several plant species (Chuck et al., 2007; Wu et al., 2009; Bhogale et al., 2014). In *Arabidopsis*, mature miR172 levels are higher under LD than SD and this difference does not seem to be driven by transcription, given that the abundance of at least two miR172 primary transcripts is lower under LD than SD (Jung et al., 2007). Rather, it probably results from regulation of miR172 processing. A *gigantea* (*gi*) mutant, which shows a reduced response to photoperiod and reduced expression of two genes involved in miRNA processing, *DCL1* and *SE*, shows lower mature miR172, but not *pri-miR172*, levels than the wild type. However, additional factors must contribute to the photoperiodic regulation of miR172, given that this miRNA still responds to photoperiod in the *gi* mutant (Jung et al., 2007). Potato and soybean also show different miR172 levels under LD and SD (Martin et al., 2009; Li et al., 2015), suggesting that the photoperiodic regulation of this miRNA may be evolutionarily conserved.

Interestingly, at least four photoreceptors, phytochrome A, phytochrome B, and cryptochrome 1 and 2, regulate miR172 levels (Jung et al., 2007; Martin et al., 2009; Zhao et al., 2015). In agreement with this, red light downregulates miR172 and blue light upregulates it in *Arabidopsis*. Therefore, both light quality and duration affect miR172 abundance. In addition, TIMING OF CAB EXPRESSION1 (*TOC1*), a component of the circadian clock, reduces mature miR172 levels, but despite this, mature miR172 levels do not show daily oscillations in *Arabidopsis* (Jung et al., 2007). Further research is still needed to determine whether the circadian clock regulates miR172.

In *Arabidopsis*, flowering is regulated by photoperiod, such that plants flower earlier under LD than SD. By contrast, soybean flowering and potato tuber formation are induced by SD. Overexpression of miR172 promotes flowering in *Arabidopsis* and tuberization in potato and reduces the photoperiodic tuberization response (Aukerman and Sakai, 2003; Chen, 2004; Jung et al., 2007; Martin et al., 2009), strongly suggesting that miR172 is involved in the regulation of photoperiodic processes. miR172 targets a subfamily of AP2-like genes, including *AP2* itself, which partially redundantly repress flowering (Zhu and Helliwell, 2011). In *Arabidopsis*, and probably in soybean, alteration of miR172-target gene expression or function leads to a reduced response of flowering to photoperiod, confirming the role of the miR172/AP2-like module in the photoperiodic regulation of flowering (Mathieu et al., 2009; Yant et al., 2010;

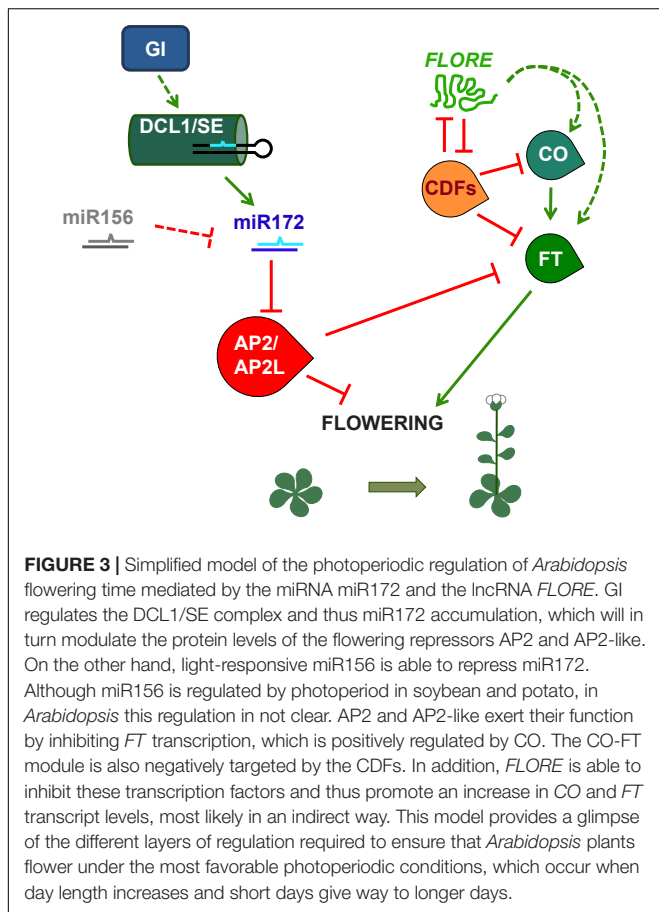
Zhao et al., 2015). In *Arabidopsis* plants grown under LD, GI, and probably several photoreceptors, upregulates miR172, which negatively regulates its target genes, which in turn directly repress the expression of several genes promoting flowering, including *FLOWERING LOCUS T* (*FT*), and promote the expression of genes repressing flowering (Figure 3) (Jung et al., 2007; Mathieu et al., 2009; Yant et al., 2010). Downregulation of miR172 target genes, thus, accelerates flowering under LD conditions. This photoperiodic flowering pathway is independent of the photoperiodic flowering regulator *CONSTANS* (*CO*) in *Arabidopsis* (Jung et al., 2007; Mathieu et al., 2009; Yant et al., 2010), although it has been proposed to involve a *CO*-like gene in soybean (Zhao et al., 2015).

Given that miR156 responds to light in several species, to photoperiod in potato and soybean, is involved in flowering control, and negatively regulates miR172 (Chuck et al., 2007; Wang et al., 2009; Wu et al., 2009; Bhogale et al., 2014; Li et al., 2015) (Table 1), it seems likely that miR156 may play a role in photoperiodic flowering. However, whether this is the case still needs to be proved. In addition to light quality and photoperiod, miR156 and miR172 levels respond to other environmental and endogenous signals and play roles in processes other than flowering and tuberization (Zhu and Helliwell, 2011; Han et al., 2013; Nova-Franco et al., 2015; Ripoll et al., 2015; Wang and Wang, 2015; Huo et al., 2016; Li et al., 2017;

Díaz-Manzano et al., 2018; Luan et al., 2018), indicating that miR156 and miR172 influence many aspects of plant biology.

The miR170/171 family has been shown to be responsive to different light conditions in several species (Table 1). In addition, the rice *OsMIR171C* promoter contains several elements putatively involved in light responses (Fan et al., 2015). Consistent with this, *OsMIR171C* transcript levels show a diurnal oscillation with a peak early in the morning, similar to the oscillation of mature miR171 in *Arabidopsis* (Fan et al., 2015). Conversely, transcript levels of four rice miR171 target genes, *OsHAM1* to *OsHAM4*, accumulate from the evening until early morning under light/dark cycles. Although it is not clear whether these rice oscillation patterns are regulated by light or by the circadian clock, miR171c levels respond to photoperiod, with higher levels under LD than SD (Fan et al., 2015). miR171c upregulation in the *delayed heading* (*dh*) mutant, which carries a T-DNA insertion in the *OsMIR171C* promoter, leads to late flowering (heading), strongly suggesting that miR171c represses flowering (Fan et al., 2015). Moreover, the upregulation of miR171c by LD, which are non-inductive conditions for rice flowering, fits with a role in delaying flowering. However, further research is still required to determine if miR171 affects the photoperiodic regulation of flowering in rice. Nevertheless, the expression pattern of miR171c and *OsHAMs* in the shoot apex of wild-type plants is consistent with a role in regulating the floral transition. This is also supported by the downregulation of three key positive regulators of flowering (*Ehd1*, *Hd3a*, and *RFT1*) in the *dh* mutant in comparison with the wild type (Fan et al., 2015). Interestingly, the *dh* mutant also shows upregulation of miR156 (Fan et al., 2015), a flowering repressor miRNA in several species (Wang, 2014), and of *OsPHYC*, which encodes a photoreceptor that also delays flowering in rice (Takano et al., 2005; Fan et al., 2015). Therefore, miR171 is regulated by light and, in turn, it may be involved in regulating light responses mediated by *OsPHYC*.

Another miRNA that responds to day length and regulates photoperiodic flowering is miR5200, a miRNA specific of the *Pooideae* family. In a *Brachypodium distachyon* accession that flowers much earlier under LD than SD, mature miR5200, as well as *pri-miR5200a* and *pri-miR5200b*, is present at much higher levels under SD than LD and its *pri*-miRNAs follow a daily rhythm under SD (Wu et al., 2013). The expression pattern of *MIR5200A* and *MIR5200B* correlates with active histone marks in these genes under SD and repressive histone marks under LD. Another five *Pooideae* species accumulate more miR5200 under SD than LD, revealing conservation of this photoperiodic regulation among these grass species (Wu et al., 2013). In *B. distachyon*, miR5200 targets two *FT*-like genes, *FTL1* and *FTL2*, whose transcripts are more abundant under long than short photoperiods. *B. distachyon* plants overexpressing this miRNA flower much later than wild-type plants under LD, whereas plants with reduced miR5200 activity (*MIM5200*) flower earlier than the wild type under SD, and like the wild type under LD, showing therefore a reduced response to photoperiod. Consistently, *FTL1* and *FTL2* mRNA levels are increased in *MIM5200* plants under SD, but unchanged under



LD in comparison with wild-type plants (Wu et al., 2013). Therefore, miR5200 plays an important role in the photoperiodic regulation of flowering in *B. distachyon* by delaying flowering under SD.

## LONG NON-CODING RNAs

Next-generation sequencing techniques have unraveled a whole new transcriptional landscape that includes lncRNAs. Several screening approaches have been designed to identify novel plant lncRNAs that would accumulate in response to certain environmental cues, in specific organs/tissues or at particular developmental stages. However, although lncRNA listings are available, the functional characterization of most of these transcripts is yet to be done. Most of the knowledge on plant lncRNA function is associated with *FLOWERING LOCUS C* (*FLC*), a negative regulator of flowering. In fact, several lncRNAs arising from this locus were shown to repress *FLC* transcript levels upon vernalization and thus promote flowering. *COLD OF WINTER-INDUCED NON-CODING RNA FROM THE PROMOTER* (*COLDWRAP*) is transcribed from the *FLC* promoter region, whereas *COLD-ASSISTED INTRONIC NON-CODING RNA* (*COLDIAIR*) is expressed from the first intron. *ANTISENSE LONG* (*ASL*) and *COLD-INDUCED LONG ANTISENSE INTRAGENIC RNA* (*COOLAIR*) both originate from the *FLC* 3'UTR region and show partial overlap. The combined action of the three lncRNAs was shown to silence *FLC* expression, and therefore to promote flowering, through the accumulation of repressive histone marks (reviewed in Chekanova, 2015; Whittaker and Dean, 2017). Although most of the research on these lncRNAs pertains to vernalization-regulated flowering, and therefore is outside the scope of this review, there are reports connecting *FLC* with photoperiod and circadian regulation. *FLC* lengthens the circadian period, especially at high temperature (Swarup et al., 1999; Edwards et al., 2006), but shortening of the circadian period by vernalization is not mediated by *FLC* (Salathia et al., 2006). Therefore, whether the regulation of *FLC* by vernalization-associated lncRNAs is linked to its role in circadian periodicity is yet to be determined.

Several *FLC* clade members were shown to regulate flowering both under LDs and SDs. Protein levels of one of them, *MAF3* oscillated throughout the day, suggesting regulation by the photoperiod or the circadian clock (Gu et al., 2013). In fact, *FLC* clade members could form multi-protein complexes and regulate flowering depending on photoperiod and temperature. However, further research is needed to determine whether vernalization-associated lncRNAs would also participate in these processes. Therefore, in this section, we will discuss the available reports focused on light and circadian-responsive lncRNAs.

## lncRNAs as Regulators of Light-Dependent Processes

One of the screening approaches that have allowed the identification of novel lncRNA transcriptional units used a

reproducibility-based tiling-array analysis strategy (RepTAS) previously described by Liu et al. (2012). Thus, Wang H. et al. (2014) identified a total of 37,238 light-responsive lncNAT pairs (NAT pairs where one of the components is an lncRNA) in *Arabidopsis*. These NAT pairs were afterward verified by different methods, including the screening of a novel ATH (*Arabidopsis thaliana*) NAT custom array, strand-specific RNA-seq (ssRNA-seq), and quantitative RT-PCR (RT-qPCR). To identify *bona fide* light-regulated lncRNAs, the expression levels of these antisense and sense transcripts were determined in etiolated seedlings and seedlings undergoing de-etiolation that were exposed to continuous white light for 1 and 6 h. Not only did this approach allow the identification of light-responsive lncNATs but also revealed their preferential accumulation in cotyledons compared to hypocotyls, and roots. Interestingly, the number of light-responsive lncNAT pairs was much higher after 6 h of treatment than at 1 h, indicating spatial-, developmental-, and temporal-specific patterns of accumulation. When the expression levels of both the sense and antisense components of each NAT pair was determined, these lncNAT pairs were classified into two groups: light-responsive concordant NAT pairs and light-responsive discordant NAT pairs. In order to further characterize their expression pattern, a parallel analysis was performed matching specific histone modifications with their transcriptional changes. This approach showed that transcription changes of these light-regulated NATs were preferentially correlated with histone H3 acetylation (H3K9ac and/or H3K27ac), H3K9ac being the predominant positive histone mark, suggesting its involvement in regulating light-responsive NAT pairs (Wang H. et al., 2014). Although the functional characterization of most of these candidates is yet to be done, these results provide the first evidence that light is a major shaper of the plant non-coding transcriptional landscape, including both long and small non-coding RNAs.

*HIDDEN TREASURE 1* was the first lncRNA described as a positive regulator of photomorphogenesis under continuous red light. This is achieved by *HID1* ability to downregulate the expression levels of *PIF3*, a negative regulator of this process (Wang Y. et al., 2014). In agreement with this, *PIF3* mRNA levels were increased in *hid1* mutants that displayed elongated hypocotyls under red light, a typical phenotype of PIF accumulation. Functional characterization of *HID1* revealed that this lncRNA assembles in nuclear protein–RNA complexes where it is able to associate with *PIF3* 5'UTR and repress its expression. Therefore, *HID1* would act through *PIF3* to modulate hypocotyl elongation. Sequence homology searches revealed that *HID1* is present in other plant species. Confirming its conservation, *Arabidopsis hid1* mutants expressing *OsHID1*, a rice homolog, could be rescued of their elongated hypocotyl phenotype (Wang Y. et al., 2014). These results suggest that lncRNAs may be conserved between species and regulate similar biological processes, similar to the families of conserved miRNAs.

Interestingly, besides its role as a positive regulator of photomorphogenesis in red light, *HID1* can also act as a



negative regulator of cotyledon greening during de-etiolation of *Arabidopsis* in white light (Wang et al., 2018). Within the several steps of this process, the reduction of protochlorophyllide (Pchl<sub>id</sub>) to chlorophyllide *via* POR is fundamental for cotyledon greening to occur. Confirming its repressor role in this process, dark-grown *hid1* mutants transferred to white light presented a higher rate of cotyledon greening compared to *Arabidopsis* wild-type plants. Moreover, *HID1* knockdown led to a reduction of Pchl<sub>id</sub> content in the dark, as well as higher *POR* mRNA levels, indicating that *HID1* plays a role in cotyledon greening by promoting Pchl<sub>id</sub> accumulation and repressing *POR* transcription (Wang et al., 2018). To uncover the *HID1/PIF3* relationship in cotyledon greening, double *hid1 pif3* mutants were generated. *PIF3* was previously shown to regulate seedling greening by preventing Pchl<sub>id</sub> over-accumulation (Leivar and Monte, 2014). However, *hid1 pif3* double mutants greening rate was comparable to that of *hid1*, suggesting that *HID1* would act downstream of *PIF3* (Wang et al., 2018). Since *PIF3* does not regulate *HID1* expression, the exact mechanism by which *HID1* and *PIF3* regulate greening still remains to be determined (Wang et al., 2018). Together, these results suggest that depending on the light-mediated developmental process, *HID1* could associate with different factors allowing it to perform different functions.

## Circadian-Regulated lncRNAs in Photoperiod-Dependent Flowering

Long non-protein coding RNAs can also display circadian regulation. This is characterized by their ability to maintain an oscillatory pattern of expression even in the absence of environmental cues (e.g., under constant conditions). Several screening approaches allowed the identification of circadian ncRNAs in fungi, mammals, and *Arabidopsis*, some of them being lncRNAs that were also NATs (Kramer et al., 2003; Hazen et al., 2009; Coon et al., 2012; Xue et al., 2014). However, even if non-coding NATs were reported for components of the *Arabidopsis* oscillator (Hazen et al., 2009), their function had not been investigated. Recently, we characterized a circadian-regulated *Arabidopsis* lncRNA, *FLORE*, which is a NAT to the *CYCLING DOF FACTOR 5* (*CDF5*) transcriptional regulator (Henriques et al., 2017). *FLORE* and *CDF5* displayed an antiphasic pattern of expression that resulted in part from a mutual inhibitory function, as well as an opposite biological function (Figure 3). CDFs have been characterized as repressors of photoperiodic flowering due to their ability to repress *CO* and *FT* transcription (Song et al., 2015). *FLORE* inhibits *CDF5* accumulation and contributes to maintain its proper oscillatory pattern. Therefore, due to downregulation of *CDF5*, *FLORE* overexpression in the vascular tissue promotes flowering. On the other hand, *CDF5* will also regulate *FLORE* transcript levels, thus contributing to its oscillation. These findings indicate that the circadian clock is able to control the expression of both coding and non-coding transcripts, in this case, *CDF5* and *FLORE* (Henriques et al., 2017). However, the mutual regulation within this NAT

pair seems to be required to ensure the timely accumulation of each transcript. Further research is still necessary to uncover the molecular mechanisms that account for this regulation.

## FINAL CONCLUSIONS

Both small and lncRNAs have been associated with specific biological processes that can occur at particular developmental stages, often in very precise locations within certain organs and tissues. Light is one of the major environmental cues for plants and essential for their survival. Therefore, it is not surprising that light modulates the expression, processing, and activity of these non-coding RNAs. It became apparent that, in the case of specific miRNAs, a correlation pattern emerges where the accumulation of specific families associates with particular light conditions in different species. Moreover, this seems to be a two-way regulation since miRNAs can also target basic components of light and photoperiod signaling, thus being also required for the proper functioning of these pathways. Although some of the miRNAs that are consistently revealed as light-regulated in diverse plant species have not been shown to play functional roles in light responses so far, they are excellent candidates for performing such roles. lncRNAs have recently become the subject of scrutiny, but the few reports available prevent the establishment of any specific trend. Sequencing and bioinformatics efforts led to the construction of a plant-specific lncRNA database (Jin et al., 2013). However, other approaches combining transcriptomics, epigenetic, and functional assays would provide further insights into the biological role of plant lncRNAs. Nevertheless, it is noteworthy mentioning that, from the reduced group of studied lncRNAs, two of them associate with light-dependent processes. These findings do show, however, that light is a massive re-programmer of non-coding transcription able to set in motion-specific fine-tuning mechanisms with the ultimate goal of adjusting plant development to their surroundings.

## AUTHOR CONTRIBUTIONS

All the authors read the cited publications, prepared summaries, and wrote the manuscript.

## FUNDING

RH group was supported by grants from the Spanish “Ministerio de Economía y Competitividad” (MINECO) (BIO2015-72161-EXP, BIO2015-70812-ERC, and RYC-2011-09220) and by the European Commission (PCIG2012-GA-2012-334052). PS-L acknowledges financial support from the MINECO (Grant BFU2015-64409-P). We also acknowledge the financial support of the “Severo Ochoa Program for Centres of Excellence in R&D” 2016-2019, SEV-2015-0533) and the Generalitat de Catalunya (CERCA Program).



## REFERENCES

- Andrés, F., and Coupland, G. (2012). The genetic basis of flowering responses to seasonal cues. *Nat. Rev. Genet.* 13, 627–639. doi: 10.1038/nrg3291
- Aukerman, M. J., and Sakai, H. (2003). Regulation of flowering time and floral organ identity by a microRNA and its *APETALA2*-like target genes. *Plant Cell* 15, 2730–2741. doi: 10.1105/tpc.016238
- Ballaré, C. L., and Pierik, R. (2017). The shade-avoidance syndrome: multiple signals and ecological consequences. *Plant Cell Environ.* 40, 2530–2543. doi: 10.1111/pce.12914
- Bhogale, S., Mahajan, A. S., Natarajan, B., Rajabhoj, M., Thulasiram, H. V., and Banerjee, A. K. (2014). MicroRNA156: a potential graft-transmissible microRNA that modulates plant architecture and tuberization in *Solanum tuberosum* ssp. *andigena*. *Plant Physiol.* 164, 1011–1027. doi: 10.1104/pp.113.230714
- Borges, F., and Martienssen, R. A. (2015). The expanding world of small RNAs in plants. *Nat. Rev. Mol. Cell Biol.* 16, 727–741. doi: 10.1038/nrm4085
- Cao, D., Li, Y., Wang, J., Nan, H., Wang, Y., Lu, S., et al. (2015). GmmiR156b overexpression delays flowering time in soybean. *Plant Mol. Biol.* 89, 353–363. doi: 10.1007/s11103-015-0371-5
- Casadevall, R., Rodriguez, R. E., Debernardi, J. M., Palatnik, J. F., and Casati, P. (2013). Repression of growth regulating factors by the MicroRNA396 inhibits cell proliferation by UV-B radiation in *Arabidopsis* leaves. *Plant Cell* 25, 3570–3583. doi: 10.1105/tpc.113.117473
- Casati, P. (2013). Analysis of UV-B regulated miRNAs and their targets in maize leaves. *Plant Signal. Behav.* 8:e26758. doi: 10.4161/psb.26758
- Casati, P., Stapleton, A. E., Blum, J. E., and Walbot, V. (2006). Genome-wide analysis of high-altitude maize and gene knockdown stocks implicates chromatin remodeling proteins in response to UV-B. *Plant J.* 46, 613–627. doi: 10.1111/j.1365-313X.2006.02721.x
- 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, X. (2004). A microRNA as a translational repressor of *APETALA2* in *Arabidopsis* flower development. *Science* 303, 2022–2025. doi: 10.1126/science.1088060
- Cho, S. K., Chaabane, S. B., Shah, P., Poulsen, C. P., and Yang, S. W. (2014). COP1 E3 ligase protects HYL1 to retain microRNA biogenesis. *Nat. Commun.* 5:5867. doi: 10.1038/ncomms5867
- Chuck, G., Cigan, A. M., Saetern, K., and Hake, S. (2007). The heterochronic maize mutant *Corngress1* results from overexpression of a tandem microRNA. *Nat. Genet.* 39, 544–549. doi: 10.1038/ng2001
- Chuck, G., Meeley, R., and Hake, S. (2008). Floral meristem initiation and meristem cell fate are regulated by the maize *AP2* genes *ids1* and *sid1*. *Development* 135, 3013–3019. doi: 10.1242/dev.024273
- Chung, P. J., Park, B. S., Wang, H., Liu, J., Jang, I.-C., and Chua, N.-H. (2016). Light-inducible miR163 targets *PXMT1* transcripts to promote seed germination and primary root elongation in *Arabidopsis*. *Plant Physiol.* 170, 1772–1782. doi: 10.1104/pp.15.01188
- Coon, S. L., Munson, P. J., Cherukuri, P. F., Sugden, D., Rath, M. F., Müller, M., et al. (2012). Circadian changes in long noncoding RNAs in the pineal gland. *Proc. Natl. Acad. Sci. U.S.A.* 109, 13319–13324. doi: 10.1073/pnas.1207748109
- Cui, L.-G., Shan, J.-X., Shi, M., Gao, J.-P., and Lin, H.-X. (2014). The miR156-SPL9-DFR pathway coordinates the relationship between development and abiotic stress tolerance in plants. *Plant J.* 80, 1108–1117. doi: 10.1111/tjp.12712
- Cuperus, J. T., Fahlgren, N., and Carrington, J. C. (2011). Evolution and functional diversification of MIRNA genes. *Plant Cell* 23, 431–442. doi: 10.1105/tpc.110.082784
- Díaz-Manzano, F. E., Cabrera, J., Ripoll, J.-J., del Olmo, I., Andrés, M. F., Silva, A. C., et al. (2018). A role for the gene regulatory module *microRNA172/TARGET OF EARLY ACTIVATION TAGGED 1/FLOWERING LOCUS T (miRNA172/TOE1/FT)* in the feeding sites induced by *Meloidogyne javanica* in *Arabidopsis thaliana*. *New Phytol.* 217, 813–827. doi: 10.1111/nph.14839
- Dotto, M., and Casati, P. (2017). Developmental reprogramming by UV-B radiation in plants. *Plant Sci.* 264, 96–101. doi: 10.1016/j.plantsci.2017.09.006
- Dubos, C., Le Gourrierec, J., Baudry, A., Hupé, G., Lanet, E., Debeaujon, I., et al. (2008). MYBL2 is a new regulator of flavonoid biosynthesis in *Arabidopsis thaliana*. *Plant J.* 55, 940–953. doi: 10.1111/j.1365-313X.2008.03564.x
- Edwards, K. D., Anderson, P. E., Hall, A., Salathia, N. S., Locke, J. C. W., Lynn, J. R., et al. (2006). *FLOWERING LOCUS C* mediates natural variation in the high-temperature response of the *Arabidopsis* circadian clock. *Plant Cell* 18, 639–650. doi: 10.1105/tpc.105.038315
- Fan, T., Li, X., Yang, W., Xia, K., Ouyang, J., and Zhang, M. (2015). Rice osa-miR171c mediates phase change from vegetative to reproductive development and shoot apical meristem maintenance by repressing four OsHAM transcription factors. *PLoS One* 10:e0125833. doi: 10.1371/journal.pone.0125833
- Gou, J.-Y., Felippes, F. F., Liu, C.-J., Weigel, D., and Wang, J.-W. (2011). Negative regulation of anthocyanin biosynthesis in *Arabidopsis* by a miR156-targeted SPL transcription factor. *Plant Cell* 23, 1512–1522. doi: 10.1105/tpc.111.084525
- 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, Y., Zhang, X., Wang, Y., and Ming, F. (2013). The Suppression of WRKY44 by GIGANTEA-miR172 pathway is involved in drought response of *Arabidopsis thaliana*. *PLoS One* 8:e73541. doi: 10.1371/journal.pone.0073541
- Hazen, S., Naef, F., Quisel, T., Gendron, J., Chen, H., Ecker, J., et al. (2009). Exploring the transcriptional landscape of plant circadian rhythms using genome tiling arrays. *Genome Biol.* 10:R17. doi: 10.1186/gb-2009-10-2-r17
- Hegedus, D., Yu, M., Baldwin, D., Gruber, M., Sharpe, A., Parkin, I., et al. (2003). Molecular characterization of *Brassica napus* NAC domain transcriptional activators induced in response to biotic and abiotic stress. *Plant Mol. Biol.* 53, 383–397. doi: 10.1023/B:PLAN.0000006944.61384.11
- Henriques, R., Wang, H., Liu, J., Boix, M., Huang, L.-F., and Chua, N.-H. (2017). The antiphasic regulatory module comprising *CDF5* and its antisense RNA *FLORE* links the circadian clock to photoperiodic flowering. *New Phytol.* 216, 854–867. doi: 10.1111/nph.14703
- Hernando, C. E., Garcia, C., and Mateos, J. L. (2017). Casting away the shadows: elucidating the role of light-mediated posttranscriptional control in plants. *Photochem. Photobiol.* 93, 656–665. doi: 10.1111/php.12762
- Hornitschek, P., Kohnen, M. V., Lorrain, S., Rougemont, J., Ljung, K., López-Vidriero, I., et al. (2012). Phytochrome interacting factors 4 and 5 control seedling growth in changing light conditions by directly controlling auxin signaling. *Plant J.* 71, 699–711. doi: 10.1111/j.1365-313X.2012.05033.x
- Huo, H., Wei, S., and Bradford, K. J. (2016). DELAY OF GERMINATION1 (DOG1) regulates both seed dormancy and flowering time through microRNA pathways. *Proc. Natl. Acad. Sci. U.S.A.* 113, E2199–E2206. doi: 10.1073/pnas.1600558113
- Jang, I., Henriques, R., Seo, H., Nagatani, A., and Chua, N. (2010). *Arabidopsis* PHYTOCHROME INTERACTING FACTOR proteins promote phytochrome B polyubiquitination by COP1 E3 ligase in the nucleus. *Plant Cell* 22, 2370–2383. doi: 10.1105/tpc.109.072520
- Jia, X., Ren, L., Chen, Q.-J., Li, R., and Tang, G. (2009). UV-B-responsive microRNA in *Populus tremula*. *J. Plant Physiol.* 166, 2046–2057. doi: 10.1016/j.jplph.2009.06.011
- Jia, X., Shen, J., Liu, H., Li, F., Ding, N., Gao, C., et al. (2015). Small tandem target mimic-mediated blockage of microRNA858 induces anthocyanin accumulation in tomato. *Planta* 242, 283–293. doi: 10.1007/s00425-015-2305-5
- Jin, J., Liu, J., Wang, H., Wong, L., and Chua, N.-H. (2013). PLncDB: plant long non-coding RNA database. *Bioinformatics* 29, 1068–1071. doi: 10.1093/bioinformatics/btt107
- Jones-Rhoades, M. W., and Bartel, D. P. (2004). Computational identification of plant microRNAs and their targets, including a stress-induced miRNA. *Mol. Cell* 14, 787–799. doi: 10.1016/j.molcel.2004.05.027
- Juarez, M. T., Kui, J. S., Thomas, J., Heller, B. A., and Timmermans, M. C. P. (2004). microRNA-mediated repression of *rolled leaf1* specifies maize leaf polarity. *Nature* 428, 84–88. doi: 10.1038/nature02363
- Jung, J.-H., Ju, Y., Seo, P. J., Lee, J.-H., and Park, C.-M. (2012). The SOC1-SPL module integrates photoperiod and gibberellic acid signals to control flowering time in *Arabidopsis*. *Plant J.* 69, 577–588. doi: 10.1111/j.1365-313X.2011.04813.x
- Jung, J.-H., Seo, Y.-H., Seo, P. J., Reyes, J. L., Yun, J., Chua, N.-H., et al. (2007). The GIGANTEA-regulated microRNA172 mediates photoperiodic flowering independent of CONSTANS in *Arabidopsis*. *Plant Cell* 19, 2736–2748. doi: 10.1105/tpc.107.054528

- Kim, J. Y., Lee, H. J., Jung, H. J., Maruyama, K., Suzuki, N., and Kang, H. (2010). Overexpression of microRNA395c or 395e affects differently the seed germination of *Arabidopsis thaliana* under stress conditions. *Planta* 232, 1447–1454. doi: 10.1007/s00425-010-1267-x
- Kooke, R., and Keurentjes, J. J. B. (2012). Multi-dimensional regulation of metabolic networks shaping plant development and performance. *J. Exp. Bot.* 63, 3353–3365. doi: 10.1093/jxb/err373
- Kramer, C., Loros, J. J., Dunlap, J. C., and Crosthwaite, S. K. (2003). Role for antisense RNA in regulating circadian clock function in *Neurospora crassa*. *Nature* 421, 948–952. doi: 10.1038/nature01427
- Lauter, N., Kampani, A., Carlson, S., Goebel, M., and Moose, S. P. (2005). *microRNA172* down-regulates *glossy15* to promote vegetative phase change in maize. *Proc. Natl. Acad. Sci. U.S.A.* 102, 9412–9417. doi: 10.1073/pnas.0503927102
- Leivar, P., and Monte, E. (2014). PIFs: systems integrators in plant development. *Plant Cell* 26, 56–78. doi: 10.1105/tpc.113.120857
- Leivar, P., Tepperman, J. M., Cohn, M. M., Monte, E., Al-Sady, B., Erickson, E., et al. (2012). Dynamic antagonism between phytochromes and PIF family basic helix-loop-helix factors induces selective reciprocal responses to light and shade in a rapidly responsive transcriptional network in *Arabidopsis*. *Plant Cell* 24, 1398–1419. doi: 10.1105/tpc.112.095711
- Li, H., Liang, Z., Ding, G., Shi, L., Xu, F., and Cai, H. (2016). A natural light/dark cycle regulation of Carbon-Nitrogen metabolism and gene expression in rice shoots. *Front. Plant Sci.* 7:1318. doi: 10.3389/fpls.2016.01318
- Li, S., Castillo-González, C., Yu, B., and Zhang, X. (2017). The functions of plant small RNAs in development and in stress responses. *Plant J.* 90, 654–670. doi: 10.1111/tjp.13444
- Li, W., Wang, P., Li, Y., Zhang, K., Ding, F., Nie, T., et al. (2015). Identification of MicroRNAs in response to different day lengths in soybean using high-throughput sequencing and qRT-PCR. *PLoS One* 10:e0132621. doi: 10.1371/journal.pone.0132621
- Li, Y., Varala, K., and Hudson, M. E. (2014). A survey of the small RNA population during far-red light-induced apical hook opening. *Front. Plant Sci.* 5:156. doi: 10.3389/fpls.2014.00156
- Lin, M.-C., Tsai, H.-L., Lim, S.-L., Jeng, S.-T., and Wu, S.-H. (2017). Unraveling multifaceted contributions of small regulatory RNAs to photomorphogenic development in *Arabidopsis*. *BMC Genomics* 18:559. doi: 10.1186/s12864-017-3937-6
- Liu, J., Jung, C., Xu, J., Wang, H., Deng, S., Bernad, L., et al. (2012). Genome-wide analysis uncovers regulation of Long Intergenic Noncoding RNAs in *Arabidopsis*. *Plant Cell* 24, 4333–4345. doi: 10.1105/tpc.112.102855
- López-Juez, E., Dillon, E., Magyar, Z., Khan, S., Hazeldine, S., de Jager, S. M., et al. (2008). Distinct light-initiated gene expression and cell cycle programs in the shoot apex and cotyledons of *Arabidopsis*. *Plant Cell* 20, 947–968. doi: 10.1105/tpc.107.057075
- Lorrain, S., Allen, T., Duek, P. D., Whitelam, G. C., and Fankhauser, C. (2008). Phytochrome-mediated inhibition of shade avoidance involves degradation of growth-promoting bHLH transcription factors. *Plant J.* 53, 312–323. doi: 10.1111/j.1365-3113.2007.03341.x
- Loudet, O., Michael, T. P., Burger, B. T., Le Metté, C., Mockler, T. C., Weigel, D., et al. (2008). A zinc knuckle protein that negatively controls morning-specific growth in *Arabidopsis thaliana*. *Proc. Natl. Acad. Sci. U.S.A.* 105, 17193–17198. doi: 10.1073/pnas.0807264105
- Luan, Y., Cui, J., Li, J., Jiang, N., Liu, P., and Meng, J. (2018). Effective enhancement of resistance to *Phytophthora infestans* by overexpression of *miR172a* and *b* in *Solanum lycopersicum*. *Planta* 247, 127–138. doi: 10.1007/s00425-017-2773-x
- Ma, Z., Hu, X., Cai, W., Huang, W., Zhou, X., Luo, Q., et al. (2014). *Arabidopsis* miR171-targeted scarecrow-like proteins bind to GT cis-elements and mediate gibberellin-regulated chlorophyll biosynthesis under light conditions. *PLoS Genet.* 10:e1004519. doi: 10.1371/journal.pgen.1004519
- Martin, A., Adam, H., Díaz-Mendoza, M., Żurczak, M., González-Schain, N. D., and Suárez-López, P. (2009). Graft-transmissible induction of potato tuberization by the microRNA miR172. *Development* 136, 2873–2881. doi: 10.1242/dev.031658
- Mathieu, J., Yant, L. J., Mürdter, F., Küttner, F., and Schmid, M. (2009). Repression of flowering by the miR172 Target SMZ. *PLoS Biol.* 7:e1000148. doi: 10.1371/journal.pbio.1000148
- Matsui, K., Umemura, Y., and Ohme-Takagi, M. (2008). AtMYB12, a protein with a single MYB domain, acts as a negative regulator of anthocyanin biosynthesis in *Arabidopsis*. *Plant J.* 55, 954–967. doi: 10.1111/j.1365-3113.2008.03565.x
- Michael, T., Mockler, T., Breton, G., McEntee, C., Byer, A., Trout, J., et al. (2008). Network discovery pipeline elucidates conserved time-of-day-specific cis-regulatory modules. *PLoS Genet.* 4:e14. doi: 10.1371/journal.pgen.0040014
- Nova-Franco, B., Íñiguez, L. P., Valdés-López, O., Alvarado-Affantranger, X., Leija, A., Fuentes, S. I., et al. (2015). The Micro-RNA172c-APETALA2-1 node as a key regulator of the common Bean-*Rhizobium etli* nitrogen fixation symbiosis. *Plant Physiol.* 168, 273–291. doi: 10.1104/pp.114.255547
- Qian, M., Ni, J., Niu, Q., Bai, S., Bao, L., Li, J., et al. (2017). Response of miR156-SPL module during the red peel coloration of bagging-treated Chinese sand pear (*Pyrus pyrifolia* Nakai). *Front. Physiol.* 8:550. doi: 10.3389/fphys.2017.00550
- Qian, M., Zhang, D., Yue, X., Wang, S., Li, X., and Teng, Y. (2013). Analysis of different pigmentation patterns in ‘Mantianhong’ (*Pyrus pyrifolia* Nakai) and ‘Cascade’ (*Pyrus communis* L.) under bagging treatment and postharvest UV-B/visible irradiation conditions. *Sci. Hortic.* 151, 75–82. doi: 10.1016/j.scienta.2012.12.020
- Qiao, Y., Zhang, J., Zhang, J., Wang, Z., Ran, A., Guo, H., et al. (2017). Integrated RNA-seq and sRNA-seq analysis reveals miRNA effects on secondary metabolism in *Solanum tuberosum* L. *Mol. Genet. Genomics* 292, 37–52. doi: 10.1007/s00438-016-1253-5
- Qu, D., Yan, F., Meng, R., Jiang, X., Yang, H., Gao, Z., et al. (2016). Identification of microRNAs and their targets Associated with fruit-bagging and subsequent sunlight re-exposure in the “Granny Smith” apple exocarp using high-throughput sequencing. *Front. Plant Sci.* 7:27. doi: 10.3389/fpls.2016.00027
- Ripoll, J. J., Bailey, L. J., Mai, Q.-A., Wu, S. L., Hon, C. T., Chapman, E. J., et al. (2015). microRNA regulation of fruit growth. *Nat. Plants* 1:15036. doi: 10.1038/nplants.2015.36
- Salathia, N., Davis, S. J., Lynn, J. R., Michaels, S. D., Amasino, R. M., and Millar, A. J. (2006). FLOWERING LOCUS C-dependent and -independent regulation of the circadian clock by the autonomous and vernalization pathways. *BMC Plant Biol.* 6:10. doi: 10.1186/1471-2229-6-10
- Sharma, D., Tiwari, M., Pandey, A., Bhatia, C., Sharma, A., and Trivedi, P. K. (2016). MicroRNA858 is a potential regulator of phenylpropanoid pathway and plant development. *Plant Physiol.* 171, 944–959. doi: 10.1104/pp.15.01831
- Siré, C., Moreno, A. B., García-Chapa, M., López-Moya, J. J., and Segundo, B. S. (2009). Diurnal oscillation in the accumulation of *Arabidopsis* microRNAs, miR167, miR168, miR171 and miR398. *FEBS Lett.* 583, 1039–1044. doi: 10.1016/j.febslet.2009.02.024
- 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
- Sorin, C., Bussell, J. D., Camus, I., Ljung, K., Kowalczyk, M., Geiss, G., et al. (2005). Auxin and light control of adventitious rooting in *Arabidopsis* require ARGONAUTE1. *Plant Cell* 17, 1343–1359. doi: 10.1105/tpc.105.031625
- Sun, W., Xu, X. H., Wu, X., Wang, Y., Lu, X., Sun, H., et al. (2015). Genome-wide identification of microRNAs and their targets in wild type and phyB mutant provides a key link between microRNAs and the phyB-mediated light signaling pathway in rice. *Front. Plant Sci.* 6:372. doi: 10.3389/fpls.2015.00372
- Sun, Z., Li, M., Zhou, Y., Guo, T., Liu, Y., Zhang, H., et al. (2018). Coordinated regulation of *Arabidopsis* microRNA biogenesis and red light signaling through Dicer-like 1 and phytochrome-interacting factor 4. *PLoS Genet.* 14:e1007247. doi: 10.1371/journal.pgen.1007247
- Swarup, K., Alonso-Blanco, C., Lynn, J. R., Michaels, S. D., Amasino, R. M., Koornneef, M., et al. (1999). Natural allelic variation identifies new genes in the *Arabidopsis* circadian system. *Plant J.* 20, 67–77. doi: 10.1046/j.1365-3113.1999.00577.x
- Takano, M., Inagaki, N., Xie, X., Yuzurihara, N., Hihara, F., Ishizuka, T., et al. (2005). Distinct and cooperative functions of phytochromes A, B, and C in the control of deetiolation and flowering in rice. *Plant Cell* 17, 3311–3325. doi: 10.1105/tpc.105.035899
- 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
- Voinnet, O. (2009). Origin, biogenesis, and activity of plant microRNAs. *Cell* 136, 669–687. doi: 10.1016/j.cell.2009.01.046

- Wang, B., Sun, Y., Song, N., Wang, X., Feng, H., Huang, L., et al. (2013). Identification of UV-B-induced microRNAs in wheat. *Genet. Mol. Res.* 12, 4213–4221. doi: 10.4238/2013.October.7.7
- Wang, H., Chung, P. J., Liu, J., Jang, I.-C., Kean, M. J., Xu, J., et al. (2014). Genome-wide identification of long noncoding natural antisense transcripts and their responses to light in *Arabidopsis*. *Genome Res.* 24, 444–453. doi: 10.1101/gr.165555.113
- Wang, H., and Wang, H. (2015). The miR156/SPL module, a regulatory hub and versatile toolbox, gears up crops for enhanced agronomic traits. *Mol. Plant* 8, 677–688. doi: 10.1016/j.molp.2015.01.008
- 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
- Wang, J.-W., Czech, B., and Weigel, D. (2009). miR156-regulated SPL transcription factors define an endogenous flowering pathway in *Arabidopsis thaliana*. *Cell* 138, 738–749. doi: 10.1016/j.cell.2009.06.014
- Wang, L., Mai, Y.-X., Zhang, Y.-C., Luo, Q., and Yang, H.-Q. (2010). MicroRNA171c-Targeted *SCL6-II*, *SCL6-III*, and *SCL6-IV* Genes regulate shoot branching in *Arabidopsis*. *Mol. Plant* 3, 794–806. doi: 10.1093/mp/ssq042
- 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
- Wang, Y., Li, J., Deng, X.-W., and Zhu, D. (2018). *Arabidopsis* noncoding RNA modulates seedling greening during deetiolation. *Sci. China Life Sci.* 61, 199–203. doi: 10.1007/s11427-017-9187-9
- Wang, Y., Wang, Y., Song, Z., and Zhang, H. (2016). Repression of *MYB12* by both microRNA858a and *HY5* leads to the activation of anthocyanin biosynthetic pathway in *Arabidopsis*. *Mol. Plant* 9, 1395–1405. doi: 10.1016/j.molp.2016.07.003
- 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
- Wu, G., Park, M. Y., Conway, S. R., Wang, J.-W., Weigel, D., and Poethig, R. S. (2009). The sequential action of miR156 and miR172 regulates developmental timing in *Arabidopsis*. *Cell* 138, 750–759. doi: 10.1016/j.cell.2009.06.031
- Wu, L., Liu, D., Wu, J., Zhang, R., Qin, Z., Liu, D., et al. (2013). Regulation of *FLOWERING LOCUS T* by a microRNA in *Brachypodium distachyon*. *Plant Cell* 25, 4363–4377. doi: 10.1105/tpc.113.118620
- Xie, K., Wu, C., and Xiong, L. (2006). Genomic organization, differential expression, and interaction of *SQUAMOSA* Promoter-Binding-Like transcription factors and microRNA156 in rice. *Plant Physiol.* 142, 280–293. doi: 10.1104/pp.106.084475
- Xie, Y., Liu, Y., Wang, H., Ma, X., Wang, B., Wu, G., et al. (2017). Phytochrome-interacting factors directly suppress *MIR156* expression to enhance shade-avoidance syndrome in *Arabidopsis*. *Nat. Commun.* 8:348. doi: 10.1038/s41467-017-00404-y
- Xue, Z., Ye, Q., Anson, S. R., Yang, J., Xiao, G., Kowbel, D., et al. (2014). Transcriptional interference by antisense RNA is required for circadian clock function. *Nature* 514, 650–653. doi: 10.1038/nature13671
- Yant, L., Mathieu, J., Dinh, T. T., Ott, F., Lanz, C., Wollmann, H., et al. (2010). Orchestration of the floral transition and floral development in *Arabidopsis* by the bifunctional transcription factor *APETALA2*. *Plant Cell* 22, 2156–2170. doi: 10.1105/tpc.110.075606
- Yu, B., Yang, Z., Li, J., Minakhina, S., Yang, M., Padgett, R. W., et al. (2005). Methylation as a crucial step in plant microRNA biogenesis. *Science* 307, 932–935. doi: 10.1126/science.1107130
- Yu, Y., Jia, T., and Chen, X. (2017). The ‘how’ and ‘where’ of plant microRNAs. *New Phytol.* 216, 1002–1017. doi: 10.1111/nph.14834
- Zhang, H., He, H., Wang, X., Wang, X., Yang, X., Li, L., et al. (2011). Genome-wide mapping of the *HY5*-mediated genenetworks in *Arabidopsis* that involve both transcriptional and post-transcriptional regulation. *Plant J.* 65, 346–358. doi: 10.1111/j.1365-313X.2010.04426.x
- Zhang, H., Zhao, X., Li, J., Cai, H., Deng, X. W., and Li, L. (2014). MicroRNA408 is critical for the *HY5-SPL7* gene network that mediates the coordinated response to light and copper. *Plant Cell* 26, 4933–4953. doi: 10.1105/tpc.114.127340
- Zhao, H., Chen, D., Peng, Z., Wang, L., and Gao, Z. (2013). Identification and characterization of MicroRNAs in the leaf of Ma Bamboo (*Dendrocalamus latiflorus*) by deep sequencing. *PLoS One* 8:e78755. doi: 10.1371/journal.pone.0078755
- Zhao, X., Cao, D., Huang, Z., Wang, J., Lu, S., Xu, Y., et al. (2015). Dual functions of GmTOE4a in the regulation of photoperiod-mediated flowering and plant morphology in soybean. *Plant Mol. Biol.* 88, 343–355. doi: 10.1007/s11103-015-0322-1
- Zhou, B., Fan, P., Li, Y., Yan, H., and Xu, Q. (2016). Exploring miRNAs involved in blue/UV-A light response in *Brassica rapa* reveals special regulatory mode during seedling development. *BMC Plant Biol.* 16:111. doi: 10.1186/s12870-016-0799-z
- Zhou, X., Wang, G., and Zhang, W. (2007). UV-B responsive microRNA genes in *Arabidopsis thaliana*. *Mol. Syst. Biol.* 3:103. doi: 10.1038/msb4100143
- Zhu, Q.-H., and Helliwell, C. A. (2011). Regulation of flowering time and floral patterning by miR172. *J. Exp. Bot.* 62, 487–495. doi: 10.1093/jxb/erq295

**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 © 2018 Sánchez-Retuerta, Suárez-López and Henriques. 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.



# Transposon-Derived Non-coding RNAs and Their Function in Plants

**Junnam Cho\***

*The Sainsbury Laboratory, University of Cambridge, Cambridge, United Kingdom*

## OPEN ACCESS

### Edited by:

Dora Szakonyi,  
Instituto Gulbenkian de Ciência (IGC),  
Portugal

### Reviewed by:

German Martinez,  
Swedish University of Agricultural  
Sciences, Sweden

Deqiang Zhang,  
Beijing Forestry University, China

### \*Correspondence:

Junnam Cho  
junnam.cho@slcu.cam.ac.uk

### Specialty section:

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

**Received:** 28 February 2018

**Accepted:** 16 April 2018

**Published:** 03 May 2018

### Citation:

Cho J (2018) Transposon-Derived  
Non-coding RNAs and Their Function  
in Plants. *Front. Plant Sci.* 9:600.  
doi: 10.3389/fpls.2018.00600

Transposable elements (TEs) are often regarded as harmful genomic factors and indeed they are strongly suppressed by the epigenetic silencing mechanisms. On the other hand, the mobilization of TEs brings about variability of genome and transcriptome which are essential in the survival and evolution of the host species. The vast majority of such controlling TEs influence the neighboring genes in *cis* by either promoting or repressing the transcriptional activities. Although TEs are highly repetitive in the genomes and transcribed in specific stress conditions or developmental stages, the *trans*-acting regulatory roles of TE-derived RNAs have been rarely studied. It was only recently that TEs were investigated for their regulatory roles as a form of RNA. Particularly in plants, TEs are ample source of small RNAs such as small interfering (si) RNAs and micro (mi) RNAs. Those TE-derived small RNAs have potentials to affect non-TE transcripts by sequence complementarity, thereby generating novel gene regulatory networks including stress resistance and hybridization barrier. Apart from the small RNAs, a number of long non-coding RNAs (lncRNAs) are originated from TEs in plants. For example, a retrotransposon-derived lncRNA expressed in rice root acts as a decoy RNA or miRNA target mimic which negatively controls miRNA171. The post-transcriptional suppression of miRNA171 in roots ensures the stabilization of the target transcripts encoding SCARECROW-LIKE transcription factors, the key regulators of root development. In this review article, the recent discoveries of the regulatory roles of TE-derived RNAs in plants will be highlighted.

**Keywords:** transposable elements, domestication, small RNA, long non-coding RNA, microRNA target mimic

## INTRODUCTION

Transposable elements (TEs) are the major constituent of many eukaryotic genomes. Especially in the cereal crops (e.g., barley, wheat, and maize), more than 80% of their genomes are made up of transposons (Tenaillon et al., 2010). TEs are classified to two major classes depending on their modes of transposition; class I and class II (Feschotte et al., 2002; Wicker et al., 2007). Class I TEs, also known as retrotransposons, move through RNA intermediates that are later converted to cDNAs, creating extra copies in the genome. The long terminal repeat (LTR) retrotransposons and the long interspersed nuclear elements (LINEs) are the two main types of retrotransposon. Both LTR retrotransposons and LINEs are autonomous elements since they encode for the proteins necessary for transposition, while those that depend on the autonomous elements such as large retrotransposon derivatives (LARDs), terminal repeat retrotransposons in miniature (TRIMs) and short interspersed nuclear elements (SINEs) are non-autonomous retrotransposons. Unlike class I, class II transposons, or DNA TEs, are excised from one location and insert to another genomic



position by the transposase protein which is encoded within DNA TEs. Class II TEs include another subclass, Helitrons, which replicate through rolling circle amplification. In many plant genomes, retrotransposons are more abundant compared to class II elements. Particularly, the LTR retrotransposons are the predominant families of TEs in many plants (Vitte et al., 2017). The replication cycle of the LTR retrotransposons initiates with transcription of genomic copy by the host's RNA polymerase (Pol) II. The mRNAs of LTR retrotransposons are subjected to both translation and reverse-transcription (Grandbastien, 1998). Autonomous LTR retrotransposons produce multiple proteins including GAG, aspartic protease, reverse-transcriptase, RNase H and integrase which are required for the completion of retrotransposition cycle. As a result of reverse-transcription, the linear and double-stranded DNA is produced which is known as extrachromosomal DNA (ecDNA). The ecDNAs are then transported back to the nucleus and integrate to genomic chromosomal DNA by the integrase protein.

Since TE mobilization can be mutagenic, the host genomes have evolved elaborate mechanisms to suppress their activities (Matzke and Mosher, 2014). TEs are primarily repressed by the epigenetic silencing pathways including histone modification and DNA methylation. In plants, the RNA-directed DNA methylation (RdDM) pathway plays a central role in TE silencing. Genomic regions marked by DNA methylation are recognized by the plant-specific RNA polymerase, RNA PolIV, which transcribes relatively short stretches of RNAs (Blevins et al., 2015; Li et al., 2015; Zhai et al., 2015). The transcribed RNAs are then duplexed by the RNA-dependent RNA polymerase (RDR) 2 and subsequently sliced to 24 nucleotide (nt) small interfering (si) RNAs by the DICER-like (DCL) 3. These 24 nt-siRNAs are bound by the ARGONAUTE (AGO) 4 proteins and interact with the nascent RNA transcribed by the RNA PolV. AGO4 then recruits multiple proteins including SU(VAR)3-9 HOMOLOG (SUVH) 4/5/6 and DOMAINS REARRANGED METHYLASE (DRM) 1/2 that mediate repressive histone modification (H3K9me2) and DNA methylation, respectively, thus contributing to reinforcement of the silenced state of TE chromatin (Zilberman et al., 2004; Tran et al., 2005; Zhong et al., 2014). TEs escaped from silencing or newly introduced to the genome are recognized by the RDR6-RdDM pathway that post-transcriptionally degrades TE mRNAs. RNA PolIII-transcribed TE mRNAs are processed to 21 or 22 nt-siRNAs by the RDR6 and DCL2/4 (Creasey et al., 2014). These 21 or 22 nt-siRNAs associate with AGO1 and target TE mRNAs for degradation. Intriguingly, TE-associated siRNAs can also interact with non-TE target transcripts exerting certain regulatory roles in various biological processes. In mammals, PIWI-interacting RNAs regulate a large number of mRNAs and long non-coding RNAs (lncRNAs) in testis, suggesting widespread regulatory roles of TE-derived small RNAs in both plants and animals (Watanabe et al., 2015).

In addition to siRNAs, many plant miRNAs have been suggested to be evolved from TEs (Piriyapongsa and Jordan, 2008; Li et al., 2011). Although miRNAs are distinct from siRNAs in origin and biogenesis by definition (Borges and Martienssen, 2015), the categorization of small RNAs identified

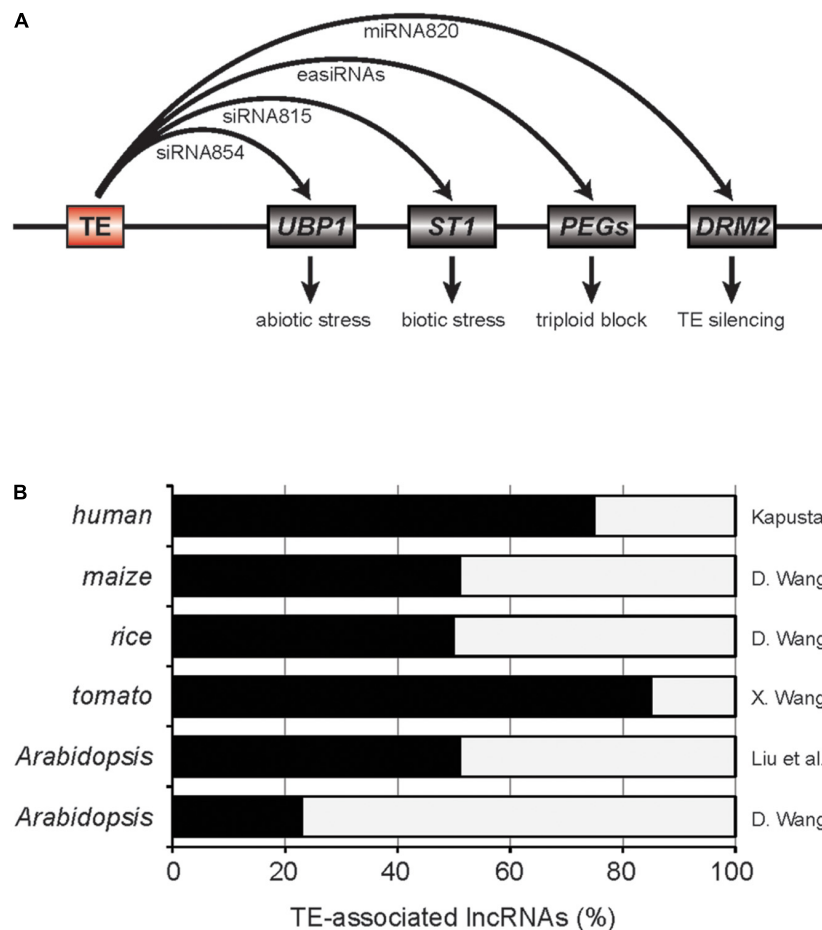
by deep sequencing has not been done with sufficient precision. In fact, considerable number of siRNAs are mis-annotated to miRNAs (McCue et al., 2012). Nonetheless, multiple lines of evidence indicate that TEs have co-opted to miRNAs since the repeated sequences associated with TEs can readily form RNA hairpin structures that can be subsequently processed by miRNA biogenesis pathways (Piriyapongsa and Jordan, 2008; Li et al., 2011). Moreover, the vast majority of lncRNAs are originated from TEs in mammalian as well as plant genomes (Kelley and Rinn, 2012; Liu et al., 2012; Kapusta et al., 2013), suggesting dynamic evolutionary exaptation of TEs in the form of RNA. In the following two sections, several examples of TE domestication to functionally relevant regulatory RNAs in plants will be explained.

## TRANSPOSON-DERIVED SMALL RNAs

### TE-siRNA in Stress Response

In the mutant of *Decreased DNA methylation 1 (DDM1)*, a gene encoding ATP-dependent chromatin remodeler in *Arabidopsis*, global DNA methylation level is dramatically reduced, thereby numerous TEs are reactivated (Vongs et al., 1993; Creasey et al., 2014). A large fraction of those reactivated TEs are accompanied with the production of 21 or 22 nt-siRNAs, known as epigenetically activated siRNAs (easiRNAs) (Creasey et al., 2014). The easiRNAs target TE transcripts for cleavage ensuring the silencing of TEs at the post-transcriptional step. Interestingly, a subset of the easiRNAs in *ddm1* mutants can interact with genic mRNAs reducing their expression levels. For example, siRNA854 is one of the easiRNAs generated in *ddm1* mutant and produced from *Athila6A* TE. It targets 3' UTR of *UBP1* transcript which encodes a stress granule protein (Figure 1A; McCue et al., 2012). Using the multiple reporter gene constructs containing the 3' UTR of *UBP1*, it was demonstrated that siRNA854 represses non-TE targets as well. The expression levels of *Athila6A* and siRNA854 are increasingly upregulated in multiple generations of *ddm1* mutation. For example, the plants with *ddm1* homozygous mutation for six generations (*ddm1* F6) have higher levels of siRNA854 compared to *ddm1* F2. The *ubp1* mutants show strong susceptibility to osmotic stress and similar phenotype was also observed in *ddm1* F6 but not in *ddm1* F2 plants. Therefore, the targeting of *Athila6A*-derived siRNA854 to *UBP1* transcript and alteration of resistance to abiotic stresses provides insight into how TEs adapted to changing environment in plants. In addition to *UBP1*, *Athila6A*-derived easiRNAs can target other genic mRNAs including *AMS* and *HHP2* (McCue et al., 2013). Several of those targets were experimentally validated for the easiRNA-mediated repression by the short tandem target mimic (STTM) transgenic approach, however, the biological relevance of this regulation is yet to be answered.

A more recent paper by Zhang et al. (2016) suggested that TE-siRNA815 in rice can induce *de novo* DNA methylation in the target gene loci through RdDM pathway (Figure 1A). Two allelic transcription factor genes, *WRKY45-1* and *WRKY45-2*, were previously shown to have opposing effects in the resistance to



**FIGURE 1 |** TE-associated small RNAs and long non-coding RNAs. **(A)** Roles of TE-derived small RNAs in plants. Red and gray boxes are TE and genes, respectively. PEGs, paternally expressed imprinted genes. **(B)** The fractions of lncRNAs associated to TEs. Related references are provided on the right.

*Xanthomonas oryzae* pv. *oryzae* (*Xoo*). *WRKY45-1* allele produces TE-siRNA815 from WANDERER\_OS-type DNA TE located in the intron. TE-siRNA815 then recognizes the complementary sequence and deposits DNA methylation through RdDM pathway in the intron of *ST1* locus, which is critical in the resistance against *Xoo*. On the other hand, *WRKY45-2* allele lacks such siRNA producing region that ensures stable *ST1* expression contributing to the pathogen resistance.

### TE-siRNA in Hybridization Barrier

Since easiRNAs are mainly produced in the epigenetic mutants, it has been questioned if easiRNAs have a function in natural conditions. Two recent studies answered this question by demonstrating the roles of easiRNAs in the hybridization barrier in *Arabidopsis* (Borges et al., 2018; Martinez et al., 2018). The vegetative nuclei of pollen grains have reduced activity of *DDM1* and numerous TEs are reactivated (Slotkin et al., 2009). Pollen-specific miRNA845b recognizes the conserved sequence [primer-binding site (PBS)] in the LTR retrotransposons activated in pollen and triggers the initial cleavage of LTR-TE mRNAs followed by the production of easiRNAs (Borges et al., 2018).

Higher dosage of the paternal genome brings higher amount of easiRNAs in fertilization which gives rise to the unbalanced gametic siRNAs and ultimately seed failure (triploid block, Martinez et al., 2018). When the paternal easiRNA levels were suppressed by *poliv* mutation, the triploid blockage phenotype was partly restored, indicating a critical role of the paternal easiRNAs in the hybridization barrier (Martinez et al., 2018). The exact mechanism for the easiRNA-mediated triploid block is still unclear, however, it was suggested that the excess paternal easiRNAs might interfere with DNA methylation establishment around the paternally expressed imprinted genes by hijacking PolV-transcribed nascent transcripts (Figure 1A).

### TE-Small RNA as Anti-silencing Factor

Transposable elements have often domesticated to miRNA genes in plants (Li et al., 2011). One example is miRNA820 of rice. miRNA820 is 22 or 24 nt in size and is originated from the internal region of CACTA DNA TE (Nosaka et al., 2012). Interestingly, miRNA820 targets the transcripts of *DRM2* that encodes a *de novo* DNA methyltransferase (Figure 1A). The targeting of miRNA820 to *DRM2* is evolutionarily conserved

in the *Oryza* genus and the repression of *DRM2* gene results in strong reduction in DNA methylation and transcriptional upregulation of many TEs. Therefore, miRNA820 can be seen as an anti-silencing factor encoded within a TE that works at the post-transcriptional level. Similarly, UBP1 is the *Arabidopsis* homolog of TIA-1 in mammals which is known to suppress the viral translation of Tick-Borne Encephalitis Virus (Albornoz et al., 2014). McCue et al. (2013) have also shown that UBP1 protein forms the cytoplasmic stress granules in abiotic stress condition or when heterochromatic TE silencing is released, for instance in *ddm1* mutant. As is the case in mammals that TIA-1 inhibits the viral translation, the level of GAG protein encoded in *Athila6A* was elevated in *ddm1 rdr6 ubp1* triple mutants (McCue et al., 2013). This data supports the notion that UBP1 acts on the activated TE mRNAs to suppress their translation and therefore TE-siRNAs are the repressors of the host TE silencing mechanism.

## TRANSPOSON-DERIVED lncRNAs

There is emerging evidence of TE domestication to lncRNAs in both mammalian and plant genomes (Figure 1B; Kelley and Rinn, 2012; Kapusta et al., 2013; Johnson and Guigo, 2014; Liu et al., 2015; Wang et al., 2015, 2016; Quattro et al., 2017). lncRNA can be defined as a transcript of at least 200 bp in size but has low protein-coding potential (Liu et al., 2015). It is well-studied that lncRNAs perform various cellular function including the recruitment of the epigenetic regulators to target chromatin or the sequestration of miRNAs (Franco-Zorrilla et al., 2007; Heo and Sung, 2011; Csorba et al., 2014). In the past decade, transcriptomic analyses have dramatically expanded the catalog of the lncRNAs in various tissues and stress conditions of many plant species. Despite the large number of plant lncRNAs identified so far, however, the biological roles are still largely unexplored. In this section, the current status of plant TE-lncRNA studies will be discussed.

## TE-lncRNA in Stress Response

Since many TEs in plants possess stress-responsive *cis*-acting elements within them (Paszowski, 2015), TE-lncRNAs often appear in specific stress conditions (Liu et al., 2012; Quattro et al., 2017; Wang et al., 2017). In a recent report, Wang et al. (2017) interrogated the lncRNAs in *Arabidopsis*, rice and maize under various abiotic stresses. There was a large discrepancy in TE families that TE-lncRNAs are originated from; RC/Helitron in *Arabidopsis*, MITEs in rice and LTR retrotransposons in maize were predominantly overrepresented in TE-lncRNAs (Wang et al., 2017). Particularly, TE-lncRNA11195 in *Arabidopsis* contains an LTR-type retrotransposon and is activated after abiotic stresses or ABA treatment. The deletion of the LTR sequence compromised the ABA responsiveness, suggesting that LTR sequence conferred the stress responsiveness to TE-lncRNA11195 (Wang et al., 2017). TE-lncRNA11195 was tested for its role in stress response using T-DNA insertional mutants. Interestingly, two independent mutant lines showed marked increase in resistance to abscisic acid (ABA) in

root elongation and shoot fresh weight (Wang et al., 2017). TE-lncRNAs in tomato were also described to be responsive for both abiotic and biotic stresses (Wang et al., 2015), however, the biological function has not been investigated as yet.

## TE-lncRNA and Development

In mammals, the majority of lncRNAs are derived from TEs and exhibit strong tissue-specific expression pattern (Kelley and Rinn, 2012; Kapusta et al., 2013). Similarly, TE activation in plants is associated with specific development stages potentiating the emergence of tissue-specific TE-lncRNA (Hsieh et al., 2009; Slotkin et al., 2009; Baubec et al., 2014; Cho and Paszowski, 2017). Recently, Cho and Paszowski (2017) have investigated the expression pattern of TEs in various rice tissues and identified a retrotransposon-derived transcript called *MIKKI* which is specifically transcribed in rice roots. *MIKKI* contains multiple introns and has low coding potential, which is a strong sign of domestication to lncRNA. Intriguingly, the fourth intron of *MIKKI* is derived from an independent family of retrotransposon and the splicing of this intron generates a binding site for miR171 in the exon-exon junction. Despite the miR171-binding sequence, *MIKKI* mRNAs are not cleaved by miR171. The miR171-binding site of *MIKKI* does not perfectly base-pair with its cognate miRNA but has two mismatches at the positions where the cleavage is supposed to occur. It is very well-known that mismatches in the cleavage positions attenuate the cleavage activity of miRNA and is regarded as the signature of miRNA target mimic (Franco-Zorrilla et al., 2007; Yan et al., 2012; Reichel et al., 2015). Indeed, the knock-out mutants of *MIKKI* that had lost the target mimicking sequence showed higher levels of miR171, while overexpression of *MIKKI* resulted in the downregulation of miR171. miR171 targets the mRNAs encoding SCARECROW-Like (SCL) transcription factors which are critical regulators of root development (Wang et al., 2010). Therefore, *MIKKI* evolved from retrotransposons in rice and was positively selected to suppress miR171 in root. This in turn stabilizes the mRNAs of SCLs which are essential in the root development. Similarly, Quattro et al., 2017 also identified multiple TE-derived lncRNAs from *Brachypodium* genome that are able to interact with miRNAs, however, their target mimicry activities are yet to be confirmed.

## CONCLUDING REMARKS

Transposon-derived RNAs have been underestimated for a long time and their biological function have just started to be unveiled. The fact that TEs are repeated in the genome has significantly hampered the investigation of transposons so far. For example, the short length of the next-generation sequencing reads causes drastic ambiguity and imprecision in mapping the TE reads. Recent advance of the long read sequencing of PacBio (Disdero and Filée, 2017) and Oxford Nanopore (Debladis et al., 2017) is expected to overcome this shortcoming. In addition, due to the multiplicity of TEs in the genome and possible redundancy

between them, the genetic analyses of TEs have been challenging. CRISPR-mediated mutagenesis has become more efficient and even a large deletion of an entire TE can be made by triggering the double-strand breaks in the flanking regions of the targeted TE (Gao et al., 2016; Ordon et al., 2017). Another option worth to consider is the population genetics approach. There is increasing number of available genome sequences and the number will grow exponentially as the sequencing cost drops. A large scale genome resequencing data analysis performed in the *Arabidopsis* natural accessions revealed that the TE landscape is very dynamic and the transcriptomic, epigenomic as well as phenotypic variations are attributed to TEs (Quadrona et al., 2016; Stuart et al., 2016). Taken all together, it seems that now is the best time to explore the hidden roles of TEs in plants by applying the new technologies developed recently.

## REFERENCES

- Albornoz, A., Carletti, T., Corazza, G., and Marcello, A. (2014). The stress granule component TIA-1 binds tick-borne encephalitis virus RNA and is recruited to perinuclear sites of viral replication to inhibit viral translation. *J. Virol.* 88, 6611–6622. doi: 10.1128/JVI.03736-13
- Baubec, T., Finke, A., Mittelsten Scheid, O., and Pecinka, A. (2014). Meristem-specific expression of epigenetic regulators safeguards transposon silencing in *Arabidopsis*. *EMBO Rep.* 15, 446–452. doi: 10.1002/embr.201337915
- Blevins, T., Podicheti, R., Mishra, V., Marasco, M., Wang, J., Rusch, D., et al. (2015). Identification of Pol IV and RDR2-dependent precursors of 24 nt siRNAs guiding de novo DNA methylation in *Arabidopsis*. *Elife* 4:e09591. doi: 10.7554/eLife.09591
- Borges, F., and Martienssen, R. A. (2015). The expanding world of small RNAs in plants. *Nat. Rev. Mol. Cell Biol.* 16, 727–741. doi: 10.1038/nrm4085
- Borges, F., Parent, J. S., Van Ex, F., Wolff, P., Martinez, G., Köhler, C., et al. (2018). Transposon-derived small RNAs triggered by miR845 mediate genome dosage response in *Arabidopsis*. *Nat. Genet.* 50, 186–192. doi: 10.1038/s41588-017-0032-5
- Cho, J., and Paszkowski, J. (2017). Regulation of rice root development by a retrotransposon acting as a microRNA sponge. *Elife* 6:e30038. doi: 10.7554/eLife.30038
- Creasey, K. M., Zhai, J., Borges, F., Van Ex, F., Regulski, M., Meyers, B. C., et al. (2014). miRNAs trigger widespread epigenetically activated siRNAs from transposons in *Arabidopsis*. *Nature* 508, 411–415. doi: 10.1038/nature13069
- Csorba, T., Questa, J. I., Sun, Q., and Dean, C. (2014). Antisense COOLAIR mediates the coordinated switching of chromatin states at FLC during vernalization. *Proc. Natl. Acad. Sci. U.S.A.* 111, 16160–16165. doi: 10.1073/pnas.1419030111
- Debladis, E., Llauro, C., Carpentier, M. C., Mirouze, M., and Panaud, O. (2017). Detection of active transposable elements in *Arabidopsis thaliana* using Oxford Nanopore Sequencing technology. *BMC Genomics* 18:537. doi: 10.1186/s12864-017-3753-z
- Disdero, E., and Filée, J. (2017). LoRTE: detecting transposon-induced genomic variants using low coverage PacBio long read sequences. *Mob. DNA* 8:5. doi: 10.1186/s13100-017-0088-x
- Feschotte, C., Jiang, N., and Wessler, S. R. (2002). Plant transposable elements: where genetics meets genomics. *Nat. Rev. Genet.* 3, 329–341. doi: 10.1038/nrg793
- Franco-Zorrilla, J. M., Valli, A., Todesco, M., Mateos, I., Puga, M. I., Rubio-Somoza, I., et al. (2007). Target mimicry provides a new mechanism for regulation of microRNA activity. *Nat. Genet.* 39, 1033–1037. doi: 10.1038/ng2079
- Gao, X., Chen, J., Dai, X., Zhang, D., and Zhao, Y. (2016). An effective strategy for reliably isolating heritable and Cas9-free *Arabidopsis* mutants generated by CRISPR/Cas9-mediated genome editing. *Plant Physiol.* 171, 1794–1800. doi: 10.1104/pp.16.00663
- Grandbastien, M. A. (1998). Activation of plant retrotransposons under stress conditions. *Trends Plant Sci.* 3, 181–187. doi: 10.1016/S1360-1385(98)01232-1
- Heo, J. B., and Sung, S. (2011). Vernalization-mediated epigenetic silencing by a long intronic noncoding RNA. *Science* 331, 76–80. doi: 10.1126/science.1197349
- Hsieh, T. F., Ibarra, C. A., Silva, P., Zemach, A., Eshed-Williams, L., Fischer, R. L., et al. (2009). Genome-wide demethylation of *Arabidopsis* endosperm. *Science* 324, 1451–1454. doi: 10.1126/science.1172417
- Johnson, R., and Guigo, R. (2014). The RIDL hypothesis: transposable elements as functional domains of long noncoding RNAs. *RNA* 20, 959–976. doi: 10.1261/rna.044560.114
- Kapusta, A., Kronenberg, Z., Lynch, V. J., Zhuo, X., Ramsay, L., Bourque, G., et al. (2013). Transposable elements are major contributors to the origin, diversification, and regulation of vertebrate long noncoding RNAs. *PLoS Genet.* 9:e1003470. doi: 10.1371/journal.pgen.1003470
- Kelley, D. R., and Rinn, J. L. (2012). Transposable elements reveal a stem cell-specific class of long noncoding RNAs. *Genome Biol.* 13:R107. doi: 10.1186/gb-2012-13-11-r107
- Li, S., Vandivier, L. E., Tu, B., Gao, L., Won, S. Y., Li, S., et al. (2015). Detection of Pol IV/RDR2-dependent transcripts at the genomic scale in *Arabidopsis* reveals features and regulation of siRNA biogenesis. *Genome Res.* 25, 235–245. doi: 10.1101/gr.182238.114
- Li, Y., Li, C., Xia, J., and Jin, Y. (2011). Domestication of transposable elements into microRNA genes in plants. *PLoS One* 6:e19212. doi: 10.1371/journal.pone.0019212
- Liu, J., Jung, C., Xu, J., Wang, H., Deng, S., Bernad, L., et al. (2012). Genome-wide analysis uncovers regulation of long intergenic noncoding RNAs in *Arabidopsis*. *Plant Cell* 24, 4333–4345. doi: 10.1105/tpc.112.102855
- Liu, J., Wang, H., and Chua, N. H. (2015). Long noncoding RNA transcriptome of plants. *Plant Biotechnol. J.* 13, 319–328. doi: 10.1111/pbi.12336
- Martinez, G., Wolff, P., Wang, Z., Moreno-romero, J., Santos-gonzález, J., Conze, L. L., et al. (2018). Paternal easiRNAs regulate parental genome dosage in *Arabidopsis*. *Nat. Genet.* 50, 193–198. doi: 10.1038/s41588-017-0033-4
- Matzke, M. A., and Mosher, R. A. (2014). RNA-directed DNA methylation: an epigenetic pathway of increasing complexity. *Nat. Rev. Genet.* 15, 394–408. doi: 10.1038/nrg3683
- McCue, A. D., Nuthikattu, S., Reeder, S. H., and Slotkin, R. K. (2012). Gene expression and stress response mediated by the epigenetic regulation of a transposable element small RNA. *PLoS Genet.* 8:e1002474. doi: 10.1371/journal.pgen.1002474
- McCue, A. D., Nuthikattu, S., Reeder, S. H., and Slotkin, R. K. (2013). Genome-wide identification of genes regulated in trans by transposable element small interfering RNAs. *RNA Biol.* 10, 1379–1395. doi: 10.4161/rna.25555

## AUTHOR CONTRIBUTIONS

The author confirms being the sole contributor of this work and approved it for publication.

## FUNDING

This work was supported by the European Research Council (EVOBREED) [322621] and the Gatsby Charitable Foundation [AT3273/GLE].

## ACKNOWLEDGMENTS

I thank Dr. Jayne Griffiths for critical reading.



- Nosaka, M., Itoh, J. I., Nagato, Y., Ono, A., Ishiwata, A., and Sato, Y. (2012). Role of transposon-derived small RNAs in the interplay between genomes and parasitic DNA in rice. *PLoS Genet.* 8:e1002953. doi: 10.1371/journal.pgen.1002953
- Ordon, J., Gantner, J., Kemna, J., Schwalgun, L., Reschke, M., Streubel, J., et al. (2017). Generation of chromosomal deletions in dicotyledonous plants employing a user-friendly genome editing toolkit. *Plant J.* 89, 155–168. doi: 10.1111/tpj.13319
- Paszkowski, J. (2015). Controlled activation of retrotransposition for plant breeding. *Curr. Opin. Biotechnol.* 32, 200–206. doi: 10.1016/j.copbio.2015.01.003
- Piriyapongsa, J., and Jordan, I. K. (2008). Dual coding of siRNAs and miRNAs by plant transposable elements. *RNA* 14, 814–821. doi: 10.1261/rna.916708.ferred
- Quadana, L., Silveira, A. B., Mayhew, G. F., LeBlanc, C., Martienssen, R. A., Jeddeloh, J. A., et al. (2016). The *Arabidopsis thaliana* mobilome and its impact at the species level. *Elife* 5:e15716. doi: 10.7554/eLife.15716
- Quattro, C. D., Enrico Pè, M., and Bertolini, E. (2017). Long noncoding RNAs in the model species *Brachypodium distachyon*. *Sci. Rep.* 7:11252. doi: 10.1038/s41598-017-11206-z
- Reichel, M., Li, Y., Li, J., and Millar, A. A. (2015). Inhibiting plant microRNA activity: molecular SPONGEs, target MIMICs and STTMs all display variable efficacies against target microRNAs. *Plant Biotechnol. J.* 13, 915–926. doi: 10.1111/pbi.12327
- Slotkin, R. K., Vaughn, M., Borges, F., Tanurdžić, M., Becker, J. D., Feijó, J. A., et al. (2009). Epigenetic reprogramming and small RNA silencing of transposable elements in pollen. *Cell* 136, 461–472. doi: 10.1016/j.cell.2008.12.038
- Stuart, T., Eichten, S. R., Cahn, J., Karpievitch, Y. V., Borevitz, J. O., and Lister, R. (2016). Population scale mapping of transposable element diversity reveals links to gene regulation and epigenomic variation. *Elife* 5:e20777. doi: 10.7554/eLife.20777
- Tenaillon, M. I., Hollister, J. D., and Gaut, B. S. (2010). A triptych of the evolution of plant transposable elements. *Trends Plant Sci.* 15, 471–478. doi: 10.1016/j.tplants.2010.05.003
- Tran, R. K., Zilberman, D., De Bustos, C., Ditt, R. F., Henikoff, J. G., Lindroth, A. M., et al. (2005). Chromatin and siRNA pathways cooperate to maintain DNA methylation of small transposable elements in *Arabidopsis*. *Genome Biol.* 6:R90. doi: 10.1186/gb-2005-6-11-r90
- Vitte, C., Fustier, M. A., Alix, K., and Tenaillon, M. I. (2017). The bright side of transposons in crop evolution. *Brief. Funct. Genomics* 13, 276–295. doi: 10.1093/bfpg/elu002
- Vongs, A., Kakutani, T., Martienssen, R. A., and Richardst, E. J. (1993). *Arabidopsis thaliana* DNA methylation mutants. *Science* 260, 1926–1928. doi: 10.1126/science.8316832
- Wang, D., Qu, Z., Yang, L., Zhang, Q., Liu, Z. H., Do, T., et al. (2017). Transposable elements (TEs) contribute to stress-related long intergenic noncoding RNAs in plants. *Plant J.* 90, 133–146. doi: 10.1111/tpj.13481
- Wang, L., Mai, Y. X., Zhang, Y. C., Luo, Q., and Yang, H. Q. (2010). MicroRNA171c-targeted SCL6-II, SCL6-III, and SCL6-IV genes regulate shoot branching in *Arabidopsis*. *Mol. Plant* 3, 794–806. doi: 10.1093/mp/ssq042
- Wang, X., Ai, G., Zhang, C., Cui, L., Wang, J., Li, H., et al. (2015). Expression and diversification analysis reveals transposable elements play important roles in the origin of *Lycopersicon*-specific lncRNAs in tomato. *New Phytol.* 209, 1442–1455. doi: 10.1111/nph.13718
- Wang, Z., Schwacke, R., and Kunze, R. (2016). DNA damage-induced transcription of transposable elements and long non-coding RNAs in *Arabidopsis* is rare and ATM-dependent. *Mol. Plant* 9, 1142–1155. doi: 10.1016/j.molp.2016.04.015
- Watanabe, T., Cheng, E. C., Zhong, M., and Lin, H. (2015). Retrotransposons and pseudogenes regulate mRNAs and lncRNAs via the piRNA pathway in the germline. *Genome Res.* 25, 368–380. doi: 10.1101/gr.180802.114
- Wicker, T., Sabot, F., Hua-Van, A., Bennetzen, J. L., Capy, P., Chalhoub, B., et al. (2007). A unified classification system for eukaryotic transposable elements. *Nat. Rev. Genet.* 8, 973–982. doi: 10.1038/nrg2165
- Yan, J., Gu, Y., Jia, X., Kang, W., Pan, S., Tang, X., et al. (2012). Effective small RNA destruction by the expression of a short tandem target mimic in *Arabidopsis*. *Plant Cell* 24, 415–427. doi: 10.1105/tpc.111.094144
- Zhai, J., Bischof, S., Wang, H., Feng, S., Lee, T. F., Teng, C., et al. (2015). A one precursor one siRNA model for pol IV-dependent siRNA biogenesis. *Cell* 163, 445–455. doi: 10.1016/j.cell.2015.09.032
- Zhang, H., Tao, Z., Hong, H., Chen, Z., Wu, C., Li, X., et al. (2016). Transposon-derived small RNA is responsible for modified function of WRKY45 locus. *Nat. Plants* 2:16016. doi: 10.1038/nplants.2016.16
- Zhong, X., Du, J., Hale, C. J., Gallego-Bartolome, J., Feng, S., Vashisht, A. A., et al. (2014). Molecular mechanism of action of plant DRM de novo DNA methyltransferases. *Cell* 157, 1050–1060. doi: 10.1016/j.cell.2014.03.056
- Zilberman, D., Cao, X., Johansen, L. K., Xie, Z., Carrington, J. C., and Jacobsen, S. E. (2004). Role of *Arabidopsis* ARGONAUTE4 in RNA-directed DNA methylation triggered by inverted repeats. *Curr. Biol.* 14, 1214–1220. doi: 10.1016/j.cub.2004.06.055

**Conflict of Interest Statement:** The author declares 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 © 2018 Cho. 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 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.



# Cold-Dependent Expression and Alternative Splicing of Arabidopsis Long Non-coding RNAs

Cristiane P. G. Calixto<sup>1</sup>, Nikoleta A. Tzioutziou<sup>1</sup>, Allan B. James<sup>2</sup>, Csaba Hornyik<sup>3</sup>, Wenbin Guo<sup>1,4</sup>, Runxuan Zhang<sup>4</sup>, Hugh G. Nimmo<sup>2</sup> and John W. S. Brown<sup>1,3\*</sup>

<sup>1</sup> Plant Sciences Division, School of Life Sciences, University of Dundee, Dundee, United Kingdom, <sup>2</sup> Institute of Molecular, Cell and Systems Biology, College of Medical, Veterinary and Life Sciences, University of Glasgow, Glasgow, United Kingdom, <sup>3</sup> Cell and Molecular Sciences, The James Hutton Institute, Dundee, United Kingdom, <sup>4</sup> Information and Computational Sciences, The James Hutton Institute, Dundee, United Kingdom

## OPEN ACCESS

### Edited by:

Mathew G. Lewsey,  
La Trobe University, Australia

### Reviewed by:

Alice Pajoro,  
Max Planck Institute for Plant  
Breeding Research, Germany  
Anthony Gobert,  
UPR2357 Institut de Biologie  
Moléculaire des Plantes (IBMP),  
France  
Yuichiro Watanabe,  
The University of Tokyo, Japan

### \*Correspondence:

John W. S. Brown  
j.w.s.brown@dundee.ac.uk

### Specialty section:

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

**Received:** 23 November 2018

**Accepted:** 12 February 2019

**Published:** 28 February 2019

### Citation:

Calixto CPG, Tzioutziou NA,  
James AB, Hornyik C, Guo W,  
Zhang R, Nimmo HG and Brown JWS  
(2019) Cold-Dependent Expression  
and Alternative Splicing  
of Arabidopsis Long Non-coding  
RNAs. *Front. Plant Sci.* 10:235.  
doi: 10.3389/fpls.2019.00235

Plants re-program their gene expression when responding to changing environmental conditions. Besides differential gene expression, extensive alternative splicing (AS) of pre-mRNAs and changes in expression of long non-coding RNAs (lncRNAs) are associated with stress responses. RNA-sequencing of a diel time-series of the initial response of *Arabidopsis thaliana* rosettes to low temperature showed massive and rapid waves of both transcriptional and AS activity in protein-coding genes. We exploited the high diversity of transcript isoforms in AtRTD2 to examine regulation and post-transcriptional regulation of lncRNA gene expression in response to cold stress. We identified 135 lncRNA genes with cold-dependent differential expression (DE) and/or differential alternative splicing (DAS) of lncRNAs including natural antisense RNAs, sORF lncRNAs, and precursors of microRNAs (miRNAs) and *trans*-acting small-interfering RNAs (tasiRNAs). The high resolution (HR) of the time-series allowed the dynamics of changes in transcription and AS to be determined and identified early and adaptive transcriptional and AS changes in the cold response. Some lncRNA genes were regulated only at the level of AS and using plants grown at different temperatures and a HR time-course of the first 3 h of temperature reduction, we demonstrated that the AS of some lncRNAs is highly sensitive to small temperature changes suggesting tight regulation of expression. In particular, a splicing event in *TAS1a* which removed an intron that contained the miR173 processing and phased siRNAs generation sites was differentially alternatively spliced in response to cold. The cold-induced reduction of the spliced form of *TAS1a* and of the tasiRNAs suggests that splicing may enhance production of the siRNAs. Our results identify candidate lncRNAs that may contribute to the regulation of expression that determines the physiological processes essential for acclimation and freezing tolerance.

**Keywords:** long non-coding RNA, primary microRNA, alternative splicing, diel time-course, high-resolution RNA-seq, cold transcriptome

## INTRODUCTION

Non-coding RNAs (ncRNAs) are a diverse set of RNAs which do not generally code for proteins. They include families of house-keeping ncRNAs and their precursors such as small nuclear ribonucleoprotein particle RNAs (snRNAs), small Cajal body RNAs (scaRNAs), ribosomal RNAs (rRNAs), transfer RNAs (tRNAs), and small nucleolar RNA (snoRNAs) (Shaw and Brown, 2012; Cech and Steitz, 2014). The regulatory ncRNAs include small RNAs such as microRNAs (miRNAs), short-interfering RNAs (siRNAs), and long non-coding RNAs (lncRNAs) which are expressed from intergenic regions or introns and include natural antisense transcripts (NATs). Some lncRNAs are precursors of small RNA production: primary transcripts are processed to miRNAs or siRNAs such as trans-acting siRNAs (tasiRNAs) or natural antisense siRNAs (nat-siRNAs) derived from double-stranded RNA molecules. The regulatory ncRNAs function in a wide range of cellular processes, from the regulation of transcription and splicing, to chromatin modification, gene inactivation, and translation (Liu et al., 2015). In plants and animals, tens of thousands of lncRNAs are now routinely detected in RNA-seq analyses demonstrating a new level of complexity of gene expression (Liu et al., 2012, 2015; Zhang and Chen, 2013; Yuan et al., 2016; Lu et al., 2017; Wang et al., 2017; Severing et al., 2018; Zhao et al., 2018).

In plants, the majority of lncRNAs are transcribed by RNA polymerase II (Pol II) but some are generated with Pol IV and/or Pol V (Liu et al., 2015). Like mRNAs, they are usually capped at the 5' end, can be spliced and they form two classes which are either polyadenylated or non-polyadenylated (Liu et al., 2012; Di et al., 2014; Liu et al., 2015). Genome-wide analyses have identified thousands of plant lncRNAs in different plant species such as *Arabidopsis*, rice, maize, tomato, wheat, cucumber, poplar, cassava, and cotton and they are shown to be differentially expressed in response to stresses caused by abiotic environmental factors such as cold, heat, drought, and salt (Ben Amor et al., 2009; Liu et al., 2012; Di et al., 2014; Shuai et al., 2014; Zhang et al., 2014; Hao et al., 2015; He et al., 2015; Wang et al., 2015a,b; Chen et al., 2016; Forestan et al., 2016; Khemka et al., 2016; Lu et al., 2016; Yuan et al., 2016; Zhao et al., 2016; Li et al., 2017; Shumayla et al., 2017; Wang et al., 2017; Severing et al., 2018; Zhao et al., 2018). Infection of different plant species by pathogens also leads to differential expression (DE) of lncRNAs: for example, wheat with powdery mildew and stripe rust (Xin et al., 2011; Zhang et al., 2013), *Arabidopsis* with *Fusarium oxysporum* (Zhu et al., 2014), *Brassica napus* with stem rot (Joshi et al., 2016), and tomato with Tomato Yellow Curly Leaf Virus (Wang et al., 2018). The function of the vast majority of lncRNAs is unknown but, in general, they are primarily associated with regulation of gene expression via a range of different mechanisms and ultimately are important in differentiation, development, and abiotic and biotic stress responses (Ben Amor et al., 2009; Liu et al., 2012, 2015; Zhang and Chen, 2013; Wang et al., 2017; Severing et al., 2018; Zhao et al., 2018). The functions of specific lncRNAs illustrate the breadth of regulation of different plant processes and their modes of action (Kim and Sung, 2012; Liu et al., 2012, 2015; Zhang and Chen, 2013; Qin et al., 2017;

Wang et al., 2017; Kindgren et al., 2018; Zhao et al., 2018). For example, lncRNAs are involved in vernalization-mediated regulation of flowering (Swiezewski et al., 2009; Ietswaart et al., 2012), circadian regulation of flowering (Henriques et al., 2017), low temperature suppression of flowering (Zhao et al., 2018), lateral root development (Bardou et al., 2014), gametophyte development (Wunderlich et al., 2014), photoperiod-sensitive male sterility (Ding et al., 2012), phosphate homeostasis (Franco-Zorrilla et al., 2007), cold acclimation (Kindgren et al., 2018), and drought (Qin et al., 2017).

Low temperatures negatively affect plant growth and development and, in general, plants from temperate climatic regions can tolerate chilling temperatures (0–15°C) and increase their freezing tolerance by prior exposure to low, non-freezing temperatures. This process of acclimation involves complex physiological, biochemical, and molecular changes that protect cell integrity during exposure to freezing. The multiple, cold-responsive changes reflect reprogramming of gene expression involving chromatin modification, transcription, post-transcriptional processing, post-translational modification, and protein turnover (Thomashow, 2010; Knight and Knight, 2012; Zhu, 2016). For many years, the major focus of transcriptome reprogramming in *Arabidopsis* in response to cold has been at the level of transcription and identifying gene targets of specific transcription factors (e.g., C-repeat binding factors – CBFs and the CBF regulon; Thomashow, 2010; Knight and Knight, 2012; Zhu, 2016). High-throughput RNA-sequencing now enables investigation of other aspects of expression control including alternative splicing (AS) and lncRNAs. AS of protein-coding pre-mRNAs is a regulated process and a major level of post-transcriptional regulation of expression; it has been associated with many different stress responses including cold (Leviatan et al., 2013; Staiger and Brown, 2013; Calixto et al., 2016; Hartmann et al., 2016; Laloum et al., 2017; Pajoro et al., 2017). Recently, the scale and dynamics of the contribution of AS to the cold response has been demonstrated by coincident waves of transcription and AS occurring in the first few hours of exposure to cold and the speed and sensitivity of some genes to small temperature changes (Calixto et al., 2018). The importance of AS in both stress responses and development is to increase proteome diversity by generating AS isoforms which encode different functional protein variants. However, AS also can regulate expression of genes by modulating the proportion of protein-coding transcripts vs transcripts which contain premature termination codons (PTCs), many of which are targeted for degradation by non-sense-mediated decay (NMD; Kalyna et al., 2012; Drechsel et al., 2013; Fu and Ares, 2014; Lee and Rio, 2015).

Little is known about the extent of AS of plant lncRNAs *per se* or how AS can affect the levels or function of lncRNAs. Many lncRNAs contain introns which are spliced during lncRNA biogenesis (Liu et al., 2012) and a number of annotated (TAIR) lncRNAs have splice variants. The best-characterized AS of a plant lncRNA is the differential AS and alternative polyadenylation of *COOLAIR* (antisense to *FLOWERING LOCUS C*), which determine the levels of FLC (Swiezewski et al., 2009; Marquardt et al., 2014). The lncRNA,

*FLORE*, is regulated by the circadian clock and promotes flowering by repressing *CYCLING DOF FACTOR 5* (*CDF5*) to allow expression of FT (Henriques et al., 2017). *FLORE* is differentially expressed and alternatively spliced in response to cold but the function of *FLORE* AS is unknown. On the other hand, lncRNAs can affect the splicing/AS of target genes by binding and sequestering specific splicing factors (Bardou et al., 2014); reviewed in Romero-Barrios et al. (2018). Primary miRNA transcripts (pri-miRNAs) are a class of lncRNAs where splicing and AS have been well characterized and shown to be important to the production of some mature miRNAs (Brown et al., 2008; Yan et al., 2012; Bielewicz et al., 2013; Barciszewska-Pacak et al., 2015; Knop et al., 2017; Stepień et al., 2017). Recently, in human, deep RNA-seq targeting very low expressed RNAs found that virtually all lncRNAs are alternatively spliced generating massive lncRNA diversity which may be important to the evolution of regulatory modules of expression (Deveson et al., 2018). The intrinsic potential of splicing/AS of lncRNAs to modify their function, and of lncRNAs to influence splicing/AS and expression of target genes through a variety of mechanisms suggests a need to increase our understanding of the interplay between lncRNAs and AS.

We recently performed an ultra-deep RNA-seq analysis of a diel time-series of Arabidopsis plants transferred to cold (Calixto et al., 2018; **Supplementary Figure S1**). The analysis used AtRTD2, a new Arabidopsis transcriptome with greatly increased numbers of transcripts compared to the Arabidopsis TAIR10 and Araport11 databases (Zhang et al., 2017). The greater number and diversity of transcripts allowed the identification of nearly 2500 cold-regulated differentially alternatively spliced protein-coding genes, the majority of which were novel to the cold response. The time-series showed (1) massive peaks of transcription and AS of thousands of protein-coding genes in the first few hours of exposure to cold, (2) defined early and late, transient, and adaptive changes in expression and AS, and (3) demonstrated the speed and sensitivity of AS of particular genes (Calixto et al., 2018). We have interrogated this extensive time-series dataset to examine the transcriptional and post-transcriptional regulation of expression of lncRNAs and the relative contributions and dynamics of transcriptional and AS. We identify cold-responsive lncRNAs in terms of both significant DE and AS.

## MATERIALS AND METHODS

### Plant Material and Growth Conditions

Details of the plant material and growth conditions have been described previously (Calixto et al., 2018). Briefly, *Arabidopsis thaliana* Col-0 seeds were grown hydroponically for 5 weeks in Microclima environment-controlled cabinets (Snijders Scientific), maintaining 20°C, 55% relative humidity, and 12 h light:12 h dark (James et al., 2012). Arabidopsis rosettes (9–13) were harvested and pooled at each sampling time forming one biological replicate. Harvesting occurred every 3 h beginning with the last 24 h at 20°C, and on days 1 and 4 after transfer to 4°C giving 26 time-points in the time-series (**Supplementary Figure S1A**). Day 1 at 4°C represents the “transition” from 20

to 4°C when plants first begin to experience the temperature decrease; day 4 at 4°C represents “acclimation” where plants have been exposed to 4°C for 4 days (**Supplementary Figure S1A**). Three biological replicates were generated for each time-point in separate experiments (78 samples in total). The switch to 4°C from 20°C was initiated at dusk. In a temperature reduction, the cabinet used here typically takes 2 h to reach 4°C air temperature. To analyze the speed of change in expression and AS, three biological replicates each consisting of 9–13 of 5-week-old rosettes were harvested just before reducing the temperature (0 min, 20°C) and after 90 min (5°C) and 120 and 180 min (4°C). Tissue was rapidly frozen in liquid N<sub>2</sub> and stored at –80°C until isolation of RNA and preparation of cDNA.

### RNA Extraction, Library Preparation, and Sequencing

Total RNA was extracted from Arabidopsis tissue using the RNeasy Plant Mini Kit (Qiagen), followed by either on-column DNase treatment (for HR RT-PCR, see below), or the TURBO DNA-free™ Kit (Ambion) (for library preparation and RT-qPCR, see below). RNA-seq libraries were constructed by following instructions for a TruSeq RNA library preparation (Illumina protocol 15026495 Rev. B). In these preparations, polyA+ selection was used to enrich for mRNA, RNA was fragmented for 8 min at 94°C, and random hexamers were used for first-strand cDNA synthesis. The libraries had an average insert size of approximately 280 bp and each library was sequenced on Illumina HiSeq 2500 platform generating 100 bp paired-end reads. Residual adaptor sequences at both 5′ and 3′ ends were removed from raw reads using cutadapt version 1.4.2<sup>1</sup> with quality score threshold set at 20 and minimum length of the trimmed read kept at 20. The “–paired-output” option was used to keep the two paired read files synchronized and avoid unpaired reads. The sequencing files before and after the trimming were examined using fastQC version 0.10.0.

### Quantification of Transcripts and AS

Arabidopsis transcript expression from the RNA-seq data was carried out using Salmon version 0.82 (Patro et al., 2017) in conjunction with AtRTD2-QUASI augmented by eight genes that were not originally present (Zhang et al., 2017). For indexing, we used the quasi-mapping mode to build an auxiliary k-mer hash over k-mers of length 31 (–type quasi –k 31). For quantification, the option to correct for the sequence specific bias (“–seqBias”) was used. The number of bootstraps was set to 30 and all other parameters were on default settings. Transcript expression results are provided in Calixto et al. (2018) and expression profiles in the time-series data are available at [https://wyguo.shinyapps.io/atrt2d2\\_profile\\_app/](https://wyguo.shinyapps.io/atrt2d2_profile_app/).

### Differential Gene Expression (DE) and AS (DAS) Analysis of the RNA-seq Data

The pre-processing of the read data and DE and differential AS analyses are described in detail in Calixto et al. (2018). Briefly,

<sup>1</sup><https://pypi.python.org/pypi/cutadapt/1.4.2>



transcript and gene read counts were generated from transcripts per million (TPM) data and low expressed transcripts that did not have  $\geq 1$  counts per million (CPM) in three or more samples were removed. At the gene level, if any transcript passed the expression level filtering step, the gene was included as an expressed gene and then the normalization factor was estimated using the weighted trimmed mean of  $M$  values method using edgeR version 3.12.1 (Robinson et al., 2010). Batch effects between biological replicates were estimated using RUVSeq R package version 1.4.0 with the residual RUVr approach (Risso et al., 2014). Normalized read counts in CPM were then log<sub>2</sub> transformed and mean-variance trends were estimated and weights of variance adjustments were generated using the voom function in limma version 3.26.9 (Law et al., 2014, 2016; Ritchie et al., 2015). General linear models to determine DE at both gene and transcript levels were established using time and batch effects of biological replicates as factors and 18 contrast groups were set up where corresponding time-points in the day 1 and day 4 at 4°C blocks were compared to those of the 20°C block (e.g., block2.T1 vs block1.T1, block2.T2 vs block1.T2; Calixto et al., 2018). Genes were significantly DE at the gene level if they had at least two contrast groups at consecutive time-points with adjusted  $p < 0.01$  and greater than twofold change in expression in each contrast group. Genes with significant DAS had at least two consecutive contrast groups with adjusted  $p < 0.01$  and with these contrast groups having at least one transcript with  $\geq 10\%$  change in expression.

## Identification of lncRNA Genes

For accurate DE and differential AS analyses of lncRNAs, only those genes in AtRTD2 were selected. AtRTD2 is a new Arabidopsis transcriptome containing over 82k unique transcripts and thereby far greater diversity of alternatively spliced isoforms than TAIR10 and Araport11 (Zhang et al., 2017). lncRNA gene lists from Liu et al. (2012), PLncDB:Plant Long non-coding RNA Database, lncRNAs from TAIR, and Araport and antisense lncRNAs from Araport were compared to AtRTD2 giving 379 lncRNA genes.

## Quantitative Reverse Transcription RT-PCR (RT-qPCR)

Real-time RT-PCR was performed essentially as described previously (James et al., 2012). Complementary DNA (cDNA) was synthesized from 2  $\mu$ g of total RNA using oligo dT primers and SuperScriptII reverse transcriptase (ThermoFisher Scientific). Each reaction (1:100 dilution of cDNA) was performed with Brilliant III SYBR Green QPCR Master Mix (Agilent) on a StepOnePlus (Fisher Scientific-UK Ltd., Loughborough, United Kingdom) real-time PCR system. The average Ct values for *IPP2* (AT3G02780) were used as internal control expression levels. The delta-delta Ct algorithm (Livak and Schmittgen, 2001; James et al., 2012) was used to determine relative changes in gene expression. Primers TAS1a-ex1-fwd 5'-CTAAGCGGCTAAGCCTGACGTCA-3' and TAS1a-ex2-ex1-rev 5'-CACCCATTACAAGCCTTTCTATCAGACAAGAC-3' targeted spliced *TAS1a* transcripts where the latter primer bridged the spliced intron (between exonic nucleotides

underlined in the primer sequence). Primers amplifying total *TAS1a* transcripts comprised the aforementioned TAS1a-ex1-fwd primer in combination with primer TAS1a-ex1-rev 5'-CAGACAAGACCATGACTCGATCTAAAGGC-3'.

## High-Resolution (HR) RT-PCR

High-resolution (HR) RT-PCR reactions were conducted as described previously (Simpson et al., 2008). Gene-specific primer pairs (**Supplementary Table S3**) were used for analyzing the expression and AS of different genes. For each primer pair, the forward primer was labeled with 6-carboxyfluorescein (FAM). cDNA was synthesized from 4  $\mu$ g of total RNA using the Sprint RT Complete – Double PrePrimed Kit following manufacturer's instructions (Clontech Laboratories, Takara Bio Company, United States). The PCR reaction usually contained 3  $\mu$ L of diluted cDNA (1:10) as a template, 0.1  $\mu$ L of each of the forward and reverse primers (100 mM), 2  $\mu$ L of 10 X PCR Buffer, 0.2  $\mu$ L of Taq Polymerase (5 U/ $\mu$ L, Roche), 1  $\mu$ L of 10 mM dNTPs (Invitrogen, Life Technologies), and RNase-free water (Qiagen) up to a final volume of 20  $\mu$ L. For each reaction, an initial step at 94°C for 2 min was used followed by 24–26 cycles of (1) denaturation at 94°C for 15 s, (2) annealing at 50°C for 30 s, and (3) elongation at 70°C for either 1 min (for fragments smaller than 1000 bp) or 1.5 min (for fragments between 1000 and 1200 bp) and a final extension cycle of 10 min at 70°C. To separate the RT-PCR products, 1.5  $\mu$ L of PCR product was mixed with 8.5  $\mu$ L of Hi-Di<sup>TM</sup> formamide (Applied Biosystems) and 0.01  $\mu$ L of GeneScan<sup>TM</sup> 500 LIZ<sup>TM</sup> dye or 0.04  $\mu$ L of GeneScan<sup>TM</sup> 1200 LIZ<sup>TM</sup> dye size standard and run on a 48-capillary ABI 3730 DNA Analyser (Applied Biosystems, Life Technologies). PCR products were separated to single base-pair resolution and the intensity of fluorescence was measured and used for quantification in relative fluorescent units (RFUs). The different PCR products and their peak levels of expression were calculated using the Genemapper<sup>®</sup> software (Applied Biosystems, Life Technologies).

## Small RNA Extraction and Detection of miRNAs and siRNAs

Total RNA was extracted from rosette material using TRI Reagent reagents (Sigma-Aldrich, United States) according to the manufacturer's instructions. For RNA gel blot hybridization of small RNAs 10  $\mu$ g of total RNA was separated on a 15% polyacrylamide (19:1) gel with 8 M urea and 1  $\times$  MOPS (20 mM MOPS/NaOH, pH7) buffer. RNA markers (Decade RNA markers, Ambion, United States) were end-labeled by [<sup>32</sup>P] $\gamma$ -ATP according to the manufacturer's instructions. RNA was blotted onto Hybond-N membrane (Amersham, GE Healthcare, United Kingdom) using a Panther<sup>TM</sup> Semi-dry Electrobloetter, HEP-1 (Thermo Scientific Owl Separation Systems, United States) and cross-linked by *N*-(3-dimethylaminopropyl)-*N'*-ethylcarbodiimide hydrochloride (EDC, Sigma-Aldrich, United States; Pall et al., 2007). DNA oligos (20 pmol) were end-labeled by [<sup>32</sup>P] $\gamma$ -ATP using T4 polynucleotide kinase (NEB, United States) to visualize small RNAs. Hybridization was performed in PerfectHyb<sup>TM</sup> Plus

Hybridization Buffer (Sigma–Aldrich, United States). After overnight incubation at 37°C, the membrane was washed twice in 2 X SSC and 0.1% SDS for 15 min at 37°C. After washing, signals were detected by phosphorimager plate visualized by FLA-7000 Fluorescent Image Analyzing System (Fujifilm, United States). The scanned images were quantified by AIDA Image Analyzer software (Fujifilm, United States). The same membrane was re-hybridized with the different probes; it was stripped after each hybridization using 0.1 × SSC, 0.1% SDS at 65°C for 30 min and the efficiency was checked by overnight exposure. Student's *t*-test ( $p < 0.05$ ) was used to identify differentially expressed small RNAs.

## RESULTS

### Identification of Cold-Induced Changes in Expression and Alternative Splicing

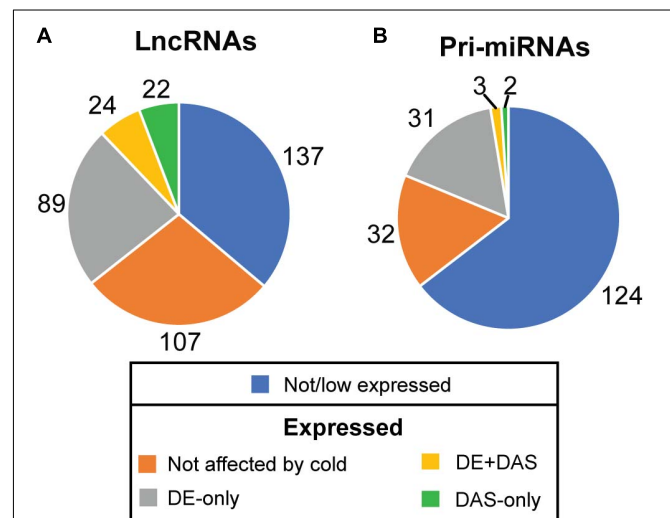
To examine changes in gene expression and AS in response to low temperature, we previously performed deep RNA-seq on a diel time-series of 5-week-old Arabidopsis Col-0 rosettes grown at 20°C and transferred to 4°C (Supplementary Figure S1A; Calixto et al., 2018). Briefly, rosettes were sampled at 3 h intervals for the last day at 20°C, the first day at 4°C, and the fourth day at 4°C (Supplementary Figure S1A) and each time-point consisted of three biological replicates. Over 360 million paired-end reads were generated for each of the 26 time-points and transcript abundances were quantified using Salmon (Patro et al., 2017) and AtRTD2-QUASI as the reference transcriptome (Zhang et al., 2017). The time-series data was analyzed at both the gene and individual transcript levels to identify genes with significant DE and significant differential alternative splicing (DAS). Briefly, this was achieved by generating read counts data using tximport (Soneson et al., 2015), normalizing data across samples with edgeR (Robinson et al., 2010), transforming using the voom function in limma (Law et al., 2014; Ritchie et al., 2015; Law et al., 2016), and establishing contrast groups in limma (for details see section “Materials and Methods” and Calixto et al., 2018). The experimental design allowed direct comparisons between equivalent time-points at 20°C and those in day 1 or day 4 at 4°C. This controlled for any effects of time-of-day variation in expression so that the changes detected were due to reduction in temperature. The statistical criteria for a DE gene were that it must have a log2-fold change  $\geq 1$  ( $\geq 2$ -fold change) in expression in at least two consecutive contrast groups with an adjusted *p*-value of  $<0.01$ . To detect DAS genes, the consistency of expression changes between the total expression of the gene and individual transcripts was examined using *F*-tests. For a gene to be significantly differentially alternatively spliced, a log2-fold change value of at least one of the transcripts must differ significantly from the gene log2-fold change value with an adjusted *p*-value of  $<0.01$ , and show a  $\Delta$  percent spliced ( $\Delta$ PS) of  $\geq 0.1$  in at least two consecutive contrast groups (for details, see Calixto et al., 2018). Using these stringent criteria, we identified a total of 7302 DE genes and 2442 DAS genes whose expression was significantly differentially expressed or alternatively spliced, respectively. The overlap between DE and

DAS genes was 795 genes being both significantly DE and DAS (regulated by both transcription and AS) in response to low temperature (Supplementary Figure S1B; Calixto et al., 2018).

### Cold-Induced Expression and AS of lncRNAs

The RNA-seq time-series data was analyzed using AtRTD2, a new transcriptome dataset for Arabidopsis (Zhang et al., 2017). DE and DAS analyses were therefore applied to the 379 lncRNA genes in AtRTD2. In general, for the majority of these lncRNA genes, AtRTD2 contained novel alternatively spliced transcripts or extended some of the shorter/truncated transcript models currently in the TAIR10/Araport11 thereby providing increased diversity of AS isoforms of the lncRNA genes.

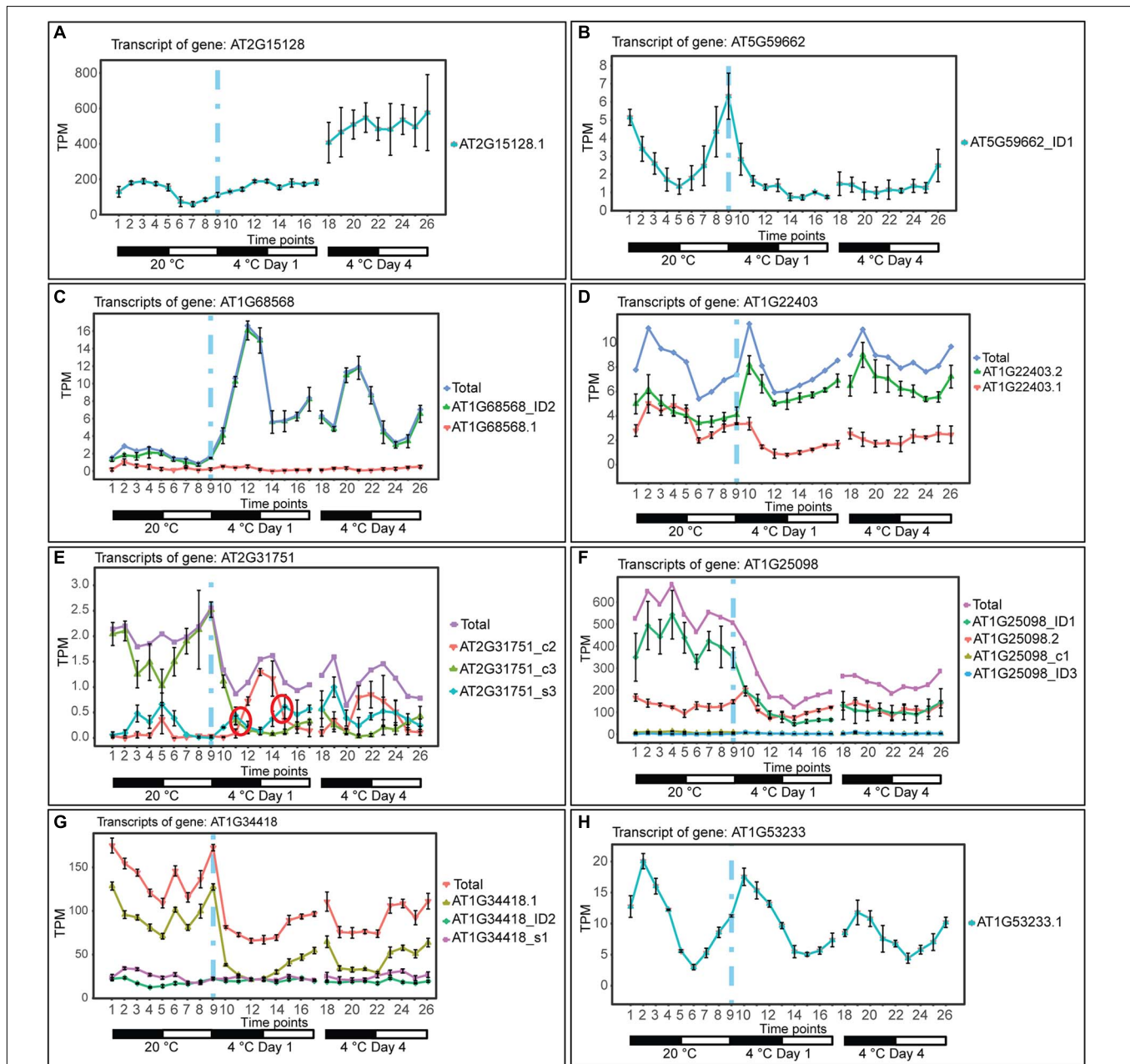
The depth of sequencing and the resolution of the RNA-seq time-course here allowed transcript-specific expression profiles of non-protein-coding genes to be analyzed. To examine the effect of low temperature on the expression and AS of lncRNAs, we searched the DE, DE+DAS, and DAS gene lists for the 379 lncRNA genes (see section “Materials and Methods”). Nearly a third of these genes (135) exhibited significant DE and/or AS in response to cold with 89 DE-only, 24 DE+DAS, and 22 DAS-only lncRNAs (Figure 1A and Supplementary Table S1A). Of these, 82 were NATs, eight encoded small open reading frames (sORFs; Hsu et al., 2016), two coded for tasiRNA precursors, and a third tasiRNA, TAS4, also encoded sORF27 (Supplementary Table S1B). The gene descriptions of the DE only, DE+DAS, and DAS only lncRNAs and of the genes overlapping the NAT lncRNAs are given in Supplementary Tables S1C–H. The expression profiles of DE lncRNAs showed a range of behaviors including cold-induced increase or decrease of expression, transient changes mainly seen in day 1 at 4°C, adaptive changes where the



**FIGURE 1** | Expression and alternative splicing of lncRNAs and pri-miRNAs in response to cold. The number of (A) lncRNA and (B) pri-miRNA genes in RTD2 (Zhang et al., 2017) according to their expression in the RNA-seq time course and identifying the number which are differentially regulated by cold at the expression (DE) and/or alternative splicing level (DAS).

change in expression after transfer to cold persisted throughout the cold treatment, and late expression where changes were mainly observed in day 4 at 4°C (**Figure 2**). For example, AT2G15128 (DE) showed an increase in expression in day 4

at 4°C (**Figure 2A**) while AT5G59662 (DE) showed rhythmic expression at 20°C and expression decreased dramatically within the first 6–12 h after lowering the temperature (**Figure 2B**). AT2G15128 had one of the highest levels of expression increasing



**FIGURE 2 |** Expression of lncRNA transcripts in response to cold. Transcripts below 1.85 TPM in all time-points are not shown (except in **E**, where we used a 0.5 TPM cut-off value). Transition to cold is represented by a vertical dashed blue line at time-point 9. **(A)** The single transcript of AT2G15128 is DE showing a significant upregulation upon long exposure to cold. **(B)** AT5G59662\_ID1 transcript is DE showing a significant downregulation upon cold and loss of a high amplitude rhythm is also observed. **(C)** The AT1G68568\_ID2 transcript is DE showing a significant upregulation throughout the cold treatment and a gain of a high amplitude rhythm is also observed. **(D)** AT1G22403 (DAS-only) undergoes splicing regulation rapidly in the cold and the relative abundance of the AT1G22403.2 transcript is maintained in day 4 at 4°C (adaptive) whereas the total gene level is not significantly affected by cold. **(E)** AT2G31751 and **(F)** AT1G25098 are DE and DAS showing a significant downregulation upon cold. **(G)** AT1G34418 (DAS-only) undergoes only significant splicing regulation showing a rapid decrease in abundance of AT1G34418.1 and increasing expression during the day. **(H)** Rhythmic expression of AT1G53233 (not affected by cold). Rapid and significant isoform switches detected by TSIS (Guo et al., 2017) are labeled with a red circle. NAT lncRNA genes (**B,C,E,F,H**); sORF (**G**); other RNA (**A,D**).



from around 200 TPM at 20°C to 450 TPM in day 4 at 4°C (**Figure 2A**). The DE gene, AT1G68568, had two transcripts which were expressed at relatively low levels at 20°C but the AT1G68568\_ID2 transcript increased immediately upon onset of cold and showed rhythmic expression with the maximal level of expression phased to around the middle of the dark period (**Figure 2C**). For DAS genes (which had no significant DE at the gene level), significant changes in AS again showed a variety of expression profiles at the level of transcripts. For example, AT1G22403 had two transcripts which were expressed at similar levels at 20°C but the AT1G22403.2 transcript increased with cold with a concomitant decrease in the levels of AT1G22403.1 (**Figure 2D**). AT2G31751 is a DE+DAS gene with three main transcripts of which AT2G31751\_c3 is the most abundant at 20°C peaking around dusk. At 4°C, its expression is drastically reduced and replaced by AT2G31750\_c2 which becomes the most abundant transcript and has an altered expression pattern, peaking at dawn (**Figure 2E**). Two other lncRNAs (AT1G25098 – DE+DAS and AT1G34418 – DAS-only – encoding the sORF15 lncRNA) had two and three main transcripts, respectively (**Figures 2F,G**). In both cases, the most abundant transcript at 20°C decreased rapidly after the start of the cold treatment. Some lncRNA expression profiles were clearly rhythmic, peaking at different times of the day. For example, although not differentially expressed or differentially alternatively spliced, AT1G53233 had clear rhythmic profile across cooling with maximal expression shortly after the onset of dark (**Figure 2H**). Interestingly, in some cases, rhythmic expression was either dampened or amplified in the cold (**Figures 2B,C**, respectively).

To investigate possible functions of the NAT lncRNAs, a GO enrichment analysis of the gene descriptions of the potential protein-coding targets of the 84 DE, DE+DAS, and DAS NATs was performed. The only enriched terms were biological process: flavonoid glucuronidation and molecular function: UDP-glycosyltransferase activity, quercetin 7-O-glucosyltransferase and quercetin 3-O-glucosyltransferase activity (FDR cut-off < 0.05). The time-course also allowed the identification of genes which showed the largest and quickest changes in expression or AS (Calixto et al., 2018). Thirteen of the NAT targets were in this group and included three UDP-glycosyltransferase and a flavone-3-hydroxylase (consistent with the GO analysis) and three transcription factors: AT1G69570 – *CDF5*, AT5G18240 – *MYB-RELATED PROTEIN 1* (*MYR1*), and AT5G15850 – *CONSTANS-LIKE 1* (*COL1*). We also identified AS in three of the sORF-encoding lncRNA transcripts (AT3G344184 – *sORF15*; AT3G26612 – *sORF28*, and AT5G24735 – *sORF31*). Translation of the AS isoforms of all three showed that AS did not affect the presence or integrity of the sORF (not shown).

## Cold-Induced Expression and AS of pri-miRNAs

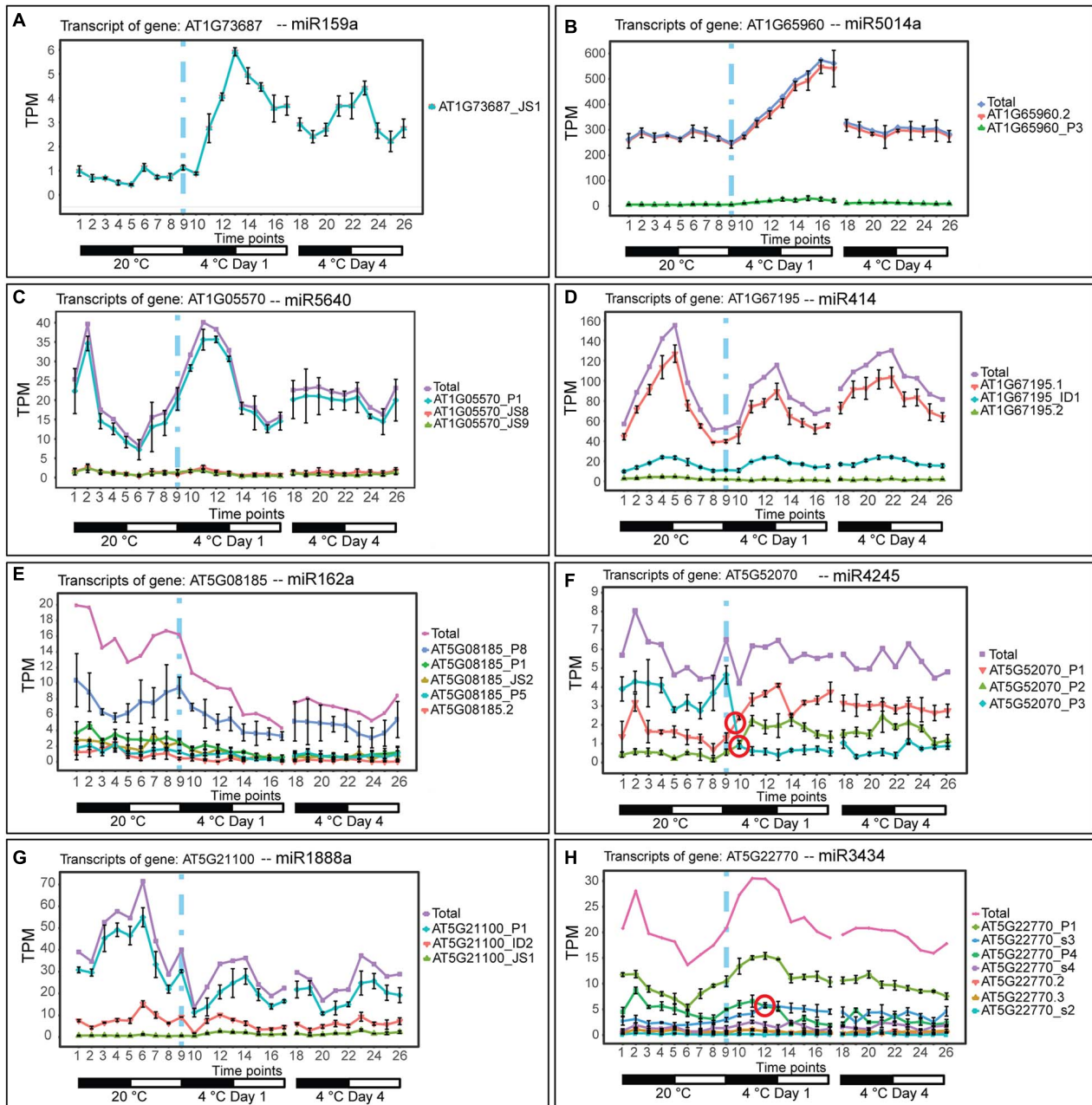
To investigate the effect of low temperature on the expression of pri-miRNAs, we identified the miRNA host genes that were DE-only, DE+DAS, and DAS-only. Using the miREx miRNA gene list (Bielewicz et al., 2012; Zielezinski et al., 2015), 192

pri-miRNA genes were present in AtRTD2 (**Supplementary Table S2A**). For the majority of these genes, AtRTD2 again contained novel alternatively spliced transcripts increasing the number of AS isoforms or extended some of the shorter/truncated transcript models currently in the TAIR10/Araport11. We detected expression of 68/192 pri-miRNA genes (**Supplementary Table S2B**). Of these, 31 were DE-only, three were DE+DAS, and two genes were DAS-only (**Figure 1B**). As with the lncRNAs, reducing the temperature caused increased or decreased expression at different time-points of the cold treatment. For example, the DE pri-miRNA gene, AT1G73687 (encoding miR159a) showed a concomitant rapid increase in expression and a diurnal waveform with a peak toward the end of the dark period in day 1 and day 4 at 4°C (**Figure 3A**). In contrast, the highly expressed pri-miRNA AT1G65960 (miR5014a) showed a transient increase in expression; levels doubled during Day 1 at 4°C but returned to the 20°C levels by day 4 at 4°C (**Figure 3B**). The main transcript of the DE pri-miRNA gene AT1G05570 (miR5640) showed rhythmic expression peaking in the night but appeared to lose rhythmicity by day 4 at 4°C (**Figure 3C**). Although not significantly affected by cold, the transcripts of pri-miRNA AT1G67195 (miR414) illustrate rhythmic expression (**Figure 3D**). Two DE+DAS pri-miRNA genes, AT5G08185 (miR162a) and AT5G21100 (miR1888a), showed reduction in the expression levels of their main transcripts with cold (**Figures 3E,G**). The DAS-only pri-miRNA gene, AT5G52070, encodes miR4245 (**Figure 3F**). At 20°C, AT5G52070\_P3 is the most abundant transcript but shows a rapid decrease after onset of cold with a concomitant increase in the other two transcripts, AT5G52070\_P1 and \_P2 with isoform switches in the first 3 h of cold (**Figure 3F**). Finally, the DAS-only pri-miRNA gene, AT5G22770 (miR3434), had three main transcripts; AT5G22770\_P1 and AT5G22770\_s3 increased their expression slightly with cold while AT5G22770\_P4 was decreased (**Figure 3H**). Of the 36 pri-miRNA genes with DE and/or differential AS, 13 were protein-coding genes containing an intronic miRNA (**Supplementary Table S2B**).

## Cold-Induced Alternative Splicing of *TAS1a*

*TAS1a* is an lncRNA that is regulated only by AS in response to cold (no significant DE at the gene level; **Figure 4B**). *TAS1a* RNA is initially targeted by miR173, converted to double-stranded RNA and subsequently cleaved into 21 nt phased small-interfering RNAs (phasiRNAs; Allen et al., 2005; Allen and Howell, 2010; **Figure 4A**). *TAS1a* and *TAS2* have both been reported to contain an intron (Vazquez et al., 2004; Yoshikawa et al., 2005). In AtRTD2, we confirmed the presence of the introns in *TAS1a* and *TAS2* and identified an intron in *TAS1c* which all contained the miRNA binding site and phasiRNAs (**Supplementary Figures S2–S5**). *TAS1a* produces two transcript variants: an unspliced transcript (AT2G27400.1) and a transcript where an intron is removed (AT2G27400\_ID1). The intron contains the entire region of *TAS1a* which has the miR173 binding site and tasiRNAs (**Figure 4A**). Splicing of the intron was confirmed by RT-PCR using primers in exons 1 and 2

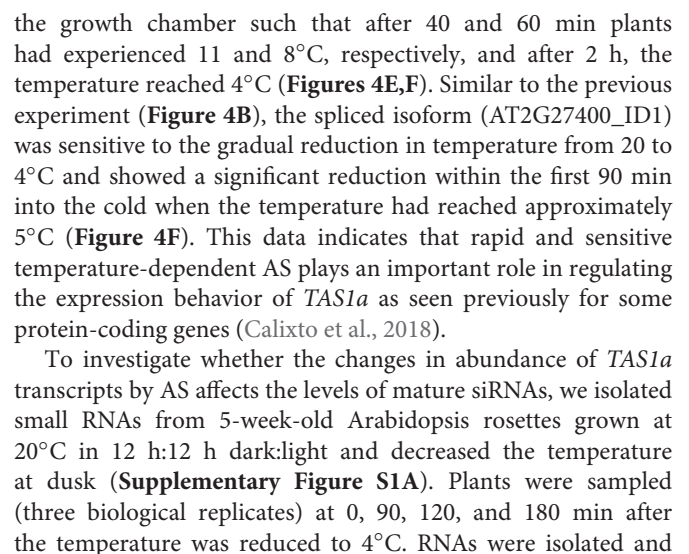




**FIGURE 3 |** Expression of pri-miRNA transcripts in response to cold. Transcripts below 1.85 TPM in all time-points are not shown (except in **H**, where we used a 1 TPM cut-off value). Transition to cold is represented by a vertical dashed blue line at time-point 9. **(A)** pri-miR159a is DE showing a significant upregulation upon cold and a gain of a high amplitude rhythm is also observed. **(B)** pri-miR5014a, is an intronic miRNA in the *GLUTAMATE DECARBOXYLASE 2* gene (AT1G65960); the profile of the pre-mRNA is DE showing a significant upregulation only in the first day of cold treatment. **(C)** pri-miR5640 is an intronic miRNA in *CALLOSE SYNTHASE 1* (AT1G05570) which is DE in the cold, while upon longer exposure to cold (day 4) rhythmicity is lost/dampened. **(D)** pri-miR414 is rhythmically expressed and not significantly affected by cold. **(E)** pri-miR162a is DE and DAS showing a significant downregulation throughout the cold treatment. **(F)** pri-miR4245 is intronic in an *AGENET DOMAIN-CONTAINING PROTEIN* gene (AT5G52070) which undergoes splicing regulation with no significant expression change at the gene level. **(G)** pri-miR1888a is an intronic miRNA in an *L-ASCORBATE OXIDASE* gene (AT5G21100) that is downregulated at the transcriptional level and undergoes differential alternative splicing. **(H)** pri-miR3434 is an intronic miRNA in an *ALPHA-ADAPTIN* gene (AT5G22770) which undergoes only splicing regulation. Rapid and significant isoform switches detected by TSIS (Guo et al., 2017) are labeled with a red circle.

(Supplementary Figure S2A). In RNA-seq data, both transcripts were expressed at similar levels at 20°C with higher levels of expression during the night (Figure 4B). Upon temperature

reduction, there was a rapid decrease of the spliced isoform (AT2G27400\_ID1, Figure 4B) in the first 6 h after the start of cold application, while the unspliced AT2G27400.1 increased in



separated on denaturing polyacrylamide gels, transferred to membrane, and hybridized with a series of different [ $^{32}$ P]-labeled oligonucleotides specific to miR173 and siRNAs: siRNA752, siRNA255, and siRNA477 (Figures 5A–D, respectively, and Supplementary Figure S6) and U6snRNA as control. The intensity of hybridization signals was quantified using AIDA Image Analyzer software using U6snRNA as a loading control. Over the short cold exposure treatments, there was a significant decrease in the abundance of the miR173 and the siRNAs over the first 2 h (Figure 5). Thus, that decreasing temperature causes a reduction in the abundance of siRNAs derived from *TAS1a*.

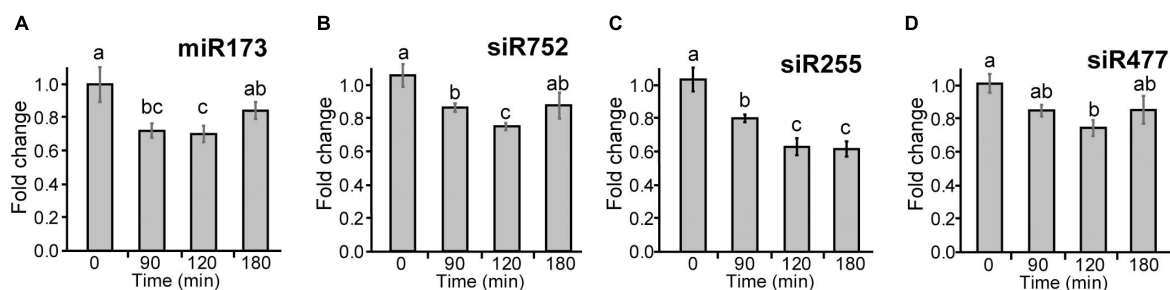
## Dynamic Expression and AS of lncRNAs and pri-miRNAs

Besides *TAS1a*, other lncRNAs and pri-miRNA transcripts also showed rapid changes in expression and AS. As for *TAS1a*, the speed and sensitivity of such changes was also investigated using RNA from leaf samples collected during the first 3 h of cold treatment (see Figure 4E). The RNA-seq expression profile of the two main transcript isoforms of the NAT lncRNA, AT1G34844 (Figure 6A), showed that AT1G34844\_ID2 decreased and AT1G34844.1 increased rapidly in the cold with an isoform switch in the first 3 h of cold (Figure 6C). In the expanded time-course covering the first 3 h when the temperature decreased gradually to 4°C, the AT1G34844.1 transcript (introns 2 and 3 retained) and AT1G34844\_ID2 (fully spliced, protein-coding) showed significant increase and decrease, respectively, after only 40 min (11°C; Figure 6E, right panel) in comparison to the constant 20°C control (Figure 6E, left panel). AT3G26612 lncRNA (encoding sORF28) had three AS isoforms (Figure 6B) where AT3G26612.1 and AT3G26612\_c1 switched in response to cold (Figure 6D). In the 3 h time-course, the two main transcripts of the AT3G26612 lncRNA (encoding sORF28) showed little difference in the 20°C control (Figure 6E, left panel) but significant changes in their abundance with temperature reduction. AT3G26612.1 decreased and AT3G26612\_c1 increased rapidly with the first significant differences being detected at 40 min when the temperature had reached only 11°C (Figure 6F).

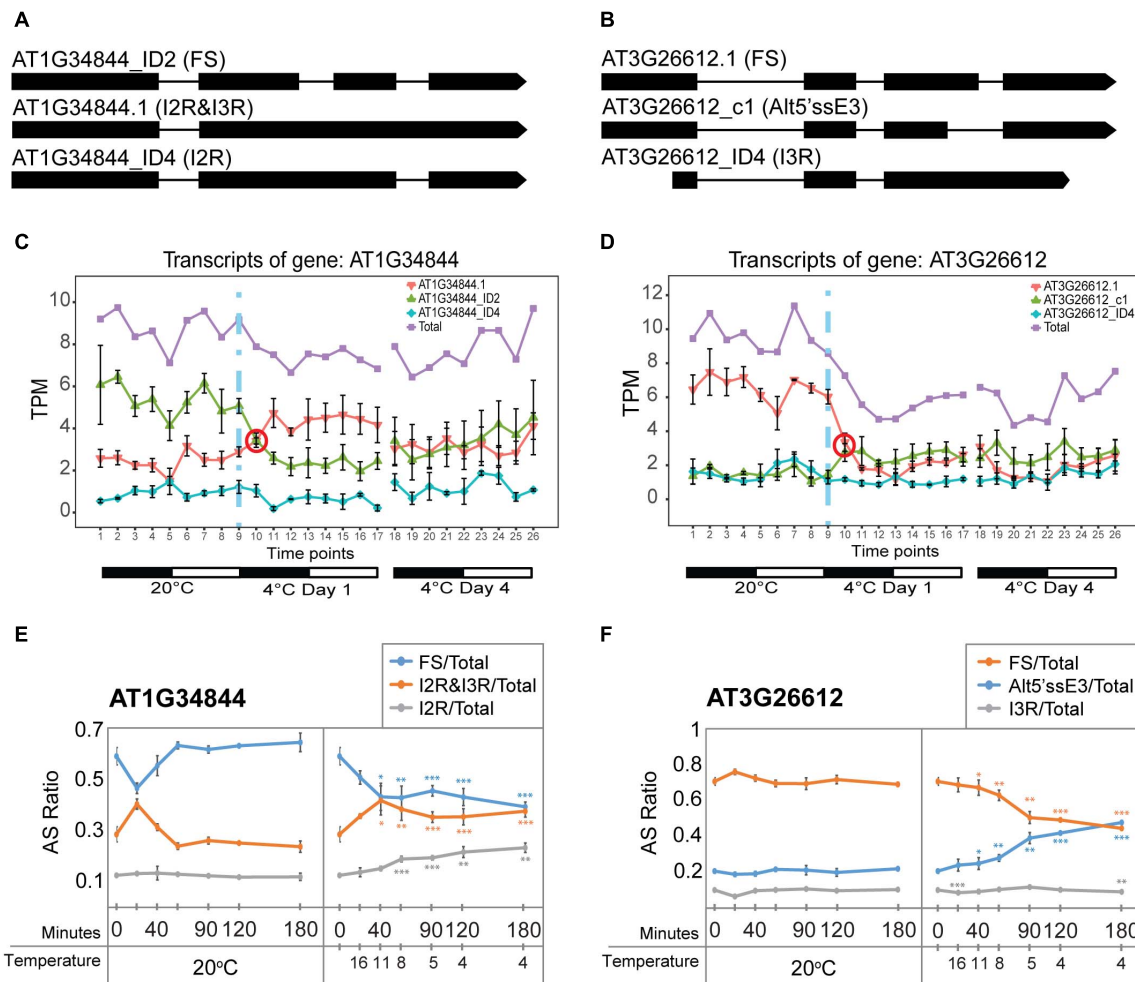
## DISCUSSION

Dynamic changes in expression at both the gene and transcript levels occur in response to lowering temperature in Arabidopsis (Calixto et al., 2018). AS makes a major contribution to changes in the transcriptome with over a quarter of genes whose expression changes significantly undergoing AS in the cold. The dynamic contribution of AS was illustrated by the rapid cold-induced wave of AS activity accompanying the transcriptional response and further AS changes throughout the period of cold exposure. The analysis of the RNA-seq cold response time-course at the transcript level identified hundreds of protein-coding genes which were only regulated by AS with no significant changes in expression at the gene level: the majority of these genes were novel cold-responsive genes (Calixto et al., 2018). Here, we have focussed on DE and differential AS of lncRNAs and pri-miRNAs. We identified 135 cold-responsive lncRNAs which were significantly differentially expressed and/or differentially alternatively spliced. Of these, a third involved changes in AS which have not been described previously. In particular, the transcript level expression analysis identified cold-responsive DAS-only lncRNAs which would not be detected by microarrays or gene level RNA-seq analyses. Different lncRNAs showed different responses to cold with, for example, transient or adaptive changes in expression/AS and the AS of some lncRNAs responded extremely rapidly and to small reductions in temperature showing that the AS of these genes is temperature-sensitive. Finally, we identified cold-induced AS of *TAS1a* and that the processing of siRNAs from the primary transcript may be splicing dependent. Therefore, cold-induced AS occurs in many lncRNAs/pri-miRNAs but there is little knowledge of the molecular mechanisms by which AS modulates levels of lncRNAs or miRNAs or their downstream biological functions.

Pre-mRNA processing such as splicing, AS, and alternative polyadenylation can affect gene expression levels by controlling mRNA export and stability, and the production of different functional variants (particularly at the protein level). The interactions between these typical pre-mRNA processing steps, environmental stress, and lncRNA production have been relatively well studied in Arabidopsis pri-miRNAs (Kruszka et al.,



**FIGURE 5 |** Cold-induced reduction of miR173 and *TAS1a*-derived siRNAs. Small RNA gel blot analysis was performed using RNA from rosette leaves harvested at different time-points after application of cold. Samples were hybridized with radioactively labeled probes specific to miR173 and the different siRNAs and bands were quantified. (A) miR173, (B) siR752, (C) siR255, and (D) siR477. The temperature at the different time-points was 20°C (0 min), 5°C (90 min), and 4°C (120 and 180 min). The letters on the bars indicate significantly different expression levels of small RNAs using Student's *t*-test ( $p < 0.05$ ); different letters mean significant differences between the samples.



**FIGURE 6 |** Rapid cold-induced alternative splicing of lncRNAs. **(A)** Transcript structures of AT1G34844 isoforms showing the fully spliced transcript (AT1G34844\_ID2), AT1G34844.1 which retains introns 2 and 3, and AT1G34844\_ID4 which retains intron 2. **(B)** Transcript structures of AT3G26612 isoforms showing the fully spliced transcript (AT3G26612.1), AT3G26612\_c1 which has an alternative 5' splice site in exon 3, and AT3G26612\_ID4 which retains intron 3. **(C)** AT1G34844 is regulated by alternative splicing only. AT1G34844\_ID2 and AT1G34844.1 show a transient decrease and increase, respectively, between 20°C and day 1 at 4°C while AT1G34844\_ID4 remains unchanged throughout the experiment. **(D)** AT3G26612 is regulated by alternative splicing only. AT3G26612.1 significantly decreases in the first 6 h (day 1) and throughout the cold treatment, while the other two transcripts have similar levels of expression, with AT3G26612\_c1 slightly increasing and AT3G26612\_ID4 remaining unchanged. **(E,F)** High-resolution RT-PCR analysis of splicing ratios at 20°C (control – left panels) and decreasing to 4°C (right panels). **(E)** The fully spliced and I2R&I3R transcripts of AT1G34844 respond rapidly to changes in temperature within 40 min when the temperature reaches 11°C and onward as the temperature decreases further. **(F)** The fully spliced and Alt5'ssE3 transcripts of AT3G26612 respond rapidly to changes in temperature within 40 min when the temperature reaches 11°C and onward as the temperature decreases further, while I3R is mostly unresponsive. Student's *t*-tests were performed to compare each temperature reduction results against 20°C control. Significant differences are labeled with asterisks (\*\**p* < 0.01; \*\*\**p* < 0.001).

2012; Yan et al., 2012; Bielewicz et al., 2013; Szweykowska-Kulińska et al., 2013; Barciszewska-Pacak et al., 2016; Knop et al., 2017; Stepień et al., 2017). Plant miRNAs are mostly exonic and encoded by independent transcription units of which > 50% contain introns (Szczepanowska et al., 2009). Both splicing of introns downstream of the miRNA and the integrity of proximal 5' splice sites have been shown to be required for efficient miRNA production (Bielewicz et al., 2013; Knop et al., 2017; Stepień et al., 2017). Mutations in various splicing factor genes including STA1 and the cold-induced GRP7 impact miRNA biogenesis from intron-containing pri-miRNAs (Ben Chaabane et al., 2013; Köster et al., 2014). GRP7, in particular, caused a reduction in the

levels of many miRNAs along with accumulation of pri-miRNAs; direct binding of GRP7 was required to inhibit pri-miRNA processing. GRP7 also directly affected the splicing/AS of two pri-miRNAs (Köster et al., 2014). Furthermore, in Arabidopsis, 29 miRNAs are encoded within introns in protein-coding or non-coding host genes (Brown et al., 2008; Yan et al., 2012) and miRNA biogenesis is affected by splicing/AS and alternative polyadenylation via various mechanisms. For example, AS can remove regions of the pre-miRNA affecting its ability to fold correctly and be processed, and conversely, pre-miRNA secondary structure can affect splice site choice (Brown et al., 2008; Barciszewska-Pacak et al., 2016; Stepień et al.,



2017). Production of miR400, an intron-located miRNA, was dependent on splice site choice: an AS event which removed part of the intron and left the miR400 in the mRNA of the host gene was induced by heat and led to an increase of the host gene mRNA containing the miRNA and a decrease in abundance of mature miR400 (Yan et al., 2012). The AS event effectively changed the position of the pre-miRNA from intronic to exonic (Yan et al., 2012; Szweykowska-Kulińska et al., 2013) suggesting that miRNA production was splicing-dependent and regulated by temperature. On the other hand, the intronic miRNA, miR402, is also induced by heat and correlates with selection of an intronic alternative polyadenylation site which competes with splicing and the miRNA is processed from the alternatively polyadenylated transcript (Knop et al., 2017). Thus, splicing and AS of pri-miRNAs impact the processing and levels of mature miRNAs and the AS of some pri-miRNAs is affected by cold.

Far less is known about AS in plant lncRNAs. Here, we exploited the increased number of transcript isoforms in ATRTD2 to identify novel AS events in Arabidopsis lncRNAs in the cold. In the analysis of the dynamics of AS in response to cold, we identified genes which showed the largest ( $\Delta PS > 0.25$ ) and quickest (0–6 h of cold) changes in cold-induced AS (Calixto et al., 2018). One of these genes encoded the tasiRNA lncRNA, *TAS1a*, which showed a rapid decrease in the level of the spliced isoform which correlated with a reduction in siRNAs. This suggested that splicing was required for the generation of the *TAS1a*-encoded siRNAs and was consistent with the earlier observation that a T-DNA insertion into the *TAS1a* intron led to greatly decreased siRNA production (Vazquez et al., 2004). Previously, reduced levels of *TAS1a*-derived siRNAs were observed in the cold while the levels of *TAS1a* were unaffected (Kume et al., 2010). Here, we also find reduced levels of the siRNAs in the cold but the transcript level analysis is able to distinguish the underlying AS. *TAS1c* and *TAS2* also have introns that contain the miRNA binding site and siRNAs but while the *TAS2* intron is efficiently spliced, *TAS1c* is spliced at very low frequency (Supplementary Figures S2–S5). The different organization of the TAS RNAs suggests that splicing may be important for efficient production of siRNAs for some genes but what determines why some genes undergo AS and the balance between splicing and processing are unknown.

The functions of the majority of Arabidopsis lncRNAs and, in particular, those with AS are still to be determined. The levels of *TAS1a*-derived siRNAs are affected in different environmental conditions and lower levels are detected in salt, drought, and cold stress (Sunkar and Zhu, 2004; Kume et al., 2010). The siRNAs derived from *TAS1a* have five gene targets of largely unknown function. However, the *TAS1a* siRNA target genes *AT1G51670*, *AT5G18040*, and *AT4G29760* are upregulated at 4°C with the increase in *AT1G51670* due to reduced silencing by tasiRNAs (Kume et al., 2010). Two of the targets, *AT4G29770* and *AT5G18040* (*HEAT-INDUCED TAS1 TARGET1* and 2 – *HTT1* and *HTT2*), are upregulated in heat stress and mediate thermotolerance (Li et al., 2014). Plants over-expressing *TAS1a* produced higher levels of *TAS1a* siRNAs, reduced expression of *HTT1* and *HTT2*, and showed weaker thermotolerance than wild-type (Li et al., 2014). Here, we show cold-induced

reduction in *TAS1a*-derived siRNAs and significant upregulation of *HTT2*. Thus, *TAS1a*-derived siRNAs may modulate expression of genes regulating the response of plants to both increased and decreased temperatures.

Other lncRNAs include natural antisense RNAs which generate dsRNA for silencing their target mRNAs, or lncRNAs which interact with protein splicing factors or chromatin modification factors to affect AS and expression of downstream genes (Romero-Barrios et al., 2018). We have identified many cold-responsive NAT lncRNAs with altered expression/AS and around half of the protein-coding genes that overlap these NATs were also differentially expressed and/or differentially alternatively spliced suggesting that the regulation of at least some protein-coding targets involve NAT lncRNAs. The GO enrichment analysis of the NAT targets of the DE/DAS lncRNAs identified genes involved in biosynthesis of flavonoids, secondary metabolites with a range of functions involved in growth, physiology, detoxification, and possibly acting as scavengers of reactive oxygen species (Pourcel et al., 2007). Some of these genes also had the earliest detectable DE which may reflect the upregulation of many genes involved in tolerance to reactive oxygen species early in response to chilling temperatures (Knight and Knight, 2012). In addition, two transcription factor NAT targets with the largest and most rapid changes in expression were involved in flowering control: *CDF5* and *MYR1*. The lncRNA, *FLORE*, is a natural antisense RNA which oscillates, has antiphasic expression to its *CDF5* mRNA target, and regulates flowering time (Henriques et al., 2017). Interestingly, *FLORE* is alternatively spliced into four isoform variants but the role of this AS is unknown. *MYR1* is a G2-like protein containing an N-terminal MYB-like domain, a central coiled-coil domain, and a C-terminal transactivation domain. *MYR1* is a negative regulator of flowering time under low light (Zhao and Beers, 2013). It has conserved alternative 3' splice sites events which affect a highly conserved sequence in the coiled coil domain and homo-dimerisation properties or interactions with other transcription factors (Zhao and Beers, 2013). The *MYR1* isoforms increase rapidly over the first 6 h of cold treatment and then decrease 12–15 h after onset of cold. The *MYR1* NAT (AT5G18245) peaks in the dark at 20°C and rapidly decreased in the first 3–6 h of cold. Thus, the *MYR1* NAT and *MYR1* AS transcripts appear to be anti-phasic and the increase of isoforms may produce transcription factor complexes which act to suppress flowering. The regulation of *CDF5* and *MYR1*, both involved in flowering, by NATs, parallels the recent report of the NAT lncRNA, *MAS*, which is induced by cold and required for activation of *MAF4* expression and suppression of precocious flowering (Zhao et al., 2018). Therefore, one function of the changes in expression and/or AS of NAT lncRNAs in response to cold is suppression of flowering. The expression profiles of some Arabidopsis pre-mRNAs and lncRNAs showed altered patterns of rhythmic expression in response to cold (Calixto et al., 2018; Figure 2). A third NAT-associated transcription factor with the largest and fastest changes in expression, *COL1* is involved in regulation of specific circadian rhythms (Ledger et al., 2001). Over-expression of *COL1* shortened the period of by 2–3 h.

Here, both the *COL1 NAT* and *COL1* showed a large transient increase in expression in day 1 at 4°C with the peak of expression advanced by around 3 h compared to 20°C. Thus, the increase in expression and altered timing of the NAT correlates with the cold response of *COL1* and may contribute to altered circadian control of expression of other genes in response to cold. Finally, although not included in this study, transcription of the lncRNA *SVALK1-asCBF1* which overlaps *CBF1* suppresses *CBF1* expression and impacts cold acclimation (Kindgren et al., 2018).

Plants experience constantly changing temperatures on hourly, daily, and seasonal scales and must have flexible regulatory systems to modulate gene expression. Gene regulation is complex involving the interplay between transcription and various post-transcriptional processes. AS is a major contributor to the changes in expression in the cold response (Calixto et al., 2018). In this paper, we demonstrate cold-induced changes in expression and AS of pri-miRNAs and lncRNAs and that AS of some lncRNA occurs very rapidly and are highly temperature sensitive. It is therefore likely that AS impacts the expression of target genes to contribute to both short term temperature responses and cold acclimation. A key question is the role of AS in regulating the processing, levels, and function of lncRNAs. The recent upsurge of interest in the role of intron retention in regulation of expression in plants and animals may provide possible explanations for splicing/AS of some lncRNAs. Intron retention can cause transcripts to be retained in the nucleus to wait for a splicing signal or be degraded while splicing can activate and enhance export of the transcripts to the cytoplasm where they can interact with target mRNAs to affect translational efficiency or be degraded (reviewed by Jacob and Smith, 2017). Thus, splicing/AS of lncRNAs may be a mechanism for regulating the levels of lncRNAs in the nucleus or cytoplasm or the stability of the lncRNAs which will impact the expression of target or downstream genes. Rapid changes in splicing factor activity, levels, or nuclear localization in response to cold can affect the efficiency of splicing/AS and determine whether different transcript isoforms are in the nucleus or cytoplasm affecting their processing or translation (e.g., sORF lncRNAs). Such mechanisms will require a thorough analysis of the transcript diversity of the thousands of plant lncRNAs to allow further studies on how the structure and

processing pathways of different types of lncRNAs influence their localization, stability, and function.

## DATA AVAILABILITY

Publicly available datasets were analyzed in this study. This data can be found here: <https://www.ebi.ac.uk/ena/data/search?query=prjeb19974>.

## AUTHOR CONTRIBUTIONS

JB, HN, and RZ obtained the funding support. JB and CC conceived and designed the experiments. CC, AJ, and NT collected and prepared samples. CC, AJ, NT, and CH performed the experiments. JB, CC, NT, AJ, and CH interpreted the main findings. CC and JB drafted the manuscript. All authors engaged in discussions during the project and revised and approved the final manuscript.

## FUNDING

This work was supported by funding from the Biotechnology and Biological Sciences Research Council (BBSRC) (BB/K006568/1, BB/P009751/1, BB/N022807/1 to JB; BB/K006835/1 to HN) and the Scottish Government Rural and Environment Science and Analytical Services division (RESAS) (to JB and RZ).

## ACKNOWLEDGMENTS

We thank Janet Laird (University of Glasgow) for technical assistance. We wish to apologize to all the authors whose relevant work was not cited in this article due to space limitations.

## SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fpls.2019.00235/full#supplementary-material>

## REFERENCES

- Allen, E., and Howell, M. D. (2010). miRNAs in the biogenesis of trans-acting siRNAs in higher plants. *Semin. Cell Dev. Biol.* 21, 798–804. doi: 10.1016/j.semcdb.2010.03.008
- Allen, E., Xie, Z., Gustafson, A. M., and Carrington, J. C. (2005). microRNA-directed phasing during trans-acting siRNA biogenesis in plants. *Cell* 121, 207–221. doi: 10.1016/j.cell.2005.04.004
- Barciszewska-Pacak, M., Knop, K., Jarmolowski, A., and Szweykowska-Kulińska, Z. (2016). *Arabidopsis thaliana* microRNA162 level is posttranscriptionally regulated via splicing and polyadenylation site selection. *Acta Biochim. Polon.* 63, 811–816. doi: 10.18388/abp.2016\_1349
- Barciszewska-Pacak, M., Milanowska, K., Knop, K., Bielewicz, D., Nuc, P., Plewka, P., et al. (2015). *Arabidopsis* microRNA expression regulation in a wide range of abiotic stress responses. *Front. Plant Sci.* 6:410. doi: 10.3389/fpls.2015.00410
- Bardou, F., Ariel, F., Simpson, C. G., Romero-Barrios, N., Laporte, P., Balzerque, S., et al. (2014). Long noncoding RNA modulates alternative splicing regulators in *Arabidopsis*. *Dev. Cell* 30, 166–176. doi: 10.1016/j.devcel.2014.06.017
- Ben Amor, B., Wirth, S., Merchan, F., Laporte, P., D'aubenton-Carafa, Y., Hirsch, J., et al. (2009). Novel long non-protein coding RNAs involved in *Arabidopsis* differentiation and stress responses. *Genome Res.* 19, 57–69. doi: 10.1101/gr.080275.108
- Ben Chaabane, S., Liu, R., Chinnusamy, V., Kwon, Y., Park, J. H., Kim, S. Y., et al. (2013). STA1, an *Arabidopsis* pre-mRNA processing factor 6 homolog, is a new player involved in miRNA biogenesis. *Nucleic Acids Res.* 41, 1984–1997. doi: 10.1093/nar/gks1309
- Bielewicz, D., Dolata, J., Zielezinski, A., Alaba, S., Szarzynska, B., Szczesniak, M. W., et al. (2012). mirEX: a platform for comparative exploration of plant

- pri-miRNA expression data. *Nucleic Acids Res.* 40, D191–197. doi: 10.1093/nar/gkr878
- Bielewicz, D., Kalak, M., Kalyna, M., Windels, D., Barta, A., Vazquez, F., et al. (2013). Introns of plant pri-miRNAs enhance miRNA biogenesis. *EMBO Rep.* 14, 622–628. doi: 10.1038/embor.2013.62
- Brown, J. W., Marshall, D. F., and Echeverria, M. (2008). Intronic noncoding RNAs and splicing. *Trends Plant Sci.* 13, 335–342. doi: 10.1016/j.tplants.2008.04.010
- 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., 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
- Cech, T. R., and Steitz, J. A. (2014). The noncoding RNA revolution-trashing old rules to forge new ones. *Cell* 157, 77–94. doi: 10.1016/j.cell.2014.03.008
- Chen, M., Wang, C., Bao, H., Chen, H., and Wang, Y. (2016). Genome-wide identification and characterization of novel lncRNAs in *Populus* under nitrogen deficiency. *Mol. Genet. Genom.* 291, 1663–1680. doi: 10.1007/s00438-016-1210-3
- Deveson, I. W., Brunck, M. E., Blackburn, J., Tseng, E., Hon, T., Clark, T. A., et al. (2018). Universal alternative splicing of noncoding exons. *Cell Syst.* 6, 245.e245–255.e245. doi: 10.1016/j.cels.2017.12.005
- Di, C., Yuan, J., Wu, Y., Li, J., Lin, H., Hu, L., et al. (2014). Characterization of stress-responsive lncRNAs in *Arabidopsis thaliana* by integrating expression, epigenetic and structural features. *Plant J.* 80, 848–861. doi: 10.1111/tpj.12679
- Ding, J., Lu, Q., Ouyang, Y., Mao, H., Zhang, P., Yao, J., et al. (2012). A long noncoding RNA regulates photoperiod-sensitive male sterility, an essential component of hybrid rice. *Proc. Natl. Acad. Sci. U.S.A.* 109, 2654–2659. doi: 10.1073/pnas.1121374109
- 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
- Forestan, C., Aiese Cigliano, R., Farinati, S., Lunardon, A., Sanseverino, W., and Varotto, S. (2016). Stress-induced and epigenetic-mediated maize transcriptome regulation study by means of transcriptome reannotation and differential expression analysis. *Sci. Rep.* 6:30446. doi: 10.1038/srep30446
- Franco-Zorrilla, J. M., Valli, A., Todesco, M., Mateos, I., Puga, M. I., Rubio-Somoza, I., et al. (2007). Target mimicry provides a new mechanism for regulation of microRNA activity. *Nat. Genet.* 39, 1033–1037. doi: 10.1038/ng2079
- Fu, X. D., and Ares, M. J. (2014). Context-dependent control of alternative splicing by RNA-binding proteins. *Nat. Rev. Genet.* 15, 689–701. doi: 10.1038/nrg3778
- Guo, W., Calixto, C. P. G., Brown, J. W. S., and Zhang, R. (2017). TSIS: an R package to infer alternative splicing isoform switches for time-series data. *Bioinformatics* 33, 3308–3310. doi: 10.1093/bioinformatics/btx411
- Hao, Z., Fan, C., Cheng, T., Su, Y., Wei, Q., and Li, G. (2015). Genome-wide identification, characterization and evolutionary analysis of long intergenic noncoding RNAs in cucumber. *PLoS One* 10:e0121800. doi: 10.1371/journal.pone.0121800
- 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
- He, F., Liu, Q., Zheng, L., Cui, Y., Shen, Z., and Zheng, L. (2015). RNA-seq analysis of rice roots reveals the involvement of post-transcriptional regulation in response to cadmium stress. *Front. Plant Sci.* 6:1136. doi: 10.3389/fpls.2015.01136
- Henriques, R., Wang, H., Liu, J., Boix, M., Huang, L. F., and Chua, N. H. (2017). The antiphasic regulatory module comprising CDF5 and its antisense RNA FLORE links the circadian clock to photoperiodic flowering. *New Phytol.* 216, 854–867. doi: 10.1111/nph.14703
- Hsu, P. Y., Calviello, L., Wu, H. L., Li, F. W., Rothfels, C. J., Ohler, U., et al. (2016). Super-resolution ribosome profiling reveals unannotated translation events in *Arabidopsis*. *Proc. Natl. Acad. Sci. U.S.A.* 113, E7126–E7135. doi: 10.1073/pnas.1614788113
- Ietswaart, R., Wu, Z., and Dean, C. (2012). Flowering time control: another window to the connection between antisense RNA and chromatin. *Trends Genet.* 28, 445–453. doi: 10.1016/j.tig.2012.06.002
- 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
- Joshi, R. K., Megha, S., Basu, U., Rahman, M. H., and Kav, N. N. (2016). Genome wide identification and functional prediction of long non-coding rnas responsive to sclerotinia sclerotiorum infection in *Brassica napus*. *PLoS One* 11:e0158784. doi: 10.1371/journal.pone.0158784
- 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
- Khemka, N., Singh, V. K., Garg, R., and Jain, M. (2016). Genome-wide analysis of long intergenic non-coding RNAs in chickpea and their potential role in flower development. *Sci. Rep.* 6:33297. doi: 10.1038/srep33297
- Kim, E. D., and Sung, S. (2012). Long noncoding RNA: unveiling hidden layer of gene regulatory networks. *Trends Plant Sci.* 17, 16–21. doi: 10.1016/j.tplants.2011.10.008
- Kindgren, P., Ard, R., Ivanov, M., and Marquardt, S. (2018). Transcriptional read-through of the long non-coding RNA SVALKA governs plant cold acclimation. *Nat. Commun.* 9:4561. doi: 10.1038/s41467-018-07010-6
- Knight, M. R., and Knight, H. (2012). Low-temperature perception leading to gene expression and cold tolerance in higher plants. *New Phytol.* 195, 737–751. doi: 10.1111/j.1469-8137.2012.04239.x
- Knop, K., Stepien, A., Barciszewska-Pacak, M., Taube, M., Bielewicz, D., Michalak, M., et al. (2017). Active 5' splice sites regulate the biogenesis efficiency of *Arabidopsis* microRNAs derived from intron-containing genes. *Nucleic Acids Res.* 45, 2757–2775.
- Köster, T., Meyer, K., Weinholdt, C., Smith, L. M., Lummer, M., Speth, C., et al. (2014). Regulation of pri-miRNA processing by the hnRNP-like protein AtGRP7 in *Arabidopsis*. *Nucleic Acids Res.* 42, 9925–9936. doi: 10.1093/nar/gku716
- Kruszka, K., Pieczynski, M., Windels, D., Bielewicz, D., Jarmolowski, A., Szweykowska-Kulinska, Z., et al. (2012). Role of microRNAs and other sRNAs of plants in their changing environments. *J. Plant Physiol.* 169, 1664–1672. doi: 10.1016/j.jplph.2012.03.009
- Kume, K., Tsutsumi, K., and Saitoh, Y. (2010). TAS1 trans-acting siRNA targets are differentially regulated at low temperature, and TAS1 trans-acting siRNA mediates temperature-controlled Atlg51670 expression. *Biosci. Biotechnol. Biochem.* 74, 1435–1440. doi: 10.1271/bbb.100111
- Laloum, T., Martin, G., and Duque, P. (2017). Alternative splicing control of abiotic stress responses. *Trends Plant Sci.* 23, 140–150. doi: 10.1016/j.tplants.2017.09.019
- Law, C. W., Alhamdoosh, M., Su, S., Smyth, G. K., and Ritchie, M. E. (2016). RNA-seq analysis is easy as 1-2-3 with limma, Glimma and edgeR. *F1000 Res.* 5:ISCB Comm J-1408.
- Law, C. W., Chen, Y., Shi, W., and Smyth, G. K. (2014). voom: precision weights unlock linear model analysis tools for RNA-seq read counts. *Genome Biol.* 15:R29. doi: 10.1186/gb-2014-15-2-r29
- Ledger, S., Strayer, C., Ashton, F., Kay, S. A., and Putterill, J. (2001). Analysis of the function of two circadian-regulated CONSTANS-LIKE genes. *Plant J.* 26, 15–22. doi: 10.1046/j.1365-3113x.2001.01003.x
- 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
- 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, S., Liu, J., Liu, Z., Li, X., Wu, F., and He, Y. (2014). Heat-induced tas1 target1 mediates thermotolerance via heat stress transcription factor ala-directed pathways in *Arabidopsis*. *Plant Cell* 26, 1764–1780. doi: 10.1105/tpc.114.124883



- Li, S., Yu, X., Lei, N., Cheng, Z., Zhao, P., He, Y., et al. (2017). Genome-wide identification and functional prediction of cold and/or drought-responsive lncRNAs in cassava. *Sci. Rep.* 7:45981. doi: 10.1038/srep45981
- Liu, J., Jung, C., Xu, J., Wang, H., Deng, S., Bernard, L., et al. (2012). Genome-wide analysis uncovers regulation of long intergenic noncoding RNAs in *Arabidopsis*. *Plant Cell* 24, 4333–4345. doi: 10.1105/tpc.112.102855
- Liu, J., Wang, H., and Chua, N. H. (2015). Long noncoding RNA transcriptome of plants. *Plant Biotechnol. J.* 13, 319–328. doi: 10.1111/pbi.12336
- Livak, K. J., and Schmittgen, T. D. (2001). Analysis of relative gene expression data using real-time quantitative PCR and the 2(-Delta Delta C(T)) Method. *Methods* 25, 402–408. doi: 10.1006/meth.2001.1262
- Lu, X., Chen, X., Mu, M., Wang, J., Wang, X., Wang, D., et al. (2016). Genome-wide analysis of long noncoding rnas and their responses to drought stress in cotton (*Gossypium hirsutum* L.). 11:e0156723. doi: 10.1371/journal.pone.0156723
- Lu, Z., Xia, X., Jiang, B., Ma, K., Zhu, L., Wang, L., et al. (2017). Identification and characterization of novel lncRNAs in *Arabidopsis thaliana*. *Biochem. Biophys. Res. Commun.* 488, 348–354. doi: 10.1016/j.bbrc.2017.05.051
- 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
- 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
- Pall, G. S., Codony-Servat, C., Byrne, J., Ritchie, L., and Hamilton, A. (2007). Carbodiimide-mediated cross-linking of RNA to nylon membranes improves the detection of siRNA, miRNA and piRNA by northern blot. *Nucleic Acids Res.* 35:e60. doi: 10.1093/nar/gkm112
- 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
- Pourcel, L., Routaboul, J. M., Cheynier, V., Lepiniec, L., and Debeaujon, I. (2007). Flavonoid oxidation in plants: from biochemical properties to physiological functions. *Trends Plant Sci.* 12, 29–36. doi: 10.1016/j.tplants.2006.11.006
- Qin, T., Zhao, H., Cui, P., Albeshar, N., and Xiong, L. (2017). A nucleus-localized long non-coding rna enhances drought and salt stress tolerance. *Plant Physiol.* 175, 1321–1336. doi: 10.1104/pp.17.00574
- Risso, D., Nqai, J., Speed, T. P., and Dudoit, S. (2014). Normalization of RNA-seq data using factor analysis of control genes or samples. *Nat. Biotechnol.* 32, 896–902. doi: 10.1038/nbt.2931
- Ritchie, M. E., Phipson, B., Wu, D., Hu, Y., Law, C. W., Shi, W., et al. (2015). limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res.* 43, e47. doi: 10.1093/nar/gkv007
- 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
- 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
- Severing, E., Faino, L., Jamge, S., Busscher, M., Kuijter-Zhang, Y., Bellinzano, F., et al. (2018). *Arabidopsis thaliana* ambient temperature responsive lncRNAs. *BMC Plant Biol.* 18:145. doi: 10.1186/s12870-018-1362-x
- Shaw, P., and Brown, J. (2012). Nucleoli: composition, function, and dynamics. *Plant Physiol.* 158, 44–51. doi: 10.1104/pp.111.188052
- Shuai, P., Liang, D., Tang, S., Zhang, Z., Ye, C. Y., Su, Y., et al. (2014). Genome-wide identification and functional prediction of novel and drought-responsive lincRNAs in *Populus trichocarpa*. *J. Exp. Bot.* 65, 4975–4983. doi: 10.1093/jxb/eru256
- Shumayla, Sharma, S., Taneja, M., Tyagi, S., Singh, K., and Upadhyay, S. K. (2017). Survey of high throughput RNA-Seq data reveals potential roles for lncRNAs during development and stress response in bread wheat. *Front. Plant Sci.* 8:1019. doi: 10.3389/fpls.2017.01019
- 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.1365-313X.2007.03392.x
- Soneson, C., Love, M. I., and Robinson, M. D. (2015). Differential analyses for RNA-seq: transcript-level estimates improve gene-level inferences. *F1000 Res.* 4:1521. doi: 10.12688/f1000research.7563.1
- 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
- Stepien, A., Knop, K., Dolata, J., Taube, M., Bajczyk, M., Barciszewska-Pacak, M., et al. (2017). Posttranscriptional coordination of splicing and miRNA biogenesis in plants. *Wiley Interdiscipl. Rev. RNA* 8.\*p\* doi: 10.1002/wrna.1403
- Sunkar, R., and Zhu, J. K. (2004). Novel and stress-regulated microRNAs and other small RNAs from *Arabidopsis*. *Plant Cell* 16, 2001–2019. doi: 10.1105/tpc.104.022830
- Swiezewski, S., Liu, F., Magusin, A., and Dean, C. (2009). Cold-induced silencing by long antisense transcripts of an *Arabidopsis* Polycomb target. *Nature* 462, 799–802. doi: 10.1038/nature08618
- Szarzynska, B., Sobkowiak, L., Pant, B. D., Balazadeh, S., Scheible, W. R., Mueller-Roeber, B., et al. (2009). Gene structures and processing of *Arabidopsis thaliana* HYL1-dependent pri-miRNAs. *Nucleic Acids Res.* 37, 3083–3093. doi: 10.1093/nar/gkp189
- Szweykowska-Kulińska, Z., Jarmolowski, A., and Vazquez, F. (2013). The crosstalk between plant microRNA biogenesis factors and the spliceosome. *Plant Signal. Behav.* 8:e26955. doi: 10.4161/psb.26955
- Thomashow, M. F. (2010). Molecular basis of plant cold acclimation: insights gained from studying the CBF cold response pathway. *Plant Physiol.* 154, 571–577. doi: 10.1104/pp.110.161794
- Vazquez, F., Vaucheret, H., Rajagopalan, R., Lepers, C., Gascioli, V., Mallory, A. C., et al. (2004). Endogenous trans-acting siRNAs regulate the accumulation of *Arabidopsis* mRNAs. *Mol. Cell* 16, 69–79. doi: 10.1016/j.molcel.2004.09.028
- Wang, J., Meng, X., Dobrovolskaya, O. B., Orlov, Y. L., and Chen, M. (2017). Non-coding RNAs and their roles in stress response in plants. *Genom. Proteom. Bioinform.* 15, 301–312. doi: 10.1016/j.gpb.2017.01.007
- Wang, J., Yu, W., Yang, Y., Li, X., Chen, T., Liu, T., et al. (2015a). 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, T. Z., Liu, M., Zhao, M. G., Chen, R., and Zhang, W. H. (2015b). Identification and characterization of long non-coding RNAs involved in osmotic and salt stress in *Medicago truncatula* using genome-wide high-throughput sequencing. *BMC Plant Biol.* 15:131. doi: 10.1186/s12870-015-0530-5
- Wang, Y., Zou, W., Xiao, Y., Cheng, L., Liu, Y., Gao, S., et al. (2018). MicroRNA1917 targets CTR4 splice variants to regulate ethylene responses in tomato. *J. Exp. Bot.* 69, 1011–1025. doi: 10.1093/jxb/erx469
- Wunderlich, M., Gross-Hardt, R., and Schöffl, F. (2014). Heat shock factor HSF2a involved in gametophyte development of *Arabidopsis thaliana* and its expression is controlled by a heat-inducible long non-coding antisense RNA. *Plant Mol. Biol.* 85, 541–550. doi: 10.1007/s11103-014-0202-0
- Xin, M., Wang, Y., Yao, Y., Song, N., Hu, Z., Qin, D., et al. (2011). Identification and characterization of wheat long non-protein coding RNAs responsive to powdery mildew infection and heat stress by using microarray analysis and SBS sequencing. *BMC Plant Biol.* 11:61. doi: 10.1186/1471-2229-11-61
- Yan, K., Liu, P., Wu, C. A., Yang, G. D., Xu, R., Guo, Q. H., et al. (2012). Stress-induced alternative splicing provides a mechanism for the regulation of microRNA processing in *Arabidopsis thaliana*. *Mol. Cell* 48, 521–531. doi: 10.1016/j.molcel.2012.08.032
- Yoshikawa, M., Peragine, A., Park, M. Y., and Poethig, R. S. (2005). A pathway for the biogenesis of trans-acting siRNAs in *Arabidopsis*. *Genes Dev.* 19, 2164–2175. doi: 10.1101/gad.1352605
- Yuan, J., Zhang, Y., Dong, J., Sun, Y., Lim, B. L., Liu, D., et al. (2016). Systematic characterization of novel lncRNAs responding to phosphate starvation in *Arabidopsis thaliana*. *BMC Genomics* 17:655. doi: 10.1186/s12864-016-2929-2
- Zhang, H., Chen, X., Wang, C., Xu, Z., Wang, Y., Liu, X., et al. (2013). Long non-coding genes implicated in response to stripe rust pathogen stress in wheat (*Triticum aestivum* L.). *Mol. Biol. Rep.* 40, 6245–6253. doi: 10.1007/s11033-013-2736-7
- 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, W., Han, Z., Guo, Q., Liu, Y., Zheng, Y., Wu, F., et al. (2014). Identification of maize long non-coding RNAs responsive to drought stress. *PLoS One* 9:e98958. doi: 10.1371/journal.pone.0098958
- Zhang, Y. C., and Chen, Y. Q. (2013). Long noncoding RNAs: new regulators in plant development. *Biochem. Biophys. Res. Commun.* 436, 111–114. doi: 10.1016/j.bbrc.2013.05.086
- Zhao, C., and Beers, E. (2013). Alternative splicing of Myb-related genes MYR1 and MYR2 may modulate activities through changes in dimerization, localization, or protein folding. *Plant Signal. Behav.* 8:e27325. doi: 10.4161/psb.27325
- Zhao, J., He, Q., Chen, G., Wang, L., and Jin, B. (2016). Regulation of Non-coding RNAs in heat stress responses of plants. *Front. Plant Sci.* 7:1213. doi: 10.3389/fpls.2016.01213
- Zhao, X., Li, J., Lian, B., Gu, H., Li, Y., and Qi, Y. (2018). Global identification of *Arabidopsis* lncRNAs reveals the regulation of MAF4 by a natural antisense RNA. *Nat. Commun.* 9:5056. doi: 10.1038/s41467-018-07500-7
- Zhu, J. K. (2016). Abiotic stress signaling and responses in plants. *Cell* 167, 313–324. doi: 10.1016/j.cell.2016.08.029
- Zhu, Q. H., Stephen, S., Taylor, J., Helliwell, C. A., and Wang, M. B. (2014). Long noncoding RNAs responsive to *Fusarium oxysporum* infection in *Arabidopsis thaliana*. *New Phytol.* 201, 574–584. doi: 10.1111/nph.12537
- Zielezinski, A., Dolata, J., Alaba, S., Kruszka, K., Pacak, A., Swida-Barteczka, A., et al. (2015). mirEX 2.0 – An integrated environment for expression profiling of plant microRNAs. *BMC Plant Biol.* 15:144. doi: 10.1186/s12870-015-0533-2

**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 Calixto, Tzioutziou, James, Hornyik, Guo, Zhang, Nimmo and Brown. 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.



# Nuclear Speckle RNA Binding Proteins Remodel Alternative Splicing and the Non-coding Arabidopsis Transcriptome to Regulate a Cross-Talk Between Auxin and Immune Responses

Jérémy Bazin<sup>1\*</sup>, Natali Romero<sup>1</sup>, Richard Rigo<sup>1</sup>, Celine Charon<sup>1</sup>, Thomas Blein<sup>1</sup>, Federico Ariel<sup>1,2</sup> and Martin Crespi<sup>1\*</sup>

<sup>1</sup> CNRS, INRA, Institute of Plant Sciences Paris-Saclay IPS2, Univ Paris Sud, Univ Evry, Univ Paris-Diderot, Sorbonne Paris-Cite, Université Paris-Saclay, Orsay, France, <sup>2</sup> Instituto de Agrobiotecnología del Litoral, CONICET, Universidad Nacional del Litoral, Santa Fe, Argentina

## OPEN ACCESS

### Edited by:

Dorothee Staiger,  
Bielefeld University, Germany

### Reviewed by:

Rossana Henriques,  
University College Cork, Ireland  
Heike Lange,  
UPR2357 Institut de Biologie  
Moléculaire des Plantes (IBMP),  
CNRS, France

### \*Correspondence:

Jérémy Bazin  
jeremie.bazin@free.fr  
Martin Crespi  
martin.crespi@u-psud.fr

### Specialty section:

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

**Received:** 16 May 2018

**Accepted:** 27 July 2018

**Published:** 21 August 2018

### Citation:

Bazin J, Romero N, Rigo R,  
Charon C, Blein T, Ariel F and  
Crespi M (2018) Nuclear Speckle  
RNA Binding Proteins Remodel  
Alternative Splicing  
and the Non-coding Arabidopsis  
Transcriptome to Regulate  
a Cross-Talk Between Auxin  
and Immune Responses.  
Front. Plant Sci. 9:1209.  
doi: 10.3389/fpls.2018.01209

Nuclear speckle RNA binding proteins (NSRs) act as regulators of alternative splicing (AS) and auxin-regulated developmental processes such as lateral root formation in *Arabidopsis thaliana*. These proteins were shown to interact with specific alternatively spliced mRNA targets and at least with one structured lncRNA, named *Alternative Splicing Competitor* RNA. Here, we used genome-wide analysis of RNAseq to monitor the NSR global role on multiple tiers of gene expression, including RNA processing and AS. NSRs affect AS of 100s of genes as well as the abundance of lncRNAs particularly in response to auxin. Among them, the *FPA* floral regulator displayed alternative polyadenylation and differential expression of antisense *COOLAIR* lncRNAs in *nsra/b* mutants. This may explain the early flowering phenotype observed in *nsra* and *nsra/b* mutants. GO enrichment analysis of affected lines revealed a novel link of NSRs with the immune response pathway. A RIP-seq approach on an NSRa fusion protein in mutant background identified that lncRNAs are privileged direct targets of NSRs in addition to specific AS mRNAs. The interplay of lncRNAs and AS mRNAs in NSR-containing complexes may control the crosstalk between auxin and the immune response pathway.

**Keywords:** RNA binding proteins, RNP complexes, alternative splicing, immune response, auxin

## INTRODUCTION

RNA binding proteins (RBPs) have been shown to affect all steps of post-transcriptional gene expression control, including alternative splicing (AS), silencing, RNA decay, and translational control (Bailey-Serres et al., 2009). The *Arabidopsis thaliana* genome encodes for more than 200 proteins predicted to bind RNAs. The picture becomes even more complex since over 500 proteins were found to bind polyA<sup>+</sup> RNA in a recent study attempting to define the RNA interactome using affinity capture and proteomics (Marondedze et al., 2016). However, only a small subset of RBPs has been functionally assigned in plants. The versatility of RBPs on gene expression regulation has been recently highlighted by the identification of several among them acting at multiple steps of post-transcriptional gene regulation (Lee and Kang, 2016; Oliveira et al., 2017). During mRNA

maturation, the transcript acquires a complex of proteins at each exon–exon junction during pre-mRNA splicing that influences the subsequent steps of mRNA translation and decay (Maquat, 2004). Although all RBPs bind RNA, they exhibit different RNA-sequence specificities and affinities. As a result, cells are able to generate diverse ribonucleoprotein complexes (RNPs) whose composition is unique to each mRNA and these complexes are further remodeled during the life of the mRNA in order to determine its fate. One approach to determine RBP function consisted in the identification of all interacting molecules (the so-called RNPome) of a specific RNP and the conditions of their association. The ribonucleoprotein immunopurification assay facilitates the identification and quantitative comparison of RNA association to specific proteins under different experimental conditions. This approach has been successfully used to elucidate the genome-wide role of a number of plant RBPs involved in pre-mRNA splicing, stress granule formation or translational control (Sorenson and Bailey-Serres, 2014; Gagliardi and Matarazzo, 2016; Foley et al., 2017; Köster and Meyer, 2018).

The nuclear speckle RNA binding proteins (NSRs) are a family of RBPs that act as regulators of AS and auxin regulated developmental processes such as lateral root formation in *Arabidopsis thaliana*. These proteins were shown to interact with some of their alternatively spliced mRNA targets and at least with one structured lncRNA, named *Alternative Splicing Competitor RNA* (ASCO) (Bardou et al., 2014). Overexpression of ASCO was shown to affect AS of a subset of mRNA regulated by NSRs, similar to *nsra/b* double mutants, and ASCO was also shown to compete *in vitro* with the binding of one AS mRNA target. This study suggested that plant lncRNAs are able to modulate AS of mRNA by hijacking RBPs, such as NSRs, involved in splicing (Romero-Barrios et al., 2018). In addition, transcriptome analysis using microarrays and specific AS analysis on a subset of mRNAs suggested a role of NSR in transcriptome remodeling in response to auxin (Bardou et al., 2014).

Here we used genome wide analysis to monitor the NSR global role on multiple tiers of gene expression, including RNA processing and AS. This allowed us to find a new role of NSR in the control of flowering time regulators as well as to suggest that NSRs control the crosstalk between auxin and the immune response pathway.

## RESULTS

### Auxin Regulation of Gene Expression Is Altered in *nsra/b* Double Mutant

To characterize the role of NSRs in the control of auxin regulated gene expression, we performed paired-end strand specific RNA sequencing on the *nsra/nsrb* (*nsra/b*) double mutant and wild type (Col-0) seedlings treated for 24 h with the synthetic auxin NAA (100 nM) or a mock solution (Bardou et al., 2014; Tran et al., 2016) (Figure 1A).

In mock treated samples, 63 and 41 genes were found to be differentially up and down-regulated between mutant and wild type seedlings (Supplementary Table S1B). Remarkably, in response to auxin, we identified 709 and 465 genes

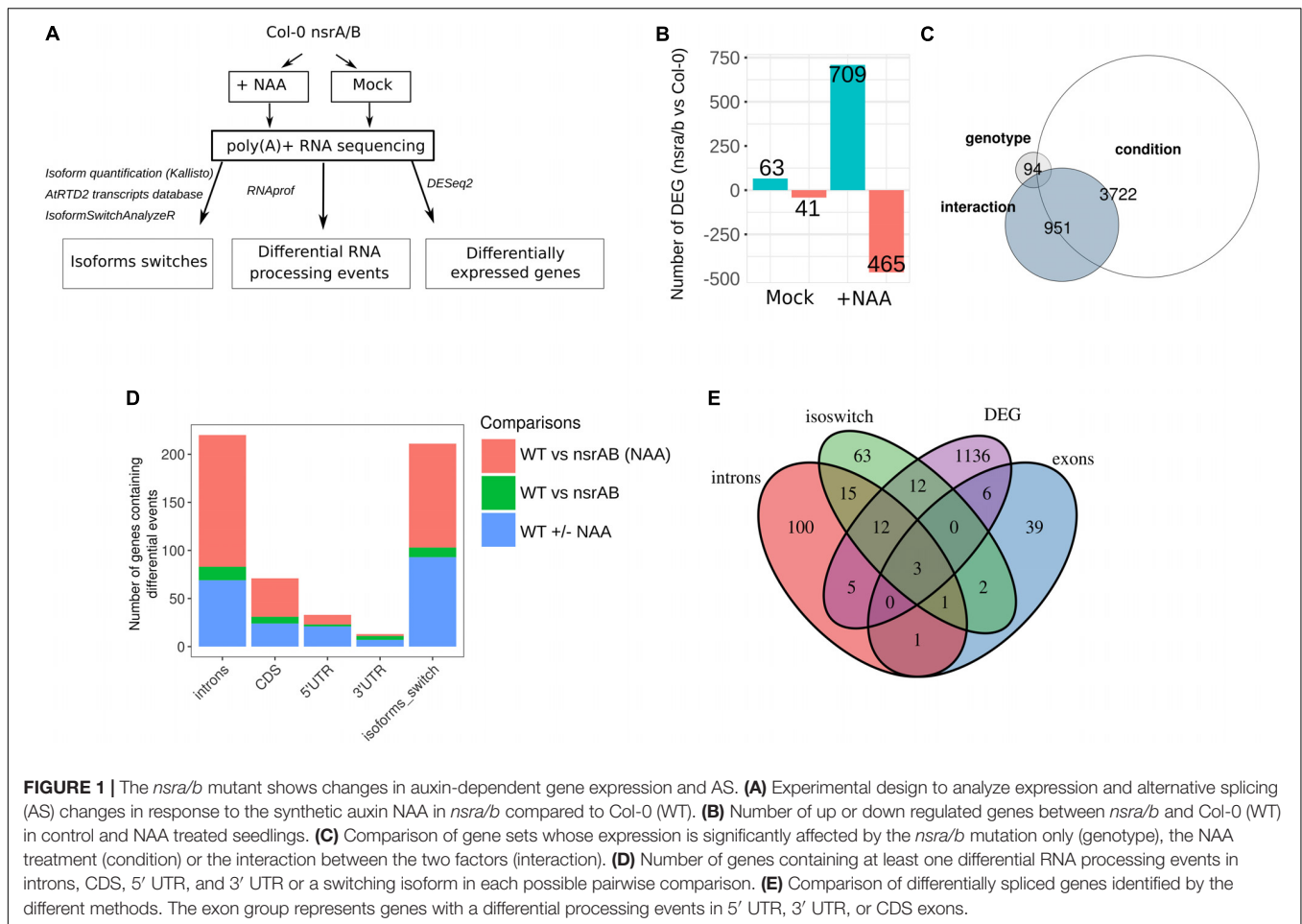
significantly up and down-regulated in *nsra/b*, compared to wild type (Figure 1B and Supplementary Table S1B). Principal component analysis (PCA) showed a dispersion of the data compatible with statistical comparisons between groups (Supplementary Figure S1). Multifactor analysis of differential gene expression further showed that *nsra/b* mutation has a major effect on auxin-regulated gene expression. Indeed, a set of 951 genes showed significant interaction between genotype and auxin regulation (Figure 1C and Supplementary Table S1B). This is in agreement with our previous findings indicating that NSRs mediate auxin regulation of gene expression (Bardou et al., 2014).

We have previously shown that NSRs modulate auxin-induced AS of a particular subset of genes using specific qRT-PCR assays (Bardou et al., 2014). We use now our RNA-seq dataset to characterize genome-wide effects of NSRs on AS and more generally on RNA processing (Figure 1A). To this end, we made use of the RNAprof software, which implements a gene-level normalization procedure and can compare RNA-seq read distributions on transcriptional units to detect significant profile differences. This approach allows *de novo* identification of RNA processing events independently of any gene feature or annotation independently of gene expression differences (Tran et al., 2016). RNAprof results were parsed to retain only highly significant differential RNA processing events ( $p_{\text{adj}} < 10e-4$ ) and further crossed with gene annotation in order to classify them according to their gene features. The majority of events overlapped with intronic regions (Figure 1D and Supplementary Table S1C), which is in accordance with data showing that intron retention is the major event of AS in plants (Ner-Gaon et al., 2004). The effect of *nsra/b* on RNA processing and splicing is enhanced in response to NAA. In other words the vast majority of differential events between *nsra/b* and wild type plants were identified essentially in presence of auxin.

To further support the results from RNAprof and to gain knowledge on the functional consequences of NSR mediated AS events, we quantified mRNA transcript isoforms of the AtRTD2 database (Brown et al., 2017) using kallisto (Bray et al., 2016). Then, we searched for marked changes in isoform usage using IsoformSwitchAnalyzeR package (Vitting-Seerup and Sandelin, 2017), which allows statistical detection and visualization and prediction of functional consequences of isoform switching events. As a result, we identified 118 NSR-dependent isoform switching events including 108 only detected in NAA-treated samples (Figure 1E and Supplementary Table S1D). Comparison of gene sets affected in their steady state abundance, containing differential RNA processing or isoforms switching events in *nsra/b* highlighted the fact that most differentially spliced genes are not differentially expressed. In addition, over 35% of genes predicted with isoforms switching events were also found using RNAprof (Figure 1E).

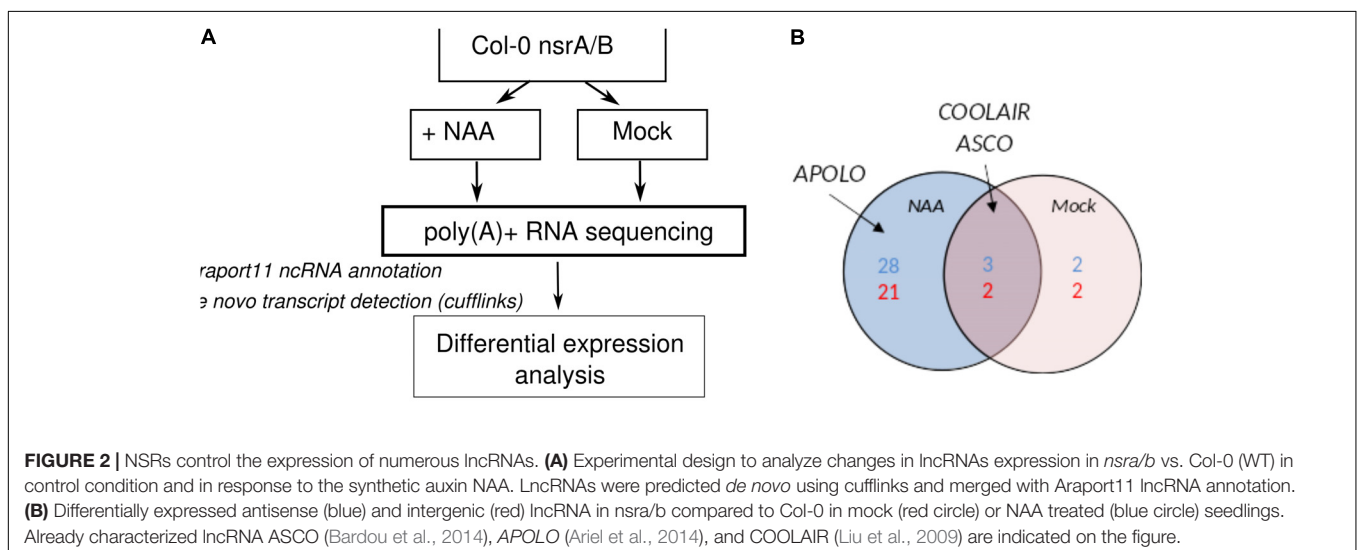
### NSRs Affect the Abundance of Numerous lncRNAs

The activity of NSR proteins on AS is modulated by the lncRNA ASCO and the abundance of ASCO RNA is increased in *nsra/b* mutant (Bardou et al., 2014). Therefore, we conducted a global



analysis of lncRNAs detection and expression in our RNA-seq datasets. Annotated lncRNA (Araport11) were combined with *de novo* predicted transcripts and further classified based on their location in intergenic and antisense regions of coding genes

(Figure 2A). More than 2440 lncRNAs were detected in our RNAseq data with more than 1 TPM (Supplementary Table S1A) in at least three samples. In mock conditions, differential expression analysis served to identify five antisense and four





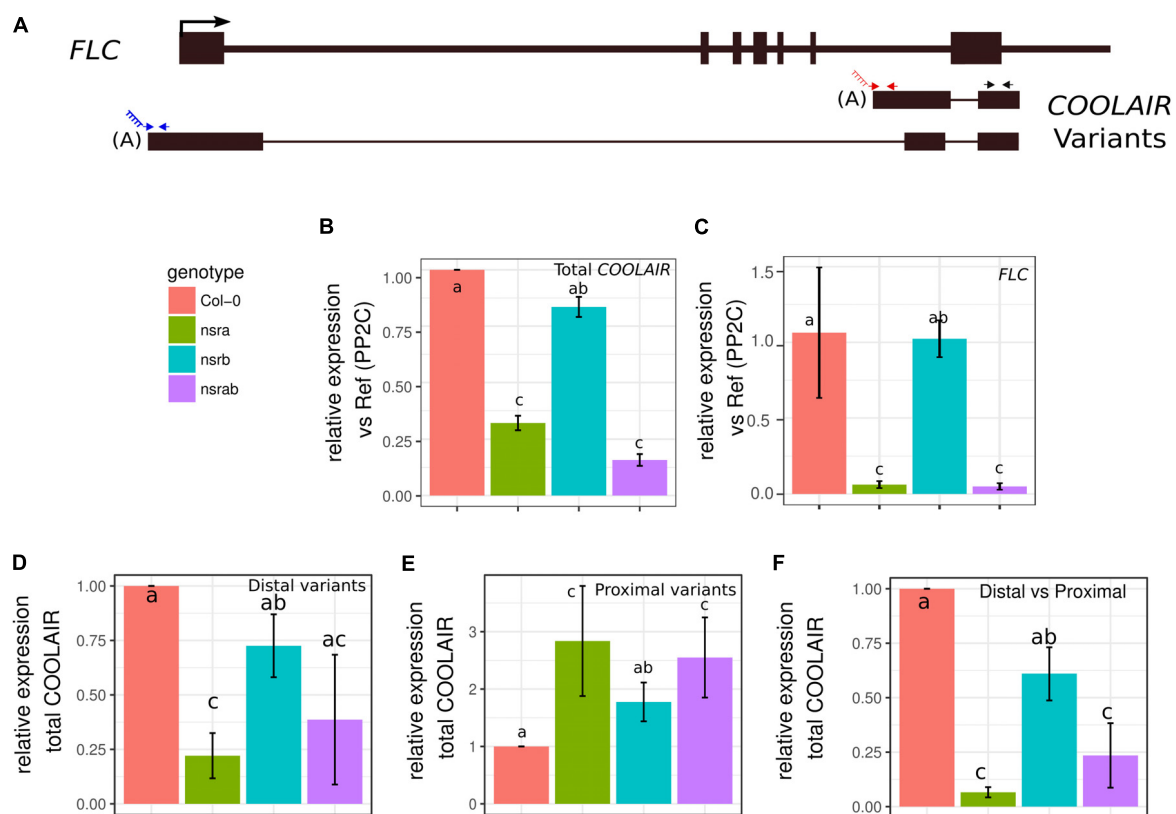
intergenic lncRNAs differentially expressed between mutant and wild type seedlings, whereas 31 intergenic and 23 antisense lncRNAs were found to be differentially regulated between mutant and wild type in the presence of auxin (**Figure 2B**). Differentially expressed lncRNAs included a number of well-characterized lncRNAs such as *APOLO*, which has been shown to influence root gravitropism in response to auxin via its action on PINOID protein kinase expression dynamics. In addition, the expression of lncRNA *ASCO*, shown to interact with NSR to modulate AS of its mRNA targets, was also affected in *nsra/b* suggesting a feedback regulation of NSR on *ASCO* lncRNA (**Figure 2B**).

### NSRa Is Involved in the Control Flowering Time Through the Modulation of the *COOLAIR*/*FLC* Module

Interestingly, we also identified the lncRNA *COOLAIR* as down regulated in *nsra/b*, both in mock or NAA treated samples (**Figure 2B**). *COOLAIR* designate a set of transcripts expressed in antisense orientation of the locus encoding the floral repressor

*FLC* (Whittaker and Dean, 2017). Two main classes of *COOLAIR* lncRNAs are produced by AS and polyadenylation of antisense transcripts generated from the *FLC* locus. One uses a proximal splice site and a polyadenylation site located in intron 6 of *FLC*, whereas the distal one results from the use of a distal splice and polyadenylation sites located in the *FLC* promoter (reviewed in Whittaker and Dean, 2017) (**Figure 3A**).

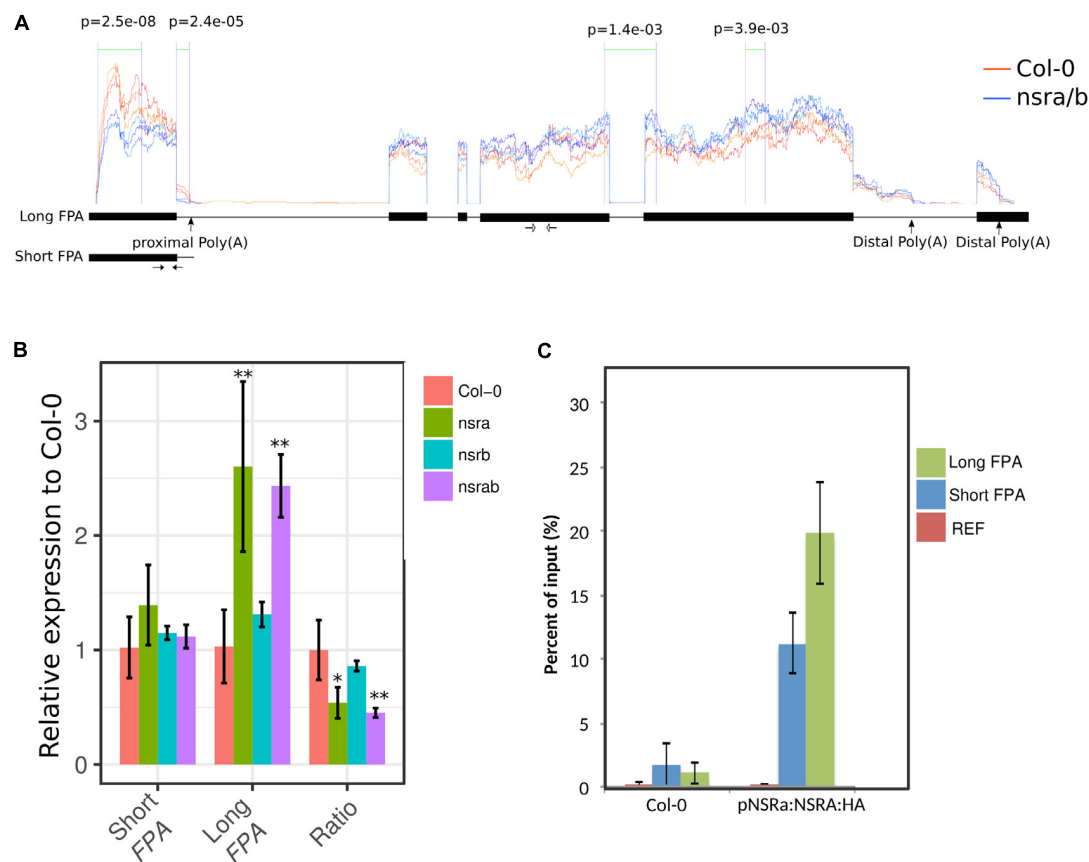
Strikingly, *FLC* is one of most deregulated genes in *nsra/b* mutants in control and NAA-treated samples. Notably, it was shown that a number of splicing and RNA processing factors control *FLC* expression by modulating the ratio of *COOLAIR* proximal and distal variants (Liu et al., 2009; Marquardt et al., 2014; Whittaker and Dean, 2017). Therefore, we determined the abundance and the ratio of *COOLAIR* variants in wild type, single *nsra*, *nsrb* and the double *nsra/b* mutants in control and NAA treated conditions using a dedicated strand-specific RT-qPCR assay (Marquardt et al., 2014). First, we confirmed that total *COOLAIR* and *FLC* abundance was decreased in *nsra* and *nsra/b* but not *nsrb* (**Figures 3B,C**). More importantly, we found that relative usage of the short (proximal) variant of *COOLAIR* increased by twofold in *nsra* and *nsra/b* but not in



**FIGURE 3 | NSRs modulate the relative abundance of lncRNA *COOLAIR* variants. (A)** Schematic representation of transcripts from the *FLC*/*COOLAIR* locus. *COOLAIR* isoforms are shown including positions of primers (arrows) used to measure distal (blue arrows) and proximal (red arrows) and total (black arrows) *COOLAIR* variant abundance. Black rectangles and black lines denote exons and introns, respectively. **(B)** *COOLAIR* and **(C)** *FLC* abundance measured by RT-qPCR in *nsra*, *nsrb*, *nsra/b* and Col-0 in seedlings. **(D)** Proximal and **(E)** distal variant usage normalized to the total amount of *COOLAIR*. **(F)** Distal vs. proximal variant usage ratio. Data represent the mean of three biological replicates  $\pm$  standard error. Results were analyzed by one-way analysis of variance (ANOVA) followed by Tukey's *post-hoc* test: groups with different letters are statistically different ( $p \leq 0.05$ ) and groups with the same letters are statistically equal ( $p \leq 0.05$ ). Significance was determined using an ANOVA coupled with a Tukey pairwise test ( $p$ -value  $< 0.05$ ).

*nsrb* leading to an increase of the ratio of distal vs. proximal COOLAIR isoforms in the same genotypes (Figures 3D–F). When analyzing the relative abundance of both variants against a housekeeping gene, we determined the decrease of total COOLAIR transcripts associated with a specific decrease of the distal variants. In contrast, proximal variant abundance remains stable (Supplementary Figure S2), leading to a change in relative variant usage (Figures 3D,E). Interestingly, the proximal COOLAIR variant was associated with a down-regulation of *FLC* and an early flowering phenotype (Marquardt et al., 2014). Together, these results suggest that the modulation of COOLAIR polyadenylation and/or splicing in *nsra* mutants contributes to the control of *FLC* expression. In addition, RNAprof also identified that the mRNA coding for the FPA protein (Horniyk et al., 2010) was differentially processed in *nsra/b* seedling treated with NAA (Figure 4A). The differential RNA processing event

occurred at the end of intron 1, which has been shown to contain an alternative polyadenylation site necessary for FPA negative autoregulation (Horniyk et al., 2010). RNAprof analysis hinted a significant reduction of the short FPA variant in *nsra/b* mutant compared to Col-0 (Figure 4A). RT-qPCR analysis using isoform specific primers (Figure 4A) showed that the long isoform accumulated in *nsra* and *nsra/b* but not in *nsrb* whereas the short isoform remained unaffected (Figure 4B). Hence, our data suggested that the use of the proximal polyA site is reduced in *nsra* and *nsra/b* mutant, which is predicted to lead to an increase of the full-length functional FPA. Interestingly, FPA was shown to favor proximal COOLAIR variants forms (Horniyk et al., 2010), suggesting that the effect of NSR mutation on COOLAIR variant ratio may be mediated by changes in FPA polyadenylation site usage. To address this potential mechanism, we checked whether COOLAIR or FPA are direct targets of



**FIGURE 4 | FPA is differentially processed in *nsra/b* plants. (A)** The RNA processing event detected in FPA by RNAprof from the comparison of WT (in orange) and *nsra/b* (blue). Significant differential events are delimited by green lines and labeled with their  $p$ -value ( $p$ ). The Y-axis shows the normalized RNA-seq coverage from RNAprof. Section between two purple lines with  $p$ -values indicated denote significant differences between nucleotide based coverage. Orange and blue traces correspond triplicate samples of Col-0 and *nsra/b* treated with a mock solution, respectively. The X-axis represents gene coordinates (boxes and lines representing exons and introns, respectively). Positions of polyadenylation sites identified in Horniyk et al. (2010) are shown on the gene model as well as the two transcript variants deriving. Positions of primer pairs used to amplify the short and long FPA variant are indicated as black and with arrows (respectively). **(B)** Isoforms specific RT-qPCR analysis of short and long FPA variant and their abundance ratio in *nsra*, *nsrb*, and *nsra/b*. Depicted data is the mean of fold change compared to Col-0  $\pm$  standard deviation of three biological replicates. Significance was determined according to a Student's  $t$ -test ( $*p < 0.05$ ;  $**p < 0.01$ ). **(C)** RIP assays using ProNSRa::NSRa::HA (NSRa), Col-0 (w/o: without tag) plants on total cell lysates of 10-day-old seedlings treated with 10 mM NAA for 24 h. Results of RT-qPCR are expressed as mean of the percentage of the respective INPUT signal (total signal before RIP) from three independent replicates  $\pm$  standard error. Genes analyzed are a housekeeping gene (At1g13320) named here REF and FPA (AT2G43410) short and long isoforms.

NSRa by RNA immunoprecipitation (RIP) using transgenic lines expressing a tagged version of the NSRa protein. Although we did not find *COOLAIR* binding to NSR, both the long and the short *FPA* variant were enriched in the RIP assay supporting the idea that NSRa directly influences the processing of *FPA* mRNA (Figure 4C). Given the critical role of *FPA*, *COOLAIR*, and *FLC* in flowering, we hypothesized that NSRa may be involved in the control of flowering time. Indeed, we observed that *nsra/b* mutant displays an early flowering phenotype (Figure 5A). We then quantified this phenotype by counting the number of rosette leaves when the flower stem emerged from the plants. Data showed that *nsra* and *nsra/b* but not *nsrb* display an early flowering phenotype (Figure 5B), which is consistent with a lower expression of *FLC* in *nsra* and *nsra/b* mutants only (Figure 4C). Altogether, our results indicate that NSRa-dependent modulation of *FPA* polyadenylation may impacts the activity of the *COOLAIR/FLC* module, affecting flowering time in Arabidopsis.

### NSRs Affect Auxin-Dependent Expression of Biotic Stress Response Genes

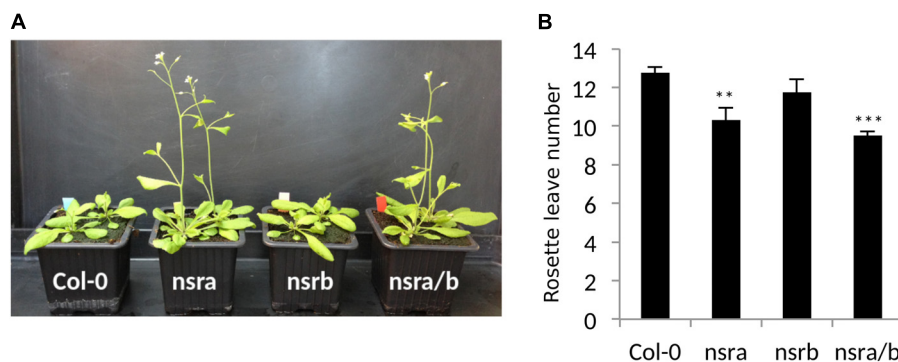
To extend our understanding on the genome-wide roles of NSRs in the control of auxin-dependent gene expression, we searched the putative function of differentially expressed and/or spliced gene groups using clustering and Gene Ontology (GO) enrichment analyses. Hierarchical clustering of differentially expressed genes determine two clusters of genes showing opposite expression patterns in response to NAA in *nsra/b* as compared to wild type plants (Figure 6A). GO analyses revealed that cluster 2 (Figure 6B), e.g., genes up-regulated by NAA in wild type plants but down-regulated by NAA in *nsra/b* is significantly enriched for genes belonging to GO categories such as “response to hormone” (FDR < 1e-6); “response to water deprivation” (FDR < 5e-9). On the other hand cluster 3 genes (Figure 6C), e.g., down-regulated or not affected by NAA in wild type but up-regulated in the mutant are highly significantly enriched for GO categories related to pathogen responses such as “response to biotic stimulus”

(FDR < 5e-16); “response to chitin” (FDR < 1e-26). We then confirmed the results of RNA-seq datasets (Figure 7) by RT-qPCR analysis of a small subset of genes belonging to clusters 2 and 3.

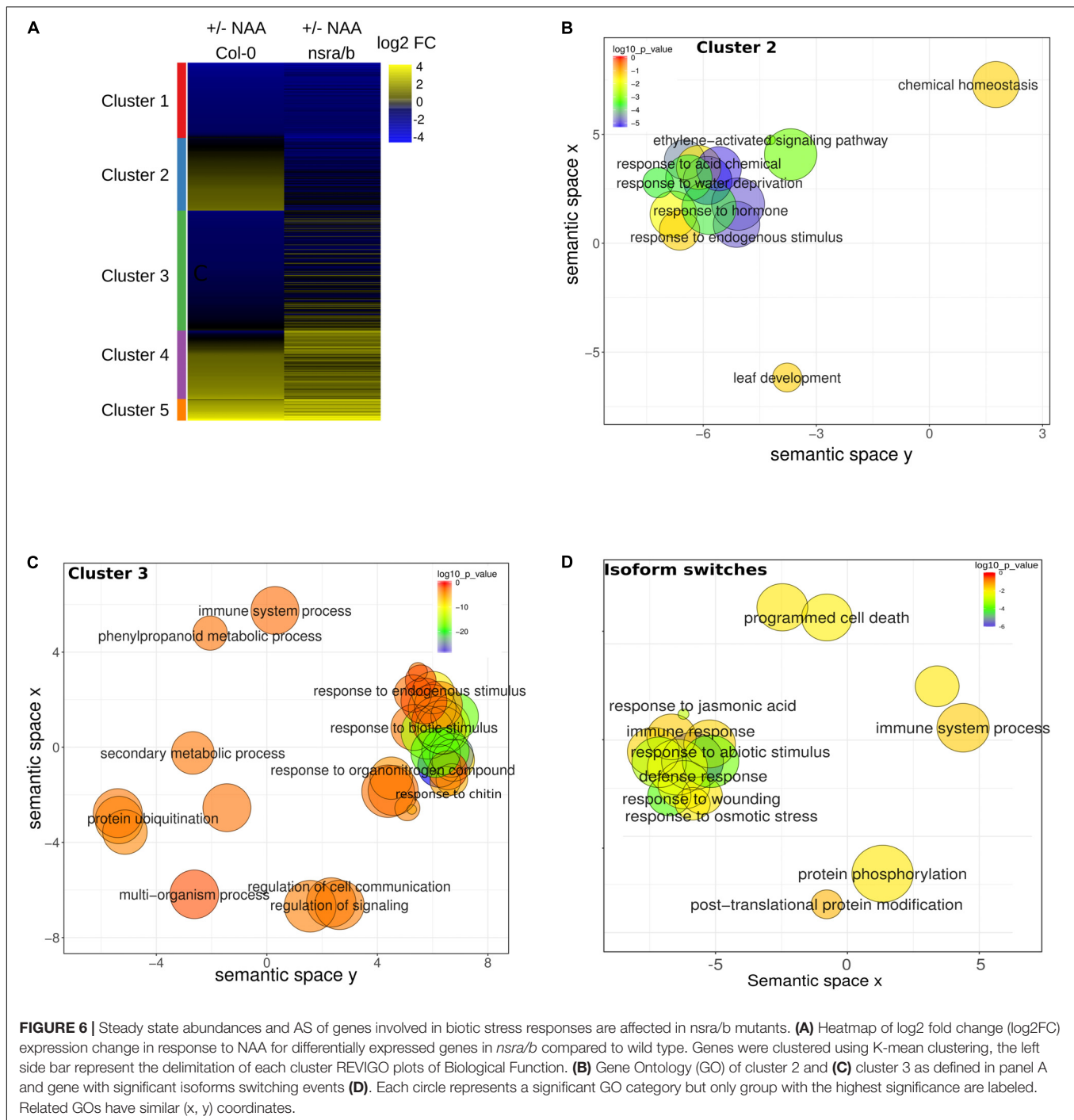
Given the important effect of NSRs on AS regulation, we also examined the putative function of differentially spliced genes having a switch in isoform usage. Strikingly, we identified a number of AS proteins located upstream of the immune response pathway. They include the MKP2 phosphatase (Lumbreras et al., 2010), the Toll/interleukin receptor (TIR) domain-containing protein TN1 and three members of the jasmonate co-receptor family (JAZ7, JAZ6, and JAZ2). In agreement, GO enrichment analysis of genes predicted to have significant isoforms switching events between *nsra/b* and Col-0 revealed a strong enrichment toward biological functions related to biotic stress responses (Figure 6D).

### NSRa Directly Recognizes Transcripts Involved in Biotic Stress Responses

To address the question whether these targets are directly related to NSR function and/or indirectly affected by other proteins, we aimed to identify direct targets of NSRs using a genome-wide RIP-seq approach. We focused our analysis on NSRa as it is globally more highly expressed than NSRb (Bardou et al., 2014). Transgenic lines expressing an epitope tagged version of NSRa under its native promoter in the *nsra* mutant genetic background were used to avoid interference with the endogenous version of NSRa. Ten days-old seedlings treated for 24 h with NAA were used to match the transcriptome analysis. Immunoprecipitation was performed on UV cross-linked tissue using HA antibodies and mouse IgG as negative control (Figure 8A). NSRa-HA was detected from the input sample as well as from the eluate of the immunoprecipitation when it was performed with an HA antibody but not when mouse IgG were used (Figure 8B) qRT-PCR analysis of previously identified targets and a randomly selected abundant housekeeping gene confirmed the specific enrichment of target genes in the RIP sample compared to the input (Figure 8C). In addition, RNA extracted from mock IP eluate did not give



**FIGURE 5 | NSRa affects flowering time in Arabidopsis. (A)** Representative picture of Col-0, *nsra*, *nsrb*, and *nsra/b* at 21 days after germination. **(B)** Mean number of rosette leaves at bolting in Col-0, *nsra*, *nsrb*, and *nsra/b*. Data is mean of 12 plants  $\pm$  standard deviation. Significance was determined using a Student's *t*-test (\*\**p*-value < 0.01; \*\*\**p*-value < 0.001).

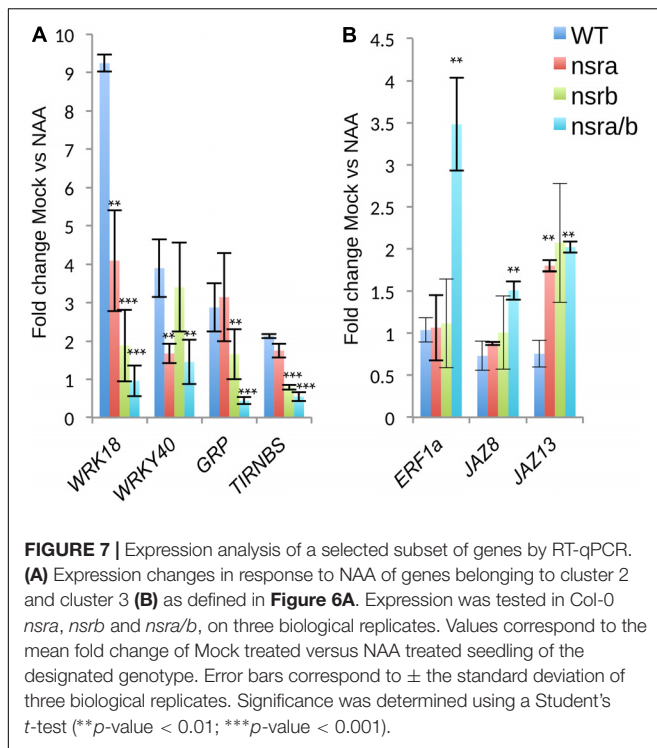


detectable amount of RNA supporting the specificity of this assay. Total RNA-seq libraries were prepared in duplicate from input, RIP and Mock samples. PCA and correlation analysis showed a dispersion of the data compatible with statistical comparisons between groups (**Supplementary Figure S3**). To detect putative NSRa targets, we used a multi-factor differential expression analysis using DESeq2 in order to identify transcripts significantly enriched in RIP as compared to the input (FDR < 0.01; log<sub>2</sub>FC > 2) that were depleted from Mock samples. After

filtering out all transcripts with less than two TPM in RIP libraries, we finally identified 342 putative targets of NSRa (**Figure 9A**).

Comparing this list of genes with those differentially expressed in *nsra/b* in mock or NAA treated seedling, we found that 33% of putative target genes were also deregulated in *nsra/b* (**Figure 9B**). Further examination of putative targets genes revealed that the large majority of these genes are up-regulated in *nsra/b* suggesting that NSRs are negatively controlling their transcript abundance





*in vivo* (Figure 9D). GO enrichment analysis revealed that putative NSRa targets (Figure 9E) are enriched for genes involved in biological processes associated with defense responses such as “response to chitin” (FDR < 1.76e-9), “response to wounding” (FDR < 2.6e-3) or “immune system processes” (FDR < 1.7e-3). Interestingly, NSR target genes were also enriched for the GO category “regulation of transcription, DNA-templated” (FDR < 1.6e-8). Further examination of targets genes belonging to this GO category revealed that 56 transcription factors (TFs) are likely to be direct targets of NSRa (Supplementary Table S1E). Among them, we found the mRNA encoding the MYC2 TF, a key regulator of immune responses (Kazan and Manners, 2013) as well as nine *WRKY* and seven *ERF* TF transcripts, which both classes have been associated with the regulation of the plant immune response (Pandey and Somssich, 2009; Huang et al., 2016). Ten putative target genes were selected for RT-qPCR validation of the RIP assay. Among them, seven showed a significant enrichment over the input samples (Figure 9C) further supporting the genome-wide approach of NSRa target identification. Together, these results suggest that direct recognition of a subset of defense response genes by NSRa may affect their steady state abundance during auxin response.

## LncRNAs Are Overrepresented Among NSRa Targets

It was previously demonstrated that a direct interaction between NSR and the lncRNA *ASCO* is able to modulate NSR function (Bardou et al., 2014). Thus, we thoroughly analyzed global lncRNA abundance in RIP-seq datasets. Interestingly, lncRNAs appeared among the most highly enriched transcripts within the

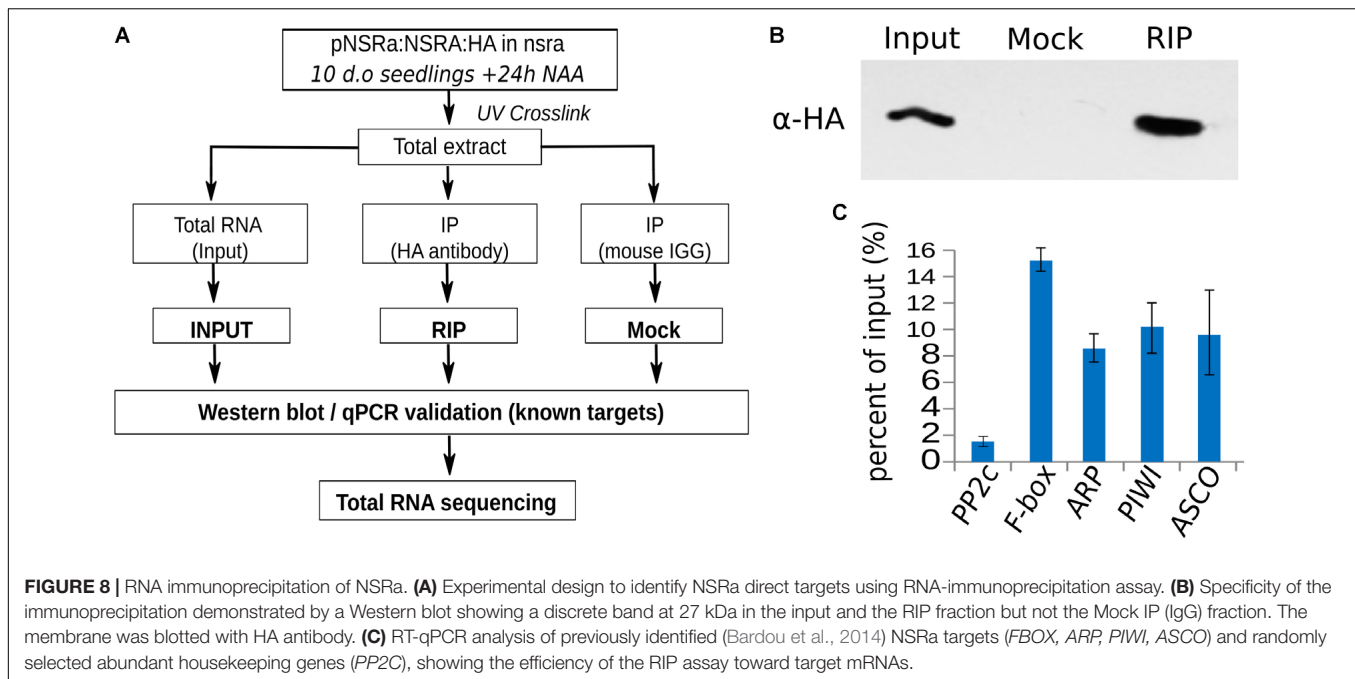
putative targets of NSRa. We found that, out of the 342 putative NSRa targets, 53 were lncRNA including 20 and 33 intergenic and antisense lncRNA, respectively (Figure 10A). In fact, relatively to the total number of lncRNAs detected in the input, lncRNA were significantly enriched over mRNA in the set of putative targets transcripts (hypergeometric test: 1.9 fold, *p*-value < 4.06e-4) (Figure 10B).

We further validated the NSR-lncRNA interaction by RIP-qPCR. We found four out of five lncRNA enriched over the input RNA in NSRa RIP samples (Figure 10C). Analyses of target lncRNA expression in *nsra/b* revealed that, similarly to the behavior of *ASCO*, seven target lncRNA are significantly upregulated in the *nsra/b* mutant (Figures 10D,E). Together, these results suggest that lncRNAs are overrepresented among targets of NSRa and that NSRs might control the accumulation of lncRNA *in vivo*. Future works on the interplay between lncRNA and mRNAs in NSR-containing complexes should shed light on their global impact over the transcriptome.

## DISCUSSION

In agreement with our previous study based on microarrays, a novel thorough analysis of *nsra/b* transcriptome using RNA-seq has revealed an important role of these RBPs in the control of auxin-responsive genes. A previous study monitoring AS changes of a subset of 288 genes using high-resolution real-time PCR, first uncovered the important roles of NSR in auxin-driven AS changes and targeted RIP-qPCR showed that both NSR proteins were able to bind AS mRNA targets *in planta* (Bardou et al., 2014). Our global AS analysis further confirmed this function of NSRs on AS modulation and demonstrated the impact of these proteins at genome-wide level. However, our RIP-seq global analysis of NSR targets did not show a strong enrichment toward AS modulated transcripts. Instead, a large fraction of NSR targets were transcriptionally upregulated in *nsra/b*, suggesting that NSR may play a direct role in controlling their stability or transcription. Several splicing factors have been shown to affect transcription by interacting with the transcriptional machinery and to modulate Pol II elongation rates (Kornblihtt et al., 2004). In addition, specific RBPs deposited during pre-mRNA splicing at exon–exon splicing junctions, can influence their mRNA decay (Lumbreras et al., 2010; Nishtala et al., 2016). Further dissection of the NSR recognition sites on mRNAs may support a role of NSRs on mRNA decay.

The combination of our RNA-seq and RIP-seq approaches revealed that lncRNAs are privileged targets of NSRa and that a significant fraction of the auxin-responsive non-coding transcriptome is deregulated in the *nsra/b* genetic background. This is in accordance with our previous results showing that the specific interaction of NSR with the *ASCO* lncRNA is able to modify AS pattern of a subset of NSR-target genes. Our study suggests that NSRs may play a broader role in lncRNA biology. In particular, we found that a large majority of lncRNA targeted by NSRa are upregulated in *nsra/b*, suggesting a new role of these proteins in the control of lncRNA transcription and/or stability. So far, very little is known about lncRNA biogenesis,



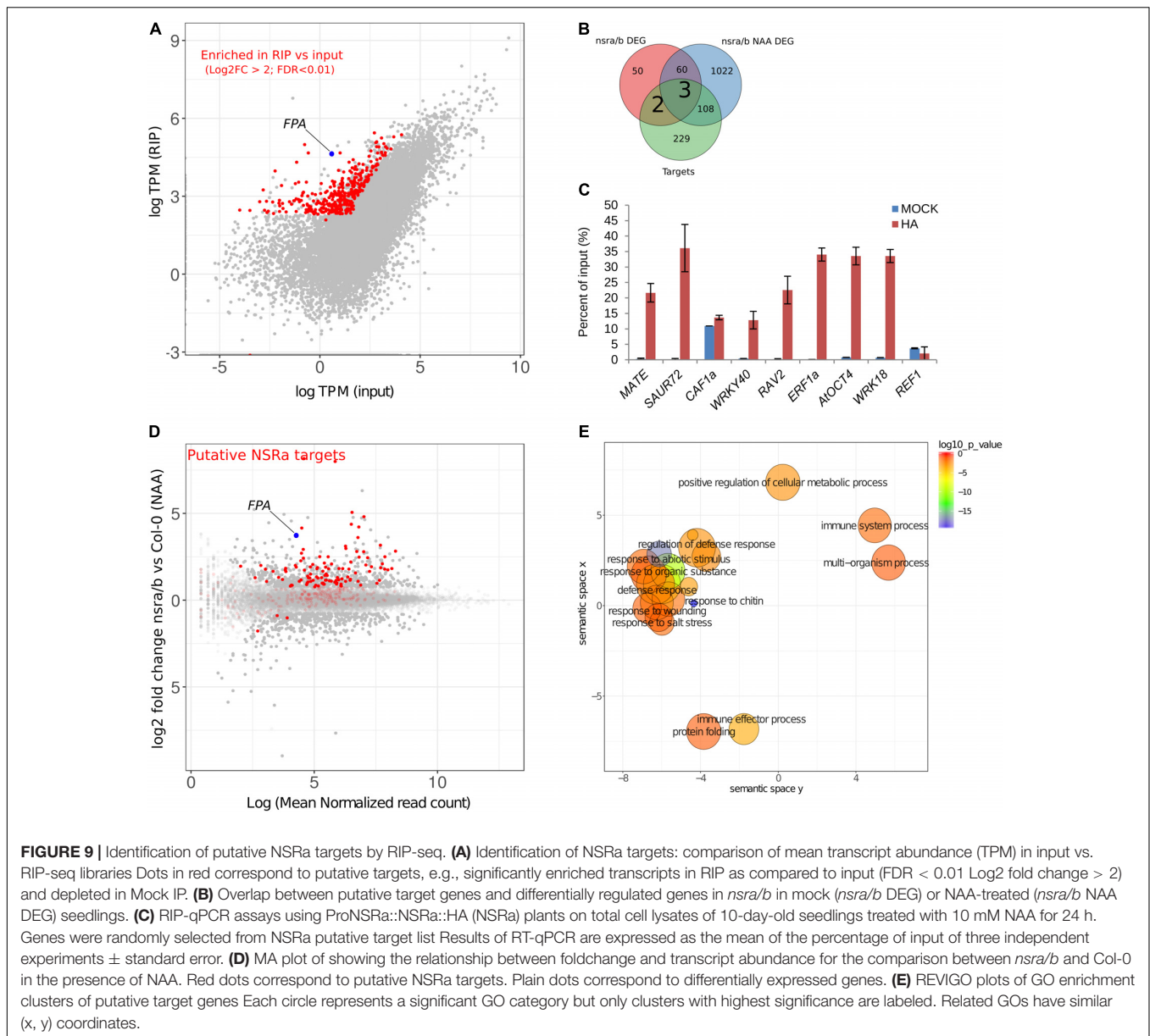
especially in plants. Other RBPs have been shown to affect lncRNA abundance. For instance several members of the cap binding complex such as CBP20, CBP80, and SERRATE have been shown to co-regulate the abundance of a large subset of lncRNAs in *Arabidopsis* seedlings (Liu et al., 2012). Interestingly, these three proteins, like NSRs, have also been associated with major roles in the control of AS patterns (Raczynska et al., 2010, 2014). This suggests that the splicing machinery might be used to control lncRNAs abundance in the nucleus and that the interplay between lncRNA and mRNAs may be an emerging mechanism in splicing regulation. Further genetic dissection is required to determine whether NSRs are involved in the same pathway that CBP20, CPB80, and SERRATE.

The strong deregulation of the *FLC/COOLAIR* module in *nsra/b* led us to identify a new role of NSRa in the control of flowering time. A number of forward genetic screenings aiming to identify new genes controlling flowering time through *FLC* expression modulation have consistently identified RNA processing and splicing factors that promote formation of the short *COOLAIR* isoforms, such as *FCA*, *FPA*, *HLP1*, *GRP7* and the core spliceosome component *PRP8a* (Deng and Cao, 2017). Loss of function mutants of these factors lead to a reduced usage of *COOLAIR* proximal polyadenylation site and an increase of *FLC* transcription which is associated with late flowering phenotypes (Deng and Cao, 2017). Interestingly, our analysis of the *FLC/COOLAIR* module in *nsr* mutants revealed an opposite role of NSRa in *COOLAIR* polyadenylation site usage, leading to the increased use of *COOLAIR* proximal polyadenylation site, and reduced *FLC* levels associated with an early flowering phenotype.

We also identified a new role of NSRs in the regulation of auxin-mediated expression and AS of transcripts related to biotic stress response. Interestingly, it has been shown for

several years that natural (i.e., IAA) and synthetic (i.e., NAA) auxins can promote pathogen virulence of *P. syringae* (Mutka et al., 2013). More recently, a conserved pathway of auxin biosynthesis was demonstrated in *Pseudomonas* as contributing to pathogen virulence in *Arabidopsis thaliana* (McClerkin et al., 2018). However, little is known on the specific plant factors that modulate immune responses upon endogenous or pathogen produced auxins. Our work shows that NSRs do not affect the global auxin responses but rather have an impact on the abundance of mRNAs coding for proteins involved in plant immune response, suggesting that these RBPs may participate in the regulation of plant defense by endogenous or pathogen-produced auxins.

In higher plants, AS plays a key role in gene expression as shown by the fact that 60–70% of intron-containing genes undergoes alternative processing. Several genome-wide studies of AS has shown that this mechanism may represent a way to enhance the ability for plant cells to cope with stress via the modulation of transcriptome plasticity. Here we show that among the genes with significant isoforms switching events in *nsra/b* mutant treated with auxin, we identified several genes involved in the modulation of the MAPK kinase modules, a core regulator of defense responses. They included MKP2 phosphatase which functionally interacts with MPK3 and MPK6 to mediate disease response in *Arabidopsis* (Lumbreras et al., 2010) and PTI-4 kinase which was found in MPK6 containing complexes *in vivo* and was shown to function in the MPK6 signaling cascade (Forzani et al., 2011). As activation of MAPK signaling cascades regulate the expression of 1000s of downstream targets genes, we can speculate that a large fraction of the transcriptome change observed in *nsra/b* mutant could be a consequence of AS defect of genes involved in such early phase of the defense response pathway.



## MATERIALS AND METHODS

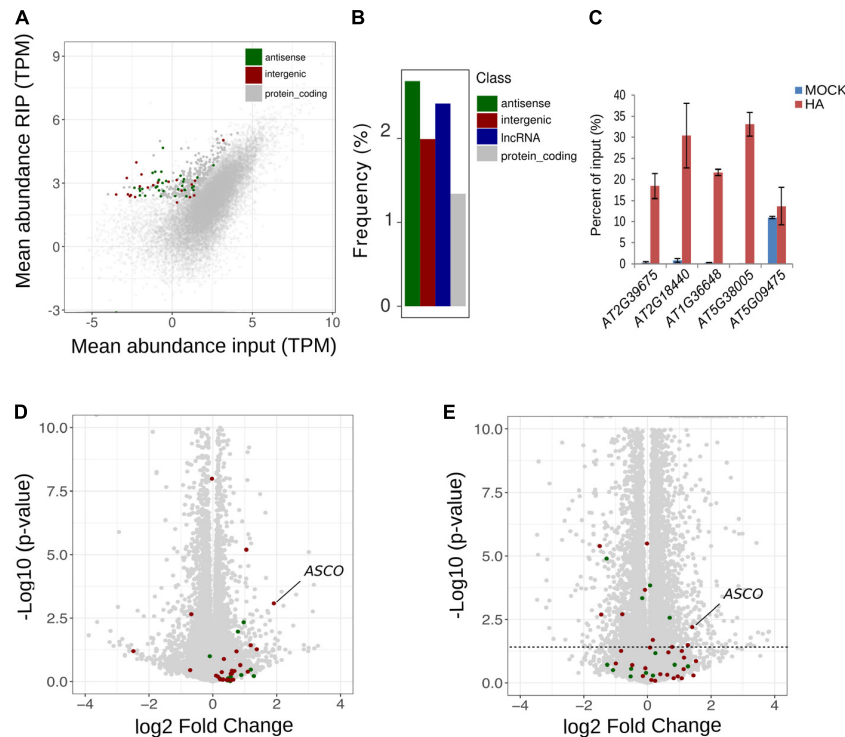
### Plant Material and Treatments

All mutants were in the Columbia-0 (Col-0) background. *Atnsra* (SALK\_003214) and *Atnsrb* (Sail\_717) were from the SALK and SAIL T-DNA collections, respectively. For RIP, lines expressing pNSRa::NRSa-HA in *Atnsra* or pNSRa::NRSb-HA in *Atnsrb* were used (Bardou et al., 2014). Plants were grown on soil in long day (16 h light/8 h dark) conditions at 23°C. For RNA-sequencing and RIP-seq WT and *nsra/nsrb* were grown on nylon membrane (Nitex 100  $\mu$ m) in plate filled with  $\frac{1}{2}$ MS medium for 10 days and then transferred for 24 h to  $\frac{1}{2}$ MS medium containing 100 nM NAA or a mock solution before the whole seedlings were harvested. For flowering time analysis, plants were grown under long day conditions and the number of rosette

leave were counted from 12 plants when the flower stem was 1 cm tall.

### RNA Sequencing Analysis

Stranded mRNA sequencing libraries were performed on three biological replicate of Col-0, *nsra/b* treated with a 100 nM NAA or a mock solution. One  $\mu$ g of total RNA from Col-0 and *nsra/b* seedlings was used for library preparation using the Illumina TruSeq Stranded mRNA library prep kit according to the manufacturer instruction. Libraries were sequenced on an HiSeq2000 sequencer using 150 nt pair-end read mode. A minimum 28 Million of were obtained for each sample, quality filtered using fastqc (Andrews, 2010) with default parameters and aligned using tophat (Trapnell et al., 2012) with the following arguments: -g 1 -i 5 -p



**FIGURE 10 |** NSRa binds to numerous lncRNA. **(A)** Comparison of mean transcript abundance (TPM) in input vs. RIP-seq libraries. Dots in gray; red and green correspond to protein coding, intergenic and antisense lncRNA transcripts, respectively. Plain dots correspond to significantly enriched genes in RIP vs. Input, e.g., putative targets. **(B)** Frequency of all lncRNAs, antisense lncRNA, intergenic lncRNA and protein coding genes among the NSRa targets: blue, red, green, and gray bars, respectively. Frequency was calculated compared the number detected genes in the input for each class. **(C)** RIP-qPCR assays using proNSRa::NSRa::HA (NSRa) plants on total cell lysates of 10-day-old seedlings treated with 10mM NAA for 24 h. lncRNA were randomly selected from NSRa putative target list. Results of RT-qPCR are expressed as the mean of the percentage of input of three independent experiments  $\pm$  standard error. Volcano plots of showing the relationship between the fold change and  $p$ -value of the comparison between *nsra/b* and Col-0 in **(D)** mock or **(E)** NAA treated samples. Plain colored dots correspond to intergenic (red) and antisense (blue) lncRNA which are putative targets of NSRa. The dotted line delineates a  $p$ -value of 0.05.

6 -I 2000 -segment-mismatches 2 -segment-length 20 -library-type fr-firststrand. Read were counted using SummarizeOverlap function from the GenomicRange R package (Lawrence et al., 2013) using strand specific and Union mode. Differential gene expression analysis was done one pairwise comparison using DESeq2 (Love et al., 2014) with FDR correction of the  $p$ -value. K-mean clustering analysis was performed in R on scaled log2 fold change data and the optimal number of cluster was determined using the elbow method. Heatmap was plotted using heatmap.2 function of the gplots package (Warnes et al., 2009). Sequence files have been submitted to the NCBI GEO database under accession GSE65717 and GSE116923.

## Gene Ontology Analysis

Gene ontology enrichment analysis was done using the AgriGO server<sup>1</sup> using default parameters. Lists of GO terms were visualized using REVIGO<sup>2</sup> and plotted in R. Only GO terms with a dispensability factor over 0.5 were printed in REVIGO plots.

<sup>1</sup><http://bioinfo.cau.edu.cn/agriGO/>

<sup>2</sup><http://revigo.irb.hr/>

## AS Analysis

RNAprof (v1.2.6) was used on BAM alignment files with the following parameters: LIBTYPE = fr-unstranded, SEQTYPE = “-Pair”, MIS = 1000. All possible pairwise comparisons were computed. Overlap of differential events ( $pval < 1e-04$ ) with gene annotation was done using *findOverlaps* of the GenomicRanges Package in R and custom in house scripts. Only events that were completely included in gene feature (e.g., intron, exons, 3' UTR, and 5' UTR) were kept for further analysis.

For isoforms switching identification, transcript isoforms abundance was quantified with pseudo alignment read count with kallisto (Bray et al., 2016), on all isoforms of the AtRTD2 database (Zhang et al., 2017). Then the IsoformSwitchAnalyzeR package was used to detect significant changes in isoform usage. Only significant switches ( $p_{adj} < 0.1$ ) were kept for further analyses (Vitting-Seerup and Sandelin, 2017).

## RNA Immunoprecipitation and Sequencing (RIP-Seq)

NSRa protein tagged with HA was immunoprecipitated from the *nsra* mutant background expressing the pNSRa::NSRa-HA



construct (Bardou et al., 2014). Briefly, 10 day old seedlings treated with 100 nM NAA for 24 h were irradiated three times with UV using a UV crosslinker CL-508 (Uvitec) at 0.400 J/cm<sup>2</sup>. Plants were ground in liquid nitrogen and RNA-IP was performed as in Sorenson and Bailey-Serres (2014) with the following modification: immunoprecipitation (IP) was performed using anti mouse HA-7 monoclonal antibody (Sigma) and the negative IP (Mock) was done using anti mouse IgG (Millipore). RNA was eluted from the beads with 50 U proteinase K (RNase grade, Invitrogen) in 2 µl of RNase inhibitor at 55°C for 1 h in wash buffer and extracted using Trizol according to manufacturer instructions. A 10th of the input fraction was saved for RNA and protein extraction. For western blot analysis, proteins were extracted from the beads and input fraction with 2X SDS-loading Buffer for 10 min at 75°C, directly loaded on SDS PAGE, transferred onto Nitrocellulose membranes and blotted with HA-7 antibody. For RT-qPCR analysis, RNA was reverse transcribed with Maxima Reverse Transcriptase (Thermo) using random Hexamer priming. cDNA from input, IP and Mock were amplified with primers listed in **Supplementary Table S2**. Results were analyzed using the percentage of input method. First, Ct values of input sample (10% of volume) were adjusted to 100% as follows: Adjusted Ct input = Raw Ct input - log<sub>2</sub>(10). Percentage of input was calculated as follow: 100\*2<sup>-(Adjusted Ct input - Ct IP)</sup>. Results are mean of three independent experiments. Student's *t*-test was performed to determine significance. For RNA-seq : input mock and IP RNA were depleted of rRNA using the plant leaf ribozero kit (Illumina) and libraries were prepared using the Illumina TruSeq Stranded mRNA library prep kit according to the manufacturer instruction but omitting the polyA RNA purification step and sequenced on a NextSeq500 sequencer (Illumina) using single-end 75 bp reads mode. Sequence files have been submitted to the NCBI GEO database under accession GSE116914.

## Analysis of RIP-Seq Data

Reads were mapped using STAR (Dobin et al., 2013) and TPM was calculated using RSEM (Li and Dewey, 2011). Read were counted using SummarizeOverlap function from the GenomicRange R package (Lawrence et al., 2013) using strand specific and *Union* mode. To identify putative NSRa targets we used pairwise comparison with DESeq2 package. Only genes significantly enriched in IP with anti HA as compared with the anti-mouse IgG (mock) IP were kept for further analysis (logFC ≥ 1; FDR < 0.01). Putative targets genes were defined as gene highly enriched in the IP with anti HA compared to their global level in input used for the IP (logFC > 2; FDR < 0.01). To reduce noise associated with low read counts, we excluded from this list any gene with less than two TPM in at least one of the RIP-seq libraries.

## REFERENCES

Andrews, S. (2010). *FastQC: A Quality Control Tool for High Throughput Sequence Data*. Available at: <http://www.bioinformatics.babraham.ac.uk/projects/fastqc>

## Measuring Distal and Proximal COOLAIR Variants

This was performed essentially as in Marquardt et al. (2014). 5 µg of total RNA was reverse transcribed with and oligo(dT) primer. qPCR was performed with set of primers specific to distal and proximal COOLAIR described in Marquardt et al. (2014). qPCR reactions were performed in triplicates for each sample. Average values of the triplicates were normalized to the expression of total COOLAIR quantified in the same sample.

## AUTHOR CONTRIBUTIONS

JB designed study, performed the experiments, analyzed the data, and wrote the article. NR, FA, RR, and CC performed the experiments and participated in writing. TB analyzed the data. MC designed the study and wrote the paper.

## FUNDING

This work was supported by grants of The King Abdulla University of Science and Technology (KAUST) International Program OCRF-2014-CRG4 and The LIA (Associated International Laboratory) of CNRS NOCOSYM and 'Laboratoire d'Excellence (LABEX)' Saclay Plant Sciences (SPS; ANR-10-LABX-40) and the ANR grant SPLISIL, France.

## SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fpls.2018.01209/full#supplementary-material>

**FIGURE S1 | (A)** Pearson correlation matrix heatmap with dendograms showing the relative distance between each poly(A)+ RNA-seq samples. **(B)** PCA analysis showing the effect of auxin and genotype on the variance between samples.

**FIGURE S2 | (A)** Proximal and **(B)** distal variant relative abundance normalized to an housekeeping transcript (PP2C). Error bars correspond to ± the standard deviation of three biological replicates. Significance was determined using a Student's *t*-test (\*\*\**p*-value < 0.001).

**FIGURE S3 | (A)** Pearson correlation matrix heatmap with dendograms showing the relative distance between each sample of the RIP-seq experiments. **(B)** PCA analysis showing the effect the variance between samples.

**TABLE S1 |** Summary of RNA-seq and RIP-seq data analysis. **(A)** Description of spreadsheet tab. **(B)** Differential gene expression analysis. **(C)** RNA prof analysis. **(D)** Expression and usage of all isoforms from genes containing at least one isoforms switching event. **(E)** NSRa targets identified by RIP. **(F)** Transcription Factor identified in NSRa targets.

**TABLE S2 |** Sequence of primers used in this study.

Ariel, F., Jegu, T., Latrasse, D., Romero-Barrios, N., Christ, A., Benhamed, M., et al. (2014). Noncoding transcription by alternative RNA polymerases dynamically regulates an auxin-driven chromatin loop. *Mol. Cell* 55, 383–396. doi: 10.1016/j.molcel.2014.06.011

- Bailey-Serres, J., Sorenson, R., and Juntawong, P. (2009). Getting the message across: cytoplasmic ribonucleoprotein complexes. *Trends Plant Sci.* 14, 443–453. doi: 10.1016/j.tplants.2009.05.004
- Bardou, F., Ariel, F., Simpson, C. G., Romero-Barrios, N., Laporte, P., Balzergue, S., et al. (2014). Long noncoding RNA modulates alternative splicing regulators in *Arabidopsis*. *Dev. Cell* 30, 166–176. doi: 10.1016/j.devcel.2014.06.017
- Bray, N. L., Pimentel, H., Melsted, P., and Pachter, L. (2016). Near-optimal probabilistic RNA-seq quantification. *Nat. Biotechnol.* 34, 525–527. doi: 10.1038/nbt.3519
- Brown, J. W., Calixto, C. P., and Zhang, R. (2017). High-quality reference transcript datasets hold the key to transcript-specific RNA-sequencing analysis in plants. *New Phytol.* 213, 525–530. doi: 10.1111/nph.14208
- 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
- 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
- Foley, S. W., Kramer, M. C., and Gregory, B. D. (2017). RNA structure, binding, and coordination in *Arabidopsis*. *Wiley Interdiscip. Rev. RNA* 8:e1426. doi: 10.1002/wrna.1426
- Forzani, C., Carreri, A., de la Fuente van Bentem, S., Lecourieux, D., Lecourieux, F., and Hirt, H. (2011). The *Arabidopsis* protein kinase Pto-interacting 1-4 is a common target of the oxidative signal-inducible 1 and mitogen-activated protein kinases. *FEBS J.* 278, 1126–1136. doi: 10.1111/j.1742-4658.2011.08033.x
- Gagliardi, M., and Matarazzo, M. R. (2016). RIP: RNA immunoprecipitation. *Methods Mol. Biol.* 1480, 73–86.
- Horniyk, C., Terzi, L. C., and Simpson, G. G. (2010). The spen family protein FPA controls alternative cleavage and polyadenylation of RNA. *Dev. Cell* 18, 203–213. doi: 10.1016/j.devcel.2009.12.009
- Huang, P.-Y., Catinot, J., and Zimmerli, L. (2016). Ethylene response factors in *Arabidopsis* immunity. *J. Exp. Bot.* 67, 1231–1241. doi: 10.1093/jxb/erv518
- Kazan, K., and Manners, J. M. (2013). MYC2: the master in action. *Mol. Plant* 6, 686–703. doi: 10.1093/mp/sss128
- Kornblihtt, A. R., de la Mata, M., Fededa, J. P., Munoz, M. J., and Nogues, G. (2004). Multiple links between transcription and splicing. *RNA* 10, 1489–1498.
- Köster, T., and Meyer, K. (2018). Plant ribonomics: proteins in search of RNA partners. *Trends Plant Sci.* 23, 352–365. doi: 10.1016/j.tplants.2018.01.004
- Lawrence, M., Huber, W., Pagès, H., Aboyoun, P., Carlson, M., Gentleman, R., et al. (2013). Software for computing and annotating genomic ranges. *PLOS Comput. Biol.* 9:e1003118. doi: 10.1371/journal.pcbi.1003118
- Lee, K., and Kang, H. (2016). Emerging roles of RNA-binding proteins in plant growth, development, and stress responses. *Mol. Cells* 39, 179–185. doi: 10.14348/molcells.2016.2359
- Li, B., and Dewey, C. N. (2011). RSEM: accurate transcript quantification from RNA-seq data with or without a reference genome. *BMC Bioinformatics* 12:323. doi: 10.1186/1471-2105-12-323
- Liu, F., Marquardt, S., Lister, C., Swiezewski, S., and Dean, C. (2009). Targeted 3' processing of antisense transcripts triggers *Arabidopsis* FLC chromatin silencing. *Science* 327, 94–97.
- Liu, J., Jung, C., Xu, J., Wang, H., Deng, S., Bernad, L., et al. (2012). Genome-wide analysis uncovers regulation of long intergenic noncoding RNAs in *Arabidopsis*. *Plant Cell* 24, 4333–4345. doi: 10.1105/tpc.112.102855
- 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: 10.1101/002832
- Lumbreras, V., Vilela, B., Irar, S., Solé, M., Capellades, M., Valls, M., et al. (2010). MAPK phosphatase MKP2 mediates disease responses in *Arabidopsis* and functionally interacts with MPK3 and MPK6. *Plant J.* 63, 1017–1030. doi: 10.1111/j.1365-3113.2010.04297.x
- Maquat, L. E. (2004). Nonsense-mediated mRNA decay: splicing, translation and mRNP dynamics. *Nat. Rev. Mol. Cell Biol.* 5, 89–99.
- 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
- 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
- McClerklin, S. A., Lee, S. G., Harper, C. P., Nwumeh, R., Jez, J. M., and Kunkel, B. N. (2018). Indole-3-acetaldehyde dehydrogenase-dependent auxin synthesis contributes to virulence of *Pseudomonas syringae* strain DC3000. *PLOS Pathog.* 14:e1006811. doi: 10.1371/journal.ppat.1006811
- Mutka, A. M., Fawley, S., Tsao, T., and Kunkel, B. N. (2013). Auxin promotes susceptibility to *Pseudomonas syringae* via a mechanism independent of suppression of salicylic acid-mediated defenses. *Plant J.* 74, 746–754. doi: 10.1111/tpj.12157
- Ner-Gaon, H., Halachmi, R., Savaldi-Goldstein, S., Rubin, E., Ophir, R., and Fluhr, R. (2004). Intron retention is a major phenomenon in alternative splicing in *Arabidopsis*. *Plant J.* 39, 877–885.
- Nishtala, S., Neelamraju, Y., and Janga, S. C. (2016). Dissecting the expression relationships between RNA-binding proteins and their cognate targets in eukaryotic post-transcriptional regulatory networks. *Sci. Rep.* 6:25711. doi: 10.1038/srep25711
- Oliveira, C., Faoro, H., Alves, L. R., and Goldenberg, S. (2017). RNA-binding proteins and their role in the regulation of gene expression in *Trypanosoma cruzi* and *Saccharomyces cerevisiae*. *Genet. Mol. Biol.* 40, 22–30. doi: 10.1590/1678-4685-GMB-2016-0258
- Pandey, S. P., and Somssich, I. E. (2009). The role of WRKY transcription factors in plant immunity. *Plant Physiol.* 150, 1648–1655.
- 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, 265–278. doi: 10.1093/nar/gkp869
- Raczynska, K. D., Stepień, A., Kierzkowski, D., Kalak, M., Bajczyk, M., McNicol, J., et al. (2014). The SERRATE protein is involved in alternative splicing in *Arabidopsis thaliana*. *Nucleic Acids Res.* 42, 1224–1244. doi: 10.1093/nar/gkt894
- 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
- Sorenson, R., and Bailey-Serres, J. (2014). Selective mRNA sequestration by oligouridylate-binding protein 1 contributes to translational control during hypoxia in *Arabidopsis*. *Proc. Natl. Acad. Sci. U.S.A.* 111, 2373–2378. doi: 10.1073/pnas.1314851111
- Tran, V. D. T., Souiai, O., Romero-Barrios, N., Crespi, M., and Gautheret, D. (2016). Detection of generic differential RNA processing events from RNA-seq data. *RNA Biol.* 13, 59–67. doi: 10.1080/15476286.2015.1118604
- 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
- Vitting-Seerup, K., and Sandelin, A. (2017). The landscape of isoform switches in human cancers. *Mol. Cancer Res.* 15, 1206–1220. doi: 10.1158/1541-7786.MCR-16-0459
- Warnes, G. R., Bolker, B., Bonebakker, L., Gentleman, R., Huber, W., Liaw, A., et al. (2009). *gplots: Various R programming Tools for Plotting Data. R Package Version 2.*
- 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
- 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 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 © 2018 Bazin, Romero, Rigo, Charon, Blein, Ariel and Crespi. 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.



# Regulation of Plant Microprocessor Function in Shaping microRNA Landscape

**Jakub Dolata, Michał Taube, Mateusz Bajczyk, Artur Jarmolowski, Zofia Szweykowska-Kulinska\* and Dawid Bielewicz\***

Department of Gene Expression, Institute of Molecular Biology and Biotechnology, Faculty of Biology, Adam Mickiewicz University in Poznań, Poznań, Poland

## OPEN ACCESS

### Edited by:

Dora Szakonyi,  
Instituto Gulbenkian de Ciência (IGC),  
Portugal

### Reviewed by:

Patricia Baldrich,  
Donald Danforth Plant Science  
Center, United States  
Sascha Laubinger,  
Center for Plant Molecular Biology,  
Germany

### \*Correspondence:

Zofia Szweykowska-Kulinska  
zofszyk@amu.edu.pl  
Dawid Bielewicz  
bieda@amu.edu.pl

### Specialty section:

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

**Received:** 14 February 2018

**Accepted:** 16 May 2018

**Published:** 05 June 2018

### Citation:

Dolata J, Taube M, Bajczyk M,  
Jarmolowski A,  
Szweykowska-Kulinska Z and  
Bielewicz D (2018) Regulation of Plant  
Microprocessor Function in Shaping  
microRNA Landscape.  
Front. Plant Sci. 9:753.  
doi: 10.3389/fpls.2018.00753

MicroRNAs are small molecules (~21 nucleotides long) that are key regulators of gene expression. They originate from long stem-loop RNAs as a product of cleavage by a protein complex called Microprocessor. The core components of the plant Microprocessor are the RNase type III enzyme Dicer-Like 1 (DCL1), the zinc finger protein Serrate (SE), and the double-stranded RNA binding protein Hyponastic Leaves 1 (HYL1). Microprocessor assembly and its processing of microRNA precursors have been reported to occur in discrete nuclear bodies called Dicing bodies. The accessibility of and modifications to Microprocessor components affect microRNA levels and may have dramatic consequences in plant development. Currently, numerous lines of evidence indicate that plant Microprocessor activity is tightly regulated. The cellular localization of HYL1 is dependent on a specific KETCH1 importin, and the E3 ubiquitin ligase COP1 indirectly protects HYL1 from degradation in a light-dependent manner. Furthermore, proper localization of HYL1 in Dicing bodies is regulated by MOS2. On the other hand, the Dicing body localization of DCL1 is regulated by NOT2b, which also interacts with SE in the nucleus. Post-translational modifications are substantial factors that contribute to protein functional diversity and provide a fine-tuning system for the regulation of protein activity. The phosphorylation status of HYL1 is crucial for its activity/stability and is a result of the interplay between kinases (MPK3 and SnRK2) and phosphatases (CPL1 and PP4). Additionally, MPK3 and SnRK2 are known to phosphorylate SE. Several other proteins (e.g., TGH, CDF2, SIC, and RCF3) that interact with Microprocessor have been found to influence its RNA-binding and processing activities. In this minireview, recent findings on the various modes of Microprocessor activity regulation are discussed.

**Keywords:** microprocessor, DCL1, SE, HYL1, miRNA biogenesis, *Arabidopsis*

## INTRODUCTION

Mature microRNAs are derived from long primary transcripts (pri-miRNAs) that are produced by RNA Polymerase II (RNAPII). They are capped at their 5' ends and possess a poly A tail at their 3' ends (Xie et al., 2005). Pri-miRNA levels are tightly regulated at the transcriptional (Zhang et al., 2013; Sun et al., 2015, 2018), co-transcriptional (Fang et al., 2015; Dolata et al., 2016)

and post-transcriptional levels (Ben Chaabane et al., 2013; Bielewicz et al., 2013; Zhang et al., 2015; Barciszewska-Pacak et al., 2016; Knop et al., 2017; Stepień et al., 2017; Yu et al., 2017). Interestingly, in many cases, changes in the level of a given pri-miRNA are not reflected in changes in the level of the mature microRNA (Barciszewska-Pacak et al., 2015; Dolata et al., 2016). This might be a consequence of regulation at the pri-miRNA or pre-miRNA (intermediate product during microRNA biogenesis) processing/degradation step. Production of microRNAs is driven by a complex called Microprocessor, which in *Arabidopsis* consists of three core proteins: the RNase type III enzyme Dicer-Like 1 (DCL1), the zinc finger protein Serrate (SE), and the double-stranded RNA binding protein Hyponastic Leaves 1 (HYL1). Microprocessor term was originally coined for a nuclear protein complex in animal cells for pre-miRNA production (Denli et al., 2004; Gregory et al., 2004).

## MICROPROCESSOR COMPONENTS LOCALIZATION

In plants, Microprocessor action is limited to the nucleus (Fang and Spector, 2007; Yu et al., 2017), whereas localization of its components is not restricted to one cellular compartment. According to current knowledge, the first steps of plant microRNA biogenesis occur in specialized nuclear foci called dicing bodies (D-bodies) (Fang and Spector, 2007). Together, DCL1 and HYL1 in the nucleus are found almost exclusively in D-bodies. However, SE is present in D-bodies as well as in nuclear speckles that contain serine/arginine-rich (SR) splicing factors (Ali et al., 2003). How the Microprocessor complex is assembled in D-bodies and how it is recruited to pri-miRNAs are still not clear. Nevertheless, several factors have been shown to be important for D-body formation and effective cleavage of microRNA precursors.

A growing amount of evidence indicates a direct link between RNAPII transcription and the biogenesis of small RNAs in *Arabidopsis* (Fang et al., 2015; Dolata et al., 2016; Liu et al., 2017). One of the overlapping elements is Elongator complex, firstly described in yeast (Otero et al., 1999) and further purified from plant cells. Elongator was described as a six-component complex involved in the regulation of transcription elongation (Nelissen et al., 2010). Fang et al. (2015) have shown that disruption of the Elongator complex results in reduced RNAPII occupancy at tested *MIR* genes and lower levels of a few pri-miRNAs. Mutants of two Elongator subunits (*elp2-2* and *elp5-1*) have disrupted DCL1 localization and a reduced number of D-bodies. Furthermore, all core Microprocessor components interact with Elongator complex subunits (ELP2, ELP4, and ELP5-Elongator complex Proteins 2/4/5). DCL1 associates with chromatin on *MIR* loci, and a functional Elongator complex is necessary for DCL1 recruitment to nascent *MIR* transcripts. These data suggest that the processing of at least some pri-miRNAs occurs co-transcriptionally.

More evidence for a connection between transcription and Microprocessor was presented by Wang et al. (2013), who

described two *Arabidopsis* NOT2 (Negative on TATA-less 2) proteins (NOT2a and NOT2b) as factors that promote microRNA production. In yeasts, NOT2 was shown to bind directly to RNAPII and to promote transcription elongation (Kruk et al., 2011). Similarly, in *Arabidopsis*, NOT2b co-precipitates with the large subunit of RNAPII and affects transcription. Moreover, NOT2b interacts with DCL1 and SE; however, it does not interact with HYL1. Furthermore, a *not2a-1 not2b-1* double mutant was shown to have an increased number of nuclear foci containing DCL1. Still, the possibility that NOT2 links *MIR* gene transcription and post-transcriptional processing for better coordination and efficiency cannot be excluded.

In *Arabidopsis*, the MOS4-Associated Complex (MAC) (Palma et al., 2007; Monaghan et al., 2009) is a counterpart of the NineTeen Complex (NTC) in yeast (Fabrizio et al., 2009) and is directly linked to transcription and microRNA biogenesis (Zhang et al., 2013, 2014). Recently, it was found that mutants of MAC subunits (*mac7-1* and double *mac3a mac3b*) have reduced number of HYL1-containing D-bodies (Jia et al., 2017; Li et al., 2018). Authors suggest that it is a consequence of the fact that Microprocessor complex assembly requires pri-miRNA (Wu et al., 2013). Therefore decreased level of pri-miRNAs in *mac* mutants may affect HYL1 localization.

These assumptions are made based on the previous paper by Wu et al. (2013) who found the RNA-binding protein MOS2 (Modifier of Snc1) is supporting D-body assembly. In a *mos2-2* mutant, HYL1 localization was found to be different than that in WT plants as it was relatively homogeneous in the nucleus; however, HYL1 interactions with DCL1 and SE were not affected. MOS2 does not interact directly with core Microprocessor components but instead binds pri-miRNAs. Additionally, in the absence of MOS2, the association of HYL1 with pri-miRNAs is significantly reduced (Zhang et al., 2005). These data suggest that microRNA precursors may serve as scaffolds for D-body formation.

The balance between pri-miRNA and Microprocessor components assembly may be disturbed by over-accumulation of pre-mRNA splicing intermediates. Non-branching intron lariats sequester dicing complexes and negatively affect pri-miRNA processing. *Arabidopsis* mutant in debranching enzyme (*dbr1-2*) shows increased number of DCL1 and HYL1 nuclear bodies (Li et al., 2016).

The distribution of HYL1-GFP indicates that HYL1 is present in the nucleus and the cytoplasm (Han et al., 2004). HYL1 degradation in the cytoplasm is regulated by the RING-finger E3 ligase COP1 (Constitutive Photomorphogenic 1) (Cho et al., 2014). During the day, COP1 moves to the cytoplasm and indirectly protects HYL1 from degradation, most likely by inhibiting an undefined protease. During the night, COP1 remobilizes to the nucleus, allowing the protease to cleave the N-terminus of HYL1. This specific cleavage inhibits HYL1 function and causes an immediate reduction in correctly processed microRNAs (Cho et al., 2014).

A connection between light signaling and microRNA biogenesis comes from studies on Phytochrome Interacting Factor 4 (PIF4). It was shown that PIF4 promotes the



destabilization of both: DCL1 and HYL1 during dark-to-red-light transition (Sun et al., 2018).

Processing of *MIR* gene transcripts occurs in the nucleus, and effective cytoplasm–nucleus trafficking of Microprocessor components is necessary for its proper function. Recently, KETCH1 (Karyopherin Enabling the Transport of the Cytoplasmic HYL1) was described as an HYL1-interacting importin- $\beta$  protein (Zhang et al., 2017). *KETCH1* null mutants are embryo-lethal, whereas the downregulation of *KETCH1* using artificial microRNAs causes a reduction in HYL1 level in the nucleus, although SE localization in the nucleus is not affected. A decreased level of *KETCH1* leads to disturbances in microRNA production (accumulation of several pri-miRNAs as well as pre-miRNAs and reduced levels of mature microRNAs). An *amiR-ketch1* and *hyl1-2* double mutant was found to resemble the *hyl1-2* phenotype morphologically as well as at the microRNA level, which indicates that both proteins act in the same pathway. It is not known if *KETCH1* functions are limited to HYL1 cytoplasm–nucleus transport; however, the regulation of HYL1 level and its accessibility in different tissues and under different growth conditions might play an important role in the regulation of proper levels of miRNAs that are HYL1-dependent (Szarzynska et al., 2009).

## POST-TRANSLATIONAL MODIFICATIONS (PTMs) OF MICROPROCESSOR COMPONENTS

Most proteins require PTMs for proper function. More than 40 different post-translational protein modifications have been identified (Beck-Sickinger and Mörl, 2006), and they can play important roles in protein folding, subcellular localization, catalytic activity, or stability. For the plant Microprocessor, the phosphorylation of HYL1 is the only PTM that has been found to be crucial for efficient microRNA production. The interplay between protein phosphorylation and dephosphorylation is known to enable the rapid and efficient tuning of protein function. Using a forward genetic screen, Manavella et al. (2012) found that a mutation in the *CPL1* (C-Terminal Domain Phosphatase-like 1) gene causes impaired processing of microRNA precursors and aberrant strand selection during RISC loading. *CPL1* encodes a phosphatase that was shown to dephosphorylate the C-terminal domain (CTD) of the RNAPII largest subunit specifically at Ser5 (Koiwa et al., 2004). *In vivo*, *CPL1* interacts with two components of the plant Microprocessor, SE and HYL1 (Jeong et al., 2013). Two serine residues of HYL1, S42 and S159, are especially important for HYL1 function, and hyperphosphorylated HYL1 is inactive (Manavella et al., 2012). Both serine residues are located within the dsRNA binding domains of HYL1. Dephosphorylation of HYL1 by *CPL1* is stimulated by a protein called RCF3 (Regulator of CBF Gene Expression 3, also known as HOS5 or SHINY1 (Jiang et al., 2013; Chen et al., 2015; Karlsson et al., 2015). *RCF3* expression is restricted to the vegetative shoot apical meristem, young leaf primordia and newly emerging leaves, which suggests that fine-tuning of HYL1 activity via phosphorylation can be

tissue specific (Karlsson et al., 2015). Moreover, the expression of *RCF3* is reduced by salt, hyperosmotic stress, and ABA. This may indicate that plants modulate the phosphorylation status of HYL1 in response to environmental changes (Jiang et al., 2013). Another protein that dephosphorylates HYL1 is PP4 (Protein Phosphatase 4, also termed PPX) (Su et al., 2017). PP4 is a highly conserved protein among eukaryotes that functions to assist specific regulatory subunits (for example, SMEK1 in plants, PP4RS in mammals, and PSY2 in yeast (Gingras et al., 2005; Kataya et al., 2017; Su et al., 2017). In *Arabidopsis thaliana*, the PP4 phosphatase is encoded by two genes (*PP4-1* and *PP4-2*), the proteins of which share 93% sequence identity and have the same expression pattern, suggesting that their biological functions might be very similar if not redundant (Pujol et al., 2000). Attempts to obtain stable *A. thaliana* knockdown/out lines for the *PP4-1/2* genes have been unsuccessful; however, knockout mutants of the PP4 regulatory subunit SMEK1 (Suppressor of MEK1) are viable (Kataya et al., 2017; Su et al., 2017). In *smek1* mutants, microRNA expression levels are reduced due to the accelerated degradation of hyperphosphorylated HYL1. Importantly, SMEK1 protects HYL1 from degradation in a COP1- and light-independent manner; therefore, the regulation of HYL1 activity by PP4 represents another regulatory network present in plants (Su et al., 2017).

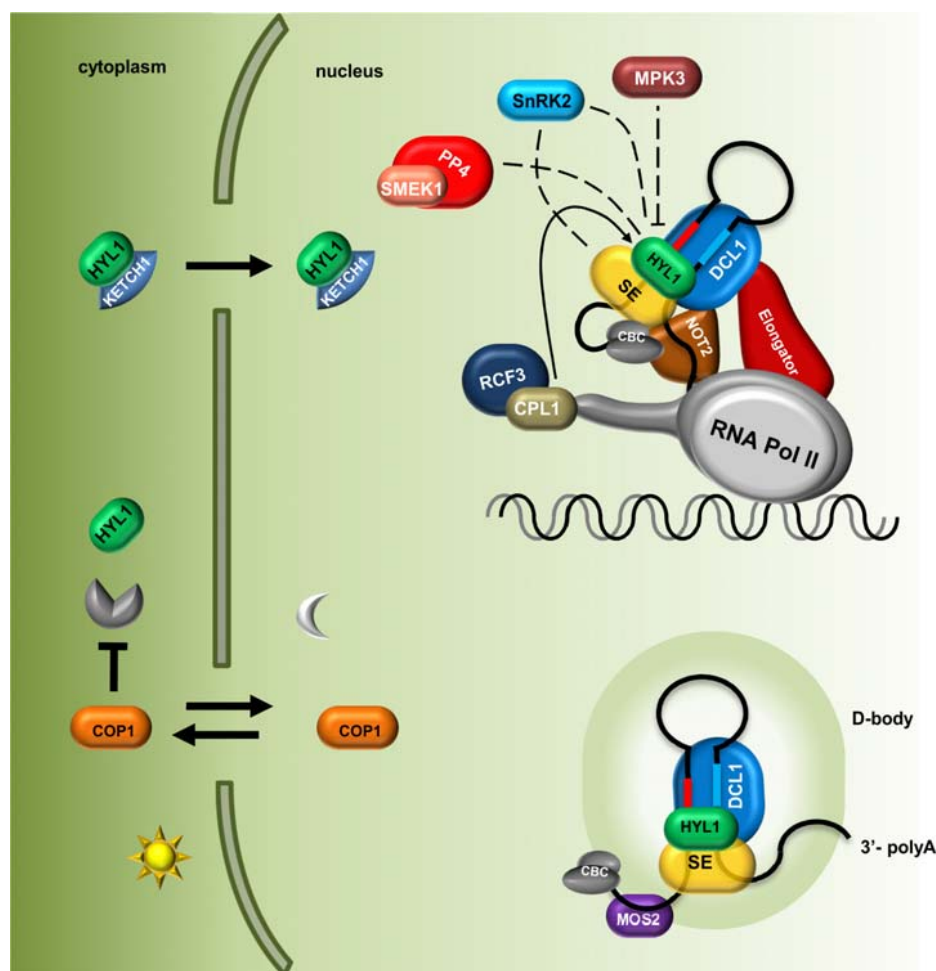
Beside phosphatases, kinases have also been found to be important for HYL1 phospho-regulation. HYL1 is phosphorylated by MPK3 (Mitogen-Activated Protein Kinase 3) and SnRK2 (SNF1-related protein kinase subfamily 2) (Raghuram et al., 2015; Yan et al., 2017). In *mpk3* mutant plants, HYL1 protein accumulates, and consequently, the levels of mature microRNAs are significantly higher than those in wild-type plants. Interestingly, the repertoires of small RNAs are affected in both *mpk3* and *cpl1* mutants, but they do not overlap with each other, which may suggest that *CPL1* (phosphatase) and MPK3 (kinase) act separately in parallel pathways (Raghuram et al., 2015). Furthermore, upon ABA treatment, MPK3 activation depends on the presence of HYL1 in the cell, which suggests that MPK3 and HYL1 are regulated in a feedback loop (Lu et al., 2002). The SnRK2 subfamily consists of 10 members (SnRK2.1–10) among which SnRK2.2/2.3/2.6 are involved in pri-miRNA processing and are strongly activated by ABA treatment. Surprisingly, in a *snrk2.2/3/6* triple mutant, the levels of HYL1 and mature microRNAs were found to be decreased (Yan et al., 2017). Yan et al. (2017) have shown that SnRK2.6 phosphorylates HYL1 *in vitro* and that SnRK2.2, SnRK2.3, and SnRK2.6 interact with HYL1 in plants. Many putative phosphorylation sites were found in the HYL1 protein (Supplementary Table S1); however, the precise amino acid residues that are phosphorylated in HYL1 by MPK3 and SnRK2.2/2.3/2.6 have not been identified. Thus, various HYL1 phosphorylation patterns might exert different functional effects on pri-miRNA biogenesis. Yan et al. (2017) showed that SnRK2.6 can phosphorylate the SE protein *in vitro* in addition to HYL1. The observation that SE can be phosphorylated in plants was reported previously (see chapter below) (Wang et al., 2013), but, currently, nothing is known about the effect of this modification on SE localization, stabilization, or activity. A model presenting

current knowledge on the role of PTMs in Microprocessor activity regulation and localization of its components is shown in the **Figure 1**.

## STRUCTURAL ASPECTS OF PTMs OF CORE MICROPROCESSOR COMPLEX PROTEINS

DCL1 is the largest protein in the Microprocessor complex. It contains several domains: a helicase domain at the N-terminus, a domain of unknown function 283, a PAZ domain, two catalytic RNase III domains and two dsRNA binding domains at the C-terminus. The HYL1 protein contains two dsRNA-binding domains (dsRBD1 and dsRBD2) at the N-terminus and six 28-amino acid imperfect repeats at the C-terminus.

The SERRATE protein possesses a core domain (195–543) that can be divided into three regions: an N-terminal alpha helical fragment, a middle domain fragment and a C2H2 zinc finger fragment. Both the N and C termini of SERRATE are predicted to be disordered in solution (Machida et al., 2011). The structures of the core domains of the *A. thaliana* SERRATE and HYL1 proteins and the second dsRBD domain of DCL1 have been determined (Rasia et al., 2010; Yang et al., 2010; Machida et al., 2011; Burdisso et al., 2012, 2014). Moreover, the structure of the dsRBD1 of HYL1 was solved as a complex containing a short 10-bp RNA duplex. Both dsRBD domains of HYL1 possess an alpha-beta-beta-beta-alpha fold, which is a signature of the dsRBD domain family (Masliah et al., 2013). Thus far, no data regarding how PTMs might interfere with the structure of Microprocessor proteins have been reported. Serine 42 is located within the loop between beta strand 1 and beta strand 2 of the HYL1 dsRBD1.

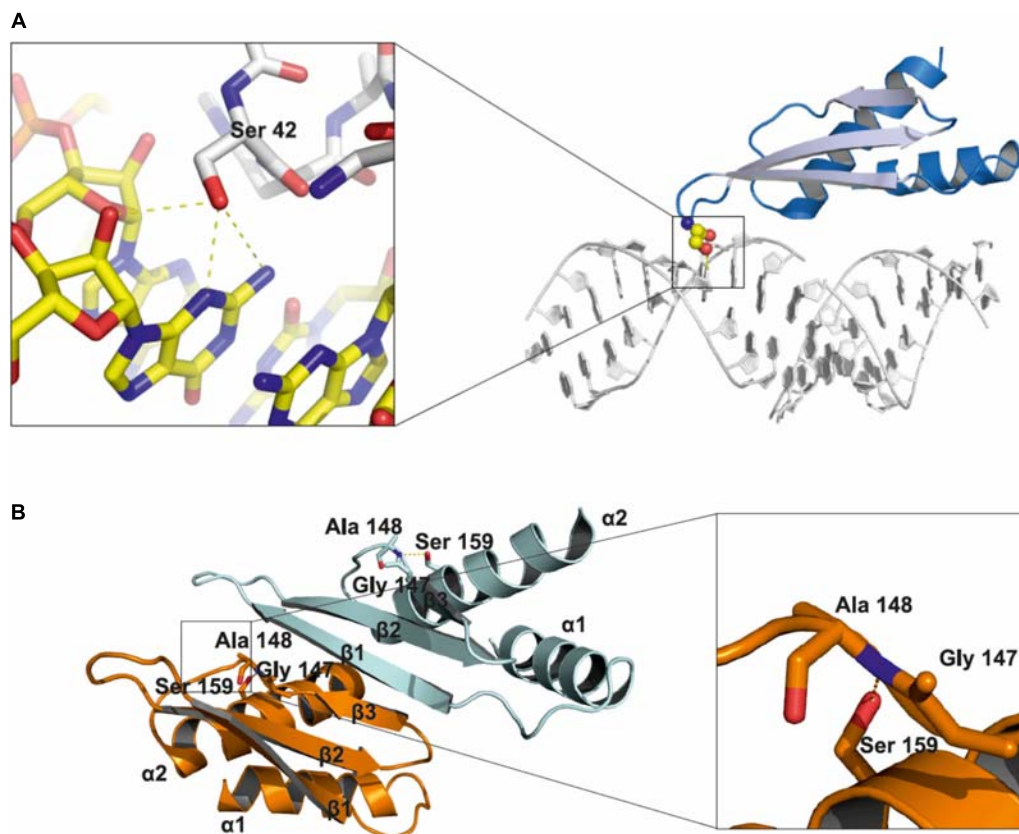


**FIGURE 1 |** Plant microRNA biogenesis is fine-tuned via regulation of Microprocessor components localization and their post-translational modifications. The level of HYL1 in the cytoplasm is indirectly regulated by COP1 in a light-dependent manner; HYL1 nuclear import is mediated by KETCH1 importin. Core components of the Microprocessor: DCL1, SE and HYL1 as well as microRNA precursors are located and interact with each other in the D-bodies. MOS2 is important for the D-bodies assembly however, does not interact directly with the Microprocessor. The Elongator complex and NOT2 proteins form a bridge coupling transcription and processing of microRNAs. Phosphorylation status of HYL1 is crucial for its efficient function and is a consequence of an interplay between kinases (MPK3 and SNRK2) and phosphatases (CPL1 and PP4). Direct influence of SnRK2 on SE and HYL activity is not known. CBC, nuclear Cap Binding Complex; dashed lines indicate that exact localization of the interactions is unknown.

Using the published structure, we noticed that the side chain of serine 42 may interact with the minor groove of the dsRNA and may form hydrogen bonds with the N2 and N3 nitrogen atoms of guanine and the 2' hydroxyl group of the ribose ring (**Figure 2A**). A bulky phosphate group attached to serine 42 could potentially interfere with the minor groove interactions and negatively regulate binding to RNA duplexes. Serine 159 is localized in alpha helix 2 of the HYL1 dsRBD2, and it may interact with the loop between beta strand 3 and alpha helix 2. In addition, this serine may form a hydrogen bond with the nitrogen from the peptide bond between glycine 147 and alanine 148 in the opposite loop of the same HYL1 molecule (**Figure 2B**). Yang et al. (2014) found that a G147E mutation significantly reduces dimerization of the HYL1 protein. Moreover, this substitution was present in a *hyl1-3* mutant (Manavella et al., 2012). Therefore, serine 159 phosphorylation could destabilize the hydrogen bond network, leading to the mislocation of beta strand 3, the disruption of the dimerization interface and, ultimately, the inactivation of HYL1.

From high-throughput studies of the *A. thaliana* phosphoproteome, several phosphorylation sites in the SERRATE protein were identified and deposited in the PhosPhAt 4.0 database (Durek et al., 2010). The phosphorylated residues are mostly

located within the N-terminal region, a low complexity domain rich in proline and serine residues, and within the middle fragment from the core structure domain. In addition, one phosphorylation site was found within the C-terminal fragment. A list of the phosphorylation sites found in the SERRATE and HYL1 proteins is shown in Supplementary Table S1. The N-terminal fragment of SERRATE (amino acids 1–194) is predicted to be disordered in solution. Additionally, the amino acid sequence conservation in this region is relatively low in comparison to that of the core domain (195–543). Nevertheless, the overall amino acid composition between SERRATE proteins from different species is very similar. Phosphorylation of low complexity domains has been shown to affect aggregation, local structure and protein-protein interactions (Kwon et al., 2013; Monahan et al., 2017). The N-terminal domain of SERRATE is responsible for its interaction with the two U1 small nuclear ribonucleoprotein particle (U1 snRNP) proteins, PRP40b and PRP40a, and deletion of the N-terminal domain causes more homogeneous nuclear localization in comparison to the speckle-like localization of the wild-type protein (Knop et al., 2017). Thus, the phosphorylation status of this domain in SERRATE may affect interactions with a large number of interacting partners in different processes in which SE is involved.



**FIGURE 2 |** Structural insights into the role of serine 42 and serine 159 residues in the HYL1 protein based on the crystal structure from Yang et al. (2010).

**(A)** Interactions of non-phosphorylated serine 42 with minor groove of dsRNA. **(B)** Stabilization of beta strand 3 by the hydrogen bond between non-phosphorylated serine 159 and peptide bond between glycine 147 and alanine 148.



## NEGATIVE FEEDBACK REGULATION OF MICROPROCESSOR

To provide fine tuning of microRNA production and to maintain balance in mRNA target degradation, Microprocessor components are regulated at the post-transcriptional level by a negative feedback loop. This feedback regulation of Microprocessor was first shown in 2003 by Carrington group. Xie et al. (2003) showed that *DCL1* mRNA level is regulated by miRNA162. In wild-type plants, the *DCL1* transcript is in relatively low-abundance because functional *DCL1* catalyzes miRNA production, and miRNA162 targets the *DCL1* transcript for degradation. However, in mutants with impaired miRNA biogenesis (for example *dcl1-7*, *hen1-1*), increased level of *DCL1* mRNA has been detected. Moreover, plants expressing the P1/HC-pro protein [Turnip mosaic virus (TuMV) RNA silencing suppressor], which inhibits the small RNA-guided cleavage of RNAs, had increased *DCL1* mRNA level (Xie et al., 2003). The abundance of the *DCL1* transcript is also regulated by the production of miRNA838, which is encoded within the 14th intron of the *DCL1* pre-mRNA. The generation of this miRNA is a consequence of *DCL1* pre-mRNA cleavage into two non-functional transcripts that are 4- and 2.5-kb in length (Rajagopalan et al., 2006). Rajagopalan et al. (2006) suggested that higher level of *DCL1* protein results in more efficient processing of *DCL1* primary transcripts by Microprocessor than its recognition by the Spliceosome, which results in a higher level of miR838 and a lower level of *DCL1* transcript in the cell. Similar to *DCL1*, the SE level is determined by a negative feedback loop that involves miR863-3p. This microRNA targets the 3' UTR of SE mRNA as well as two negative regulators of plant

defense: ARLPK1 and ARPLK2. The level of miR863-3p increases after bacterial infection and silences two negative regulators of plant defense by cleaving their mRNAs. At subsequent steps of infection, when at its highest level, miR863-3p inhibits the translation of SE mRNA, which results in lower efficiency of miRNA biogenesis and decreased miRNA levels, including those of miR863-3p (Niu et al., 2016).

## AUTHOR CONTRIBUTIONS

JD, MT, MB, AJ, ZS-K, and DB participated in preparation of draft manuscript. MT and JD prepared figures. JD, ZS-K, and DB participated in assembly and editing of the final manuscript.

## FUNDING

This work was supported by the KNOW RNA Research Centre in Poznan (Grant No. 01/KNOW2/2014) and the National Science Center projects UMO-2016/23/D/NZ1/00152 (DB), UMO-2017/25/BNZ1/00603 (JD), UMO-2013/10/A/NZ1/00557 (AJ), UMO-2016/23/B/NZ9/00862 (ZS-K), and UMO-2014/13/N/NZ1/00049 (MB).

## SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fpls.2018.00753/full#supplementary-material>

## REFERENCES

- Ali, G. S., Golovkin, M., and Reddy, A. S. (2003). Nuclear localization and in vivo dynamics of a plant-specific serine/arginine-rich protein. *Plant J.* 36, 883–893. doi: 10.1046/j.1365-3113X.2003.01932.x
- Barciszewska-Pacak, M., Knop, K., Jarmolowski, A., and Szweykowska-Kulinska, Z. (2016). *Arabidopsis thaliana* microRNA162 level is posttranscriptionally regulated via splicing and polyadenylation site selection. *Acta Biochim. Pol.* 63, 811–816. doi: 10.18388/abp.2016\_1349
- Barciszewska-Pacak, M., Milanowska, K., Knop, K., Bielewicz, D., Nuc, P., Plewka, P., et al. (2015). *Arabidopsis* microRNA expression regulation in a wide range of abiotic stress responses. *Front. Plant Sci.* 6:410. doi: 10.3389/fpls.2015.00410
- Beck-Sickinger, A. G., and Mörl, K. (2006). Posttranslational modification of proteins. Expanding Nature's Inventory. By Christopher T. Walsh. *Angew. Chem. Int. Ed.* 45, 1020–1020. doi: 10.1002/anie.200585363
- Ben Chaabane, S., Liu, R., Chinnusamy, V., Kwon, Y., Park, J. H., Kim, S. Y., et al. (2013). STA1, an *Arabidopsis* pre-mRNA processing factor 6 homolog, is a new player involved in miRNA biogenesis. *Nucleic Acids Res.* 41, 1984–1997. doi: 10.1093/nar/gks1309
- Bielewicz, D., Kalak, M., Kalyna, M., Windels, D., Barta, A., Vazquez, F., et al. (2013). Introns of plant pri-miRNAs enhance miRNA biogenesis. *EMBO Rep.* 14, 622–628. doi: 10.1038/embor.2013.62
- Burdisso, P., Milia, F., Schapire, A. L., Bologna, N. G., Palatnik, J. F., and Rasia, R. M. (2014). Structural determinants of *Arabidopsis thaliana* Hyponastic leaves 1 function in vivo. *PLoS One* 9:e113243. doi: 10.1371/journal.pone.0113243
- Burdisso, P., Suarez, I. P., Bologna, N. G., Palatnik, J. F., Bersch, B., and Rasia, R. M. (2012). Second double-stranded RNA binding domain of dicer-like ribonuclease 1: structural and biochemical characterization. *Biochemistry* 51, 10159–10166. doi: 10.1021/bi301247r
- Chen, T., Cui, P., and Xiong, L. (2015). The RNA-binding protein HOS5 and serine/arginine-rich proteins RS40 and RS41 participate in miRNA biogenesis in *Arabidopsis*. *Nucleic Acids Res.* 43, 8283–8298. doi: 10.1093/nar/gkv751
- Cho, S. K., Ben Chaabane, S., Shah, P., Poulsen, C. P., and Yang, S. W. (2014). COP1 E3 ligase protects HYL1 to retain microRNA biogenesis. *Nat. Commun.* 5:5867. doi: 10.1038/ncomms6867
- Denli, A. M., Tops, B. B., Plasterk, R. H., Ketting, R. F., and Hannon, G. J. (2004). Processing of primary microRNAs by the microprocessor complex. *Nature* 432, 231–235. doi: 10.1038/nature03049
- Dolata, J., Bajczyk, M., Bielewicz, D., Niedojadlo, K., Niedojadlo, J., Pietrykowska, H., et al. (2016). salt stress reveals a new role for ARGONAUTE1 in miRNA biogenesis at the transcriptional and posttranscriptional levels. *Plant Physiol.* 172, 297–312. doi: 10.1104/pp.16.00830
- Durek, P., Schmidt, R., Heazlewood, J. L., Jones, A., Maclean, D., Nagel, A., et al. (2010). PhosPhAt: the *Arabidopsis thaliana* phosphorylation site database. An update. *Nucleic Acids Res.* 38, D828–D834. doi: 10.1093/nar/gkp810
- Fabrizio, P., Dannenberg, J., Dube, P., Kastner, B., Stark, H., Urlaub, H., et al. (2009). The evolutionarily conserved core design of the catalytic activation step of the yeast spliceosome. *Mol. Cell* 36, 593–608. doi: 10.1016/j.molcel.2009.09.040
- Fang, X., Cui, Y., Li, Y., and Qi, Y. (2015). Transcription and processing of primary microRNAs are coupled by Elongator complex in *Arabidopsis*. *Nat. Plants* 1:15075. doi: 10.1038/nplants.2015.75
- Fang, Y., and Spector, D. L. (2007). Identification of nuclear dicing bodies containing proteins for microRNA biogenesis in living *Arabidopsis* plants. *Curr. Biol.* 17, 818–823. doi: 10.1016/j.cub.2007.04.005



- Gingras, A. C., Caballero, M., Zarske, M., Sanchez, A., Hazbun, T. R., Fields, S., et al. (2005). A novel, evolutionarily conserved protein phosphatase complex involved in cisplatin sensitivity. *Mol. Cell. Proteomics* 4, 1725–1740. doi: 10.1074/mcp.M500231-MCP200
- Gregory, R. I., Yan, K. P., Amuthan, G., Chendrimada, T., Doratotaj, B., Cooch, N., et al. (2004). The Microprocessor complex mediates the genesis of microRNAs. *Nature* 432, 235–240. doi: 10.1038/nature03120
- Han, M. H., Goud, S., Song, L., and Fedoroff, N. (2004). The Arabidopsis double-stranded RNA-binding protein HYL1 plays a role in microRNA-mediated gene regulation. *Proc. Natl. Acad. Sci. U.S.A.* 101, 1093–1098. doi: 10.1073/pnas.0307969100
- Jeong, I. S., Aksoy, E., Fukudome, A., Akhter, S., Hiraguri, A., Fukuhara, T., et al. (2013). Arabidopsis C-terminal domain phosphatase-like 1 functions in miRNA accumulation and DNA methylation. *PLoS One* 8:e74739. doi: 10.1371/journal.pone.0074739
- Jia, T., Zhang, B., You, C., Zhang, Y., Zeng, L., Li, S., et al. (2017). The Arabidopsis MOS4-associated complex promotes microRNA biogenesis and precursor messenger RNA splicing. *Plant Cell* 29, 2626–2643. doi: 10.1105/tpc.17.00370
- Jiang, J., Wang, B., Shen, Y., Wang, H., Feng, Q., and Shi, H. (2013). The Arabidopsis RNA binding protein with K homology motifs, SHINY1, interacts with the C-terminal domain phosphatase-like 1 (CPL1) to repress stress-inducible gene expression. *PLoS Genet.* 9:e1003625. doi: 10.1371/journal.pgen.1003625
- Karlsson, P., Christie, M. D., Seymour, D. K., Wang, H., Wang, X., Hagmann, J., et al. (2015). KH domain protein RCF3 is a tissue-biased regulator of the plant miRNA biogenesis cofactor HYL1. *Proc. Natl. Acad. Sci. U.S.A.* 112, 14096–14101. doi: 10.1073/pnas.1512865112
- Kataya, A. R. A., Creighton, M. T., Napitupulu, T. P., Saetre, C., Heidari, B., Ruoff, P., et al. (2017). PLATINUM SENSITIVE 2 LIKE impacts growth, root morphology, seed set, and stress responses. *PLoS One* 12:e0180478. doi: 10.1371/journal.pone.0180478
- Knop, K., Stepien, A., Barciszewska-Pacak, M., Taube, M., Bielewicz, D., Michalak, M., et al. (2017). Active 5' splice sites regulate the biogenesis efficiency of Arabidopsis microRNAs derived from intron-containing genes. *Nucleic Acids Res.* 45, 2757–2775.
- 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. *Proc. Natl. Acad. Sci. U.S.A.* 101, 14539–14544. doi: 10.1073/pnas.0403174101
- Kruk, J. A., Dutta, A., Fu, J., Gilmour, D. S., and Reese, J. C. (2011). The multifunctional Ccr4-Not complex directly promotes transcription elongation. *Genes Dev.* 25, 581–593. doi: 10.1101/gad.2020911
- Kwon, I., Kato, M., Xiang, S., Wu, L., Theodoropoulos, P., Mirzaei, H., et al. (2013). Phosphorylation-regulated binding of RNA polymerase II to fibrous polymers of low-complexity domains. *Cell* 155, 1049–1060. doi: 10.1016/j.cell.2013.10.033
- Li, S., Liu, K., Zhou, B., Li, M., Zhang, S., Zeng, L., et al. (2018). MAC3A and mac3b, two core subunits of the mos4-associated complex, positively influence miRNA biogenesis. *Plant Cell* 30, 481–494. doi: 10.1105/tpc.17.00953
- Li, Z., Wang, S., Cheng, J., Su, C., Zhong, S., Liu, Q., et al. (2016). Intron lariat RNA inhibits microRNA biogenesis by sequestering the dicing complex in Arabidopsis. *PLoS Genet.* 12:e1006422. doi: 10.1371/journal.pgen.1006422
- Liu, C., Xin, Y., Xu, L., Cai, Z., Xue, Y., Liu, Y., et al. (2017). Arabidopsis ARGONAUTE 1 binds chromatin to promote gene transcription in response to hormones and stresses. *Dev. Cell* 44, 348.e7–361.e7. doi: 10.1016/j.devcel.2017.12.002
- Lu, C., Han, M. H., Guevara-Garcia, A., and Fedoroff, N. V. (2002). Mitogen-activated protein kinase signaling in postgermination arrest of development by abscisic acid. *Proc. Natl. Acad. Sci. U.S.A.* 99, 15812–15817. doi: 10.1073/pnas.242607499
- Machida, S., Chen, H. Y., and Adam Yuan, Y. (2011). Molecular insights into miRNA processing by *Arabidopsis thaliana* SERRATE. *Nucleic Acids Res.* 39, 7828–7836. doi: 10.1093/nar/gkr428
- Manavella, P. A., Hagmann, J., Ott, F., Laubinger, S., Franz, M., Macek, B., et al. (2012). Fast-forward genetics identifies plant CPL phosphatases as regulators of miRNA processing factor HYL1. *Cell* 151, 859–870. doi: 10.1016/j.cell.2012.09.039
- Masliyah, G., Barraud, P., and Allain, F. H. (2013). RNA recognition by double-stranded RNA binding domains: a matter of shape and sequence. *Cell Mol. Life Sci.* 70, 1875–1895. doi: 10.1007/s00018-012-1119-x
- Monaghan, J., Xu, F., Gao, M., Zhao, Q., Palma, K., Long, C., et al. (2009). Two Prp19-like U-box proteins in the MOS4-associated complex play redundant roles in plant innate immunity. *PLoS Pathog.* 5:e1000526. doi: 10.1371/journal.ppat.1000526
- Monahan, Z., Ryan, V. H., Janke, A. M., Burke, K. A., Rhoads, S. N., Zerbe, G. H., et al. (2017). Phosphorylation of the FUS low-complexity domain disrupts phase separation, aggregation, and toxicity. *EMBO J.* 36, 2951–2967. doi: 10.15252/embj.201696394
- Nelissen, H., De Groeve, S., Fleury, D., Neyt, P., Bruno, L., Bitonti, M. B., et al. (2010). Plant Elongator regulates auxin-related genes during RNA polymerase II transcription elongation. *Proc. Natl. Acad. Sci. U.S.A.* 107, 1678–1683. doi: 10.1073/pnas.0913559107
- Niu, D., Lii, Y. E., Chellappan, P., Lei, L., Peralta, K., Jiang, C., et al. (2016). miRNA863-3p sequentially targets negative immune regulator ARLPKs and positive regulator SERRATE upon bacterial infection. *Nat. Commun.* 7:11324. doi: 10.1038/ncomms11324
- Otero, G., Fellows, J., Li, Y., De Bizemont, T., Dirac, A. M., Gustafsson, C. M., et al. (1999). Elongator, a multisubunit component of a novel RNA polymerase II holoenzyme for transcriptional elongation. *Mol. Cell* 3, 109–118. doi: 10.1016/S1097-2765(00)80179-3
- Palma, K., Zhao, Q., Cheng, Y. T., Bi, D., Monaghan, J., Cheng, W., et al. (2007). Regulation of plant innate immunity by three proteins in a complex conserved across the plant and animal kingdoms. *Genes Dev.* 21, 1484–1493. doi: 10.1101/gad.1559607
- Pujol, G., Baskin, T. I., Casamayor, A., Cortadellas, N., Ferrer, A., and Arino, J. (2000). The *Arabidopsis thaliana* PPX/PP4 phosphatases: molecular cloning and structural organization of the genes and immunolocalization of the proteins to plastids. *Plant Mol. Biol.* 44, 499–511. doi: 10.1023/A:1026587405656
- Raghuram, B., Sheikh, A. H., Rustagi, Y., and Sinha, A. K. (2015). MicroRNA biogenesis factor DRB1 is a phosphorylation target of mitogen activated protein kinase MPK3 in both rice and Arabidopsis. *FEBS J.* 282, 521–536. doi: 10.1111/febs.13159
- Rajagopalan, R., Vaucheret, H., Trejo, J., and Bartel, D. P. (2006). A diverse and evolutionarily fluid set of microRNAs in *Arabidopsis thaliana*. *Genes Dev.* 20, 3407–3425. doi: 10.1101/gad.1476406
- Rasia, R. M., Mateos, J., Bologna, N. G., Burdisso, P., Imbert, L., Palatnik, J. F., et al. (2010). Structure and RNA interactions of the plant MicroRNA processing-associated protein HYL1. *Biochemistry* 49, 8237–8239. doi: 10.1021/bi100672x
- Stepien, A., Knop, K., Dolata, J., Taube, M., Bajczyk, M., Barciszewska-Pacak, M., et al. (2017). Posttranscriptional coordination of splicing and miRNA biogenesis in plants. *Wiley Interdiscip. Rev. RNA* 8:e1403. doi: 10.1002/wrna.1403
- Su, C., Li, Z., Cheng, J., Li, L., Zhong, S., Liu, L., et al. (2017). The protein phosphatase 4 and SMEK1 complex dephosphorylates HYL1 to promote miRNA biogenesis by antagonizing the MAPK cascade in *Arabidopsis*. *Dev. Cell* 41, 527.e5–539.e5. doi: 10.1016/j.devcel.2017.05.008
- Sun, Z., Guo, T., Liu, Y., Liu, Q., and Fang, Y. (2015). The roles of Arabidopsis CDF2 in transcriptional and posttranscriptional regulation of primary microRNAs. *PLoS Genet.* 11:e1005598. doi: 10.1371/journal.pgen.1005598
- Sun, Z., Li, M., Zhou, Y., Guo, T., Liu, Y., Zhang, H., et al. (2018). Coordinated regulation of Arabidopsis microRNA biogenesis and red light signaling through Dicer-like 1 and phytochrome-interacting factor 4. *PLoS Genet.* 14:e1007247. doi: 10.1371/journal.pgen.1007247
- Szarzynska, B., Sobkowiak, L., Pant, B. D., Balazadeh, S., Scheible, W. R., Mueller-Roeber, B., et al. (2009). Gene structures and processing of *Arabidopsis thaliana* HYL1-dependent pri-miRNAs. *Nucleic Acids Res.* 37, 3083–3093. doi: 10.1093/nar/gkp189
- Wang, L., Song, X., Gu, L., Li, X., Cao, S., Chu, C., et al. (2013). NOT2 proteins promote polymerase II-dependent transcription and interact with multiple MicroRNA biogenesis factors in *Arabidopsis*. *Plant Cell* 25, 715–727. doi: 10.1105/tpc.112.105882

- Wu, X., Shi, Y., Li, J., Xu, L., Fang, Y., Li, X., et al. (2013). A role for the RNA-binding protein MOS2 in microRNA maturation in *Arabidopsis*. *Cell Res.* 23, 645–657. doi: 10.1038/cr.2013.23
- Xie, Z., Allen, E., Fahlgren, N., Calamar, A., Givan, S. A., and Carrington, J. C. (2005). Expression of *Arabidopsis* MIRNA genes. *Plant Physiol.* 138, 2145–2154. doi: 10.1104/pp.105.062943
- Xie, Z., Kasschau, K. D., and Carrington, J. C. (2003). Negative feedback regulation of *Dicer-Like1* in *Arabidopsis* by microRNA-guided mRNA degradation. *Curr. Biol.* 13, 784–789. doi: 10.1016/S0960-9822(03)00281-1
- Yan, J., Wang, P., Wang, B., Hsu, C. C., Tang, K., Zhang, H., et al. (2017). The SnRK2 kinases modulate miRNA accumulation in *Arabidopsis*. *PLoS Genet.* 13:e1006753. doi: 10.1371/journal.pgen.1006753
- Yang, S. W., Chen, H. Y., Yang, J., Machida, S., Chua, N. H., and Yuan, Y. A. (2010). Structure of *Arabidopsis* HYPONASTIC LEAVES1 and its molecular implications for miRNA processing. *Structure* 18, 594–605. doi: 10.1016/j.str.2010.02.006
- Yang, X., Ren, W., Zhao, Q., Zhang, P., Wu, F., and He, Y. (2014). Homodimerization of HYL1 ensures the correct selection of cleavage sites in primary miRNA. *Nucleic Acids Res.* 42, 12224–12236. doi: 10.1093/nar/gku907
- Yu, Y., Jia, T., and Chen, X. (2017). The 'how' and 'where' of plant microRNAs. *New Phytol.* 216, 1002–1017. doi: 10.1111/nph.14834
- Zhang, S., Liu, Y., and Yu, B. (2014). PRL1, an RNA-binding protein, positively regulates the accumulation of miRNAs and siRNAs in *Arabidopsis*. *PLoS Genet.* 10:e1004841. doi: 10.1371/journal.pgen.1004841
- Zhang, S., Liu, Y., and Yu, B. (2015). New insights into pri-miRNA processing and accumulation in plants. *Wiley Interdiscip. Rev. RNA* 6, 533–545. doi: 10.1002/wrna.1292
- Zhang, S., Xie, M., Ren, G., and Yu, B. (2013). CDC5, a DNA binding protein, positively regulates posttranscriptional processing and/or transcription of primary microRNA transcripts. *Proc. Natl. Acad. Sci. U.S.A.* 110, 17588–17593. doi: 10.1073/pnas.1310644110
- Zhang, Y., Cheng, Y. T., Bi, D., Palma, K., and Li, X. (2005). MOS2, a protein containing G-patch and KOW motifs, is essential for innate immunity in *Arabidopsis thaliana*. *Curr. Biol.* 15, 1936–1942. doi: 10.1016/j.cub.2005.09.038
- Zhang, Z., Guo, X., Ge, C., Ma, Z., Jiang, M., Li, T., et al. (2017). KETCH1 imports HYL1 to nucleus for miRNA biogenesis in *Arabidopsis*. *Proc. Natl. Acad. Sci. U.S.A.* 114, 4011–4016. doi: 10.1073/pnas.1619755114

**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 © 2018 Dolata, Taube, Bajczyk, Jarmolowski, Szwejkowska-Kulinska and Bielewicz. 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 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.



# miRVIT: A Novel miRNA Database and Its Application to Uncover *Vitis* Responses to *Flavescence dorée* Infection

Walter Chitarra<sup>1,2†</sup>, Chiara Pagliarani<sup>1†</sup>, Simona Abbà<sup>1</sup>, Paolo Boccacci<sup>1</sup>, Giancarlo Birello<sup>3</sup>, Marika Rossi<sup>1</sup>, Sabrina Palmano<sup>1</sup>, Cristina Marzachi<sup>1</sup>, Irene Perrone<sup>1\*</sup> and Giorgio Gambino<sup>1</sup>

<sup>1</sup> Institute for Sustainable Plant Protection, National Research Council of Italy, Turin, Italy, <sup>2</sup> Viticultural and Enology Research Centre, Council for Agricultural Research and Economics, Conegliano, Italy, <sup>3</sup> Research Institute on Sustainable Economic Growth, National Research Council of Italy, Turin, Italy

## OPEN ACCESS

### Edited by:

Ana Confraria,  
Instituto Gulbenkian de Ciência (IGC),  
Portugal

### Reviewed by:

Marina Dermastia,  
National Institute of Biology (NIB),  
Slovenia  
Alfredo Pulvirenti,  
Università degli Studi di Catania, Italy

### \*Correspondence:

Irene Perrone  
irene.perrone@ipsp.cnr.it

<sup>†</sup> These authors have contributed  
equally to this work.

### Specialty section:

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

**Received:** 02 May 2018

**Accepted:** 26 June 2018

**Published:** 17 July 2018

### Citation:

Chitarra W, Pagliarani C, Abbà S,  
Boccacci P, Birello G, Rossi M,  
Palmano S, Marzachi C, Perrone I  
and Gambino G (2018) miRVIT:  
A Novel miRNA Database and Its  
Application to Uncover *Vitis*  
Responses to *Flavescence dorée*  
Infection. *Front. Plant Sci.* 9:1034.  
doi: 10.3389/fpls.2018.01034

Micro(mi)RNAs play crucial roles in plant developmental processes and in defense responses to biotic and abiotic stresses. In the last years, many works on small RNAs in grapevine (*Vitis* spp.) were published, and several conserved and putative novel grapevine-specific miRNAs were identified. In order to reorganize the high quantity of available data, we produced “miRVIT,” the first database of all novel grapevine miRNA candidates characterized so far, and still not deposited in miRBase. To this aim, each miRNA accession was renamed, repositioned in the last version of the grapevine genome, and compared with all the novel and conserved miRNAs detected in grapevine. Conserved and novel miRNAs cataloged in miRVIT were then used for analyzing *Vitis vinifera* plants infected by *Flavescence dorée* (FD), one of the most severe phytoplasma diseases affecting grapevine. The analysis of small RNAs from healthy, recovered (plants showing spontaneous and stable remission of symptoms), and FD-infected “Barbera” grapevines showed that FD altered the expression profiles of several miRNAs, including those involved in cell development and photosynthesis, jasmonate signaling, and disease resistance response. The application of miRVIT in a biological context confirmed the effectiveness of the followed approach, especially for the identification of novel miRNA candidates in grapevine. miRVIT database is available at <http://mirvit.ipsp.cnr.it>.

**Highlights:** The application of the newly produced database of grapevine novel miRNAs to the analysis of plants infected by *Flavescence dorée* reveals key roles of miRNAs in photosynthesis and jasmonate signaling.

**Keywords:** disease resistance, jasmonate, miRNAs, phytoplasma, target genes, univocal database

## INTRODUCTION

RNA silencing or RNA interference is a pivotal mechanism of gene regulation conserved in a broad range of eukaryotic organisms, including fungi, plants, and animals (Baulcombe, 2002). Small-endogenous RNAs are important effectors of RNA silencing phenomena and are classified into two major categories based on the nature of their RNA precursors: microRNAs (miRNAs)

and small-interfering RNAs (Axtell, 2013). miRNAs are single-stranded RNA molecules of approximately 21 nt length generated from endogenous *MIR* genes (Nozawa et al., 2012). The precursors of miRNAs were processed through DICER-LIKE 1 proteins (DCL1) to release a mature miRNA:miRNA\* duplex (Kurihara and Watanabe, 2004). The mature miRNA loaded into a specific ARGONAUTE (AGO)-associated RNA-induced silencing complex (RISC) guides the identification and cleavage of complementary mRNA targets in a sequence-specific manner (Mallory et al., 2008). The miRNA guide strand is generally more abundant than the miRNA star strand (miRNA\*) and is responsible for the RISC-mediated target regulation. However, increasing evidence also suggests an association between miRNA\* and AGO proteins exerting relevant biological functions (Liu et al., 2017). miRNAs are involved in several developmental processes, in genome stability maintenance, and in plant adaptation to biotic and abiotic stresses, as reviewed in many studies (Sunkar et al., 2012; Staiger et al., 2013; Borges and Martienssen, 2015).

Among pathogens infecting plants, phytoplasmas, phloem-limited bacteria belonging to the class of *Mollicutes*, cause serious yield and economic losses in many crops (Bendix and Lewis, 2018). In *Vitis vinifera*, *Flavescence dorée* (FD) and *Bois noir* (BN) are associated with grapevine yellows, the most important and damaging phytoplasma-induced diseases in Europe. In particular, the quarantine disease FD is caused by a phytoplasma (FDp) of the elm yellows group (16SrV-C and V-D; Angelini et al., 2001) transmitted in a persistent and propagative manner by the ampelophagous leafhopper *Scaphoideus titanus* Ball (*Hemiptera Cicadellidae*). Typical symptomatology observed in FDp-infected grapevines generally includes downward rolling of leaves, yellowing or reddening of the leaves according to white or red grape varieties, drying of inflorescences and bunches, shortening of internodes, and lack of cane lignification (Caudwell, 1990). In some cases, the spontaneous and stable remission of FD symptoms accompanied by the disappearance of FDp from the plants may occur, giving rise to “recovery” phenomena (Osler et al., 2004; Maggi et al., 2017). Molecular and physiological changes in grapevine infected and recovered by FDp have recently been investigated (Musetti et al., 2007; Gambino et al., 2013; Margaria et al., 2013, 2014; Vitali et al., 2013; Prezelj et al., 2016); nevertheless, the biological mechanisms underlying these processes are not yet fully understood. Although for some plant-phytoplasma systems, the role of miRNAs in disease development appeared clear (Ehya et al., 2013; Gai et al., 2014, 2018; Minato et al., 2014; Shao et al., 2016; Snyman et al., 2017), no data are currently available in the case of FDp–grapevine interaction.

In the last few years, from the first report concerning the identification of novel miRNAs in grapevine (Carra et al., 2009), several works have been published on this topic, and many conserved and putative novel miRNAs were identified. Different species from the *Vitis* genus were analyzed, the most studied being several cultivars of *Vitis vinifera* (Carra et al., 2009; Pantaleo et al., 2010, 2016; Alabi et al., 2012; Belli Kullán et al., 2015; Sun et al., 2015; Paim Pinto et al., 2016; Bester et al., 2017a,b; Pagliarani et al., 2017; Snyman et al., 2017) and the “Summer

Black” hybrid of *V. vinifera* × *Vitis labrusca* (Wang et al., 2011, 2014; Han et al., 2014; Zhao et al., 2017). In addition, some miRNAs derived from *Vitis amurensis* (Wang et al., 2012) and from a powdery mildew-resistant accession of the wild Chinese *Vitis pseudoreticulata* Baihe-35-1 (Han et al., 2016) were characterized as well. Substantially, all grapevine tissues in several developmental stages (Belli Kullán et al., 2015) and from plants subjected to biotic (Alabi et al., 2012; Singh et al., 2012; Pantaleo et al., 2016; Bester et al., 2017a; Snyman et al., 2017) or abiotic stresses were analyzed (Sun et al., 2015; Pantaleo et al., 2016; Pagliarani et al., 2017). Nevertheless, in these studies, there is no uniformity in the classification of the identified novel miRNAs: hundreds of putative novel miRNAs were indicated in different ways, and often the same miRNA was named differently in different works.

In this study, we classified all putative novel miRNAs discovered in grapevine so far, and not deposited in the public database miRBase, into a single database, namely, miRVIT (available at <http://mirvit.ipsp.cnr.it>), in order to standardize their classification by assigning a new and univocal denomination. The obtained database was then used to analyze nine small RNA libraries produced from leaf midribs collected from FDp-infected (FD), recovered (R), and healthy (H) plants of the highly susceptible *V. vinifera* cv. Barbera. The expression dynamics of conserved, novel miRNAs and their target transcripts associated with FDp infection and/or recovery processes were investigated.

## MATERIALS AND METHODS

### Construction of a Single Database of Novel miRNAs From Grapevine

The data related to all the novel miRNAs identified so far in grapevine were collected from 18 already published works (Carra et al., 2009; Pantaleo et al., 2010, 2016; Wang et al., 2011, 2012, 2014; Alabi et al., 2012; Singh et al., 2012; Han et al., 2014, 2016; Belli Kullán et al., 2015; Sun et al., 2015; Paim Pinto et al., 2016; Bester et al., 2017a,b; Pagliarani et al., 2017; Snyman et al., 2017; Zhao et al., 2017). Each novel miRNA was aligned using BLAST both against the known miRNAs deposited in miRBase (Kozomara and Griffiths-Jones, 2014) and against all the novel miRNAs retrieved by the different studies above cited. miRNAs already classified by other authors as isomiRs of conserved miRNAs were not considered. Two or more miRNAs were considered to be a single entity, i.e., unique miRNA, when: (i) the miRNA sequences were the same, (ii) they localize in the same position on the genome, and (iii) they had the same precursors. Each novel miRNA was re-named using a “vvi\_miC” code followed by the indications “5p” or “3p” according to its location within the precursor sequence. The sequences of miRNA and miRNA\* (when available) were repositioned on the last version of the grape genome 12X V1 (PN40024<sup>1</sup>). Novel miRNAs sharing the same sequences, but with different precursors located in different regions of the grape genome, were indicated with

<sup>1</sup><http://genomes.cribi.unipd.it/grape/>



the same name followed by lowercase letters (a, b, c, etc.). Different isoforms (isomiR) of the same novel miRNA shifting a few nucleotides in the same genomic region within the same precursor were indicated as versions V1, V2, etc., of the same novel miRNA.

## Database of miRNA Targets

The target transcripts of all the conserved and novel miRNAs already validated in grapevine through either degradome or 5' RACE approaches were collected from 11 works (Carra et al., 2009; Pantaleo et al., 2010, 2016; Wang et al., 2012, 2013, 2014; Han et al., 2014; Jiu et al., 2015; Sun et al., 2015; Leng et al., 2017; Zhao et al., 2017). The annotation of the targets previously identified using the old grapevine annotation, referred to as the 8X genome, was updated to the grapevine transcriptome version 12X V1<sup>2</sup>. In addition, target transcripts of novel miRNAs were searched on the degradome library of "Pinot noir" (Pantaleo et al., 2010), and in parallel using the default parameters of the web-based tool StarScan (sRNA Target Scan, Liu et al., 2015). This tool classifies the sRNA-mediated cleavage events in different categories basing on several parameters related to the cleavage of the sequence of the target gene and miRNA, only the three most reliable classes (Z, I, and II) were considered. For each identified target, the annotation 12X V1 was integrated by Gene Ontology (GO) classifications using the Blast2GO tool<sup>3</sup>.

## Plant Material and Experimental Setup

The biological study was performed on plants of the red grape cultivar Barbera (*V. vinifera*) grown in a vineyard located in Coconato (Piemonte), northwest Italy, and monitored since 2008 for phytoplasma infection (Maggi et al., 2017). FDp and BNp were detected following a phytoplasma enrichment protocol (Marzachi et al., 1999), and the viruses that commonly affect grapevine in Italy were detected by multiplex RT-PCR (Gambino, 2015). Only the plants without both BNp and the most dangerous viruses infecting grapevine were considered for the experimental trial and further divided into three groups: FDp-infected (FD), recovered (R, i.e., plants found positive to FDp infection in the past, but then resulted FDp-negative and symptomless for the last 2 years), and healthy (H). The FDp isolate present in the vineyard belongs to the 16SrV-C subgroup (Maggi et al., 2017). From each group, five plants of the same age and located on neighboring rows in the vineyard were selected, and fully expanded leaves were collected in July 2014. For diseased plants (FD), only canes showing the typical FD symptoms (Caudwell, 1990) were chosen for sampling, and leaf tissues with advanced necrotic areas were excluded. In R and H plants, leaves inserted in the central region of the shoot were randomly collected. Leaf midribs were isolated using a sterile scalpel, immediately frozen in liquid nitrogen and stored at  $-80^{\circ}\text{C}$  until RNA extraction and JA quantification. The five plants selected for each category constituted independent biological replicates; three of them were used for small RNA libraries construction, and all the five plants were used for real-time PCR (RT-qPCR) validation assays.

## Small RNA Library Construction and Sequencing

RNA was extracted from leaf midribs using the mirPremier<sup>®</sup> microRNA Isolation Kit (Sigma-Aldrich, Inc.) following the manufacturer's instructions. Nine libraries (three biological replicates for each category H, R, and FD) of small RNAs were obtained using the TrueSeq Small RNA Sample Kit (Illumina, San Diego, CA, United States) and were sequenced using the HiSeq 2000 Illumina platform by an external service. All data were processed using the UEA small RNA Workbench (Mohorianu et al., 2017) by first removing 3'-adaptor sequences and then filtering out according to the following criteria: (a) low-quality and length (minimum 16 nt; maximum 30 nt); (b) low-complexity, i.e., sequences containing less than three distinct nucleotides were discarded; (c) reads matching to known transfer and ribosomal RNA were discarded; and (d) only reads matching to the *V. vinifera* assembly 12X<sup>1</sup> were retained. Any sequences without adaptor matches were excluded from further analysis. The miRNA predictions were performed using miRcat (Moxon et al., 2008) on the above mentioned *V. vinifera* assembly. Precursor sequences were then processed using the MFOLD software (v. 2.3<sup>4</sup>; Zuker, 2003) to analyze folding of hairpin secondary structures. Detection of similarities among predicted miRNAs and annotated miRNAs was performed running the BLAST algorithm against both miRBase and our newly produced database.

Bowtie 1.1.2 software<sup>5</sup> with no mismatch allowed in the alignment was employed to establish accurate miRNA abundance profiles of the nine sequenced samples. Following alignment, the resulting miRNA counts were normalized for differences in sequencing depths to account for the technical differences across samples. Statistically significant differences ( $P < 0.05$ ) in miRNA expression were estimated by applying the Student's *t*-test between the three possible couples of combinations (R vs. H, FD vs. R; FD vs. H).

Hierarchical clustering (HCL) analysis was then applied using Pearson's correlation distance and the software MeV v. 4.9<sup>6</sup>, using as input the normalized values of the reads generated for the three categories H, R, and FD.

## Real-Time PCR Analysis

The quantification of miRNA expression levels was carried out by RT-qPCR following the protocol by Shi and Chiang (2005), with the modifications reported in Pantaleo et al. (2016). For target quantification, total RNA was extracted using the Spectrum<sup>™</sup> Plant Total RNA extraction kit (Sigma-Aldrich, Inc.) and RT-qPCR assays were performed as reported by Pantaleo et al. (2016). Expression levels of miRNAs and related target genes were quantified after normalization to either 5.8S rRNA and U6 or *VvUBI* and *VvACT* endogenous genes used as internal controls, respectively. Specific primer pairs used in RT-qPCR reactions are listed in Supplementary Table S1. Transcript and

<sup>2</sup><http://genomes.cribi.unipd.it/DATA/V1/>

<sup>3</sup><https://www.blast2go.com/>

<sup>4</sup><http://unafold.rna.albany.edu>

<sup>5</sup><https://sourceforge.net/projects/bowtie-bio/files/bowtie/1.1.2/>

<sup>6</sup><http://compbio.dfci.harvard.edu/compbio/tools/mev>

miRNA levels were expressed as the mean and standard errors were calculated over five biological replicates.

## Quantification of Jasmonate Content in Leaf Midribs

400 mg of homogenized sample was freeze dried and transferred with 0.5 mL of methanol:water (1:1 v/v) acidified with 0.1 % of formic acid in an ultrasonic bath for 1 h. Samples were centrifuged for 2 min at 4°C and 15,000 rpm, and the supernatant was analyzed by high-performance liquid chromatography (HPLC). Original standard of (±)-jasmonic acid (purity ≥ 95%, Sigma-Aldrich) was used for the identification by comparing retention time and UV spectrum. The quantification was made by external calibration method.

The HPLC apparatus was an Agilent 1220 Infinity LC system (Agilent®, Waldbronn, Germany) model G4290B equipped with gradient pump, auto-sampler, and column oven set at 30°C. A 170 Diode Array Detector (Gilson, Middleton, WI, United States) set at 280 nm was used as detector. A XTerra RP18 analytical column (150 mm × 6 mm i.d., 5 µm, Waters) was used. The mobile phases consisted in water acidified with formic acid 0.05% (A) and acetonitrile (B), at a flow rate of 0.500 mL min<sup>-1</sup> in gradient mode, 0–20 min: from 10 to 35% of B, 20–25 min: from 35 to 100% B; 20 µL was injected for each sample.

## Statistical Analyses

Significant differences among treatments were statistically analyzed by applying a one-way ANOVA test followed by the Tukey's *HSD post hoc* test ( $P \leq 0.05$ ). Significant differences of pairwise comparisons were assessed by Student's *t*-test. The SPSS statistical software package (SPSS Inc., Cary, NC, United States, v.23) was used to run statistical analyses.

## Accession Numbers

Raw sequences from the nine miRNA libraries were deposited at the NCBI Sequence Read Archive under the accession number SRP129862.

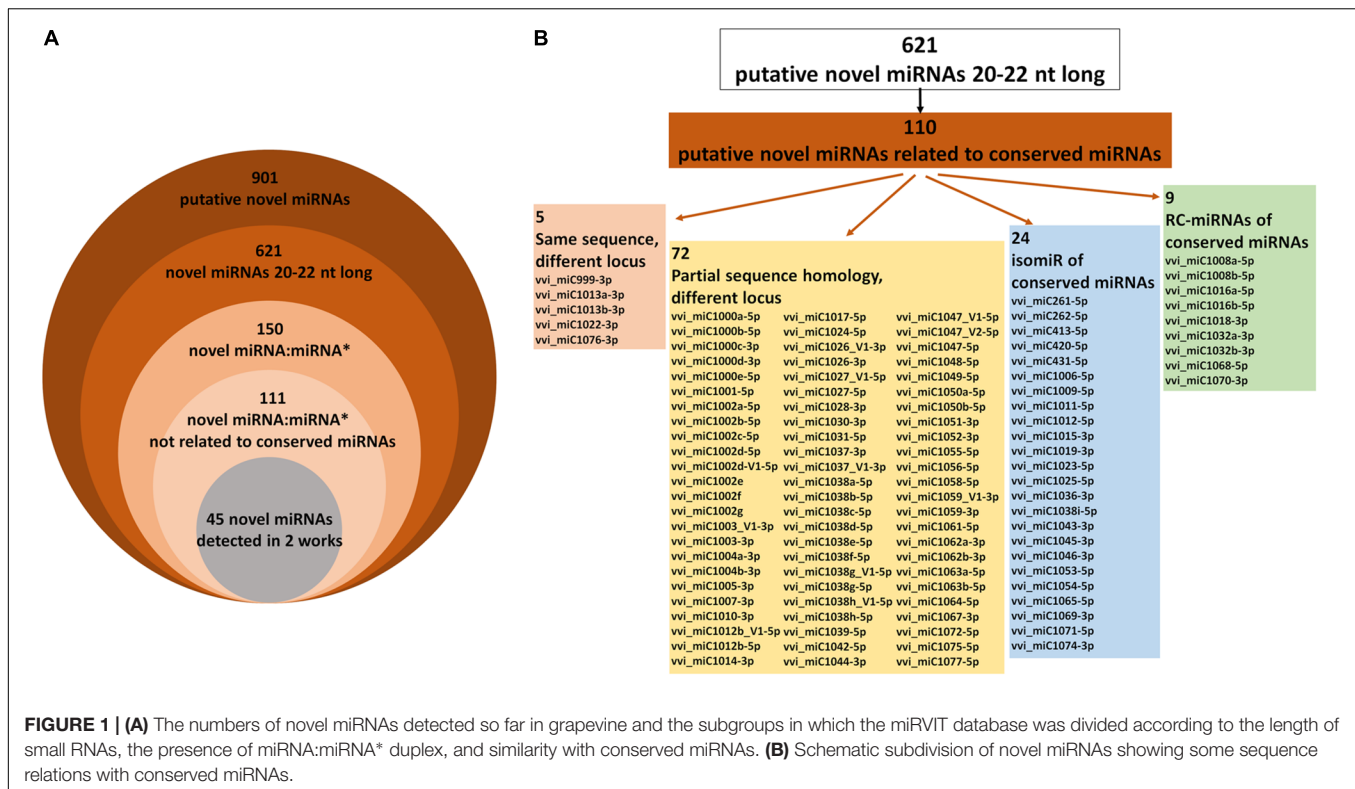
# RESULTS AND DISCUSSION

## miRVIT, a Novel Grapevine miRNA Database

Using the data reported in 18 works published until now on miRNAs in the *Vitis* genus, all putative novel miRNAs (hereafter referred to as novel miRNAs) identified and still not deposited in miRBase were collected into a single database, called “miRVIT” (available at <http://mirvit.ipsp.cnr.it>, Supplementary Table S2). Overall, 901 sequences referred to as novel miRNAs, re-named using a “vvi\_miC” code, were found; among these, 22 are miRNAs or miRNAs\* already deposited in miRBase and were excluded. In some cases, the same novel miRNA was reported using different classifications in more than 10 different works (Supplementary Table S2). For instance, 14 research groups identified vvi\_miC1039, 13 identified vvi\_miC132 (Supplementary Figure S1), and 11 identified

vvi\_miC137. The sequences of precursors were obtained by bioinformatics predictions, and different software or different versions of the same software were employed depending on the year of publication of the previous works, using in some cases different version of the grapevine genome. Consequently, some differences in the precursor sequences outside the region between miRNA and miRNA\* were identified for the same novel miRNA (Supplementary Figure S1).

Such a large number of putative grapevine-specific novel miRNAs do not appear realistic based on current knowledge, as it is conceivable that many false positives are present in this list, since in *Arabidopsis* (the most intensively studied plant species), only a few hundred miRNAs are currently known (Taylor et al., 2017; Axtell and Meyers, 2018). Starting from the criteria published for plant miRNA annotation (Meyers et al., 2008; Axtell and Meyers, 2018), we considered as novel miRNAs only the 20–22 nt long small RNAs, thus excluding small RNA sequences of all other lengths. Indeed, heterochromatic siRNAs of 23–24 nt are very common in small RNA-seq libraries, while miRNAs of 23–24 nt in length are rare and require extremely strong evidence to be classified as miRNAs (Axtell and Meyers, 2018), generally not available for grapevine novel miRNAs. The remaining 621 miRNAs 20–22 nt long were classified as 469 unique novel miRNAs (Figure 1A and Supplementary Table S3). The largest families of these novel miRNAs are vvi\_miC44-5p, with 10 members, identified by Pantaleo et al. (2010) and partially by Han et al. (2016), and vvi\_miC1038-5p, counting nine members and two isomiRs identified in several works (Supplementary Table S3). It must be taken into account that one-sixth of these novel miRNAs (110 on 621) showed either sequence similarity or some relationship with conserved miRNAs already deposited in miRBase (Figure 1B and Supplementary Table S3). This is the case of vvi\_miC999-3p, vvi\_miC1013, vvi\_miC1022, and vvi\_miC1076, which share the same sequences with vvi-miR3635-3p, vvi-miR171, vvi-miR390, and vvi-miR3634-3p, respectively, although they are located in different loci. Moreover, 72 novel miRNAs showed sequence similarity with several conserved miRNAs (particularly vvi-miR535, vvi-miR477, and vvi-miR3631), although they originated from different precursors at different genomic locations (Supplementary Table S3). Twenty-four of these are isomiRs of conserved miRNAs, which may have been not correctly classified in the original works, and nine are reverse-complementary miRNAs (RC-miRNAs, generated from the antisense strands of the miRNA precursors) of conserved miRNAs or miRNAs\* (Figure 1B and Supplementary Table S3). In addition, other RC-miRNAs were detected in three pairs of novel miRNAs not related to conserved miRNAs: vvi\_miC265-5p/vvi\_miC266-5p, vvi\_miC343-5p/vvi\_miC344, and vvi\_miC427a/vvi\_miC427b. RC-miRNAs, originally detected in animals (Tyler et al., 2008), were rarely described in plants (Shao et al., 2012; Belli Kullán et al., 2015), and their biological function is not yet completely defined, even though a potential role in transcriptional regulation through a DNA methylation-mediated pathway was hypothesized (Shao et al., 2012).



Of the 621 novel miRNAs analyzed, miRNA\* sequences were identified for only 150 accessions, and 45 of these were found in at least two different works, e.g., in at least two different small RNA libraries originating from different genotypes or tissues (Axtell and Meyers, 2018). Hypothetically, only 5% (45 of 901) of the novel miRNAs previously identified in grapevine satisfied the stringent requirements and could be really considered as novel miRNA candidates (Figure 1A and Supplementary Table S4). However, in the following sections, all the 621 20–22 nt long novel miRNAs were analyzed and their functional role was studied in a biological experiment to validate the incidence of these miRNA candidates in *V. vinifera* cv. Barbera.

## Target Genes of Grapevine miRNAs

Besides novel miRNAs, the miRVIT database groups all the target genes of conserved and novel miRNAs validated so far in grapevine using either degradome sequencing or 5' RACE. Overall, 215 targets of conserved miRNAs and 113 targets of novel miRNAs were reported in Supplementary Table S5. In addition, the database was integrated by searching target transcripts of novel miRNAs (Supplementary Table S3) in the degradome library of “Pinot noir” (Pantaleo et al., 2010) using the web-based tool StarScan (sRNA Target Scan, Liu et al., 2015). A similar analysis was previously validated by confirming the degradome predictions through 5' RACE (Pantaleo et al., 2016), thus proving to be a more reliable approach than bioinformatics prediction alone. Using StarScan, 213 potential targets (seven of which were previously validated, see Supplementary Table S5) of 120 novel miRNAs were

identified (Supplementary Table S6). Noteworthy, the novel miRNAs related to conserved miRNAs (Figure 1B) generally targeted the same transcripts of the conserved miRNA. For example, the targets of vvi\_miC1026 and vvi\_miC1027 that have sequences similar to vvi-miR396 are a number of transcripts encoding growth-regulating factors, whereas vvi\_miC1031-5p (similar to vvi-miR403) targeted *VvAGO2* and vvi\_miC1000 (similar to vvi-miR156) targeted mRNAs encoding Squamosa promoter-binding proteins (Supplementary Table S6).

## Small RNA Sequencing and Identification of miRNAs in “Barbera” Leaf Midribs

After removing adapters, redundant and low-quality sequences, reads from nine small RNA libraries of “Barbera” leaf midribs ranging from 16 to 30 nucleotides (see the metrics of the libraries in Supplementary Table S7 and Supplementary Figure S2) were aligned onto sequences of miRNAs deposited in miRBase and against miRVIT (Supplementary Table S3). As expected, we identified members of almost all known miRNAs: 137 of 168 (73.6%) deposited in miRBase. The normalized reads of each conserved miRNA were reported in Supplementary Table S8, showing that several members of vvi-miR166 family and vvi-miR3634-3p were the most expressed in “Barbera” leaf midribs. Of the 621 novel miRNAs previously identified in grapevine tissues, 308 were expressed at least in one “Barbera” small RNA library, and in 67 cases both miRNA and miRNA\* were identified (Supplementary Table S9). Interestingly, the probability of detecting previously published novel miRNAs



in “Barbera” was associated to specific factors, such as tissue-type, stress conditions, and genotype from which the novel miRNAs were originally isolated. In particular, 73.1% of the novel miRNAs originally isolated from either leaf or phloem tissues of grapevine (Alabi et al., 2012; Singh et al., 2012; Sun et al., 2015; Han et al., 2016; Pantaleo et al., 2016; Bester et al., 2017a,b; Snyman et al., 2017) were also detected in “Barbera” leaf midribs; on the contrary, only 53.8% of the novel miRNAs previously characterized in different organs and tissues (Carra et al., 2009; Pantaleo et al., 2010; Wang et al., 2011, 2012, 2014; Han et al., 2014; Belli Kullán et al., 2015; Paim Pinto et al., 2016; Pagliarani et al., 2017; Zhao et al., 2017) were found in the sequenced samples. Furthermore, 76.3% of the novel miRNAs identified in grapevine plants subjected to virus or phytoplasma infection (Alabi et al., 2012; Pantaleo et al., 2016; Bester et al., 2017a; Snyman et al., 2017) were also retrieved in FDP-infected samples of “Barbera” (Supplementary Table S9). Gene expression is notoriously influenced by tissue and environmental conditions; consequently, the probability to identify a miRNA in a small RNA library can be affected by these conditions as well. The genotype effect appeared more complicated to be clarified, as the database was biased to a higher number of works published on *V. vinifera* (Carra et al., 2009; Pantaleo et al., 2010, 2016; Alabi et al., 2012; Belli Kullán et al., 2015; Sun et al., 2015; Paim Pinto et al., 2016; Bester et al., 2017a,b; Pagliarani et al., 2017; Snyman et al., 2017) than on *V. amurensis* (Wang et al., 2012) or on *V. pseudoreticulata* (Han et al., 2016). While it was evident the low probability to identify in “Barbera” novel miRNAs originally published in *V. amurensis* (Wang et al., 2012), we did not observe a marked difference among novel miRNAs originally uncovered in *V. vinifera*, its hybrids, or *V. pseudoreticulata* (Supplementary Figure S3). Currently, the identification of novel miRNA loci in different *Vitis* spp. was biased to the use of PN40024 (Jaillon et al., 2007) or “Pinot noir” (Velasco et al., 2007) reference genomes, which show potential genomic differences relying on the cultivars or on other *Vitis* species in which the miRNAs were detected (Morgante et al., 2007; Gambino et al., 2017). Consequently, the hairpin prediction and the mapping of the reads produced for instance from *V. pseudoreticulata* on the “Pinot noir” genome (Han et al., 2016) allowed the prediction of only the miRNA loci common to *V. vinifera* cv. Pinot noir. The species-specific reads not mapping on the reference genome were generally discarded. This could be an explanation of the unexpected elevated percentage of miRNAs originally identified in *V. pseudoreticulata* and detected in “Barbera” (Supplementary Figure S3). In the near future, whether new *Vitis* genomes became available, it would be interesting to re-analyze small RNA libraries already published, in order to identify potential new species- or cultivar-specific miRNAs.

## Putative Novel miRNAs Identified for the First Time in “Barbera”

In addition to the novel miRNAs previously identified and annotated in miRVIT (Supplementary Table S2), other novel

miRNA candidates were characterized for the first time in “Barbera” using the miRCat tool and the latest release of the *V. vinifera* genome. Excluding the miRNAs already present in miRBase and miRVIT, and considering only the 20–22 nt long miRNAs and those expressed in at least three different libraries with at least 10 non-normalized reads, 13 putative novel miRNAs were selected (hereafter called “Barbera”-novel miRNAs; **Table 1** and Supplementary Figure S4). In addition, four variants of novel miRNAs already included in miRVIT were identified. However, no miRNA\* sequences were found for “Barbera”-novel miRNAs, and no significant candidate target genes were retrieved starting from the degradome library of “Pinot noir” (Pantaleo et al., 2010) and using the StarScan tool.

## Expression Profiles of miRNAs in “Barbera” Leaf Midribs

Despite the high number of miRNAs expressed in our libraries, only 25 conserved (**Table 2**) and 32 novel miRNAs (**Table 3**) showed significant differences in at least one comparison among the three considered theses (FD vs. H, R vs. H, and FD vs. R). A possible explanation for this could rely on the highly variable response of the different biological replicates constituting single “Barbera” plants. Nevertheless, by following this approach, the significant expression differences investigated in this work should be more reliable than those obtained with other sampling methodologies. An HCL analysis involving these 57 miRNAs was conducted to investigate the relationships of similarity among the three experimental categories. H and R formed a separate clade from FD, and the miRNAs were clustered in three groups: Cluster 1 included miRNAs with higher accumulation in FD, Cluster 2 counted miRNAs showing a slight peak of expression in R, and Cluster 3 contained only six miRNAs with higher expression in H (**Figure 2**).

The expression changes of the most interesting miRNAs (**Tables 2, 3**) and their already validated target genes (Supplementary Table S5) were further investigated by RT-qPCR. In Cluster 1, the higher accumulation of miRNAs induced by FDP infection was confirmed for vvi-miR156, vvi-miR167, vvi-miR2950-5p, and vvi\_miC137 (**Figure 3**) and for vvi-miR3632-3p, vvi-miR3623-5p, vvi-miR403, vvi\_miC360-3p, and vvi\_miC430-3p (Supplementary Figure S5). In addition, RT-qPCR assays conducted on the miRNAs belonging to Cluster 2, vvi-miR166, vvi-miR169, vvi-miR319, vvi-miR482, vvi\_miC1031-5p, vvi\_miC64 (**Figure 4**), vvi\_miC1038-3p, vvi\_miC132-3p, vvi\_miC197-5p, vvi\_miC281-5p, and vvi\_miC413-5p (Supplementary Figure S6), confirmed the expression trends observed in sequencing results. The “Barbera”-novel miRNAs were expressed in all groups of the considered experimental plan, and for six of them, significant differences were observed among the treatments (Supplementary Table S10). The RT-qPCR analyses conducted on a selection of these miRNAs, vvi\_miC606-3p, vvi\_miC617-3p, vvi\_miC644-3p, and vvi\_miC648-5p, were consistent with sequencing data, and vvi\_miC644-3p, which was the most expressed among the “Barbera”-novel miRNAs, was overexpressed in FD samples (Supplementary Figure S7). Further analyses are needed to prove



TABLE 1 | Putative novel miRNAs identified for the first time in “Barbera” leaf midribs (“Barbera” – novel miRNAs).

miRNA ID	Genomic position	Orientation	Sequence (5′–3′)	sRNA length (nt)	Hairpin length	Hairpin minimum free energy ( $\Delta G$ kcal mol <sup>−1</sup> )	miRNA* and Notes
wi_miC134_V1-3p	chr5: 5125249 5125229	–	CTGAGGCTTCATTTTGAAC	21	114	–61.50	Variant of wi_miC134-3p in the same locus.
wi_miC134_V1b-3p	chr15: 14464031 14464051	+	CTGAGGCTTCATTTTGAAC	21	113	–37.70	Same sequences of wi_miC134_V1-3p, different genomic location.
wi_miC443_V1-3p	chr18: 3551253 3551273	–	AAATTGGCTGTGAAATTTCT	21	142	–79.68	wi_miC443_V1-5p: AATATTACAGAGCCATTTTGG (chr18: 3551368 3551348) Variants of wi_miC443 in the same locus.
wi_miC593_V1-5p	chrUn: 26025545 26025524	–	TGGTGAACCAATACTCTGGT	22	159	–63.60	wi_miC593_V1-3p: ACCAGAGTTCTTTAGTTGACG (chrUn: 26025482 26025462) Variants of wi_miC593 in the same locus.
wi_miC597-3p	chr2: 636221 636241	–	CCAGGATTGGAACCTCAAGACC	21	118	–59.75	In the same locus and partially complementary to wi_miC49-5p
wi_miC605-3p	chr10: 10801472 10801491	+	AGAGGCTCGGTGAAATAGAC	20	98	–27.10	
wi_miC606-3p	chr10: 2522406 2522427	–	GCTAGCAATGGAGTCCTGAAC	22	85	–24.40	
wi_miC614-5p	chr14: 21339372 21339392	–	TTTAGGAAGTGTCTTCTGGGCT	21	121	–65.90	
wi_miC617-3p	chr14: 27797794 27797814	–	TCCAAACAAAACGTCGGCCT	21	158	–74.30	
wi_miC620-3p	chr16: 286248 286268	–	CCGACGTTACGCTTGAGCAGT	21	129	–63.40	
wi_miC630-3p	chr19: 3919987 3920007	+	CTTGGTTCCTGGAATAATTTGA	21	156	–50.76	
wi_miC631-3p	chr2: 15941291 15941311	–	ATGTTATGGATCTTGGTTTGT	21	94	–24.20	
wi_miC638-3p	chr4: 1959986 1960005	+	GGAAGAGCTAGAAATCTAAC	20	63	–18.80	
wi_miC644-3p	chr6: 8570952 8570972	+	TTGAGCGATCAAAACGGGCT	21	118	–32.18	
wi_miC645-5p	chr7: 1677361 1677381	+	AGAGGAACCGAAACTAGGACT	21	120	–70.60	
wi_miC648-5p	chr7: 3062319 3062339	–	TTAAGCGAGACCAATTGTGACT	21	81	–22.50	
wi_miC653-5p	chr9: 328798 328818	+	TCAACTATGAGCCCTTGAAT	21	117	–43.10	

**TABLE 2 |** Normalized number of reads (mean of three biological replicates) of conserved miRNAs in leaf midribs of FDp infected (FD), recovered (R), and healthy (H) “Barbera.”

Conserved miRNA	Normalized number of reads			Significant differences		
	H	R	FD	FD/H	R/H	FD/R
vvi-miR156d	4.76	4.12	5.51		*	
vvi-miR160e	0.15	0.47	0.47		*	
vvi-miR166c	13119.96	21411.64	19961.37		*	
vvi-miR166d	13161.60	21517.84	19907.53		*	
vvi-miR166e	12980.16	21368.43	19904.17		*	
vvi-miR166f	13019.97	21318.63	19943.36		*	
vvi-miR166g	13058.51	21486.03	19877.01		*	
vvi-miR166h	13065.48	21401.90	19950.53		*	
vvi-miR167b	0.35	0.38	0.93	*		*
vvi-miR167c	1.40	3.09	2.32		*	
vvi-miR169t	0.00	0.11	0.03		*	
vvi-miR171j	0.00	0.11	0.07		*	
vvi-miR319g	1.44	2.50	0.75			*
vvi-miR399a	0.03	0.20	0.03		*	*
vvi-miR403f	20.99	32.73	41.89	*		
vvi-miR482	96.98	143.93	128.36		*	
vvi-miR2111-5p	0.56	1.28	1.66	*		
vvi-miR2950-5p	3.17	6.80	21.33	*		*
vvi-miR3623-3p	120.73	159.46	267.45	*		*
vvi-miR3623-5p	81.04	96.61	181.92	*		*
vvi-miR3632-3p	1.72	3.20	5.94	*		
vvi-miR3633b-5p	0.38	0.32	0.80	*		*
vvi-miR3636-3p	0.18	0.19	0.03	*		
vvi-miR3636-5p	0.17	0.42	0.24		*	
vvi-miR3640-3p	0.12	0.46	0.62	*		

The asterisk denotes significant differences in the three comparisons FD/H, R/H, and FD/R using the Student's t-test ( $P \leq 0.05$ ).

their biological functions definitely. Indeed, the biological effect of miRNAs is generally mediated by their cleavage activity on specific target genes.

### Dissecting the Role of miRNAs in the Regulation of the FDp–Grapevine Interaction

miRNAs and target transcripts significantly modulated in the presence of FDp influenced the metabolic processes linked to photosynthesis, cellular development, jasmonate (JA) signaling, and defense responses. The reduction of photosynthesis efficiency caused by FDp infection, which was previously suggested by physiological experiments (Vitali et al., 2013), was likely induced by the down-regulation of specific genes, such as the photosystem I reaction center subunit II (*VvPSI*, VIT\_05s0020g03180) and the Photosystem II subunit X (*VvPSBx*, VIT\_04s0008g01730), encoding important proteins of the Photosystems I and II reaction centers, which were targeted by vvi\_miC137-3p and vvi-miR169, respectively (Figures 3, 4). These results were consistent with the observed downregulation of a chlorophyllase-encoding gene (*VvCHL*, VIT\_07s0151g00250), targeted by the grape specific miRNA vvi-miR2950 and influencing chlorophyll catabolism in infected plants (Figure 3). Accordingly, both vvi-miR2950 and the novel vvi\_miC137-3p were highly induced

upon FDp infection, showing a modulation pattern that was negatively correlated to the photosynthetic efficiency measured in diseased grapevines (Pantaleo et al., 2016).

Among highly conserved miRNA categories involved in plant development, miR156 is a well-known regulator of *Squamosa* promoter-binding protein-like genes (*SPL*, Wu and Poethig, 2006; Ferreira e Silva et al., 2014). The *SPL* transcription factor family exerts a key role in controlling the transition of floral phases, plant architecture, and in the determination of leaf cell number and size (Chen et al., 2010). In grapevine, the *VvSPL* gene (VIT\_01s0010g03910) is a target of vvi-miR156 (Pantaleo et al., 2010), together with an expansin-encoding gene (*VvEXPA14*, VIT\_13s0067g02930, Wang et al., 2014), involved in cell wall modification under stress conditions (Dal Santo et al., 2013). Our results indicated that FD presence strongly inhibited these transcripts, differently from what observed for the expression trend of vvi-miR156 upon infection (Figure 3). It is thus conceivable that the FD-mediated activation of vvi-miR156 with the consequent repression of *SPL* genes could be at the base of the developmental alterations (i.e., shortening of internodes, smaller inflorescences, etc.) that are typical symptoms associated to FDp infection in grapevine (Caudwell, 1990). These data are also supported by previous evidences attesting that, besides being a master regulator of plant development, vvi-miR156 is able to dynamically respond to environmental stresses in

**TABLE 3 |** Normalized number of reads (mean of three biological replicates) of novel miRNAs in leaf midribs of FDp infected (FD), recovered (R), and healthy (H) “Barbera.”

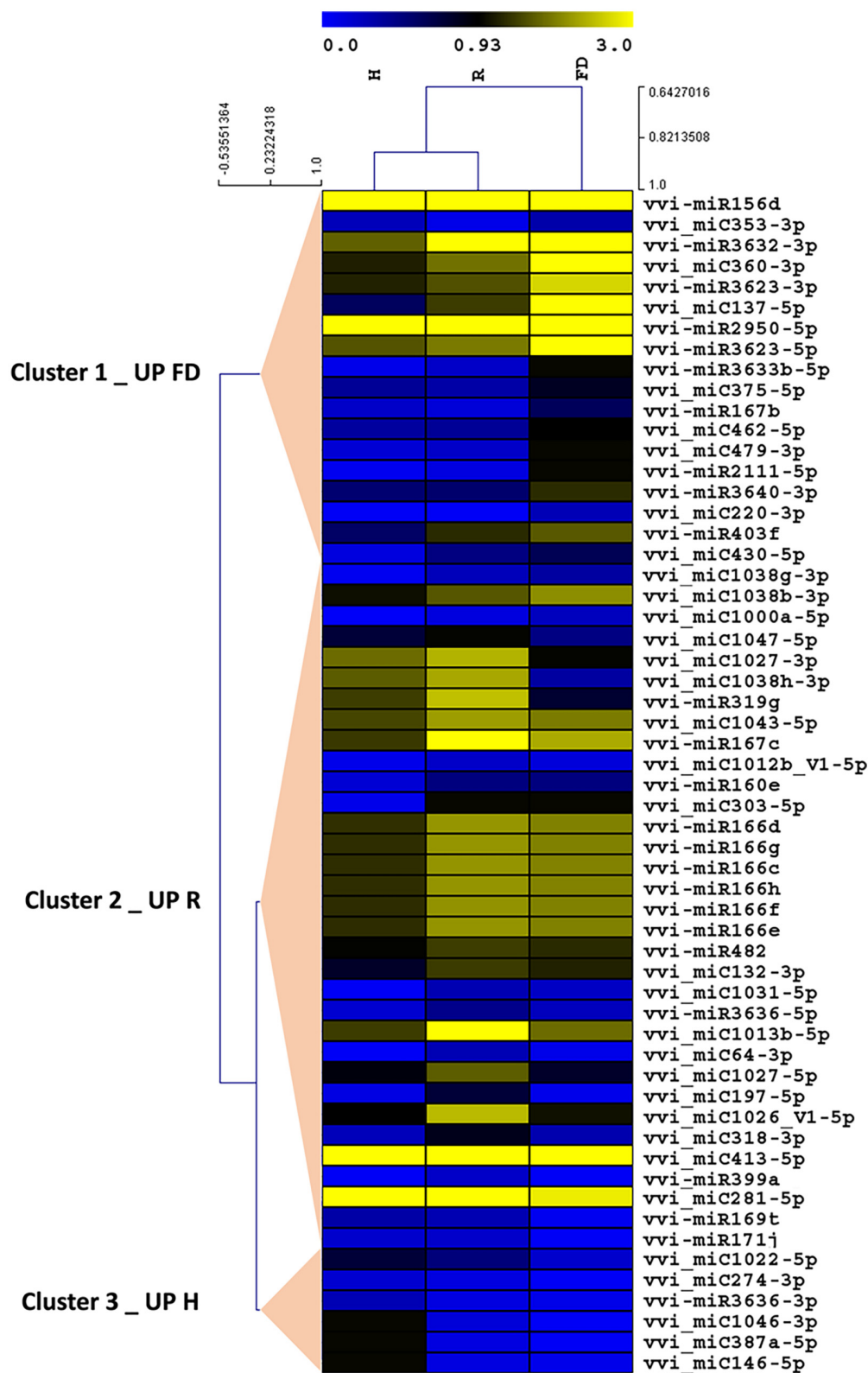
Putative novel miRNA	Normalized number of reads			Significant differences		
	H	R	FD	FD/H	R/H	FD/R
wi_miC64-3p	0.03	0.27	0.08		*	
wi_miC132-3p	15.62	28.48	24.17	*	*	
wi_miC137-5p	0.59	1.43	3.83	*	*	*
wi_miC146-5p	0.27	0.11	0.08	*		
wi_miC197-5p	0.10	0.72	0.07		*	*
wi_miC220-3p	0.05	0.25	0.34		*	
wi_miC274-3p	0.32	0.28	0.05			*
wi_miC281-5p	0.00	0.14	0.03		*	*
wi_miC303-5p	0.07	0.00	0.00	*	*	
wi_miC318-3p	0.25	0.83	0.29		*	
wi_miC353-3p	0.25	0.08	0.31		*	
wi_miC360-3p	1.19	1.86	3.75	*		
wi_miC375-5p	0.20	0.14	0.60	*		*
wi_miC387a-5p	0.17	0.11	0.03	*		
wi_miC413-5p	3.02	5.27	3.11		*	*
wi_miC430-5p	0.02	0.12	0.23	*		
wi_miC462-5p	0.52	0.53	1.29	*		
wi_miC479-3p	0.03	0.03	0.26	*		*
wi_miC1000a-5p	0.05	0.11	0.00			*
wi_miC1012b_V1-5p	0.07	0.20	0.13			*
wi_miC1013b-5p	42.82	91.15	54.42		*	
wi_miC1022-5p	4.95	4.75	2.87	*		
wi_miC1026_V1-5p	0.93	2.45	1.07		*	
wi_miC1027-3p	1.81	2.39	0.98	*		*
wi_miC1027-5p	8.76	16.95	7.65		*	*
wi_miC1031-5p	0.03	0.28	0.21		*	
wi_miC1038b-3p	0.15	0.20	0.00			*
wi_miC1038g-3p	0.07	0.20	0.00			*
wi_miC1038h-3p	1.68	2.30	0.34			*
wi_miC1043-5p	150.00	222.32	194.43		*	
wi_miC1046-3p	0.72	0.48	0.18	*	*	
wi_miC1047-5p	7.19	9.83	4.46			*

The asterisk denotes significant differences in the three comparisons FD/H, R/H, and FD/R using the Student's *t*-test ( $P \leq 0.05$ ).

different ways (Alabi et al., 2012; Sun et al., 2015; Pantaleo et al., 2016; Pagliarani et al., 2017). The strong upregulation here observed in FDp infected leaves is in contrast with data reported in grapevines infected by aster yellows (AY; Snyman et al., 2017). Nevertheless, in the associations of *Ziziphus jujuba*/Jujube witches'-broom (Shao et al., 2016), Mexican lime (*Citrus aurantifolia* L.)/Witches' broom disease (Ehya et al., 2013) and mulberry (*Morus multicaulis* Perr.)/mulberry yellow dwarf disease (Gai et al., 2014), miR156 was always overexpressed in infected plants, as in our study, thus highlighting its potential function in directing symptom development.

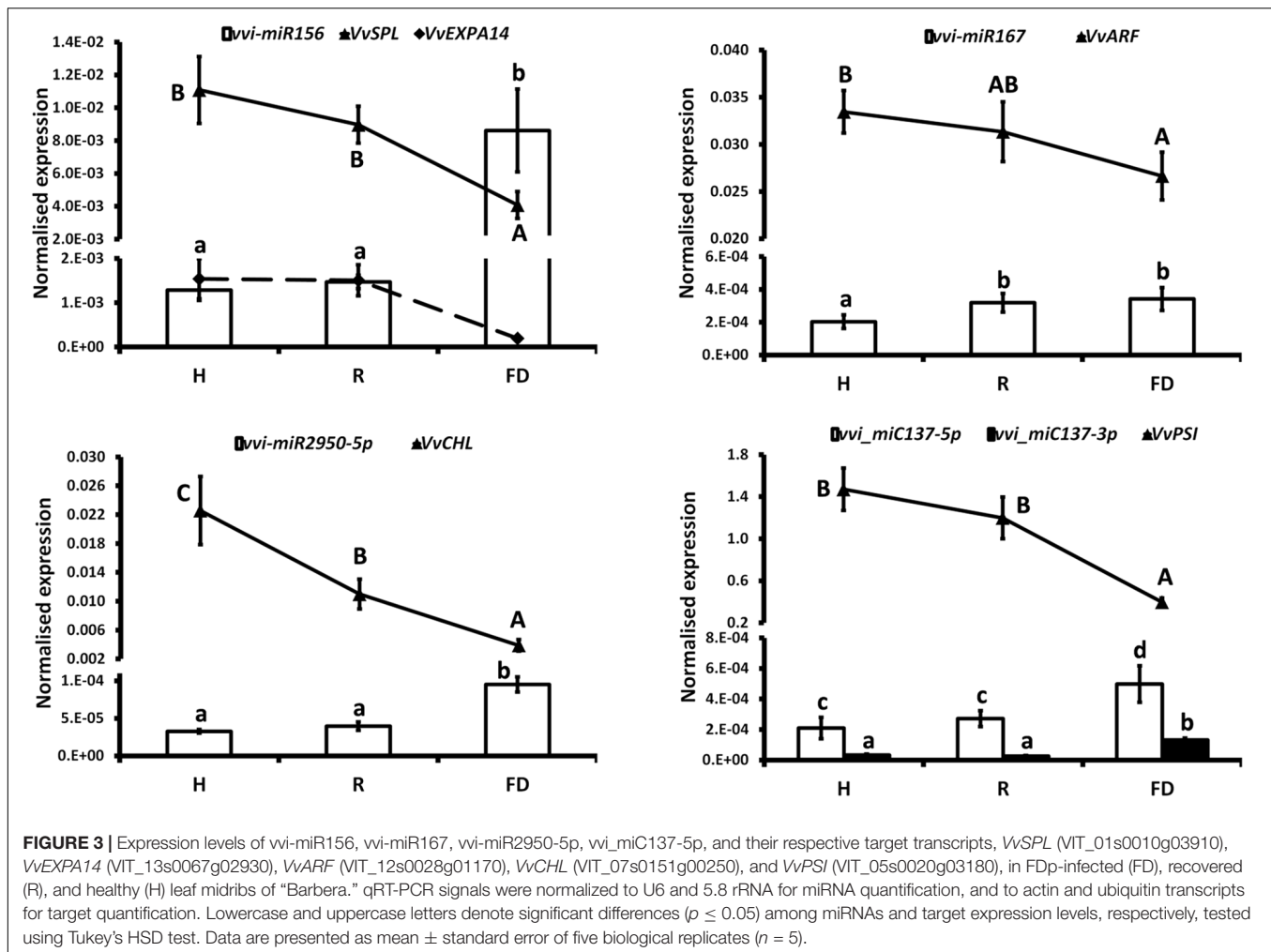
Another family of miRNAs typically tied to plant development is miR166, which regulates the broad class of III homeodomain leucine zipper (HD-ZIPIII) transcription factors. In particular, these genes were associated with *Arabidopsis* embryo and meristem development, organ polarity, and vascular development (Baucher et al., 2007 and the references therein). The levels of vvi-miR166 in FDp

infected plants seemed not to justify the high transcriptional rates of *Phabulosa* (*VvPHB*, VIT\_10s0003g04670) and *Revoluta* (*VvREV*, VIT\_13s0019g04320) transcripts (**Figure 4**). A possible explanation could be linked to a direct activation of the promoters of these genes in response to FDp that may escape the miRNA-mediated regulation, as also suggested for other miRNA-target interactions in grapevine (Pantaleo et al., 2016; Pagliarani et al., 2017). In addition, it must be considered that the complex regulatory system “AGO10-miR166-HD-ZIPIII” uncovered in *Arabidopsis*, in which AGO10 specifically sequesters miR166 and antagonizes the silencing activity mediated by AGO1-miR166 against HD-ZIP III (Zhou et al., 2015), may also occur in grapevine. This hypothesis is consistent with the observed overexpression of *VvAGO10* (VIT\_05s0020g04190) in FDp infected plants (inset in **Figure 4**), which could reduce the post-transcriptional activity of vvi-miR166 with a consequent increase in *VvPHB* and *VvREV* expression levels. In addition to development regulation, REV was also involved in leaf senescence



**FIGURE 2 |** Hierarchical clustering analysis of conserved and novel miRNAs differentially modulated among FDp-infected (FD), recovered (R), and healthy (H) leaf midribs of “Barbera” ( $p \leq 0.05$ ). The clustering was generated with the TMeV software (v. 4.9) using the average normalized number of reads of three biological replicates. The results were represented by heat map blue–yellow corresponding, respectively, to low and high miRNAs accumulation levels.

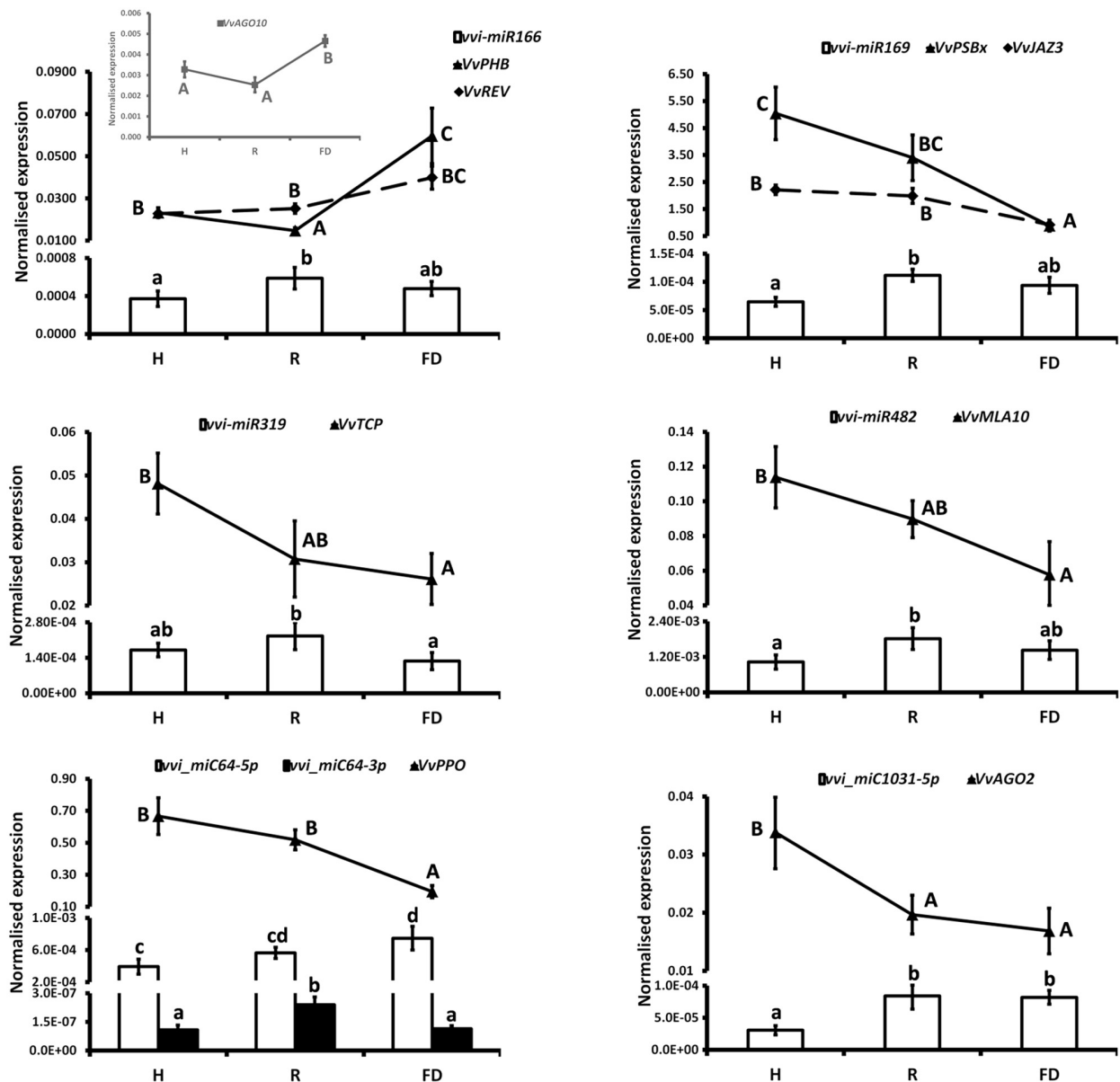




in *Arabidopsis* through the direct transcriptional induction of a WRKY transcription factor (Xie et al., 2014). The homologous WRKY gene in grapevine, *VvWRKY1*, was overexpressed in FD-infected samples (Gambino et al., 2013), in agreement with a potential role as trigger of defense responses by causing a decrease in the plant susceptibility to the pathogen presence, as indicated for some fungi (Marchive et al., 2013). The activity of *VvWRKY1* in grapevine against fungi, and hypothetically against FDp, involved the activation of distinct sets of defense-related genes. In particular, in grapevine infected by FDp, salicylic acid (SA)-mediated signaling activation and JA repression were previously suggested (Gambino et al., 2013; Prezelj et al., 2016). This type of regulation was confirmed by the present work showing that could be influenced by the activity of some miRNAs. The alteration of JA biosynthetic pathway could be promoted by tuning the TEOSINTE BRANCHED/CYCLOIDEA/PCF (*VvTCP*, VIT\_12s0028g02520)/miR319 interplay (Figure 4 and Supplementary Figure S7) through the downregulation of JA ZIM domain-containing protein (*VvJAZ3*, VIT\_01s0011g05560), target of both vvi-miR169 and the novel vvi\_miC197-5p (Figure 4), and an auxin response factor-encoding transcript (*VvARF*, VIT\_12s0028g01170), target of vvi-miR167 (Figure 3).

The downregulation of *VvTCP* occurred almost exclusively in FDp-infected plants and in parallel with the significant reduction of both *VvLOXA* and 12-oxophytodienoate reductase (*VvOPR3*) expression (Supplementary Figure S8), two of the most important genes involved in JA biosynthesis (Wasternack and Hause, 2013). Interestingly, the interplay *VvTCP*/miR319 controls the biosynthesis of JA through the regulation of lipoxygenase (*LOX*) genes (Schommer et al., 2008). Consistently, the quantification of JA content in the same tissues displayed a progressive reduction or even absence of this metabolite in R and FD samples, respectively (Supplementary Figure S8).

The ARF/miR167 system is involved in the regulation of auxin responsive genes influencing both auxin signaling and JA pathway (Gutierrez et al., 2012; Boer et al., 2014, and the references therein). Consequently, in FDp-infected grapevine, the *VvARF*/vvi-miR167 regulation, associated to the observed decrease in JA content, could be the base of the flower alterations typical of infected plants, in agreement with other reports showing that the overexpression of miR167 in tomato induces floral developmental defects in a JA-dependent manner (Liu et al., 2014). Moreover, a clear correlation among auxin, JA, and phytoplasma was previously demonstrated



**FIGURE 4 |** Expression levels of vvi-miR166, vvi-miR169, vvi-miR319, vvi-miR482, vvi-miC1031-5p, vvi-miC64, and their respective target transcripts, *VvPHB* (VIT\_10s0003g04670), *VvREV* (VIT\_13s0019g04320), *VvPSBx* (VIT\_04s0008g01730), *VvJAZ3* (VIT\_01s0011g05560), *VvTCP* (VIT\_12s0028g02520), *VvMLA10* (VIT\_04s0023g02380), *VvAGO2* (VIT\_10s0042g01180), *VvPPO* (VIT\_10s0116g00560), and *VvAGO10* (VIT\_05s0020g04190, inset), in FDp-infected (FD), recovered (R), and healthy (H) leaf midribs of “Barbera.” qRT-PCR signals were normalized to U6 and 5.8 rRNA for miRNA quantification, and to actin and ubiquitin transcripts for target quantification. Lowercase and uppercase letters denote significant differences ( $p \leq 0.05$ ) among miRNAs and target expression levels, respectively, tested using Tukey’s HSD test. Data are presented as mean  $\pm$  standard error of five biological replicates ( $n = 5$ ).

in transgenic *Arabidopsis* plants containing the phytoplasma virulence effector tengu-su inducer (TENGU) of “*Candidatus* Phytoplasma asteris,” onion yellows isolate (Hoshi et al., 2009). These *Arabidopsis* plants showed high levels of miR167 and a decreased expression of both *ARF* and *LOX* genes associated to symptom developments and reduction in JA and auxin levels (Minato et al., 2014). Similar results were obtained using transgenic *Arabidopsis* expressing the AY-WB protein 11 (SAP11), an effector produced by the Aster Yellows

Witches’ Broom (AY-WB) phytoplasma (Sugio et al., 2011). However, repression of JA pathway is not a generic response to phytoplasma infection, as in BNp-infected grapevines very different molecular and metabolic responses were reported. Indeed, JA biosynthesis was strongly activated in both BNp-recovered and diseased grapevines over the whole vegetative season (Paolacci et al., 2017). In parallel, the activation of the SA signaling pathway was also observed, as previously reported (Hren et al., 2009; Dermastia et al., 2015). The

authors hypothesized an SA-mediated repression of the genes downstream the JA biosynthesis with relative inhibition of the JA signaling in infected plants. Additionally, the activation of JA-mediated pathway in recovered grapevines could support the role of this defense pathway in the maintenance of the recovery condition in former BNp-infected grapevines (Paolacci et al., 2017). Although this hypothesis may be interesting, it could be valid only for the BNp, as FDP induces very different responses in plant (i.e., induction of SA and JA repression); thus, further insights on the specific FDP-grapevine pathosystem will be necessary to investigate this point in depth.

The regulation of miRNAs and targets linked to disease resistance induced in some cases a reduction of these responses in FDP infected plants. For example, miR482 is involved in a feedback control of NBS-LRR genes in several species, lowering the energy cost of disease resistance process by downregulating the production of these genes in the absence of pathogens (González et al., 2015). In FDP-infected grapevines, we observed an opposite regulation of this system with reduction of *VvMLA10* (VIT\_04s0023g02380) in the presence of the pathogen (Figure 4). A similar downregulation was noticed for *VvAGO2* (VIT\_10s0042g01180) in FD samples. In grapevine, this gene is targeted by the novel vvi\_miC1031-5p (with a sequence similar to vvi-miR403) and likely by vvi-miR403, as observed in other plants (Harvey et al., 2011), although this has not been confirmed in grapevine yet. AGO2 has an antiviral role against several viruses in different species (Harvey et al., 2011), while its relationship with phytoplasma infection is not currently clear. The downregulation of AGO2 here reported in FD condition (Figure 4) could suggest that phytoplasma-infected plants may be more susceptible to virus infection compared to healthy ones. This hypothesis was not confirmed in BNp infected plants, where the presence of phytoplasma did not influence the virus infection (Rotter et al., 2018). However, considering the above-reported differences in molecular and metabolic responses between plants infected by BNp and FDP, we cannot exclude that FDP-infected plants are more sensitive to virus attack, thus opening the venue for future research studies specifically addressed to explore this response.

The novel vvi\_miC64-3p, significantly induced in R (Figure 4), is the miRNA\* of vvi\_miC64-5p controlling a Polyphenol oxidase II-encoding transcript (*VvPPO*, VIT\_10s0116g00560) tied to lignin biosynthesis and to resistance responses to fungi and bacteria (Harm et al., 2011). The vvi\_miC64-5p was overall more abundant and followed a different trend of accumulation than vvi\_miC64-3p. In particular, the over-accumulation of vvi\_miC64-5p in R and FD samples occurred in parallel with the downregulation of *VvPPO* transcription (Figure 4). This downregulation may negatively influence lignification of infected tissues, thus representing another example of down-regulation of a pathway involved in response to pathogens.

Although mechanisms of plant response to the pathogen still require further deepening, even less is known on the molecular bases driving the phenomenon of symptom recovery in grapevine infected by FD. Recovered grapevines are asymptomatic and

phenotypically similar to H plants, and the expression levels of miRNAs generally showed a close relationship between R and H categories (Figure 2). However, for some combinations of miRNAs/targets, this relationship was not observed. For example, *VvCHL* and *VvPSBx*, targeted by vvi-miR2950-5p and vvi-miR169, respectively, were downregulated in R suggesting a reduced photosynthetic efficiency or a not complete recovery of the gas exchange performances in these plants, as previously reported (Vitali et al., 2013). In addition, the lower levels of JA in R grapevines, associated to the modulation of the vvi-miR319/*VvTCP* complex, point out that the recovery condition might be related to repression of JA signaling pathways even after 2 years from the disappearance of FDP. Taken together, these results suggested the persistence for long time of a “molecular memory” of the former phytoplasma infection; further investigations are needed to verify whether this “memory” could be effective to prime the recovered plants from new FDP infection events.

## CONCLUSION

In this work, we univocally cataloged for the first time all novel miRNAs identified so far in grapevine by producing a single database, miRVIT, with the final aim to put order in an intriguing research field that has been developing and evolving continuously over recent years. In addition to conserved miRNAs, some novel miRNAs detected by several authors (vvi\_miC64, vvi\_miC137) were shown to be *bona fide* miRNAs, spread among different genotypes, potentially regulated by diverse environmental conditions and likely playing important biological functions that should be further investigated in the future through *ad hoc* functional studies. miRVIT will integrate several existing bioinformatics resources designed for transcriptome, small RNAs and functional analysis in grapevine (Grimplet et al., 2009; Dereeper et al., 2011; Wong et al., 2013; Belli Kullán et al., 2015; Pulvirenti et al., 2015; Moretto et al., 2016). We also demonstrated that the application of miRVIT to the analysis of miRNAs from FDP-infected grapevines was effective for pinpointing complex interactions among miRNAs and related targets specifically linked to disease evolution. In particular, we evidenced that inhibition of cell development and photosynthetic processes (vvi-miR156/*VvSPL*, *VvAGO10*/vvi-miR166/*VvPHB*, and vvi\_miC137-3p/*VvPSI*) could be finely tuned by miRNAs in infected plants together with regulation of the cross-talk between JA signaling pathways (vvi-miR319/*VvTCP* and vvi-miR167/*VvARF*) and disease resistance response (vvi-miR482/*VvMLA10*, vvi\_miC1031-5p/*VvAGO2*, and vvi\_miC64-5p/*VvPPO*).

## AUTHOR CONTRIBUTIONS

GG, IP, and CM conceived the study. GG produced the miRVIT database. GB developed the database platform. SA performed the bioinformatic analyses. WC and CP performed most of

the molecular analyses and elaborated the corresponding data. PB, MR, SP, and IP helped with the molecular analyses and complemented the writing. GG wrote the manuscript with the help of WC, CP, and IP. All authors read and approved the manuscript.

## FUNDING

This research was supported by the project “*Inteflavi: un approccio integrato alla lotta contro la flavescenza dorata della vite*,” jointly funded by Fondazione Cassa di Risparmio di Cuneo, Fondazione Cassa di Risparmio di Torino, and Fondazione Cassa di Risparmio di Asti. IP gratefully acknowledges the financial support by the Italian Ministry of Education, University and

Research (MIUR), FIR project RBFR13GHC5: “*The epigenomic plasticity of grapevine in genotype per environment interactions*.”

## ACKNOWLEDGMENTS

The authors thank Marco Martini for conceiving and providing the miRVIT logo.

## SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fpls.2018.01034/full#supplementary-material>

## REFERENCES

- Alabi, O. J., Zheng, Y., Jagadeeswaran, G., Sunkar, R., and Naidu, R. A. (2012). High-throughput sequence analysis of small RNAs in grapevine (*Vitis vinifera* L.) affected by Grapevine leafroll disease. *Mol. Plant Pathol.* 13, 1060–1076. doi: 10.1111/j.1364-3703.2012.00815.x
- Angelini, E., Clair, D., Borgo, M., Bertaccini, A., and Boudon-Padieu, E. (2001). Flavescence dorée in France and Italy—Occurrence of closely related Phytoplasma isolates and their near relationships to Palatinate grapevine yellows and an alder yellows phytoplasma. *Vitis* 40, 79–86.
- Axtell, M. J. (2013). Classification and comparison of small RNAs from plants. *Annu. Rev. Plant Biol.* 64, 137–159. doi: 10.1146/annurev-arplant-050312-120043
- Axtell, M. J., and Meyers, B. C. (2018). Revisiting criteria for plant miRNA annotation in the era of big data. *Plant Cell* 30, 272–284. doi: 10.1105/tpc.17.00851
- Baucher, M., El Jaziri, M., and Vandeputte, O. (2007). From primary to secondary growth: origin and development of the vascular system. *J. Exp. Bot.* 58, 3485–3501. doi: 10.1093/jxb/erm185
- Baulcombe, D. (2002). RNA silencing. *Curr. Biol.* 12, R82–R84. doi: 10.1016/S0960-9822(02)00665-6
- Belli Kullán, J., Lopes Paim Pinto, D., Bertolini, E., Fasoli, M., Zenoni, S., Tornielli, G. B., et al. (2015). miRVine: a microRNA expression atlas of grapevine based on small RNA sequencing. *BMC Genomics* 16:393. doi: 10.1186/s12864-015-1610-5
- Bendix, C., and Lewis, J. D. (2018). The enemy within: phloem-limited pathogens. *Mol. Plant Pathol.* 19, 238–254. doi: 10.1111/mpp.12526
- Bester, R., Burger, J. T., and Maree, H. J. (2017a). Differential expression of miRNAs and associated gene targets in Grapevine leafroll-associated virus 3-infected plants. *Arch. Virol.* 162, 987–996. doi: 10.1007/s00705-016-3197-9
- Bester, R., Burger, J. T., and Maree, H. J. (2017b). The small RNA repertoire in phloem tissue of three *Vitis vinifera* cultivars. *Plant Gene* 10, 60–73. doi: 10.1016/j.plgene.2017.05.009
- Boer, D. R., Freire-Rios, A., van den Berg, W. A., Saaki, T., Manfield, I. W., Kepinski, S., et al. (2014). Structural basis for DNA binding specificity by the auxin-dependent ARF transcription factors. *Cell* 156, 577–589. doi: 10.1016/j.cell.2013.12.027
- Borges, F., and Martienssen, R. A. (2015). The expanding world of small RNAs in plants. *Nat. Rev. Mol. Cell Biol.* 16, 727–741. doi: 10.1038/nrm4085
- Carra, A., Mica, E., Gambino, G., Pindo, M., Moser, C., Pe, E., et al. (2009). Cloning and characterization of small noncoding RNAs from grape. *Plant J.* 59, 750–763. doi: 10.1111/j.1365-3113X.2009.03906.x
- Caudwell, A. (1990). Epidemiology and characterization of Flavescence dorée (FD) and other grapevine yellows. *Agronomie* 10, 655–663. doi: 10.1051/agro:19900806
- Chen, X., Zhang, Z., Liu, D., Zhang, K., Li, A., and Mao, L. (2010). SQUAMOSA promoter-binding protein-like transcription factors: star players for plant growth and development. *J. Integr. Plant Biol.* 52, 946–951. doi: 10.1111/j.1744-7909.2010.00987.x
- Dal Santo, S., Vannozzi, A., Tornielli, G. B., Fasoli, M., Venturini, L., Pezzotti, M., et al. (2013). Genome-wide analysis of the expansin gene superfamily reveals grapevine-specific structural and functional characteristics. *PLoS One* 8:e62206. doi: 10.1371/journal.pone.0062206
- Dereeper, A., Nicolas, S., Lecunff, L., Bacilieri, R., Doligez, A., Peros, J. P., et al. (2011). SnpPlay: a web-based tool for detection, management and analysis of SNPs. application to grapevine diversity projects. *BMC Bioinformatics* 12:134. doi: 10.1186/1471-2105-12-134
- Dermastia, M., Nikolić, P., Chersicola, M., and Gruden, K. (2015). Transcriptional profiling in infected and recovered grapevine plant responses to ‘*Candidatus* Phytoplasma solani’. *Phytopathogenic Mollicutes* 5, 123–124. doi: 10.5958/2249-4677.2015.00053.5
- Ehya, F., Monavarfeshani, A., Mohseni Fard, E., Karimi Farsad, L., Khayam Nekouei, M., Mardi, M., et al. (2013). Phytoplasma - responsive microRNAs modulate hormonal, nutritional, and stress signalling pathways in Mexican lime trees. *PLoS One* 8:e66372. doi: 10.1371/journal.pone.0066372
- Ferreira e Silva, G. F., Silva, E. M., Azevedo Mda, S., Guivin, M. A., Ramiro, D. A., Figueiredo, C. R., et al. (2014). microRNA156-targeted SPL/SBP box transcription factors regulate tomato ovary and fruit development. *Plant J.* 78, 604–618. doi: 10.1111/tpj.12493
- Gai, Y. P., Li, Y. Q., Guo, F. Y., Yuan, C. Z., Mo, Y. Y., Zhang, H. L., et al. (2014). Analysis of phytoplasma-responsive sRNAs provide insight into the pathogenic mechanisms of mulberry yellow dwarf disease. *Sci. Rep.* 4:5378. doi: 10.1038/srep05378
- Gai, Y. P., Zhao, H.-N., Zhao, Y.-N., Zhu, B.-S., Yuan, S.-S., Li, S., et al. (2018). miRNA-seq-based profiles of miRNAs in mulberry phloem sap provide insight into the pathogenic mechanisms of mulberry yellow dwarf disease. *Sci. Rep.* 8:812. doi: 10.1038/s41598-018-19210-7
- Gambino, G. (2015). “Multiplex RT-PCR method for the simultaneous detection of nine grapevine viruses,” in *Plant Virology Protocols: New Approaches to Detect Viruses and Host Responses*, 3rd Edn, Vol. 1236, eds I. Uyeda and C. Masuta (Totowa, NJ: Humana Press Inc), 39–47. doi: 10.1007/978-1-4939-1743-3\_4
- Gambino, G., Boccacci, P., Margaria, P., Palmano, S., and Gribaudo, I. (2013). Hydrogen peroxide accumulation and transcriptional changes in grapevines recovered from Flavescence dorée disease. *Phytopathology* 103, 776–784. doi: 10.1094/PHYTO-11-12-0309-R
- Gambino, G., Dal Molin, A., Boccacci, P., Minio, A., Chitarra, W., Avanzato, C. G., et al. (2017). Whole-genome sequencing and SNV genotyping of ‘Nebbiolo’ (*Vitis vinifera* L.) clones. *Sci. Rep.* 7:17294. doi: 10.1038/s41598-017-17405-y
- González, V. M., Müller, S., Baulcombe, D., and Puigdomenech, P. (2015). Evolution of NBS-LRR gene copies among dicot plants and its regulation by members of the miR482/2118 superfamily of miRNAs. *Mol. Plant* 8, 329–331. doi: 10.1016/j.molp.2014.11.013
- Grimplet, J., Cramer, G. R., Dickerson, J. A., Mathiason, K., Van Hemert, J., and Fennell, A. Y. (2009). VitisNet: “Omics” integration through grapevine molecular networks. *PLoS One* 4:e8365. doi: 10.1371/journal.pone.0008365



- Gutierrez, L., Mongelard, G., Flokova, K., Pacurar, D. I., Novák, O., Staswick, P., et al. (2012). Auxin controls *Arabidopsis* adventitious root initiation by regulating jasmonic acid homeostasis. *Plant Cell* 24, 2515–2527. doi: 10.1105/tpc.112.099119
- Han, J., Fang, J., Wang, C., Yin, Y., Sun, X., Leng, X., et al. (2014). Grapevine microRNAs responsive to exogenous gibberellin. *BMC Genomics* 15:111. doi: 10.1186/1471-2164-15-111
- Han, L., Weng, K., Ma, H., Xiang, G., Li, Z., Wang, Y., et al. (2016). Identification and characterization of Erysiphe necator-responsive microRNAs in Chinese wild *Vitis pseudoreticulata* by high-throughput sequencing. *Front. Plant Sci.* 7:621. doi: 10.3389/fpls.2016.00621
- Harm, A., Kassemeyer, H., Seibicke, T., and Regner, F. (2011). Evaluation of chemical and natural resistance inducers against downy mildew (*Plasmopara viticola*) in grapevine. *Am. J. Enol. Vitic.* 62, 184–192. doi: 10.5344/ajev.2011.09054
- Harvey, J. J., Lewsey, M. G., Patel, K., Westwood, J., Heimstadt, S., Carr, J. P., et al. (2011). An antiviral defense role of AGO2 in plants. *PLoS One* 6:e14639. doi: 10.1371/journal.pone.0014639
- Hoshi, A., Oshima, K., Kakizawa, S., Ishii, Y., Ozeki, J., Hashimoto, M., et al. (2009). A unique virulence factor for proliferation and dwarfism in plants identified from a phytopathogenic bacterium. *Proc. Natl. Acad. Sci. U.S.A.* 106, 6416–6421. doi: 10.1073/pnas.0813038106
- Hren, M., Nikolić, P., Rotter, A., Blejcek, A., Terrier, N., Ravnika, M., et al. (2009). 'Bois noir' phytoplasma induces significant reprogramming of the leaf transcriptome in the field grown grapevine. *BMC Genomics* 10:460. doi: 10.1186/1471-2164-10-460
- Jaillon, O., Aury, J. M., Noel, B., Policriti, A., Clepet, C., Casagrande, A., et al. (2007). The grapevine genome sequence suggests ancestral hexaploidization in major angiosperm phyla. *Nature* 449, 463–467. doi: 10.1038/nature06148
- Jiu, S., Zhu, X., Wang, J., Zhang, C., Mu, Q., Wang, C., et al. (2015). Genome-wide mapping and analysis of grapevine microRNAs and their potential target genes. *Plant Genome* 8, 1–16. doi: 10.3835/plantgenome2014.12.0091
- Kozomara, A., and Griffiths-Jones, S. (2014). miRBase: annotating high confidence microRNAs using sequencing data. *Nucleic Acids Res.* 42, D68–D73. doi: 10.1093/nar/gkt1181
- Kurihara, Y., and Watanabe, Y. (2004). *Arabidopsis* micro-RNA biogenesis through Dicer-like 1 protein functions. *Proc. Natl. Acad. Sci. U.S.A.* 101, 12753–12758. doi: 10.1073/pnas.0403115101
- Leng, X. P., Wang, P. P., Zhao, P. C., Wang, M. Q., Cui, L. W., Shangguan, L. D., et al. (2017). Conservation of microRNA-mediated regulatory networks in response to copper stress in grapevine. *Plant Growth Regul.* 82, 293–304. doi: 10.1007/s10725-017-0259-2
- Liu, N., Wu, S., Van Houten, J., Wang, Y., Ding, B., Fei, Z., et al. (2014). Down-regulation of AUXIN RESPONSE FACTORS 6 and 8 by microRNA 167 leads to floral development defects and female sterility in tomato. *J. Exp. Bot.* 65, 2507–2520. doi: 10.1093/jxb/eru141
- Liu, S., Li, J. H., Wu, J., Zhou, K. R., Zhou, H., Yang, J. H., et al. (2015). StarScan: a web server for scanning small RNA targets from degradome sequencing data. *Nucleic Acids Res.* 43, W480–W486. doi: 10.1093/nar/gkv524
- Liu, W.-W., Meng, J., Cui, J., and Luan, Y.-S. (2017). Characterization and function of microRNA\*s in plants. *Front. Plant Sci.* 8:2200. doi: 10.3389/fpls.2017.02200
- Maggi, F., Bosco, D., Galetto, L., Palmano, S., and Marzachi, C. (2017). Space-time point pattern analysis of Flavescence Dorée epidemic in a grapevine field: disease progression and recovery. *Front. Plant Sci.* 7:1987. doi: 10.3389/fpls.2016.01987
- Mallory, A. C., Elmayan, T., and Vaucheret, H. (2008). MicroRNA maturation and action – the expanding roles of Argonautes. *Curr. Opin. Plant Biol.* 11, 560–566. doi: 10.1016/j.pbi.2008.06.008
- Marche, C., Léon, C., Kappel, C., Coutos-Thévenot, P., Corio-Costet, M. F., Delrot, S., et al. (2013). Over-expression of VvWRKY1 in grapevines induces expression of Jasmonic acid pathway-related genes and confers higher tolerance to the downy mildew. *PLoS One* 8:e54185. doi: 10.1371/journal.pone.0054185
- Margaria, P., Abbà, S., and Palmano, S. (2013). Novel aspects of grapevine response to phytoplasma infection investigated by a proteomic and phospho-proteomic approach with data integration into functional networks. *BMC Genomics* 14:38. doi: 10.1186/1471-2164-14-38
- Margaria, P., Ferrandino, A., Caciagli, P., Kedrina, O., Schubert, A., and Palmano, S. (2014). Metabolic and transcript analysis of the flavonoid pathway in diseased and recovered Nebbiolo and Barbera grapevines (*Vitis vinifera* L.) following infection by Flavescence dorée phytoplasma. *Plant Cell Environ.* 37, 2183–2200. doi: 10.1111/pce.12332
- Marzachi, C., Alma, A., d'Aquilio, M., Minuto, G., and Boccardo, G. (1999). Detection and identification of phytoplasmas infecting cultivated and wild plants in Liguria (Italian Riviera). *J. Plant Pathol.* 81, 127–136. doi: 10.4454/jpp.v81i2.1057
- Meyers, B. C., Axtell, M. J., Bartel, B., Baulcombe, D., Bowman, J. L., Cao, X., et al. (2008). Criteria for annotation of plant microRNAs. *Plant Cell* 20, 3186–3190. doi: 10.1105/tpc.108.064311
- Minato, N., Himeno, M., Hoshi, A., Maejima, K., Komatsu, K., Takebayashi, Y., et al. (2014). The phytoplasmal virulence factor TENGU causes plant sterility by downregulating of the jasmonic acid and auxin pathways. *Sci. Rep.* 4:7399. doi: 10.1038/srep07399
- Mohorianu, I., Stocks, M. B., Applegate, C. S., Folkes, L., and Moulton, V. (2017). The UEA small RNA workbench: a suite of computational tools for small RNA analysis. *Methods Mol. Biol.* 1580, 193–224. doi: 10.1007/978-1-4939-6866-4\_14
- Moretto, M., Sonego, P., Pilati, S., Malacarne, G., Costantini, L., Grzeskowiak, L., et al. (2016). VESPUCCI: exploring patterns of gene expression in grapevine. *Front. Plant Sci.* 7:633. doi: 10.3389/fpls.2016.00633
- Morgante, M., De Paoli, E., and Radovic, S. (2007). Transposable elements and the plant pan-genomes. *Curr. Opin. Plant Biol.* 10, 149–155. doi: 10.1016/j.pbi.2007.02.001
- Moxon, S., Schwach, F., Dalmay, T., Maclean, D., Studholme, D. J., and Moulton, V. (2008). A toolkit for analysing large-scale plant small RNA datasets. *Bioinformatics* 24, 2252–2253. doi: 10.1093/bioinformatics/btn428
- Musetti, R., Marabottini, R., Badiani, M., Martini, M., Sanità di Toppi, L., Borselli, S., et al. (2007). On the role of H2O2 in the recovery of grapevine (*Vitis vinifera*, cv. Prosecco) from Flavescence dorée disease. *Funct. Plant Biol.* 34, 750–758. doi: 10.1071/FP06308
- Nozawa, M., Miura, S., and Nei, M. (2012). Origins and evolution of microRNA genes in plant species. *Genome Biol. Evol.* 4, 230–239. doi: 10.1093/gbe/evs002
- Osler, R., Loi, N., Carraro, L., Musetti, R., Loschi, A., and Ermacora, P. (2004). Spontaneous recovery in grapevines affected by Flavescence dorée phytoplasma. *Acta Physiol. Plant.* 26:134.
- Pagliarini, C., Vitali, M., Ferrero, M., Vitulo, N., Incarboni, M., Lovisolo, C., et al. (2017). The accumulation of miRNAs differentially modulated by drought stress is affected by grafting in grapevine. *Plant Physiol.* 173, 2180–2195. doi: 10.1104/pp.16.01119
- Paim Pinto, D. L., Brancadoro, L., Dal Santo, S., De Lorenzis, G., Pezzotti, M., Meyers, B. C., et al. (2016). The influence of genotype and environment on small RNA profiles in grapevine berry. *Front. Plant Sci.* 7:1459. doi: 10.3389/fpls.2016.01459
- Pantaleo, V., Szittya, G., Moxon, S., Miozzi, L., Moulton, V., Dalmay, T., et al. (2010). Identification of grapevine microRNAs and their targets using high-throughput sequencing and degradome analysis. *Plant J.* 62, 960–976. doi: 10.1111/j.0960-7412.2010.04208.x
- Pantaleo, V., Vitali, M., Boccacci, P., Miozzi, L., Cuozzo, D., Chitarra, W., et al. (2016). Novel functional microRNAs from virus-free and infected *Vitis vinifera* plants under water stress. *Sci. Rep.* 6:20167. doi: 10.1038/srep20167
- Paolacci, A. R., Cataricone, G., Ederli, L., Zadra, C., Pasqualini, S., Badiani, M., et al. (2017). Jasmonate-mediated defence responses, unlike salicylate-mediated responses, are involved in the recovery of grapevine from bois noir disease. *BMC Plant Biol.* 17:118. doi: 10.1186/s12870-017-1069-4
- Prezelj, N., Covington, E., Roitsch, T., Gruden, K., Fragner, L., Weckwerth, W., et al. (2016). Metabolic consequences of infection of grapevine (*Vitis vinifera* L.) cv. "Modra frankinja" with Flavescence Dorée phytoplasma. *Front. Plant Sci.* 7:111. doi: 10.3389/fpls.2016.00711
- Pulverenti, A., Giugno, R., Distefano, R., Pigola, G., Mongiovì, M., Giudice, G., et al. (2015). A knowledge base for *Vitis vinifera* functional analysis. *BMC Syst. Biol.* 9(Suppl. 3):S5. doi: 10.1186/1752-0509-9-S3-S5
- Rotter, A., Nikolić, P., Turnšek, N., Kogovšek, P., Blejcek, A., Gruden, K., et al. (2018). Statistical modelling of long-term grapevine response to 'Candidatus Phytoplasma solani' infection in the field. *Eur. J. Plant Pathol.* 150, 653–668. doi: 10.1007/s10658-017-1310-x

- Schommer, C., Palatnik, J. F., Aggarwal, P., Chételat, A., Cubas, P., and Farmer, E. E. (2008). Control of jasmonate biosynthesis and senescence by miR319 targets. *PLoS Biol.* 6:e230. doi: 10.1371/journal.pbio.0060230
- Shao, C., Ma, X., Xu, X., Wang, H., and Meng, Y. (2012). Genome-wide identification of reverse complementary microRNA genes in plants. *PLoS One* 7:e46991. doi: 10.1371/journal.pone.0046991
- Shao, F., Zhang, Q., Liu, H., Lu, S., and Qiu, D. (2016). Genome-wide identification and analysis of MicroRNAs involved in witches'-broom phytoplasma response in *Ziziphus jujuba*. *PLoS One* 11:e0166099. doi: 10.1371/journal.pone.0166099
- Shi, R., and Chiang, V. L. (2005). Facile means for quantifying microRNA expression by real-time PCR. *Biotechniques* 39, 519–525. doi: 10.2144/000112010
- Singh, K., Talla, A., and Qiu, W. (2012). Small RNA profiling of virus-infected grapevines: evidences for virus infection-associated and variety-specific miRNAs. *Funct. Integr. Genomics* 12, 659–669. doi: 10.1007/s10142-012-0292-1
- Snyman, M. C., Solofaharivelo, M.-C., Souza-Richards, R., Stephan, D., Murray, S., and Burger, J. T. (2017). The use of high-throughput small RNA sequencing reveals differentially expressed microRNAs in response to aster yellows phytoplasma-infection in *Vitis vinifera* cv. 'Chardonnay'. *PLoS One* 12:e0182629. doi: 10.1371/journal.pone.0182629
- Staiger, D., Korneli, C., Lummer, M., and Navarro, L. (2013). Emerging role for RNA-based regulation in plant immunity. *New Phytol.* 197, 394–404. doi: 10.1111/nph.12022
- Sugio, A., Kingdom, H. N., MacLean, A. M., Grieve, V. M., and Hogenhout, S. A. (2011). Phytoplasma protein effector SAP11 enhances insect vector reproduction by manipulating plant development and defense hormone biosynthesis. *Proc. Natl. Acad. Sci. U.S.A.* 108, E1254–E1263. doi: 10.1073/pnas.1105664108
- Sun, X., Fan, G., Su, L., Wang, W., Liang, Z., Li, S., et al. (2015). Identification of cold-inducible microRNAs in grapevine. *Front. Plant Sci.* 6:595. doi: 10.3389/fpls.2015.00595
- Sunkar, R., Li, Y. F., and Jagadeeswaran, G. (2012). Functions of microRNAs in plant stress responses. *Trends Plant Sci.* 17, 196–203. doi: 10.1016/j.tplants.2012.01.010
- Taylor, R. S., Tarver, J. E., Foroozani, A., and Donoghue, P. C. J. (2017). MicroRNA annotation of plant genomes - Do it right or not at all. *Bioessays* 39:1600113. doi: 10.1002/bies.201600113
- Tyler, D. M., Okamura, K., Chung, W. J., Hagen, J. W., Berezikov, E., Hannon, G. J., et al. (2008). Functionally distinct regulatory RNAs generated by bidirectional transcription and processing of microRNA loci. *Genes Dev.* 22, 26–36. doi: 10.1101/gad.1615208
- Velasco, R., Zharkikh, A., Troggio, M., Cartwright, D. A., Cestaro, A., Pruss, D., et al. (2007). A high quality draft consensus sequence of the genome of a heterozygous grapevine variety. *PLoS One* 2:e1326. doi: 10.1371/journal.pone.0001326
- Vitali, M., Chitarra, W., Galetto, L., Bosco, D., Marzachi, C., Gullino, M. L., et al. (2013). Flavescence dorée phytoplasma deregulates stomatal control of photosynthesis in *Vitis vinifera*. *Ann. Appl. Biol.* 162, 335–346. doi: 10.1111/aab.12025
- Wang, C., Han, J., Korir, N. K., Wang, X., Liu, H., Li, X., et al. (2013). Characterization of target mRNAs for grapevine microRNAs with an integrated strategy of modified RLM-RACE, newly developed PPM-RACE and qPCRs. *J. Plant Physiol.* 170, 943–957. doi: 10.1016/j.jplph.2013.02.005
- Wang, C., Han, J., Liu, C., Kibet, K. N., Kayesh, E., Shangguan, L., et al. (2012). Identification of microRNAs from Amur grape (*Vitis amurensis* Rupr.) by deep sequencing and analysis of microRNA variations with bioinformatics. *BMC Genomics* 13:122. doi: 10.1186/1471-2164-13-122
- Wang, C., Leng, X. P., Zhang, Y. Y., Kayesh, E., Zhang, Y. P., Sun, X., et al. (2014). Transcriptome-wide analysis of dynamic variations in regulation modes of grapevine microRNAs on their target genes during grapevine development. *Plant Mol. Biol.* 84, 269–285. doi: 10.1007/s11103-013-0132-2
- Wang, C., Wang, X., Nicholas, K. K., Song, C., Zhang, C., Li, X., et al. (2011). Deep sequencing of grapevine flower and berry short RNA library for discovery of new microRNAs and validation of precise sequences of grapevine microRNAs deposited in miRBase. *Physiol. Plant.* 143, 64–81. doi: 10.1111/j.1399-3054.2011.01481.x
- Wasternack, C., and Hause, B. (2013). Jasmonates: biosynthesis, perception, signal transduction and action in plant stress response, growth and development. An update to the 2007 review in annals of botany. *Ann. Bot.* 111, 1021–1058. doi: 10.1093/aob/mct067
- Wong, D. C. J., Sweetman, C., Drew, D. P., and Ford, C. M. (2013). VTCdb: a gene co-expression database for the crop species *Vitis vinifera* (grapevine). *BMC Genomics* 14:882. doi: 10.1186/1471-2164-14-882
- Wu, G., and Poethig, R. S. (2006). Temporal regulation of shoot development in *Arabidopsis thaliana* by miR156 and its target SPL3. *Development* 133, 3539–3547. doi: 10.1242/dev.02521
- Xie, Y., Huhn, K., Brandt, R., Potschin, M., Bieker, S., and Straub, D. (2014). REVOLUTA and WRKY53 connect early and late leaf development in *Arabidopsis*. *Development* 141, 4772–4783. doi: 10.1242/dev.117689
- Zhao, F., Wang, C., Han, J., Zhu, X., Li, X., Wang, X., et al. (2017). Characterization of miRNAs responsive to exogenous ethylene in grapevine berries at whole genome level. *Funct. Integr. Genomics* 17, 213–235. doi: 10.1007/s10142-016-0514-z
- Zhou, Y., Honda, M., Zhu, H., Zhang, Z., Guo, X., and Li, T. (2015). Spatiotemporal sequestration of miR165/166 by *Arabidopsis* Argonaute10 promotes shoot apical meristem maintenance. *Cell Rep.* 10, 1819–1827. doi: 10.1016/j.celrep.2015.02.047
- Zuker, M. (2003). Mfold web server for nucleic acid folding and hybridization prediction. *Nucleic Acids Res.* 31, 3406–3415. doi: 10.1093/nar/gkg595

**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 © 2018 Chitarra, Pagliarini, Abbà, Boccacci, Birello, Rossi, Palmano, Marzachi, Perrone and Gambino. 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.



# The MicroRNA319d/TCP10 Node Regulates the Common Bean – Rhizobia Nitrogen-Fixing Symbiosis

José Á. Martín-Rodríguez, Alfonso Leija, Damien Formey and Georgina Hernández\*

Centro de Ciencias Genómicas, Universidad Nacional Autónoma de México, Cuernavaca, Mexico

## OPEN ACCESS

### Edited by:

Ana Confraria,  
Instituto Gulbenkian de Ciência (IGC),  
Portugal

### Reviewed by:

Gary Stacey,  
University of Missouri, United States  
Ilker Buyuk,  
Ankara University, Turkey  
Carla Schommer,  
CONICET Instituto de Biología  
Molecular y Celular de Rosario (IBR),  
Argentina

### \*Correspondence:

Georgina Hernández  
gina@ccg.unam.mx

### Specialty section:

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

**Received:** 14 May 2018

**Accepted:** 23 July 2018

**Published:** 10 August 2018

### Citation:

Martín-Rodríguez JÁ, Leija A,  
Formey D and Hernández G (2018)  
The MicroRNA319d/TCP10 Node  
Regulates the Common Bean –  
Rhizobia Nitrogen-Fixing Symbiosis.  
Front. Plant Sci. 9:1175.  
doi: 10.3389/fpls.2018.01175

Micro-RNAs from legume plants are emerging as relevant regulators of the rhizobia nitrogen-fixing symbiosis. In this work we functionally characterized the role of the node conformed by micro-RNA319 (miR319) – TEOSINTE BRANCHED/CYCLOIDEA/PCF (TCP) transcription factor in the common bean (*Phaseolus vulgaris*) – *Rhizobium tropici* symbiosis. The miR319d, one of nine miR319 isoforms from common bean, was highly expressed in root and nodules from inoculated plants as compared to roots from fertilized plants. The miR319d targets *TCP10* (Phvul.005G067950), identified by degradome analysis, whose expression showed a negative correlation with miR319d expression. The phenotypic analysis of *R. tropici*-inoculated composite plants with transgenic roots/nodules overexpressing or silencing the function of miR319d demonstrated the relevant role of the miR319d/TCP10 node in the common bean rhizobia symbiosis. Increased miR319d resulted in reduced root length/width ratio, increased rhizobial infection evidenced by more deformed root hairs and infection threads, and decreased nodule formation and nitrogenase activity per plant. In addition, these plants with lower TCP10 levels showed decreased expression level of the jasmonic acid (JA) biosynthetic gene: *LOX2*. The transcription of *LOX2* by TCPs has been demonstrated for Arabidopsis and in several plants *LOX2* level and JA content have been associated with TCP levels. On this basis, we propose that in roots/nodules of inoculated common bean plants TCP10 could be the transcriptional regulator of *LOX2* and the miR319d/TCP10 node could affect nodulation through JA signaling. However, given the complexity of nodulation, the participation of other signaling pathways in the phenotypes observed cannot be ruled out.

**Keywords:** microRNAs, legume–rhizobia interaction, symbiotic nitrogen fixation, nodules, common bean, *Phaseolus vulgaris*

## INTRODUCTION

Legumes are ecologically important because of their ability to establish an efficient symbiotic association with nitrogen-fixing rhizobia, resulting in the formation of root nodules, where rhizobia can fix the atmospheric dinitrogen (N<sub>2</sub>) in forms that can be assimilated by the plant, in exchange for a carbon source. Symbiotic nitrogen fixation (SNF) reduces the cost of legume cultivation and is relevant for sustainable agriculture (Venkateshwaran et al., 2013). The evolution of this symbiosis was a key to success for the legume family that comprises 18,000 described species

with approximately 700 genera and represents one-third of the primary crop production in the world; however, legume production necessary for feed and food relies on only a few cultivated species (Doyle and Luckow, 2003). *Phaseolus vulgaris*, known as common bean, is the principal source of non-animal protein for human consumption in the developing world (Broughton et al., 2003). Besides the caloric and proteic intake, common bean grains have high contents of fiber, complex carbohydrates and other dietary elements as minerals, thiamine, folate, and a variety of flavonoids and secondary metabolites with medicinal properties (Blair et al., 2013).

In recent years, several studies have shown different classes of small non-coding RNAs (sRNA) that act as essential regulators of gene expression in plants. MicroRNAs (miRNA) are a major class of sRNA, 21 – 24 nt in length, that regulate gene expression post-transcriptionally through sequence complementarity, either via target transcript cleavage or translational inhibition. Plant miRNAs are involved in most, if not all, biological processes such as development, hormone regulation, nutrient homeostasis and interaction with pathogens and symbionts (reviewed by Jones-Rhoades and Bartel, 2004; Rogers and Chen, 2013; Li et al., 2017). Growing evidence supports the participation of miRNAs in the control of the legume-rhizobia symbiosis (Lelandais-Brière et al., 2016). Studies based on high-throughput sRNA sequencing have identified miRNA families that are expressed in nodules from different legume species (Subramanian et al., 2008; Lelandais-Brière et al., 2009; De Luis et al., 2012; Turner et al., 2012). For common bean we identified a set of 185 mature miRNAs, 106 of this, including 50 previously unpublished sequences, were present in nodules (Formey et al., 2015). Aiming to understand the role in nodules of newly identified common bean miRNAs we constructed weighted correlation networks of miRNAs with differential expression in the nodule library as compared to other libraries. The networks include miRNAs known to play regulatory roles in nodules suggesting a similar role for novel miRNAs (Formey et al., 2015). One of these weighted correlation networks included an isoform of the miR319 family, pvu-miR319d, initially described in soybean (Wong et al., 2011). In this work we analyzed the role of miR319d from common bean in symbiosis with *R. tropici*.

Though the majority of plant miRNAs have been identified by large-scale sequencing strategies and bioinformatics approaches based on the conservation of fold-back precursors (Jones-Rhoades and Bartel, 2004), Arabidopsis miR319a is an exception as it was isolated through the screening of an activation-tagging T-DNA transgenic population that generated dominant gain-of-function mutations (Weigel et al., 2000). The first described plant miRNA mutant, *jaw-D*, overexpresses miR319a that is one of the first characterized and conserved plant miRNA families. It was demonstrated that the conserved miR319 targets are the plant-specific transcription factors (TF) *TCP* (for *TEOSINTE BRANCHED/CYCLOIDEA/PCF*) (Palatnik et al., 2003). The *TCP* domain codes a DNA-binding motif that folds into a basic helix-loop-helix structure (Cubas et al., 1999). The *TCP* TFs participate in various important aspects of plant development, especially the control of cell division, expansion, and differentiation during

leaf development but also other important functions such as mitochondrial biogenesis, leaf senescence and floral development (Palatnik et al., 2003; Schommer et al., 2008, 2012; Martín-Trillo and Cubas, 2010). The *TCP*s can be subdivided in two main branches (class I and II) according to their sequence in the *TCP* domain. In Arabidopsis, the *TCP*s comprise a family of 24 members; only five of these (*TCP2*, *TCP3*, *TCP4*, *TCP10*, and *TCP24*) belonging to class II are targets of miR319 (Palatnik et al., 2003; Schommer et al., 2008).

The networks of TFs regulated by miRNAs can interact with others during plant development (Rubio-Somoza and Weigel, 2011). Several studies have revealed the interaction of the miR319/*TCP* node with miR164 and miR396 (Schommer et al., 2012). In Arabidopsis *TCP*s, belonging to class II, directly activate the transcription of *MIR396*; this miRNA targets *GROWTH-REGULATING FACTORS* (*GRF*) TFs that in turn regulate cell proliferation via the control of cell cycle genes (Rodríguez et al., 2010).

The leaf morphogenesis process that is regulated by the miR319/*TCP* node has been linked with other processes such as jasmonic acid (JA) biosynthesis and senescence (Schommer et al., 2012). JAs are lipid-derived signaling molecules in plants that regulate diverse responses to wounding, pathogen attack, reproduction, development, metabolic regulation and abiotic stress (Devoto and Turner, 2003; Howe, 2004). The participation of JA in the legume-rhizobium symbiosis has been reported in several studies (Sun et al., 2006; Seo et al., 2006; Poustini et al., 2007; Ferguson and Mathesius, 2014). However these studies are yet inconclusive, collectively these appear to indicate that JAs can act as either positive or negative regulators of nodulation and nitrogen fixation, depending on the legume species, the type of JA used, and when, where, and how the hormone is applied (Ferguson and Mathesius, 2014). The first dedicated step in the biosynthesis of JA is catalyzed by lipoxygenases encoded by the *LOX* genes. In Arabidopsis *LOX2*, and other three *LOX* genes, encode chloroplast-localized lipoxygenases that catalyze the conversion of  $\alpha$ -linolenic acid (18:3) into (13S)-hydroperoxy-linolenic acid. The *LOX2* is one of the most affected genes in the transcriptome of *tcp* loss-of-function Arabidopsis mutants (Schommer et al., 2012). It has been demonstrated that Arabidopsis *TCP*s recognize specific binding sites present in the *LOX2* promoter to directly regulate its transcription (Schommer et al., 2008; Danisman et al., 2012). Other JA biosynthetic genes also respond to miR319/*TCP* levels include the *ALLENE OXIDE SYNTHASE* (*AOS*) that catalyzes the conversion of 13-hydroperoxy-linolenic acid to an unstable allene oxide intermediate (Schommer et al., 2008; Zhang et al., 2016).

To our knowledge the participation of the miR319/*TCP* node as regulator of the legume – rhizobia symbiosis has not been reported. In this paper we analyzed the role of the common bean miR319d isoform and its target *TCP10* in the symbiosis with *Rhizobium tropici*. We confirmed the high expression of miR319d in roots/nodules of SNF common bean as compared to tissues from fertilized (non-inoculated) plants. The functional analysis of composite common bean plants with modulated expression of this miRNA revealed the effect of the miR319d/*TCP10* node



in root development, rhizobia infection, nodulation and SNF. These effects could be related with observed alterations in the expression of *LOX2*, a JA biosynthetic gene, and the participation of JA in the regulation of different stages of symbiosis with *Rhizobium*.

## MATERIALS AND METHODS

### Phylogenetic Analysis

miR319 isoform sequences from *Phaseolus vulgaris* were obtained from the small RNAseq analysis performed by Formey et al. (2015), where each miR319 isoform was referred as designated in the plant species it was discovered. The **Supplementary Table S1** shows the equivalence among the nomenclatures from this work and those from Formey et al. (2015) for each *P. vulgaris* miR319 isoform sequence.

TCP protein sequences were obtained from the *Phaseolus vulgaris* release v2.1, from Phytozome 12 database<sup>1</sup>. Sequence alignments were performed thanks to MAFFT online service v7 (Katoh et al., 2017) with L-INS-i option set. Construction of phylogenetic tree of miR319 isoforms and TCP protein sequences were based on the average linkage (UPGMA) method and Neighbor-Joining JTT model, respectively. Bootstrap values were obtained after 100 resampling.

### Plasmid Construction, Plant Transformation and Generation of Composite Plants

The overexpression and silencing of miR319d function in common bean transgenic roots were carried out using the pTDTO plasmid (Aparicio-Fabre et al., 2013). This expression plasmid contains the 35S cauliflower mosaic virus (35SCaMV) promoter and the tdTomato (red fluorescent protein) gene as a visible reporter gene. The precursor of miR319d (286-bp) was PCR-amplified using as template cDNA from common bean nodules and the specific primers Fw-pre319d (5'-ATGGATCCTGATACTAGAGTACAGGGAGA-3') and R-pre319d (5'-TCTCGAGTTGTGTGTATGTATTAATATTAATG-3'). To silence miR319 function the "Short Tandem Target Mimicry" (STTM) method (Yan et al., 2012) was employed using the specific primers Fw-STTM319d (5'-ATGGATCCGAAGGAGCTCCCTACCTTCAGTCCAGTTGTTGTTGTTATGGTCTAATTTAAATATGGTC-3') and R-STTM319d (5'-ACTCGAGTGGACTGAAGGTAGGGAGCTCCTTCATTCTTCTTCTTTAGACCATATTTAAATTAGACC-3'). The purified PCR products were cleaved by *XhoI* and *BamHI* sites and cloned into the pTDTO expression vector. The empty vector pTDTO, hereafter denominated EV, and the resulting OEmiR319d and STTMmiR319d plasmids were introduced by electroporation into *Agrobacterium rhizogenes* K599, which was then used for plant transformation as described previously (Estrada-Navarrete et al., 2007) with minor modifications (Aparicio-Fabre et al., 2013). The presence of red fluorescence

resulting from the of the tdTomato reporter gene was routinely checked in the putative transgenic roots using light microscopy.

### Plant Material and Growth Conditions

The common bean (*P. vulgaris* L.) Mesoamerican cv BAT 93 was used in this work. Seeds were surface sterilized in 10% (v/v) commercial sodium hypochlorite for 5–10 min and finally rinsed 5–6 times in sterile distilled water. Subsequently seeds were germinated on moist sterile paper towels at 30°C for 2–3 days in darkness. Germinated seedlings of similar size were planted in pots with wet sterile vermiculite. After 2 days of adaptation plants were inoculated with 1 ml saturated liquid culture of the *Rhizobium tropici* CIAT 899 strain per plant. Plants were grown in growth chambers under controlled environmental conditions (25–28°C, 16 h photoperiod) and were watered every 3 days with N-free B&D nutrient solution (Broughton and Dilworth, 1971). For fertilized and non-inoculated condition, full nutrient B&D solution was used to water the plants. Common bean composite plants with transgenic roots were generated as described below and grown in similar conditions to those for wild-type plants. Plants were harvested at different time points for analysis; tissues for RNA isolation were collected directly into liquid nitrogen and stored at –80°C.

### RNA Isolation and Analysis

Total RNA was isolated from 100 mg tissues using mirVana™ miRNA Isolation Kit (Ambion) following the supplier's recommendations. For *R. tropici*-inoculated BAT 93 plants, the tissues used for RNA isolation were roots separated from nodules and detached nodules. For *R. tropici*-inoculated composite plants RNA was isolated from transgenic nodulated root system. Three samples (biological replicates) for each tissue from different plants grown under similar conditions were analyzed.

For the quantification of mature miRNA transcript accumulation levels, cDNAs were prepared using RevertAid reverse transcriptase (Fermentas) following the stem-loop method (Kramer, 2011). Stem-loop primers for reverse transcription of miRNAs were designed as reported by Chen et al. (2005). The conditions used were: denaturation at 65°C for 5 min, then 16°C for 30 min; 60 cycles of 30°C for 30 s, and 42°C for 30 s, 50°C for 1 s followed by 70°C for 15 min. Primers for qRT-PCR amplification are listed in **Supplementary Table S2**. Resulting cDNAs were then diluted 10-fold and used to perform the qRT-PCR experiments using SYBR Green qPCR Master Mix (Fermentas) following manufacturer's instructions. The reaction mix was then dispensed in a 96 well plate and analyzed using real-time thermocycler Applied Biosystem 7300 (Foster City, CA, United States). The thermal cycler settings were as follows: 94°C for 1 min, followed by 40 cycles of 94°C for 20 s and 60°C for 60 s. Relative transcript levels for each sample were obtained using the 'comparative  $C_t$  method' and normalized with the geometrical mean of three housekeeping genes (*HSP*, *MDH*, and *UBQ9*) (Vandesompele et al., 2002) and the U6 sRNA, for the mRNA transcripts and the miRNAs, respectively. In all of our qRT-PCR analyzes a well-defined melting curve was obtained both for miRNAs and for cDNAs. A Mann-Whitney statistical test was performed to evaluate the significance of

<sup>1</sup><https://phytozome.jgi.doe.gov/pz/portal.html>

the differential expression using the mean values from three biological replicates for each condition, using the GraphPad Prism program.

## Phenotypic Analysis

Nitrogenase activity was determined in detached nodulated roots from composite plants by the acetylene reduction assay essentially as described by Hardy et al. (1968). Specific activity is expressed as nmol ethylene h<sup>-1</sup> per plant. The root fresh weight, area, length and root width were determined in composite plants grown, for 24 dpi, under symbiotic conditions. The quantification of root hair deformation and induction of infection thread upon rhizobial inoculation was performed in samples from the responsive zone of roots inoculated with *R. tropici* CIAT 899 for 2 or 6 days. Samples were collected into PBS buffer and were stained with 0.01% (w/v) Methylene Blue for 1 h and washed three times with double-distilled water. Infection events were observed in the optical microscope Axioskop 2 (Zeiss), at least 5 different root responsive zone samples (biological replicates) were used for analysis. Statistical analyses were performed using the Mann–Whitney null hypothesis statistical test.

## Prediction of Transcription Factors Binding Sites (TFBS)

To predict the transcription factors that could regulate the *LOX2* and *LOX5* genes, we performed an analysis using Clover (pre-2010 version, Frith et al., 2004) and the plant JASPAR CORE motif library (Sandelin and Wasserman, 2004) on the 4 kb upstream sequence of 5'UTR end of the corresponding genes. Predicted motifs with a *p*-value > 0.05 were discarded.

## RESULTS

### Common Bean miR319 Isoforms and Target Genes

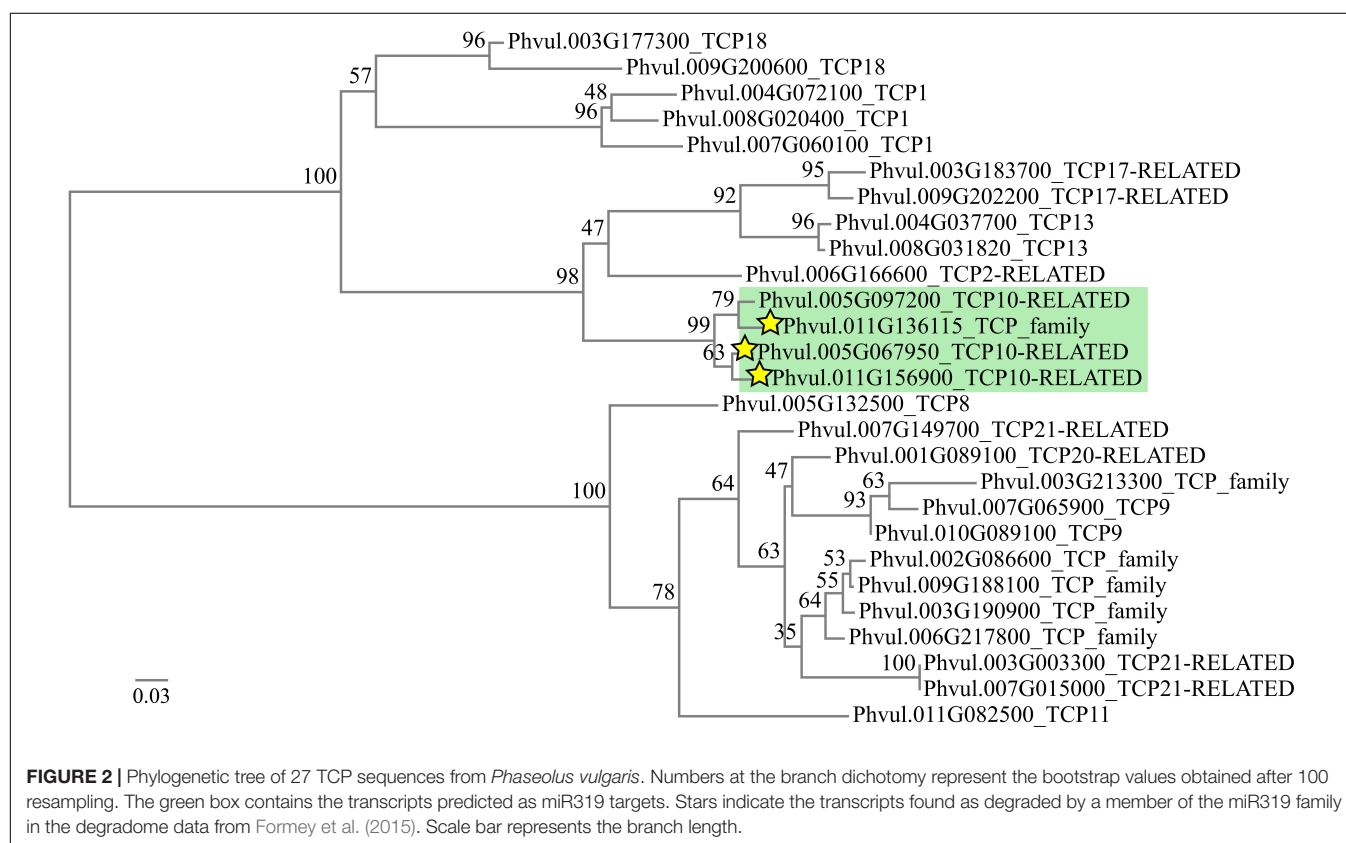
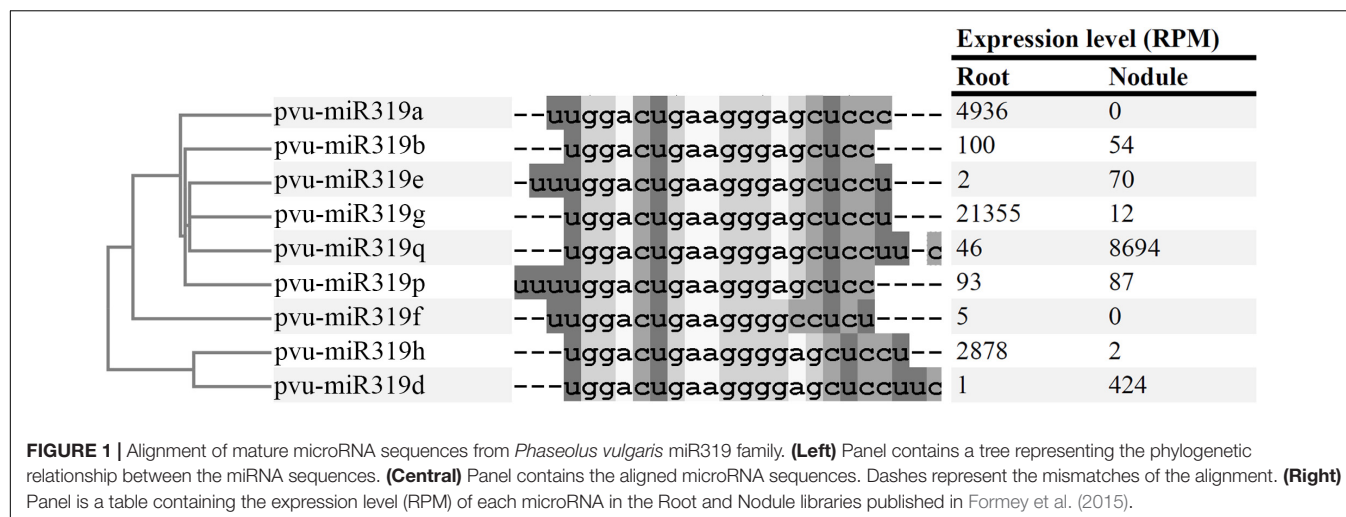
The *Arabidopsis* genome contains three loci that generate miR319 isoforms ath-miR319a to ath-miR319c, while seventeen miR319 isoforms are reported for soybean (*Glycine max*) (v.22)<sup>2</sup>. The high-throughput small RNA (sRNA) sequencing analysis by Formey et al. (2015) identified nine miR319 isoforms in common bean. Of these, five isoforms have been identified in soybean, a legume related to common bean and one was identified in grape (*Vitis vinifera*) (v.22)<sup>2</sup>, while four are new miR319 isoforms similar to the soybean isoforms gma-miR319c, d or f. Initially their nomenclature referred to the species where each miR319 isoforms were identified, while in this work we propose the pvu-miR nomenclature for the *P. vulgaris* miR319 isoforms (Figure 1 and Supplementary Table S1). The common bean miR319 isoforms, 18–22 nucleotides long, showed sequence similarities and are grouped into two well-differentiated clades (Figure 1). One clade included mature miRNAs with four guanines in their central region nucleotide sequence while in the other clade only three guanines were observed. Each clade grouped two of the novel common bean miR319 isoforms.

<sup>2</sup>www.mirbase.org

Analyses of sRNA-seq data from libraries generated from different plant organs have identified conserved and legume-specific miRNA families differentially expressed during nodule organogenesis in different legumes (De Luis et al., 2012; Turner et al., 2012; Formey et al., 2014). Our previous reports (Formey et al., 2015) revealed three isoforms of common bean miR319 with higher expression level in nodules as compared to roots: pvu-miR319e (35-fold), pvu-miR319q (189-fold) and pvu-miR319d (424-fold) (Figure 1). In addition to showing the highest nodule/root expression ratio, pvu-miR319d isoform was included by Formey et al. (2015) in a weighted correlation network of common bean miRNAs with significantly increased expression in the nodule library as compared to other libraries. These features lead us to select the common bean pvu-miR319d isoform for this study aiming to characterize its regulatory role in the rhizobia symbiosis. The pvu-miR319d, hereafter referred as miR319d, was initially identified via high-throughput sequencing data and annotated in the soybean miRBase database by Wong et al. (2011). In common bean the gene encoding miR319d was mapped to chromosome 9 (nucleotides 8534451–8534637), it generates a 187-nucleotides pre-miRNA with *bona fide* stem-loop secondary structure that give rise to the 22-nucleotides mature miRNA encoded close to the 3' end of pre-miR319d (Formey et al., 2015).

The conserved targets for miR319 in different plants are transcripts that encode transcription factors (TF) of class II subclass of the TCP family. Of the 24 *Arabidopsis* TCP TF genes, five contain a target site for miR319 (*TCP2*, *TCP3*, *TCP4*, *TCP10*, and *TCP24*) that, in every case, is located outside the TCP domain and near the 3' part of the coding region (Schommer et al., 2012). In soybean, 14 TCP TF genes have been proposed as miR319 targets (Song et al., 2011; Goettel et al., 2014). Sequence analyses from genomic and transcriptomic data (O'Rourke et al., 2014; Schmutz et al., 2014), led us to identify 27 TCP TF genes for common bean (Figure 2). From the whole set (27) we have identified 4 TCP genes with putative miR319 binding sites near their 3' part of their coding sequence (Formey et al., 2015). Interestingly, the TCP predicted targets of miR319 were organized in a single clade of the phylogenetic tree (Figure 2). From these, Phvul.011G136115, Phvul.005G067950 and Phvul.011G156900 were identified as targets in a degradome analysis (Formey et al., 2015). The base pairing of each predicted TCP target gene with the miR319d isoform as well as their expression level in roots and nodules, (v12.1.6, *Phaseolus vulgaris* v2.1)<sup>3</sup> are shown in Table 1. Phvul.005G097200 and Phvul.011G136115 transcripts showed several mismatches thus a high penalty score for miRNA:mRNA pairing (Table 1) that would not fulfill the requirements for a miR319d target according to Jones-Rhoades and Bartel (2004). Phvul.011G136115 was not expressed and Phvul.005G097200 showed similar expression in roots and nodules (Table 1). Contrastingly, Phvul.005G067950 and Phvul.011G156900 transcripts showed similar low penalty score, a perfect base pairing in the 5' miRNA region and a hybridization energy value of −41.05 Kcal/mol (Jones-Rhoades and Bartel, 2004;

<sup>3</sup>www.phytozome.net



Hammell et al., 2008). The root/nodule expression profile was different among these transcripts, Phvul.011G156900 showed low expression level and slightly higher in nodules while Phvul.005G067950 showed higher expression level in roots than in nodules (Table 1). Combined together, these results converge toward the fact that miR319d induces the cleavage of its target Phvul.005G067950. On this basis, we selected this TCP gene, hereafter denominated *TCP10*, as the target of miR319d for the analysis of this node as possible regulator in the common bean – rhizobia symbiosis. However, the targeting of miR319d to

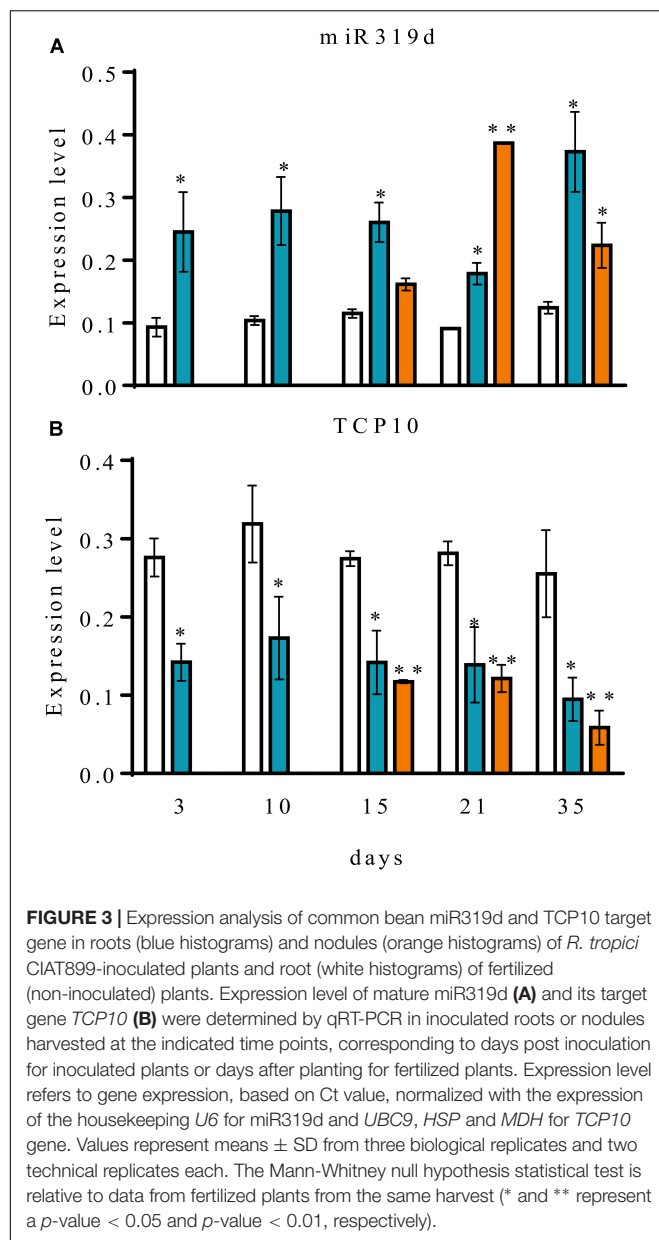
Phvul.011G156900 in other plant organs or growth conditions cannot be ruled out.

## Expression Analysis of miR319d and TCP10 During Rhizobia Symbiosis

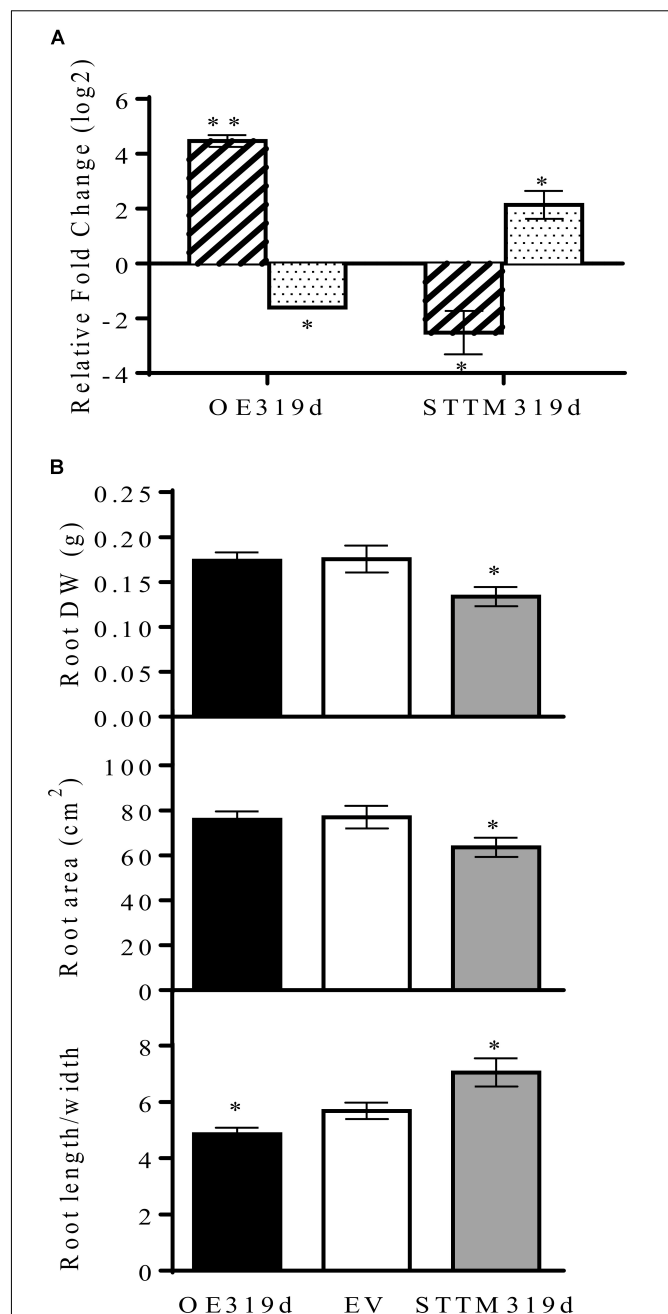
The role of miR319 in leaf/shoot development has been well documented for Arabidopsis and other plants (Schommer et al., 2012, 2014; Koyama et al., 2017). However, although miR319 isoforms have been identified in roots from different plants, its



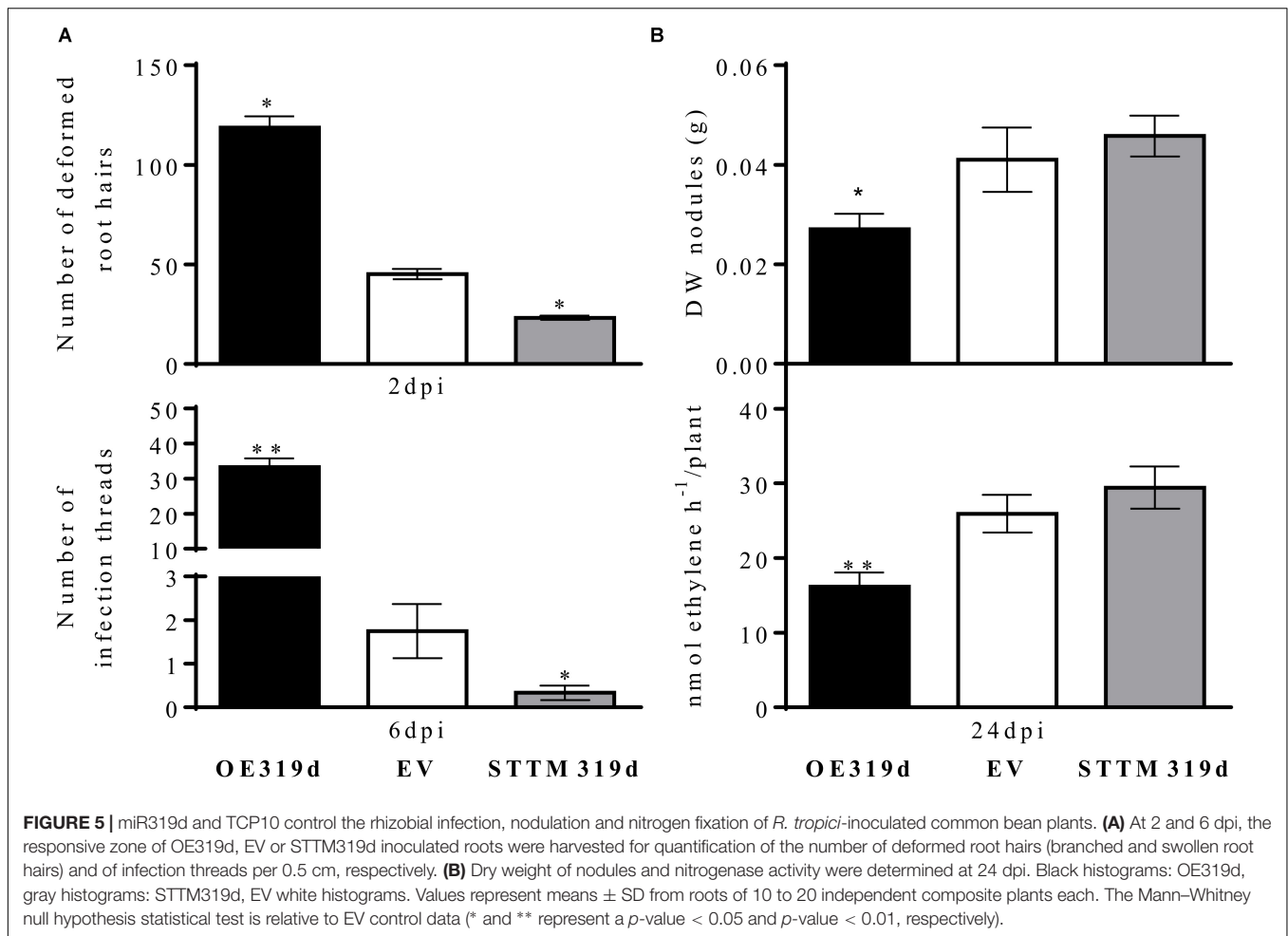




(Valdés-López et al., 2008; Naya et al., 2014; Nova-Franco et al., 2015). The construct for overexpression of the miR319d precursor (OE319d) and the function silencing function construct (STTM319d) (Yan et al., 2012) were driven by the 35SCaMV promoter. Both constructs as well as the control empty vector (EV) contained the tdTomato (red fluorescent protein) reporter gene (Naya et al., 2014). The transcript level of miR319d and *TCP10* from transgenic roots of rhizobia-inoculated composite plants transformed with OE319d and STTM319d constructs, normalized to the transcript level values from EV transgenic roots, are shown in Figure 4A. As expected, the OE319d composite plants showed very high level of miR319d and a decreased level of *TCP10* transcript. Conversely, the STTM319d plants showed low levels of miR319d and increased *TCP10* transcript levels. Also, we determined



**FIGURE 4 |** miR319d and TCP10 affect root morphology of *R. tropici*-inoculated common bean plants. (A) Transcript levels of mature miR319d (diagonally striped histograms) and *TCP10* (dotted histograms) were determined by qRT-PCR in 24 dpi inoculated roots of composite plants transformed with OE319d or STTM319d constructs. Values ( $\log_2$ ) were normalized to the value from control or EV-transformed inoculated roots that was set to 0. Values represent means  $\pm$  SD from three biological replicates and two technical replicates each. (B) Roots dry weight (DW), area and length/width ratio from composite bean plants with transgenic roots overexpressing (OE319d, black histograms) or silencing miR319d (STTM319d, gray histograms) as compared with control roots (EV-transformed, white histograms). Values represent means  $\pm$  SD from nodulated roots of sixteen to twenty independent composite plants each. The Mann-Whitney null hypothesis statistical test is relative to EV control data (\* and \*\* represent a  $p$ -value  $< 0.05$  and  $p$ -value  $< 0.01$ , respectively).



**FIGURE 5 |** miR319d and TCP10 control the rhizobial infection, nodulation and nitrogen fixation of *R. tropici*-inoculated common bean plants. **(A)** At 2 and 6 dpi, the responsive zone of OE319d, EV or STTM319d inoculated roots were harvested for quantification of the number of deformed root hairs (branched and swollen root hairs) and of infection threads per 0.5 cm, respectively. **(B)** Dry weight of nodules and nitrogenase activity were determined at 24 dpi. Black histograms: OE319d, gray histograms: STTM319d, EV white histograms. Values represent means  $\pm$  SD from roots of 10 to 20 independent composite plants each. The Mann-Whitney null hypothesis statistical test is relative to EV control data (\* and \*\* represent a  $p$ -value  $< 0.05$  and  $p$ -value  $< 0.01$ , respectively).

the transcript levels of the Phvul.011G156900 *TCP* gene, the alternative putative target of miR319d (Table 1), the values in OE319d and STTM319d transgenic roots were not significantly different from those of EV (Supplementary Figure S1). These data again indicate that *TCP10* (Phvul.005G067950), and not Phvul.011G156900, is the target gene of miR319d in common bean roots/nodules.

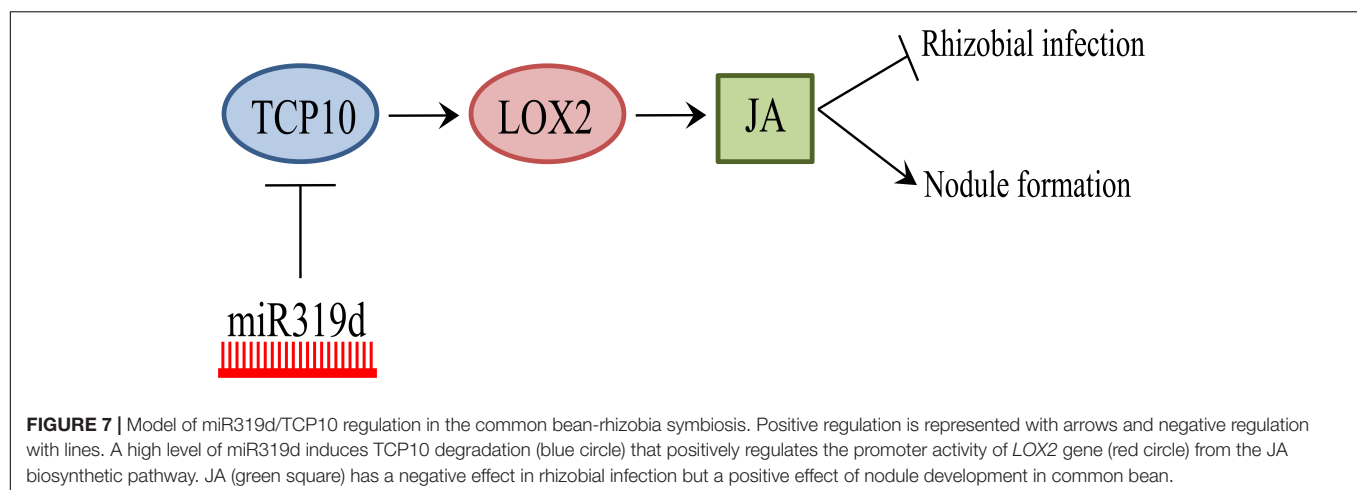
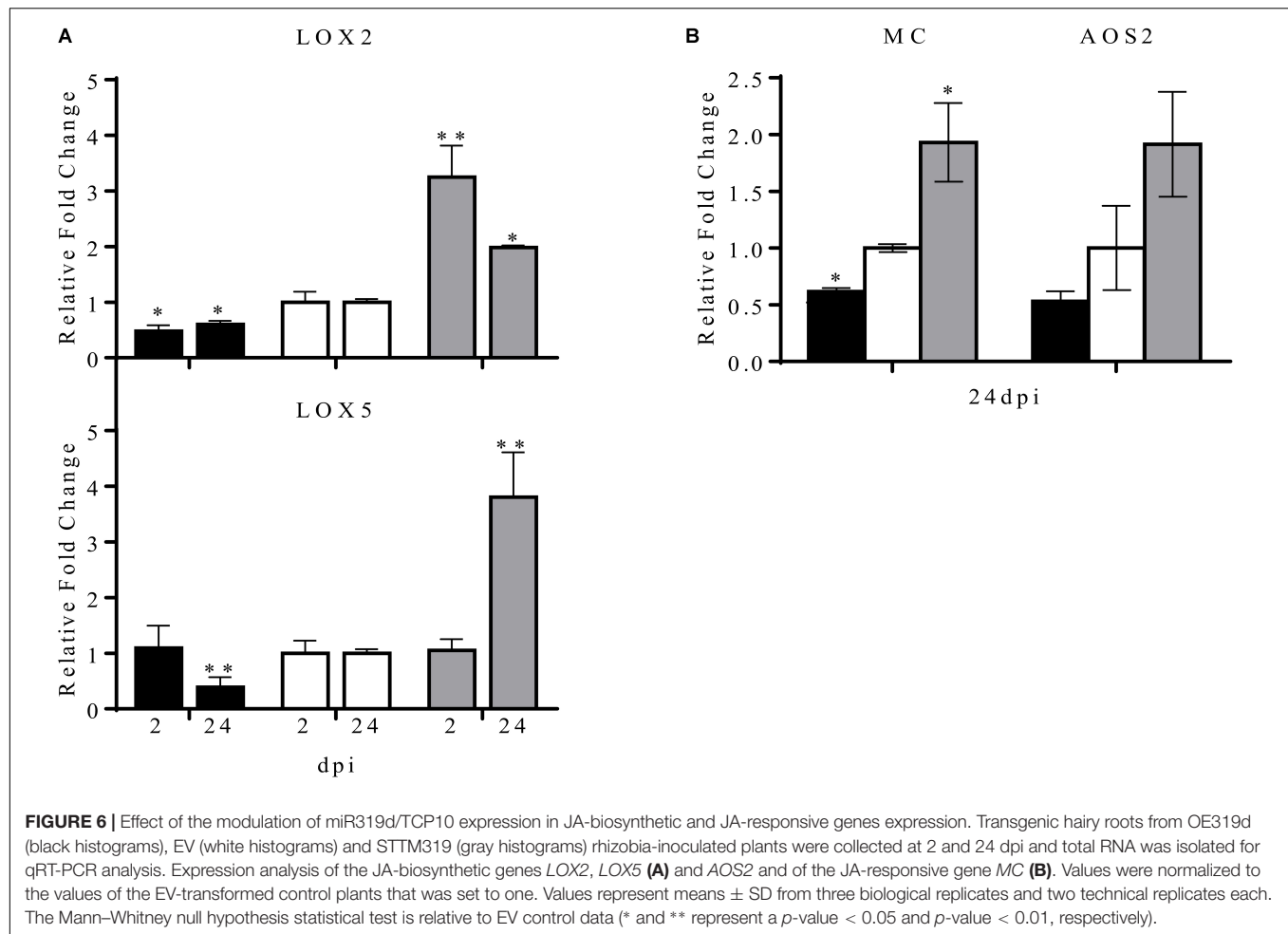
We first assessed if the modulation of miR319d expression affects the root phenotype of *R. tropici*-inoculated common bean plants. As compared to control EV roots, the roots with low miR319d (STTM319d) showed decreased root biomass and area as well as higher length/width ratio due to longer and less dense roots (Figure 4B and Supplementary Figure S2). By contrast, the OE319d roots showed lower length/width ratio and similar root biomass and area as compared to control roots (Figure 4B).

To analyze if the effect of miR319d on root development also affects rhizobial infection and SNF, we investigated the response of miR319d-modulated composite plants to *R. tropici* CIAT 899 infection, nodulation and SNF. Regarding rhizobial infection, we quantified the root hair deformation and the infection thread formation at early symbiotic stages. Notably, the amount of deformed root hairs was significantly higher in 2 dpi-inoculated roots that over-express miR319d, while the

opposite effect was observed in STTM319d roots (Figure 5A). In agreement with this result, 6 dpi-inoculated OE319d roots showed a high increase in the amount of infection threads and the opposite effect was observed in STTM319d inoculated roots (Figure 5A). In addition, earlier infection thread formation, at 2 dpi, was observed only in OE319d roots and not in the other composite plants (Supplementary Figure S3). At nodule maturation (24 dpi) OE319d composite plants showed lower nodule biomass and nitrogenase activity as compared to EV and STTM plants (Figure 5B). Nodule biomass correlated with nodule number and not with altered nodule size because the nodule perimeter was similar in nodules from the different composite plants (OE319d =  $0.43 \pm 0.011$  mm, EV =  $0.43 \pm 0.018$  mm, STTM319d =  $0.43 \pm 0.018$  mm). Overall, miR319d-overexpressing composite plants, with low TCP10 content (Figure 4A), showed a different pattern of effects in rhizobial infection vs. nodule formation/SNF (Figures 5A,B).

## Exploring Downstream TCP10 Regulation in OE319d and STTM319d SNF Plants

The TCP class II TF, targets of miR319, participate in complex regulatory networks that coordinate and balance different events



that are important for plant development and physiology. Relevant functions of TCP genes are the control of leaf and flower size and shape. The signaling pathways associated with these functions include, among others, the regulation of leaf cell proliferation by *GROWTH REGULATING FACTOR* (GRF) TFs, targets of miR396, and also the biosynthesis of JA that

regulates different processes such as senescence (Nicolas and Cubas, 2016).

On this basis, we first analyzed if *TCP10* activates the transcription of *MIR396*, resulting in the degradation of *GROWTH REGULATING FACTOR* (GRF) induced by mature miR396 to control cell proliferation in common bean

inoculated-roots. We quantified the mature miR396 transcript accumulation levels in OE319d and STTM319d nodulated roots, with diminished and increased TCP10 levels, respectively. Our data showed similar levels of miR396 in roots from the different composite plants analyzed (**Supplementary Figure S4**) thus indicating that TCP10 does not regulate miR396 transcription. We propose that miR319d/TCP10 node is not involved in the regulation of the miR396/GRF node nor in the proliferation of cells from common bean roots/nodules. However, further work is required to define if other common bean miR319 isoforms (**Figure 1**) that are highly expressed in leaves, such as pvu-miR319g and pvu-miR319h (Formey et al., 2015) and their TCP target genes participate in the miR396/GRF regulatory network to control leaf development.

The miR319/TCP regulation of leaf morphogenesis has been linked to JA biosynthesis (Schommer et al., 2012). It has been demonstrated that, two Arabidopsis TCP TFs bind to the *LOX2* gene promoter and directly regulate its transcription (Schommer et al., 2008; Danisman et al., 2012). In addition, the expression of other JA-biosynthetic and -responsive genes depends on miR319/TCP levels in several plant species (Schommer et al., 2008; Danisman et al., 2012; Hao et al., 2012; Zhang et al., 2016, 2017). On this basis, we analyzed a possible correlation of TCP10 levels and the expression level of JA-related genes in common bean roots/nodules.

In *P. vulgaris* at least five *LOX* genes have been identified, these genes differ in their expression pattern in different plant organs. Of these, the *LOX2* and *LOX5* genes are expressed during nodule development (Porta and Rocha-Sosa, 2000). In this work we analyzed the common bean *LOX2* and *LOX5* expression level in transgenic nodulated roots with overexpression or silencing of miR319d (**Figure 6**). As compared to control (EV) plants, the expression level of *LOX2* was lower in OE319d plants in contrast to its high level in STTM plants, both at early (2 dpi) and later (24 dpi) symbiotic stages (**Figure 6A**). These data indicate the correlation of *LOX2* expression level with the level of TCP10 in transgenic nodulated roots (**Figures 4, 6**). In addition, the expression levels of the JA-biosynthetic genes *LOX5* and *AOS* as well as the *MULTICYSTATIN* (*MC*) JA-responsive gene (Uppalapati et al., 2005; López-Ráez et al., 2010; Martínez-Medina et al., 2017) showed a similar trend to that observed for *LOX2*, being highly expressed in STTM319d nodulated roots at 24 dpi (**Figures 6A,B**).

## DISCUSSION

Small RNAs differentially expressed during nodule organogenesis have been identified in different legumes such as *Medicago truncatula*, soybean (*Glycine max*), *Lotus japonicus* and common bean (Lelandais-Brière et al., 2009; De Luis et al., 2012; Turner et al., 2012; Formey et al., 2014, 2015, 2016). However, only few in-depth studies that evidence the role of miRNAs in the rhizobial infection, nodulation or SNF processes have been reported (reviewed by Lelandais-Brière et al., 2016; Li et al., 2017). Our group has demonstrated the participation of common bean miR398b and miR172c in different stages of the rhizobia

symbiosis (Naya et al., 2014; Nova-Franco et al., 2015). In this work, we identified common bean miR319d as an important regulator of the rhizobial infection and nodulation.

The conserved miR319 family and its targets TCP TF have been extensively characterized in several plant models but most of these studies have been focused in their contribution to the aerial parts, especially leaf development (Schommer et al., 2012, 2014; Koyama et al., 2017). There are no previous studies about the participation of the miR319/TCP node in the control of the legume – rhizobia SNF symbiosis. The miR319d, one of the nine isoforms identified in the common bean, was highly expressed in nodules with respect to other plant organs and was included in a weighted correlation network together with other miRNAs known as regulators of the rhizobial symbiosis (Formey et al., 2015). Here we evidenced that *TCP10* (Phvul.005G067950), previously identified through degradome analysis (Formey et al., 2015), is the target gene of miR319d. Our data on *TCP10* expression profile validated those from the *P. vulgaris* Gene Expression Atlas (O'Rourke et al., 2014) and from the Phytozome data base<sup>5</sup>, regarding the negative correlation with miR319d expression in fertilized roots vs. inoculated roots/nodules at different stages of the symbiosis (**Figure 1**) and in roots/nodules from composite plants overexpressing or silencing the function of miR319d (**Figure 4**).

The regulatory role miR319/TCP in leaf development has been linked with other processes such as the control of cell proliferation by GRF TFs targets of miR396 (Nicolas and Cubas, 2016). However, our data do not support the link of TCP10 as activator of miR396 in inoculated common bean roots (**Supplementary Figure S4**). The regulation of the miR319/TCP node has also been linked to JA signaling that controls different developmental processes such as senescence (Nicolas and Cubas, 2016). JA, along with other phytohormones like ethylene and cytokinin, is a signaling molecule involved in leaf senescence and other developmental processes. The binding of the Arabidopsis TCP TFs TCP4, the target of miR319, and the class I TCP20 to specific motifs within the promoter regions of the *LOX2* JA-biosynthetic gene has been demonstrated through electrophoretic mobility shift and chromatin immunoprecipitation analyses (Schommer et al., 2008; Danisman et al., 2012). In agreement, Arabidopsis *LOX2* is one of the most affected genes depending on TCP4 levels (Schommer et al., 2008, 2012). On this basis, subsequent research in different plant species (i.e., rice, cotton) has linked the TCP transcriptional regulation to JA signaling, through the direct effect in *LOX2* expression, that results in modulation of other JA-related genes (Hao et al., 2012; Zhang et al., 2016). In this work we showed a correlation between *LOX2* and *TCP10* (target of miR319d) transcript levels in transgenic roots with miR319d over-expression or function silencing (**Figure 6**). The latter indicates that, in common bean roots, *LOX2* may be transcriptionally regulated by TCP10, a hypothesis also supported by the identification of TCP TF binding sites (TFBS) statistically over-represented ( $p$ -value < 0.05) within the *LOX2* 5'-promoter region. In addition, we observed a correlation of the

<sup>5</sup>www.phytozome.net



transcriptomic response of other JA-related genes (*LOX5*, *MC*, *AOS2*) with *TCP10* and *LOX2* levels (**Figure 6**). Based on this correlation, we propose that the effect of miR319d/TCP10 node in common bean root growth and nodulation, reported here, may be mediated by JA signaling (**Figure 7**).

The root growth inhibition was one of the first physiological effects detected for JA (Dathe et al., 1981; Staswick et al., 1992; Wasternack, 2007). Previous reports from *Arabidopsis* relate elevated JA levels with reduced root growth (Ellis et al., 2002; Wasternack, 2007). In this work we showed decreased biomass and area of common bean roots from miR319d-silenced plants with higher expression of *TCP10* and *LOX2* genes (**Figures 4, 6 and Supplementary Figure S2**).

Several studies have shown the participation of JA as a regulatory/signaling molecule in the rhizobia symbiosis with different legume species, including common bean (Poustini et al., 2007; Ferguson and Mathesius, 2014). For example, there are reports showing negative JA effects in early symbiotic stages of *L. japonicus*- and *M. truncatula* – rhizobia symbioses (Nakagawa and Kawaguchi, 2006; Sun et al., 2006) as well as positive JA effects in soybean nodulation (Seo et al., 2006; Kinkema and Gresshoff, 2008). Here, we showed that the *R. tropici*-inoculated OE319d common bean plants, with low level of *TCP10* and *LOX2*, exhibit a significant increase in the amount of root hair deformation and infection thread formation at early stages of the symbiosis but a decreased nodulation. These results indicate an arrested infection, after infection thread formation stage, that prevents nodule development (**Figures 5–7**). We propose that the regulation of common bean nodulation by miR319d/TCP10 could be mediated by JA signaling. However, because of the complex and intricate regulation of the rhizobia symbiosis we cannot rule out the participation of other signaling pathways in the affected nodulation of common bean plants modulated in the expression of miR319d/TCP10. This is the

first report about the miR319/TCP node as regulator of the rhizobial symbiosis, future in-depth studies would indicate the commonalities of such regulatory network in other legume species.

## AUTHOR CONTRIBUTIONS

JM-R and DF conceived and performed the experiments, interpreted the data and contributed to the drafting of the manuscript. AL performed the experiments. GH conceived and supervised the whole project and wrote the manuscript.

## FUNDING

This work was supported in part by Dirección General de Asuntos del Personal Académico (DGAPA) – UNAM (Grants Nos. PAPIIT IN200816 and IA203218). JM-R received a postdoctoral fellowship from DGAPA-UNAM.

## ACKNOWLEDGMENTS

We are grateful to Sara I. Fuentes, Mario Ramírez, Víctor M. Bustos, Marisel Llitas, and M. Beatriz Pérez-Morales for advice and technical assistance.

## SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fpls.2018.01175/full#supplementary-material>

## REFERENCES

- Aparicio-Fabre, R., Guillén, G., Loredó, M., Arellano, J., Valdés-López, O., Ramírez, M., et al. (2013). Common bean (*Phaseolus vulgaris* L.) PvTIFY orchestrates global changes in transcript profile response to jasmonate and phosphorus deficiency. *BMC Plant Biol.* 13:26. doi: 10.1186/1471-2229-13-26
- Blair, M., Izquierdo, P., Astudillo, C., and Grusak, M. (2013). A legume biofortification quandary: variability and genetic control of seed coat micronutrient accumulation in common beans. *Front. Plant Sci.* 4:275. doi: 10.3389/fpls.2013.00275
- Broughton, W. J., and Dilworth, M. J. (1971). Control of leghaemoglobin synthesis in snake beans. *Biochem. J.* 125, 1075–1080. doi: 10.1042/bj1251075
- Broughton, W. J., Hernández, G., Blair, M., Beebe, S., Gepts, P., and Vanderleyden, J. (2003). Beans (*Phaseolus* spp.) – model food legumes. *Plant Soil* 252, 55–128. doi: 10.1023/A:1024146710611
- Chen, C., Ridzon, D. A., Broomer, A. J., Zhou, Z., Lee, D. H., Nguyen, J. T., et al. (2005). Real-time quantification of microRNAs by stem-loop RT-PCR. *Nucleic Acids Res.* 33:e179. doi: 10.1093/nar/gni178
- Chen, L., Wang, T., Zhao, M., and Zhang, W. (2012). Ethylene-responsive miRNAs in roots of *Medicago truncatula* identified by high-throughput sequencing at whole genome level. *Plant Sci.* 184, 14–19. doi: 10.1016/j.plantsci.2011.11.007
- Cubas, P., Nick, L., John, D., and Enrico, C. (1999). The TCP domain: a motif found in proteins regulating plant growth and development. *Plant J.* 18, 215–222. doi: 10.1046/j.1365-313X.1999.00444.x
- Danisman, S., van der Wal, F., Dhondt, S., Waites, R., de Folter, S., Bimbo, A., et al. (2012). *Arabidopsis* Class I and Class II TCP transcription factors regulate jasmonic acid metabolism and leaf development antagonistically. *Plant Physiol.* 159, 1511–1523. doi: 10.1104/pp.112.200303
- Dathe, W., Rönisch, H., Preiss, A., Schade, W., Sembdner, G., and Schreiber, K. (1981). Endogenous plant hormones of the broad bean, *Vicia faba* L. (-)-jasmonic acid, a plant growth inhibitor in pericarp. *Planta* 153, 530–535. doi: 10.1007/bf00385537
- De Luis, A., Markmann, K., Cognat, V., Holt, D. B., Charpentier, M., Parniske, M., et al. (2012). Two microRNAs linked to nodule infection and nitrogen-fixing ability in the legume lotus japonicus. *Plant Physiol.* 160, 2137–2154. doi: 10.1104/pp.112.204883
- Devoto, A., and Turner, J. G. (2003). Regulation of jasmonate-mediated plant responses in *Arabidopsis*. *Ann. Bot.* 92, 329–337. doi: 10.1093/aob/mcg151
- Doyle, J. J., and Luckow, M. A. (2003). The rest of the iceberg. Legume diversity and evolution in a phylogenetic context. *Plant Physiol.* 131, 900–910. doi: 10.1104/pp.102.018150
- Ellis, C., Karafyllidis, I., Wasternack, C., and Turner, J. G. (2002). The *Arabidopsis* mutant cev1 links cell wall signaling to jasmonate and ethylene responses. *Plant Cell* 14, 1557–1566. doi: 10.1105/tpc.002022
- Estrada-Navarrete, G., Alvarado-Affantranger, X., Olivares, J.-E., Guillén, G., Díaz-Camino, C., Campos, F., et al. (2007). Fast, efficient and reproducible genetic transformation of *Phaseolus* spp. by *Agrobacterium rhizogenes*. *Nat. Protoc.* 2, 1819–1824. doi: 10.1038/nprot.2007.259

- Ferguson, B. J., and Mathesius, U. (2014). Phytohormone regulation of legume-rhizobia interactions. *J. Chem. Ecol.* 40, 770–790. doi: 10.1007/s10886-014-0472-7
- Formey, D., Iñiguez, L. P., Peláez, P., Li, Y.-F., Sunkar, R., Sánchez, F., et al. (2015). Genome-wide identification of the *Phaseolus vulgaris* sRNAome using small RNA and degradome sequencing. *BMC Genomics* 16:423. doi: 10.1186/s12864-015-1639-5
- Formey, D., Martín-Rodríguez, J. A., Leija, A., Santana, O., Quinto, C., Cárdenas, L., et al. (2016). Regulation of small RNAs and corresponding targets in nod factor-induced *Phaseolus vulgaris* root hair cells. *Int. J. Mol. Sci.* 17:887. doi: 10.3390/ijms17060887
- Formey, D., Sallet, E., Lelandais-Brière, C., Ben, C. C., Bustos-Sanmamed, P., and Niebel, A. (2014). The small RNA diversity from *Medicago truncatula* roots under biotic interactions evidences the environmental plasticity of the miRNAome. *Genome Biol.* 15:457. doi: 10.1186/s13059-014-0457-4
- Frith, M. C., Fu, Y., Yu, L., Chen, J. F., Hansen, U., and Weng, Z. (2004). Detection of functional DNA motifs via statistical over-representation. *Nucleic Acids Res.* 32, 1372–1381. doi: 10.1093/nar/gkh299
- Goettel, W., Liu, Z., Xia, J., Zhang, W., Zhao, P. X., and An, Y.-Q. (2014). Systems and evolutionary characterization of microRNAs and their underlying regulatory networks in soybean cotyledons. *PLoS One* 9:e86153. doi: 10.1371/journal.pone.0086153
- Hammell, M., Long, D., Zhang, L., Lee, A., Carmack, C. S., Han, M., et al. (2018). mirWIP: microRNA target prediction based on microRNA-containing ribonucleoprotein-enriched transcripts. *Nat. Methods* 5, 813–819. doi: 10.1038/nmeth.1247
- Hao, J., Tu, L., Hu, H., Tan, J., Deng, F., Tang, W., et al. (2012). GbTCP, a cotton TCP transcription factor, confers fibre elongation and root hair development by a complex regulating system. *J. Exp. Bot.* 63, 6267–6281. doi: 10.1093/jxb/ers278
- Hardy, R. W. F., Holsten, R. D., Jackson, E. K., and Burns, R. C. (1968). The acetylene-ethylene assay for N<sub>2</sub> fixation: laboratory and field evaluation. *Plant Physiol.* 43, 1185–1207. doi: 10.1104/pp.43.8.1185
- He, H., He, L., and Gu, M. (2014). Role of microRNAs in aluminum stress in plants. *Plant Cell Rep.* 33, 831–836. doi: 10.1007/s00299-014-1565-z
- Howe, G. A. (2004). Jasmonates as signals in the wound response. *J. Plant Growth Regul.* 23, 223–237. doi: 10.1007/s00344-004-0030-6
- Jones-Rhoades, M. W., and Bartel, D. P. (2004). Computational identification of plant microRNAs and their targets, including a stress-induced miRNA. *Mol. Cell* 14, 787–799. doi: 10.1016/j.molcel.2004.05.027
- Katoh, K., Rozewicki, J., and Yamada, K. D. (2017). MAFFT online service: multiple sequence alignment, interactive sequence choice and visualization. *Brief. Bioinform.* doi: 10.1093/bib/bbx108 [Epub ahead of print].
- Kinkema, M., and Gresshoff, P. M. (2008). Investigation of downstream signals of the soybean autoregulation of nodulation receptor kinase GmNARK. *Mol. Plant Microbe Interact.* 21, 1337–1348. doi: 10.1094/MPMI-21-10-1337
- Koyama, T., Sato, F., and Ohme-Takagi, M. (2017). Roles of miR319 and TCP transcription factors in leaf development. *Plant Physiol.* 175, 874–885. doi: 10.1104/pp.17.00732
- Kramer, M. F. (2011). Stem-Loop RT-qPCR for miRNAs. *Curr. Protoc. Mol. Biol.* 95, 15.10.1–15.10.15. doi: 10.1002/0471142727.mb1510s95
- Lelandais-Brière, C., Moreau, J., Hartmann, C., and Crespi, M. (2016). Noncoding RNAs, emerging regulators in root endosymbioses. *Mol. Plant Microbe Interact.* 29, 170–180. doi: 10.1094/MPMI-10-15-0240-FI
- Lelandais-Brière, C., Naya, L., Sallet, E., Calenge, F., Frugier, F., Hartmann, C., et al. (2009). Genome-wide *Medicago truncatula* small RNA analysis revealed novel microRNAs and isoforms differentially regulated in roots and nodules. *Plant Cell* 21, 2780–2796. doi: 10.1105/tpc.109.068130
- Li, S., Castillo-González, C., Yu, B., and Zhang, X. (2017). The functions of plant small RNAs in development and in stress responses. *Plant J.* 90, 654–670. doi: 10.1111/tpj.13444
- Li, X., Wang, X., Zhang, S., Liu, D., Duan, Y., and Dong, W. (2012). Identification of soybean microRNAs involved in soybean cyst nematode infection by deep sequencing. *PLoS One* 7:e39650. doi: 10.1371/journal.pone.0039650
- López-Ráez, J. A., Verhage, A., Fernández, L., García, J. M., Azcón-Aguilar, C., Flors, V., et al. (2010). Hormonal and transcriptional profiles highlight common and differential host responses to arbuscular mycorrhizal fungi and the regulation of the oxylipin pathway. *J. Exp. Bot.* 61, 2589–2601. doi: 10.1093/jxb/erq089
- Martínez-Medina, A., Van Wees, S. C. M., and Pieterse, C. M. J. (2017). Airborne signals from *Trichoderma* fungi stimulate iron uptake responses in roots resulting in priming of jasmonic acid-dependent defences in shoots of *Arabidopsis thaliana* and *Solanum lycopersicum*. *Plant Cell Environ.* 40, 2691–2705. doi: 10.1111/pce.13016
- Martin-Trillo, M., and Cubas, P. (2010). TCP genes: a family snapshot ten years later. *Trends Plant Sci.* 15, 31–39. doi: 10.1016/j.tplants.2009.11.003
- Nakagawa, T., and Kawaguchi, M. (2006). Shoot-applied MeJA suppresses root nodulation in lotus japonicus. *Plant Cell Physiol.* 47, 176–180. doi: 10.1093/pcp/pci222
- Naya, L., Paul, S., Valdés-López, O., Mendoza-Soto, A. B., Nova-Franco, B., and Sosa-Valencia, G. (2014). Regulation of copper homeostasis and biotic interactions by microRNA 398b in common bean. *PLoS One* 9:e84416. doi: 10.1371/journal.pone.0084416
- Nicolas, M., and Cubas, P. (2016). “Chapter 16 - The role of TCP transcription factors in shaping flower structure, leaf morphology, and plant architecture,” in *Plant Transcription Factors*, ed. D. H. Gonzalez (Boston: Academic Press), 249–267. doi: 10.1016/B978-0-12-800854-6.00016-6
- Nova-Franco, B., Iñiguez, L. P., Valdés-López, O., Alvarado-Affantranger, X., Leija, A., Fuentes, S. I., et al. (2015). The microRNA-172c-APETALA2-1 node as a key regulator of the common bean rhizobia nitrogen fixation symbiosis. *Plant Physiol.* 168, 273–290. doi: 10.1104/pp.114.255547
- O'Rourke, J. A., Iñiguez, L. P., Fu, F., Bucciarelli, B., Miller, S. S., Jackson, S. A., et al. (2014). An RNA-Seq based gene expression atlas of the common bean. *BMC Genomics* 15:866. doi: 10.1186/1471-2164-15-866
- Palatnik, J. F., Allen, E., Wu, X., Schommer, C., Schwab, R., Carrington, J. C., et al. (2003). Control of leaf morphogenesis by microRNAs. *Nature* 425, 257–263. doi: 10.1038/nature01958
- Porta, H., and Rocha-Sosa, M. (2000). A *Phaseolus vulgaris* lipoxygenase gene expressed in nodules and in *Rhizobium tropici* inoculated roots. *Biochim. Biophys. Acta* 1517, 139–142. doi: 10.1016/S0167-4781(00)00230-X
- Poustini, K., Mabood, F., and Smith, D. L. (2007). Preincubation of *Rhizobium leguminosarum* bv. phaseoli with jasmonate and genistein signal molecules increases bean (*Phaseolus vulgaris* L.) nodulation, nitrogen fixation and biomass production. *J. Agric. Sci. Technol.* 9, 107–117.
- Rodríguez, R. E., Mecchia, M. A., Debernardi, J. M., Schommer, C., Weigel, D., and Palatnik, J. F. (2010). Control of cell proliferation in *Arabidopsis thaliana* by microRNA miR396. *Development* 137, 103–112. doi: 10.1242/dev.043067
- Rogers, K., and Chen, X. (2013). Biogenesis, turnover, and mode of action of plant microRNAs. *Plant Cell* 25, 2383–2399. doi: 10.1105/tpc.113.113159
- Rubio-Somoza, I., and Weigel, D. (2011). MicroRNA networks and developmental plasticity in plants. *Trends Plant Sci.* 16, 258–264. doi: 10.1016/j.tplants.2011.03.001
- Sandelin, A., and Wasserman, W. W. (2004). Constrained binding site diversity within families of transcription factors enhances pattern discovery bioinformatics. *J. Mol. Biol.* 338, 207–215. doi: 10.1016/j.jmb.2004.02.048
- Schmutz, J., McClean, P. E., Mamidi, S., Wu, G. A., Cannon, S. B., Grimwood, J., et al. (2014). A reference genome for common bean and genome-wide analysis of dual domestications. *Nat. Genet.* 46, 707–713. doi: 10.1038/ng.3008
- Schommer, C., Bresso, E. G., Spinelli, S. V., and Palatnik, J. F. (2012). “Role of microRNA miR319 in plant development,” in *MicroRNAs Plant Dev Stress Responses*, ed. R. Sunkar (Heidelberg: Springer Berlin Heidelberg).
- Schommer, C., Debernardi, J. M., Bresso, E. G., Rodríguez, R. E., and Palatnik, J. F. (2014). Repression of cell proliferation by miR319-regulated TCP4. *Mol. Plant* 7, 1533–1544. doi: 10.1093/mp/ssu084
- Schommer, C., Palatnik, J. F., Aggarwal, P., Chételat, A., Cubas, P., Farmer, E. E., et al. (2008). Control of jasmonate biosynthesis and senescence by miR319 targets. *PLoS Biol.* 6:e230. doi: 10.1371/journal.pbio.0060230
- Seo, H. S., Li, J., Lee, S.-Y., Yu, J.-W., Kim, K.-H., Lee, S.-H., et al. (2006). The hypernodulating nts mutation induces jasmonate synthetic pathway in Soybean leaves. *Mol. Cells* 24, 185–193.
- Song, Q.-X., Liu, Y.-F., Hu, X.-Y., Zhang, W.-K., Ma, B., Chen, S.-Y., et al. (2011). Identification of miRNAs and their target genes in developing soybean seeds by deep sequencing. *BMC Plant Biol.* 11:5. doi: 10.1186/1471-2229-11-5
- Staswick, P. E., Su, W., and Howell, S. H. (1992). Methyl jasmonate inhibition of root growth and induction of a leaf protein are decreased in an *Arabidopsis thaliana* mutant. *Proc. Natl. Acad. Sci. U.S.A.* 89, 6837–6840. doi: 10.1073/pnas.89.15.6837

- Subramanian, S., Fu, Y., Sunkar, R., Barbazuk, W. B., Zhu, J.-K., and Yu, O. (2008). Novel and nodulation-regulated microRNAs in soybean roots. *BMC Genomics* 9:160. doi: 10.1186/1471-2164-9-160
- Sun, J., Cardoza, V., Mitchell, D. M., Bright, L., Oldroyd, G., and Harris, J. M. (2006). Crosstalk between jasmonic acid, ethylene and nod factor signaling allows integration of diverse inputs for regulation of nodulation. *Plant J.* 46, 961–970. doi: 10.1111/j.1365-
- Turner, M., Yu, O., and Subramanian, S. (2012). Genome organization and characteristics of soybean microRNAs. *BMC Genomics* 13:169. doi: 10.1186/1471-2164-13-169
- Uppalapati, S. R., Ayoubi, P., Weng, H., Palmer, D. A., Mitchell, R. E., Jones, W., et al. (2005). The phytotoxin coronatine and methyl jasmonate impact multiple phytohormone pathways in tomato. *Plant J.* 42, 201–217. doi: 10.1111/j.1365-313X.2005.02366.x
- Valdés-López, O., Arenas-Huertero, C., Ramírez, M., Girard, L., Sánchez, F., and Vance, C. P. (2008). Essential role of MYB transcription factor: PvPHR1 and microRNA: PvmiR399 in phosphorus-deficiency signalling in common bean roots. *Plant Cell Environ.* 31, 1834–1843. doi: 10.1111/j.1365-3040.2008.01883.x
- Valdés-López, O., Yang, S. S., Aparicio-Fabre, R., Graham, P. H., Reyes, J. L., and Vance, C. P. (2010). MicroRNA expression profile in common bean (*Phaseolus vulgaris*) under nutrient deficiency stresses and manganese toxicity. *New Phytol.* 187, 805–818. doi: 10.1111/j.1469-8137.2010.03320.x
- Vandesompele, J., De Preter, K., Pattyn, F., Poppe, B., Van Roy, N., De Paepe, A., et al. (2002). Accurate normalization of real-time quantitative RT-PCR data by geometric averaging of multiple internal control genes. *Genome Biol.* 3:research0034.1. doi: 10.1186/gb-2002-3-7-research0034
- Venkateshwaran, M., Volkening, J. D., Sussman, M. R., and Ané, J.-M. (2013). Symbiosis and the social network of higher plants. *Curr. Opin. Plant Biol.* 16, 118–127. doi: 10.1016/j.pbi.2012.11.007
- Wang, Y., Wang, L., Zou, Y., Chen, L., Cai, Z., and Zhang, S. (2014). Soybean miR172c targets the repressive AP2 transcription factor NNC1 to activate ENOD40 expression and regulate nodule initiation. *Plant Cell* 26, 4782–4801. doi: 10.1105/tpc.114.131607
- Wasternack, C. (2007). Jasmonates: an update on biosynthesis, signal transduction and action in plant stress response, growth and development. *Ann. Bot.* 100, 681–697. doi: 10.1093/aob/mcm079
- Weigel, D., Ahn, J. H., Blázquez, M. A., Borevitz, J. O., Christensen, S. K., Fankhauser, C., et al. (2000). Activation tagging in *Arabidopsis*. *Plant Physiol.* 122, 1003–1014. doi: 10.1104/pp.122.4.1003
- Wong, C. E., Zhao, Y.-T., Wang, X.-J., Croft, L., Wang, Z.-H., Haerizadeh, F., et al. (2011). MicroRNAs in the shoot apical meristem of soybean. *J. Exp. Bot.* 62, 2495–2506. doi: 10.1093/jxb/erq437
- Yan, J., Gu, Y., Jia, X., Kang, W., Pan, S., Tang, X., et al. (2012). Effective small RNA destruction by the expression of a short tandem target mimic in *Arabidopsis*. *Plant Cell* 24, 415–427. doi: 10.1105/tpc.111.094144
- Zhang, C., Ding, Z., Wu, K., Yang, L., Li, Y., Yang, Z., et al. (2016). Suppression of jasmonic acid-mediated defense by viral-inducible microRNA319 facilitates virus infection in rice. *Mol. Plant* 9, 1302–1314. doi: 10.1016/j.molp.2016.06.014
- Zhang, L., Zhang, F., Melotto, M., Yao, J., and He, S. Y. (2017). Jasmonate signaling and manipulation by pathogens and insects. *J. Exp. Bot.* 68, 1371–1385. doi: 10.1093/jxb/erw478

**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 © 2018 Martín-Rodríguez, Leija, Formey and Hernández. 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.



# Respective Contributions of URT1 and HESO1 to the Uridylation of 5' Fragments Produced From RISC-Cleaved mRNAs

Hélène Zuber\*, Hélène Scheer, Anne-Caroline Joly and Dominique Gagliardi\*

Institut de Biologie Moléculaire des Plantes (IBMP), Centre National de la Recherche Scientifique (CNRS), Université de Strasbourg, Strasbourg, France

## OPEN ACCESS

### Edited by:

Ana Confraria,  
Instituto Gulbenkian de Ciência (IGC),  
Portugal

### Reviewed by:

Tony Millar,  
Australian National University,  
Australia  
Laura Arribas-Hernández,  
University of Copenhagen, Denmark

### \*Correspondence:

Hélène Zuber  
helene.zuber@ibmp-cnrs.unistra.fr  
Dominique Gagliardi  
dominique.gagliardi@ibmp-  
cnrs.unistra.fr

### Specialty section:

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

**Received:** 31 May 2018

**Accepted:** 10 September 2018

**Published:** 09 October 2018

### Citation:

Zuber H, Scheer H, Joly A-C and  
Gagliardi D (2018) Respective  
Contributions of URT1 and HESO1  
to the Uridylation of 5' Fragments  
Produced From RISC-Cleaved  
mRNAs. *Front. Plant Sci.* 9:1438.  
doi: 10.3389/fpls.2018.01438

In plants, post-transcriptional gene silencing (PTGS) represses gene expression by translation inhibition and cleavage of target mRNAs. The slicing activity is provided by argonaute 1 (AGO1), and the cleavage site is determined by sequence complementarity between the target mRNA and the microRNA (miRNA) or short interfering RNA (siRNA) loaded onto AGO1, to form the core of the RNA induced silencing complex (RISC). Following cleavage, the resulting 5' fragment is modified at its 3' end by the untemplated addition of uridines. Uridylation is proposed to facilitate RISC recycling and the degradation of the RISC 5'-cleavage fragment. Here, we detail a 3' RACE-seq method to analyze the 3' ends of 5' fragments produced from RISC-cleaved transcripts. The protocol is based on the ligation of a primer at the 3' end of RNA, followed by cDNA synthesis and the subsequent targeted amplification by PCR to generate amplicon libraries suitable for Illumina sequencing. A detailed data processing pipeline is provided to analyze nibbling and tailing at high resolution. Using this method, we compared the tailing and nibbling patterns of RISC-cleaved *MYB33* and *SPL13* transcripts between wild-type plants and mutant plants depleted for the terminal uridylyltransferases (TUTases) HESO1 and URT1. Our data reveal the respective contributions of HESO and URT1 in the uridylation of RISC-cleaved *MYB33* and *SPL13* transcripts, with HESO1 being the major TUTase involved in uridylating these fragments. Because of its depth, the 3' RACE-seq method shows at high resolution that these RISC-generated 5' RNA fragments are nibbled by a few nucleotides close to the cleavage site in the absence of uridylation. 3' RACE-seq is a suitable approach for a reliable comparison of uridylation and nibbling patterns between mutants, a prerequisite to the identification of all factors involved in the clearance of RISC-generated 5' mRNA fragments.

**Keywords:** uridylation, TUTase, RISC, RNA silencing, Arabidopsis, RNA degradation, miRNA, Illumina

## INTRODUCTION

Small RNAs are key regulators of gene expression (Borges and Martienssen, 2015; Bartel, 2018). They are classified as two main types, microRNAs (miRNAs) and short interfering RNAs (siRNAs), because of key distinctions in their respective mode of biogenesis (Martínez de Alba et al., 2013; Borges and Martienssen, 2015). miRNAs are processed from primary transcripts that fold as a



hairpin with an imperfectly paired stem. By contrast, siRNAs are generated from near-perfect double stranded RNAs (dsRNAs) or fully paired dsRNAs when the complementary strand is synthesized by a RNA-dependent RNA polymerase (RDR), which uses the sense strand as template. miRNAs and siRNAs are loaded onto members of the argonaute (AGO) protein family to form the core of RNA induced silencing complexes (RISCs) (Vaucheret, 2008; Zhang et al., 2015). RISCs are then guided to their targets by sequence complementarity with the loaded small RNA. In plants, the base pairing of miRNAs with their targets is rather extensive, and mRNAs regulated by RISCs are repressed by AGO1-mediated cleavage, but also by translation repression (Chen, 2004; Brodersen et al., 2008; Yang et al., 2012; Li et al., 2013; Iwakawa and Tomari, 2015; Reis et al., 2015; Arribas-Hernández et al., 2016). Cleavage of mRNAs by RISC produces a 5' fragment and a 3' fragment. As detailed below, both the 5'-3' and 3'-5' RNA degradation pathways contribute to the elimination of these fragments.

In Arabidopsis, the cytosolic 5'-3' exoribonuclease XRN4 participates in the degradation of RISC 3'-cleavage fragments, as indicated by their accumulation in *xrn4* mutants (Souret et al., 2004). XRN4 was also proposed to be involved in the degradation of RISC 5'-cleavage fragments because the 5' fragment resulting from the cleavage of *MYB domain protein 33* (*MYB33*) mRNA by miR159-loaded RISC accumulates in a *xrn4* mutant (Ren et al., 2014). RISC 5'-cleavage fragments are also definitely degraded by the 3'-5' RNA degradation pathway because they accumulate in *ski2*, *ski3*, and *ski8* mutants (Branscheid et al., 2015). Together, SKI2-3-8 form the Ski complex, which is the major activator of the RNA exosome in the cytosol. Therefore, the involvement of the RNA exosome in the degradation of RISC 5'-cleavage fragments is likely in Arabidopsis, although it remains to be demonstrated using appropriate mutants. This implication of the RNA exosome would be consistent with previous findings in other organisms, such as *Drosophila melanogaster* (Orban and Izaurre, 2005). In addition, two ribonucleases were recently described in Arabidopsis as taking part in the metabolism of RISC 5'-cleavage fragments: RISC-interacting clearing 3'-5' exoribonucleases 1 and 2 (RICE-1 and -2) (Zhang et al., 2017). RICEs are homohexamers with a DnaQ-like exonuclease fold, and they interact with AGO1 and AGO10 (Zhang et al., 2017). RICEs are proposed to initiate the destabilization of RISC 5'-cleavage fragments thereby facilitating RISC dissociation. This would grant access of the 3' extremity of RISC 5'-cleavage fragments to the RNA exosome and importantly, recycle RISC, which is essential to maintain functional RISC and miRNA abundance (Zhang et al., 2017). The access of the 3' extremity of RISC 5'-cleavage fragments to the RNA exosome may also be facilitated by components of the non-stop decay (NSD) pathway when the RISC 5'-cleavage fragment is engaged in polysomes (Szádeczky-Kardoss et al., 2018). The prime function of NSD is to eliminate mRNAs lacking stop codons. Recently, orthologs of Pelota and Hbs1, two core components of NSD, were shown to participate in the elimination of RISC 5'-cleavage fragments in *Nicotiana benthamiana* and *A. thaliana*, provided that the cleavage site is within the coding region (Szádeczky-Kardoss et al., 2018). Likely, the NSD machinery promotes the

dissociation of ribosomes stalled at the extremity of a RISC 5'-cleavage fragment to promote access to the RNA exosome (Szádeczky-Kardoss et al., 2018).

Besides exoribonucleases and RNA helicases, terminal uridylyltransferases (TUTases) constitute another type of enzymatic activities involved in the clearance of RISC 5'-cleavage fragments. Indeed a striking molecular event in this process is the untemplated addition of uridines at the 3' extremity of RISC 5'-cleavage fragments (Shen and Goodman, 2004; Ren et al., 2014; Zhang et al., 2017). The uridylation of several of such fragments was originally reported in both Arabidopsis and mice (Shen and Goodman, 2004). Since then, uridylation has emerged as a conserved post-transcriptional process that shapes the coding and non-coding transcriptomes in eukaryotes (Munoz-Tello et al., 2015; Scheer et al., 2016; De Almeida et al., 2018). In Arabidopsis, two TUTases have been characterized: HEN1 SUPPRESSOR 1 (HESO1) and URIDYLTRANSFERASE 1 (URT1) (Kwak and Wickens, 2007; Ren et al., 2012; Zhao et al., 2012; Sement et al., 2013). Both HESO1 and URT1 contain the core catalytic domain (CCD) that defines proteins belonging to the terminal nucleotidyltransferase family. In addition, URT1 contains a large intrinsically disordered region (IDR) in its N-terminal region, while a shorter IDR is present in the C-terminal region of HESO1 (De Almeida et al., 2018). Those IDRs may mediate the recognition of protein partners by URT1 and HESO1, or be a key to their localization in P-bodies and/or stress granules (Sement et al., 2013; Ren et al., 2014; Wang et al., 2015). HESO1 was identified as the main TUTase uridylating miRNAs and siRNAs to trigger their degradation (Ren et al., 2012; Zhao et al., 2012). In addition, HESO1 was shown to uridylate three RISC 5'-cleavage fragments (Ren et al., 2014). Those fragments are generated from *MYB33*, *Auxin Response Factor 10* (*ARF10*), and *Lost Meristems 1* (*LOM1*) mRNAs, which are targets of miR159, miR160, and miR171, respectively. A residual uridylation of these RISC 5'-cleavage fragments is observed in *heso1* mutants (Ren et al., 2014) and, this secondary activity may be due to URT1, although experimental evidence supporting this hypothesis is lacking to date. URT1 is the main TUTase uridylating mRNAs in Arabidopsis, because mRNA uridylation is decreased by 70–80% in *urt1-1* mutants (Sement et al., 2013; Zuber et al., 2016). URT1 can also uridylate miRNAs, mostly when HESO1, the primary TUTase involved in small RNA uridylation, and HUA ENHANCER 1 (HEN1), a methyltransferase that methylate small RNA duplexes, are absent (Yu et al., 2005; Yang et al., 2006; Ren et al., 2012; Zhao et al., 2012; Tu et al., 2015; Wang et al., 2015). miRNAs are therefore the first documented example of shared RNA substrates between HESO1 and URT1 (Tu et al., 2015; Wang et al., 2015). Yet, both overlapping and distinctive roles in miRNA uridylation were attributed to each TUTase (Tu et al., 2015; Wang et al., 2015). mRNAs and RISC-cleaved transcripts could constitute other cases of shared RNA substrates between HESO1 and URT1. These possibilities remain to be experimentally addressed.

To date, the characterization of uridylated RISC 5'-cleavage fragments has relied on the use of 3' RACE PCR followed by cloning and subsequent analysis based on Sanger sequencing. Although this experimental strategy has proven useful, it has

some inherent limitations. The first one is that this method is low-throughput. It is time-consuming and the depth is usually quite limited, with often less than 20–30 clones analyzed per genotype. The second major drawback is the lack of discrimination between amplicons and independent molecules. This turns out to be a real issue when analyzing low complexity samples by PCR amplification, with the majority (up to 90% as determined here during the analysis of RISC 5'-cleavage fragments) of the final PCR products that correspond to very few independent templates. Taken together, these limitations hinder the qualitative and quantitative analysis of the uridylation of RISC-cleaved transcripts. Such an analysis is crucial to reliably compare uridylation between wild-type (WT) and mutant genetic backgrounds, and this comparison is required to identify all factors involved in the metabolism of 5' RNA fragments produced by RISC cleavage.

Here we detail a 3' RACE-seq method that has been optimized for analyzing the uridylation of 5' fragments from RISC-cleaved transcripts. Those molecules are usually low abundant within the complex mixture of all cellular RNAs, and they exhibit a rather poor diversity, with a few untemplated nucleotides usually added at a precise RISC-mediated cleavage site. We illustrate the use of 3' RACE-seq to analyze the tailing and trimming patterns of *MYB33* and *SPL13* RISC 5'-cleavage fragments by comparing WT plants and mutants lacking *HESO1* and *URT1*. This analysis revealed the respective contributions of both TUTases, and that the absence of uridylation results in the accumulation of 5'-cleavage fragments nibbled by a few nucleotides close to the site cleaved by RISC.

## MATERIALS AND METHODS

### Gene IDs and Primers

The Arabidopsis Genome Initiative (AGI) locus identifiers for the genes studied in this study are: AT2G39740 (*HESO1*), AT2G45620 (*URT1*), AT5G06100 (*MYB33*), and AT5G50570 (*SPL13A*). Please note that AT5G50570 (*SPL13A*) and AT5G50670 (*SPL13B*) have identical coding sequences and therefore cannot be discriminated in this study. For simplicity, the name *SPL13* is used thereafter. The sequence of all primers used in this study is shown in **Supplementary Table S1**.

### Plant Material

The plant material used for analyzing RISC 5'-cleavage fragments by 3' RACE-seq corresponds to Arabidopsis plantlets of Col-0 accession grown for 24 days *in vitro* on Murashige and Skoog media with 0.8% agar and 12 h light (22°C)/12 h darkness cycles (18°C). For other analyses, flowers were harvested from Arabidopsis of Col-0 accession and grown on soil with 16 h light/8 h darkness cycles. *urt1-1* (Salk\_087647C) and *heso1-1* (GK-367H02-017041) T-DNA mutant lines have been previously described (Zhao et al., 2012; Sement et al., 2013). Double mutants were obtained by down regulating *URT1* by co-suppression in *heso1-1*. For this purpose, *heso1-1* plants were transformed with a construct expressing an inactive version of *URT1* fused to YFP, which was fortuitously found to efficiently

trigger co-suppression of the endogenous *URT1* gene. The sequence encoding the inactive version of *URT1* (*URT1*<sup>D491/3A</sup>) (Sement et al., 2013) was cloned in the pEarleyGate 104 Gateway plasmid under the control of the cauliflower mosaic virus (CaMV) 35S promoter. Analyses were performed on two biological replicates of three independent *heso1-1 urt1*<sup>SIL</sup> lines. As a control, we also analyzed two biological replicates of a *urt1*<sup>SIL</sup> line obtained by co-suppressing *URT1* with a *YFP-URT1* sequence cloned in the pEarleyGate 104 Gateway plasmid.

### Protein Extraction and Western Blot Analysis

Proteins were extracted from flowers of WT, *urt1-1*, *urt1*<sup>SIL</sup> and three independent *heso1-1 urt1*<sup>SIL</sup> lines under denaturing conditions. Proteins were resolved on a 8% SDS-PAGE gel and transferred to an Immobilon-P membrane. Immunoblots were incubated with anti-URT1 antibodies raised in rabbits against the full-length recombinant *URT1*. Following incubation with horseradish peroxidase-coupled secondary antibodies and Lumi-Light Western Blotting Substrate (Roche), signals were recorded using the Fusion-FX system (Vilber Lourmat).

### RNA Extraction and Northern Blot Analysis

Total RNA was extracted from 24-day-old plantlets and flowers for 3' RACE-seq and northern blot analyses, respectively, with TRI Reagent® (Molecular Research Center) according to manufacturer's instructions. RNA was further purified by acid phenol: chloroform: isoamyl alcohol extraction followed by ethanol precipitation. For northern blot analysis of *MYB33* RISC 5'-cleavage fragments, 30 µg total RNAs from WT, *urt1-1* and *heso1-1* mutants were separated on a denaturing formaldehyde 1.5% (w/v) agarose gel and transferred onto a nylon membrane (Hybond<sup>TM</sup>-N+, GE Healthcare Life Sciences<sup>TM</sup>). Following UV-cross-link at 120 mJ/cm<sup>2</sup> for two times 30 s and incubation for 30 min in PerfectHyb Plus Hybridization buffer (Sigma), the membrane was hybridized overnight at 65°C with a probe that detects the 5' fragment of *MYB33* RISC-cleaved transcripts. The probe was prepared by PCR amplifying a 219 bp sequence upstream of the RISC cleavage site (**Supplementary Table S1**) and by random primed labeling the PCR product using [ $\alpha$ -<sup>32</sup>P]-dCTP and DecaLabel DNA labeling kit (Thermo Scientific). For northern blot analysis of miR159, 10 µg total RNA from WT, *urt1-1*, and *heso1-1* mutants were separated on 17.5% polyacrylamide/7 M urea gels and transferred onto nylon membranes (Hybond<sup>TM</sup>-NX, GE Healthcare Life Sciences<sup>TM</sup>). Following UV-cross-link at 120 mJ/cm<sup>2</sup> for two times 30 s and incubation for 30 min in PerfectHyb Plus Hybridization buffer (Sigma), membranes were hybridized overnight at 50°C with a 5' [<sup>32</sup>P]-labeled oligonucleotide to detect miR159 (**Supplementary Table S1**). The probe was labeled using [ $\gamma$ -<sup>32</sup>P] ATP and T4 PNK (NEB) according to manufacturer's instruction. Radioactive signals were detected by autoradiography and quantified using a Typhoon scanner (GE Healthcare Life Sciences) and Image Gauge software. Plant material used for

biological replicates 1 and 2 were common for both northern analyses.

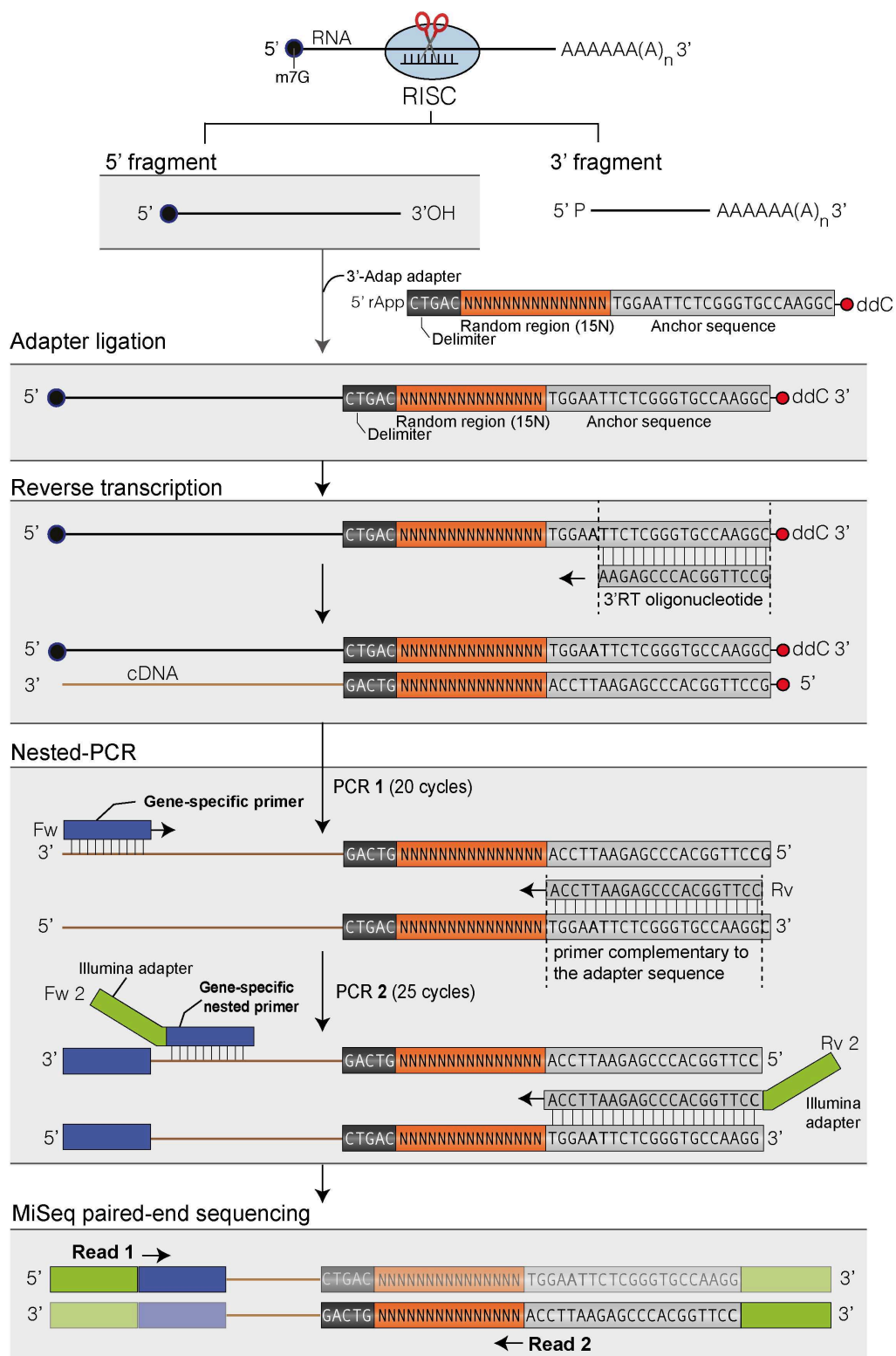
### 3' RACE-Seq Protocol

A 3' RACE-seq protocol was adapted for analyzing RISC 5'-cleavage fragments. Total RNA was extracted from 24-day-old seedlings using TRI Reagent® as described above. Twenty pmoles of a 5'-riboadenylated DNA oligonucleotide (3'-Adap, **Figure 1** and **Supplementary Table S1**) were ligated to 10 µg of total RNA using 20 U of T4 ssRNA Ligase 1 (NEB) in a final volume of 100 µl for 1 h at 37°C and 1X T4 of RNA Ligase Reaction Buffer (NEB, 50 mM Tris-HCl, 10 mM MgCl<sub>2</sub>, 1 mM DTT, pH 7.5). The ligation products were purified from reagents and non-ligated adapter molecules with Nucleospin® RNA Cleanup columns (Macherey Nagel). RNA was then precipitated with ethanol, solubilized in water and quantified. cDNA synthesis was performed in two 20 µl-reactions for each sample. Each 20 µl-reaction contained 2 µg of purified ligated RNA, 50 pmol of the 3'-RT oligonucleotide (**Supplementary Table S1**), 10 nmol of dNTP, 0.1 µmol of DTT, 40 U of RNaseOUT (Invitrogen™), 200 U of SuperScript IV reverse transcriptase (Invitrogen™) and 1X of SuperScript IV RT buffer (Invitrogen™). Reactions were incubated at 50°C for 10 min, and then at 80°C for 10 min to inactivate the reverse transcriptase. The two 20 µl-reactions for each sample were pooled, the cDNAs were extracted with phenol-chloroform, precipitated with ethanol and dissolved in 8 µl Milli-Q water. Two nested PCR amplification rounds of 20 and 25 cycles, respectively, were then performed. PCR1 was run using cDNA synthesized from 1 µg of total RNA, i.e., 2 µl of concentrated cDNA, 10 pmol of MYB33 or SPL13 gene-specific sense primer 1, 10 pmol of RACEseq\_rev1 primer (**Supplementary Table S1**), 10 nmol of dNTP, 1 U of GoTaq® DNA Polymerase (Promega) and 1X of Green GoTaq® Reaction Buffer (Promega) in a 20 µl final volume. The conditions for PCR1 were as follows: a step at 94°C for 30 s; 20 cycles at 94°C for 20 s, 50°C for 20 s and 72°C for 30 s; a final step at 72°C for 30 s. PCR2 was performed using 1 µl of PCR1 product, 10 pmol of MYB33 or SPL13 gene-specific sense primer 2, 10 pmol of a TruSeq RNA PCR index (RPI, **Supplementary Table S1**) 10 nmol of dNTP, 1 U of GoTaq® DNA Polymerase (Promega) and 1X of Green GoTaq® Reaction Buffer (Promega) in a 20 µl final volume. The conditions for PCR2 were as follows: a step at 94°C for 1 min; 25 cycles at 94°C for 30 s, 56°C for 20 s and 72°C for 30 s; a final step at 72°C for 30 s. For each sample, three to four 20 µl-reactions were run and pooled. All PCR2 products were purified using 1 volume of AMPure XP beads (Agencourt). Library concentrations were determined using a Qubit fluorometer (Invitrogen™). Libraries were analyzed on a 2100 Bioanalyzer system (Agilent) to assess quality and estimate size distribution. Library were paired-end sequenced with MiSeq (v3 chemistry) with 41 × 111 bp cycle settings. The respective numbers of sequencing cycles for read 1 and read 2 can be adjusted according to other samples that are co-analyzed. Read 1 is used to identify target transcript whereas read 2 is used to map 3' extremities and analyze 3' potential untemplated nucleotides. To compensate for the poor diversity of the amplicon libraries, 25–33% of phiX control library

(Illumina) were included. Two rounds of RACE-seq experiments were performed. For the first round, four independent biological replicates were analyzed for WT and *heso1-1* genotypes. For the second round, two independent biological replicates were analyzed for WT, *urt1-1*, *heso1-1*, *urt1*<sup>SIL</sup> line, and each of the three *heso1-1 urt1*<sup>SIL</sup> lines (i.e., six *heso1-1 urt1*<sup>SIL</sup> samples). Plant material used for biological replicates 1 and 2 was common for both rounds.

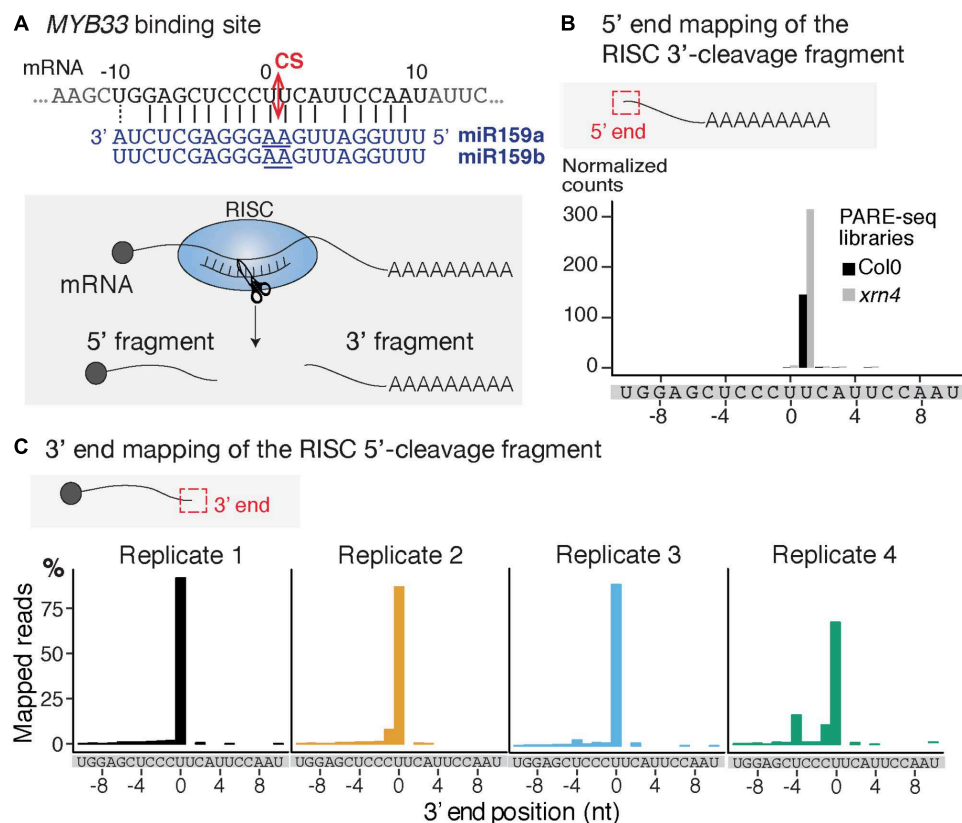
### Bioinformatic Analysis of 3' RACE-Seq Data

Sequencing run quality fit Illumina specification with more than 90% bases higher than Q30. After initial data processing by the MiSeq Control Software v 2.5 (Illumina), base calls were retrieved and further analyzed by a suite of homemade scripts (**Supplementary Data Sheets S1, S2**) using python (v2.7), biopython (v1.63), and regex (v2.4) libraries. Data processing pipeline was adapted from Sikorska et al. (2017). Reads with low quality bases (= < Q10) within the 15-base random sequence of the read 2 or within the 20 bases downstream the delimiter sequence, were filtered out. Sequences with identical nucleotides in 15-base random sequence were deduplicated. Next, the sequences AAGAATTCTCGTCGCCTGAA and GCCAGAGCTATGTTGTTGGT were searched into reads 1 to identify MYB33 and SPL13 corresponding reads, respectively. One mismatch was tolerated. Matched reads 1 and their corresponding reads 2 were extracted and annotated. Reads 2 that contain the delimiter sequence were selected and subsequently trimmed from their random and delimiter sequences. In order to map 3' extremities of target 5' RISC generated fragments, the 20 nucleotide sequences downstream the read 2 delimiter sequence were mapped to the corresponding reference sequence, which goes from the first nucleotide of the transcript that maps the forward PCR primer MYB33\_RISC\_fw2 or SPL13\_RISC\_fw2 to the last nucleotide of the miRNA binding site. To map the 3' end position of reads 2 with untemplated tails, the sequences of the unmatched reads 2 were successively trimmed from their 3' end, with a 1 nt trimming step, until successfully mapped to the reference sequence or until a maximum of 30 nt has been removed. For each successfully mapped read 2, untemplated nucleotides at the 3' end were extracted and analyzed for their size and composition. 3' modifications longer than 1 were considered only if composed of at least of 50 % of the same base (i.e., 50% U, 50% A, 50% C, or 50% G). As explained in the Results section and as illustrated in **Figures 2, 7** for MYB33 and SPL13, respectively, the sites cleaved by RISC were defined by using PARE-seq datasets to map the 5' most nucleotide of the 3' fragment. A single cleavage site was determined for MYB33, i.e., between nucleotides at position 0 and position +1, in contrast to two cleavage sites for SPL13, i.e., a major site between 0 and +1 and a minor site between +1 and +2. Because position +1 in SPL13 is a uridine, we could not determine whether this U is encoded or added post-transcriptionally and SPL13 graphs were generated by considering only U-tails > 2. Finally, a supplemental deduplication was performed to increase stringency: sequences with 13 or more identical nucleotides in



**FIGURE 1 |** Flowchart of the main steps to map the 3' ends of RISC 5'-cleavage fragments by 3' RACE-seq. Features of the 3' adapter and the principle of the main steps are indicated. The experimental workflow begins by ligating the 3' adapter to the RISC 5'-cleavage fragment. Please note that any RNA molecule with a 3' hydroxyl end in the total RNA sample is ligated to the 3' adapter. The target of interest is specifically amplified during PCR-1 and -2 due to the gene-specific sequences of the forward primers. The protocol is detailed in Methods and the scripts used to analyze data are given in **Supplementary Data Sheets S1, S2**.





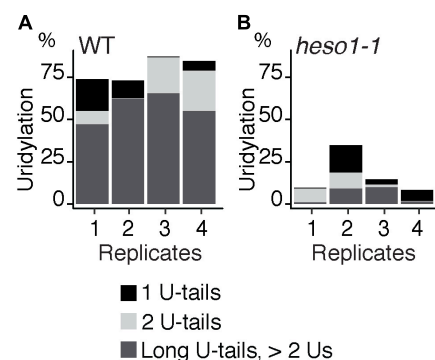
**FIGURE 2 |** Mapping of the 3' extremity of *MYB33* RISC 5'-cleavage fragments. **(A)** Schematic representation of the cleavage of *MYB33* mRNA by AGO1 loaded with miR159. The cleavage site (CS) is predicted between the positions 0 and +1, and is defined by the 10th and 11th nucleotides (underlined) of miR159. **(B)** 5' end mapping using PARE-seq data (German et al., 2008) of RISC 3'-cleavage fragments in WT and *xrn4*. **(C)** 3' end mapping by 3' RACE-seq of RISC 5'-cleavage fragments in WT within the -10/+10 window corresponding to the *MYB33* sequence recognized by miR159. For 3' RACE-seq, most reads map to a single position in each of the four WT biological replicates.

the 15-base random sequence were deduplicated. Plotting and quantitative data analysis was performed with R software (v3.3.1) and ggplot2 R (v2.2.1). Percentages of uridylated fragments were plotted for reads with 3' extremities that map at the cleavage site, with U-tails being defined as tail composed of more than 50% U. Data obtained from the two rounds of RACE-seq experiments, referred to as dataset #1 and dataset #2, have been deposited to the NCBI Gene Expression Omnibus (GEO) database with the accession code GSE115470.

## RESULTS AND DISCUSSION

### Workflow for Mapping 3' Ends of RISC 5'-Cleavage Fragments by 3' RACE-Seq

The principle of 3' RACE-seq to analyze the 3' ends of RISC 5'-cleavage fragments is shown in Figure 1. Briefly, a 5' pre-riboadenylated oligodeoxynucleotide adapter is ligated to the 3' hydroxyl end of RNA molecules using T4 RNA ligase 1 and total RNA. The sequence of the 3' adapter is identical to the one previously described for the TAIL-seq procedure (Chang et al., 2014). However, unlike for TAIL-seq, the 3' adapter does not



**FIGURE 3 |** HESO1 is the main TUTase uridylating *MYB33* RISC 5'-cleavage fragments. Percentages of uridylated *MYB33* RISC 5'-cleavage fragments in four biological replicates for **(A)** WT and **(B)** *heso1-1*. Percentages of long (>2 Us), 2 U- and 1 U-tails are indicated by dark gray, light gray, and black, respectively.

need to be biotinylated. The sequence features of the adapter are detailed in Figure 1. Five nucleotides at the 5' end of the adapter form a delimiter sequence. All reads that do not contain

this delimiter sequence are discarded during the analysis. This ensures that we accurately map the 3' extremity of a transcript that has been ligated to the 3' adapter. Untemplated nucleotides are defined as any nucleotides present between the genome-encoded sequence and the delimiter. The delimiter is followed by a random sequence of 15 bases. This random sequence is essential to remove PCR duplicates during the bioinformatic analysis. This deduplication step is crucial when using 3' RACE-seq to analyze low abundant RNA species with limited complexity, which is typically the case for RISC 5'-cleavage fragments. To further prevent amplicon biases due to the misincorporation of nucleotides in the random sequence during PCR amplification or due to sequencing errors of the random sequence, we enhance the stringency of the deduplication step by not tolerating up to two mismatches within the 15-base random sequence of deduplicated sequences.

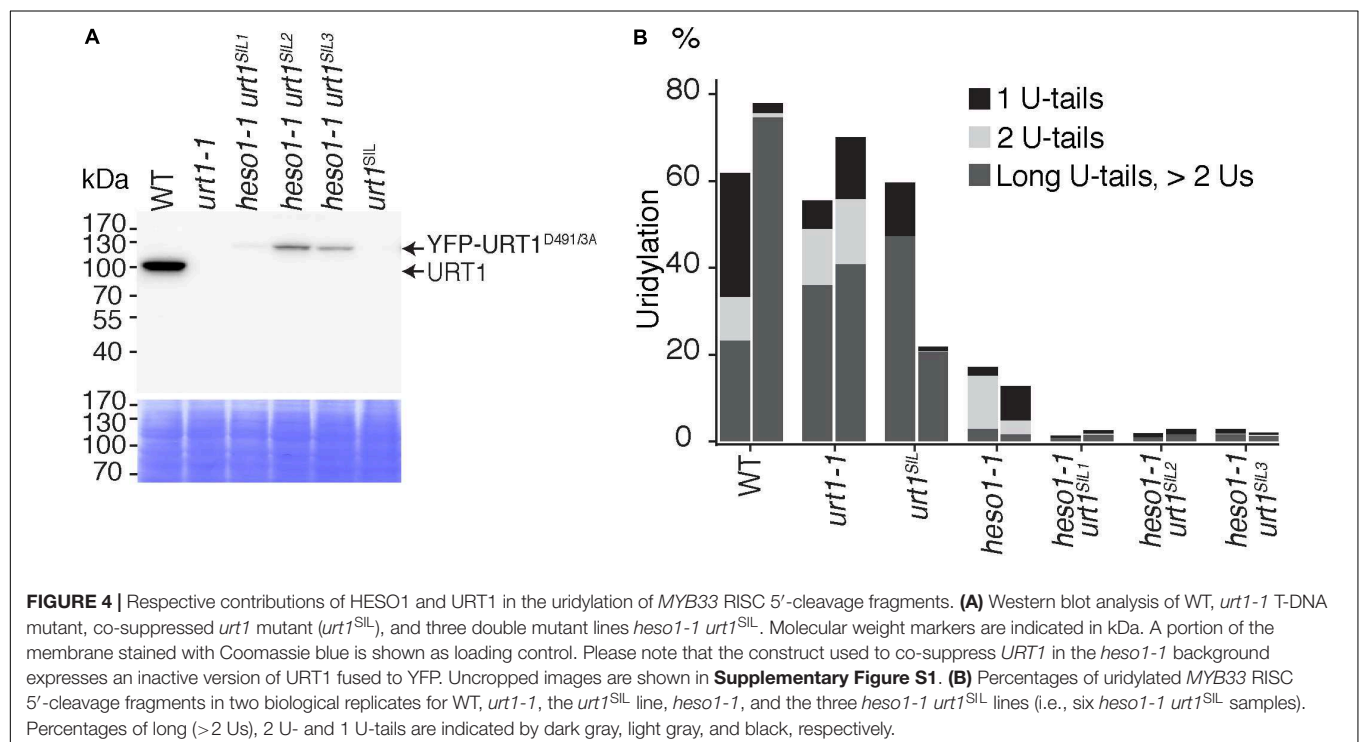
The 3' adapter then contains 22 additional bases, which provide an anchor sequence for cDNA synthesis and subsequent PCR amplification. Importantly, the primer used for cDNA synthesis is complementary to the sequence of the 3' adapter but stops five bases downstream of the random sequence (Figure 1). By using a reverse primer for PCR amplification that extends up to the random sequence, we eliminate the vast majority of cDNAs that are due to priming artifacts and specifically analyze transcripts that have the 3' adapter ligated at their 3' ends. This trick greatly enhanced the quality and depth of our libraries. Finally, the 3' adapter is terminated by a dideoxy-C to prevent self-ligation (Figure 1).

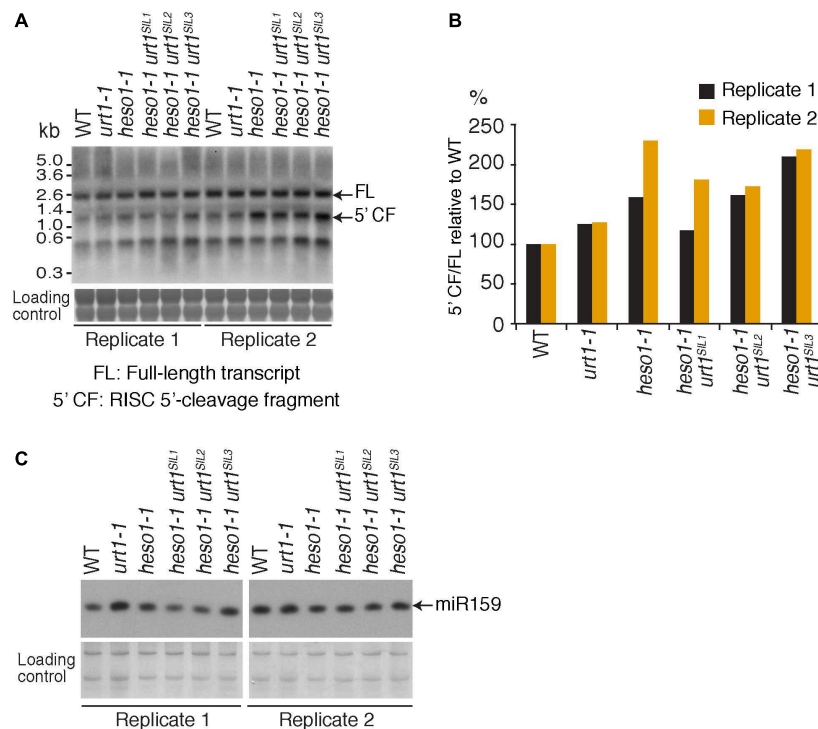
cDNAs are then subjected to two successive rounds of PCR amplification. For the first round, the forward primer is a gene-specific primer matching the sequence of a selected mRNA

and located ideally about 200–400 nucleotides upstream of the predicted RISC-mediated cleavage site. The reverse primer matches the sequence of the 3' adapter up to the random sequence (Figure 1). As mentioned above, this prevents the amplification of most cDNA priming artifacts. The second round of PCR is performed with a nested forward primer and a bar-coded reverse primer complementary to the anchor sequence (Figure 1). Typically, thirty different barcodes can be used to simultaneously analyze different genotypes or replicates. Both forward and reverse primers contain 5' extensions corresponding to the Illumina sequences that are used for flow cell hybridization and sequencing. The number of PCR cycles must be kept as low as possible for both PCRs and ideally should not exceed 20–25 per PCR. Amplicon libraries are purified using AMPure XP beads, quantified with an Invitrogen Qubit fluorometer and their size distribution is determined with a 2100 Bioanalyzer system (Agilent). Amplicon libraries are then sequenced using MiSeq paired-end sequencing for an average yield per run of 38 million of reads: 19 millions of read 1 and 19 millions of read 2.

### Mapping of the RISC-Cleavage Site in *MYB33* mRNAs by 3' RACE-Seq

We selected the *MYB33* mRNAs targeted by miR159 as a model substrate to set up the mapping of the 3' ends of RISC 5'-cleavage fragments by 3' RACE-seq. *MYB33* has been chosen in several studies to investigate uridylation of RISC 5'-cleavage fragments for two main reasons (Shen and Goodman, 2004; Ren et al., 2014; Zhang et al., 2017). First, *MYB33* RISC 5'-cleavage fragments are detectable by northern blot analysis, and therefore their accumulation can be compared between WT plants and





**FIGURE 5 |** MYB33 5'-cleavage fragments accumulate in *heso1-1* (A) Northern blot analysis of WT, *urt1-1*, and *heso1-1* mutant lines using a probe 5' of the cleavage site to detect MYB33 full-length mRNA and the RISC 5'-cleavage fragment. Molecular weight markers are indicated in kb. A portion of the membrane stained with methylene blue is shown as loading control. Uncropped images are shown in **Supplementary Figure S2**. (B) MYB33 RISC 5'-cleavage fragments accumulate mostly in the absence of HESO1. The accumulation of MYB33 RISC 5'-cleavage fragments (5' CF) relative to its full-length (FL) mRNA was determined by integrating the signals in **B** with a Phosphorimager. 5' CF/FL values for each lane were normalized to the 5' CF/FL ratio obtained for the WT control of each replicate. Two replicates are shown in black and orange, respectively, for WT, *urt1-1*, the *urt1<sup>SIL</sup>* line, *heso1-1*, and the three *heso1-1 urt1<sup>SIL</sup>* lines (i.e., six *heso1-1 urt1<sup>SIL</sup>* samples). (C) Northern blot analysis of WT, *urt1-1*, and *heso1-1* mutant lines using a probe to detect *miR159*. A portion of the membrane stained with methylene blue is shown as loading control. Uncropped images are shown in **Supplementary Figure S2**.

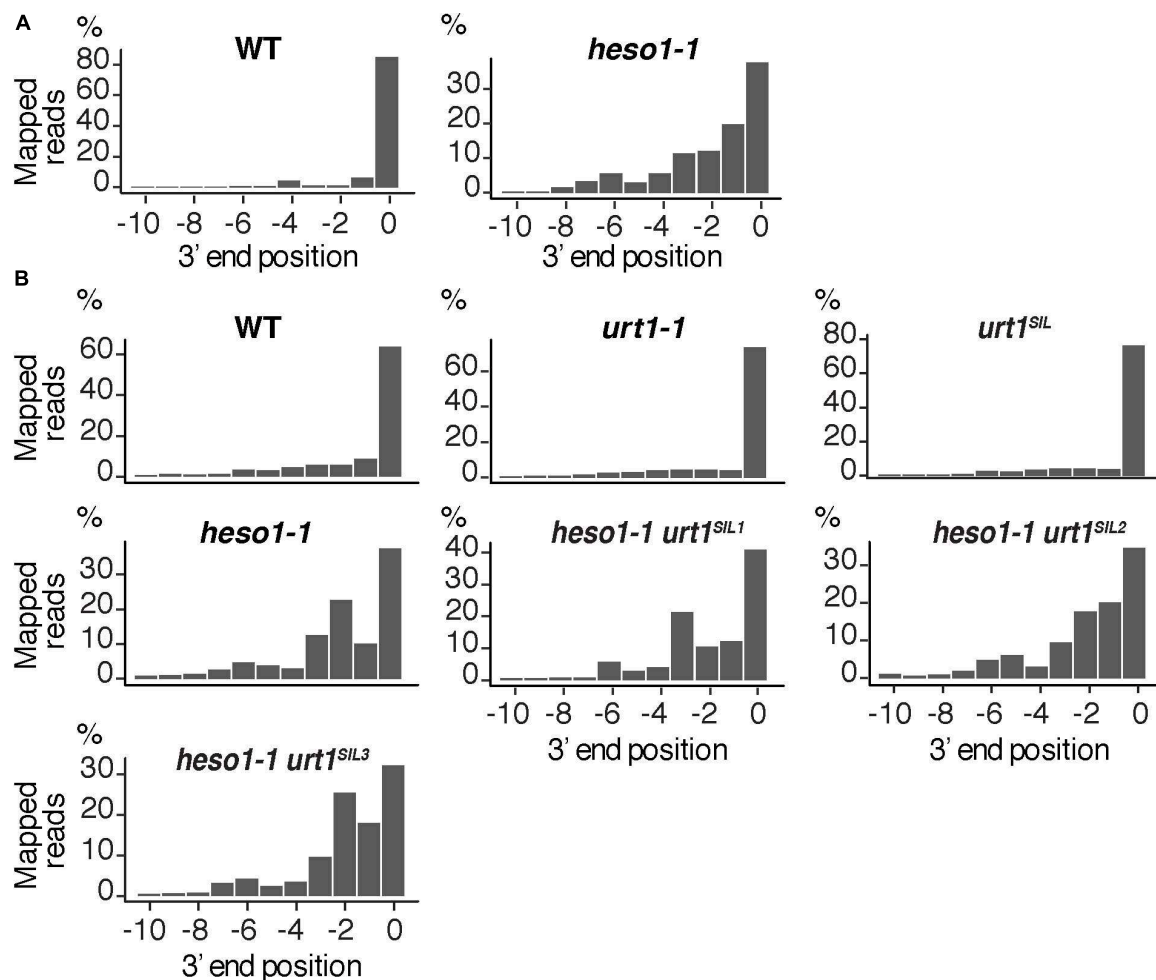
relevant mutants. Second, a high proportion of MYB33 RISC 5'-cleavage fragments is uridylated in WT plants. This proportion was in fact high enough to allow detection of uridylated MYB33 RISC 5'-cleavage fragments by sequencing of a limited number of clones (Shen and Goodman, 2004; Ren et al., 2014; Zhang et al., 2017). The high level of uridylation in WT plants is useful to monitor decrease of uridylation in mutants to identify factors that are involved in the metabolism of this RISC 5'-cleavage fragment. However, there is one drawback in choosing MYB33 to study uridylation: the predicted cleavage site, which is specified by the tenth and eleventh nucleotides of miR159, is situated between two uridines (Figure 2A). This can lead to uncertainties as to whether some 3' terminal uridines are genome-encoded or added post-transcriptionally by TUTases. To solve this issue, we took advantage of previous data generated using parallel analysis of RNA ends (PARE)-seq. PARE-seq is one of the sequencing methods designed to map 5' hydroxylated end of RNAs and used to map small RNA cleavage sites (German et al., 2008). PARE-seq unambiguously identifies the position defined here as +1 as the 5' nucleotide of the RISC-generated 3' fragment of MYB33 (Figure 2B). Therefore, cleavage of MYB33 by miR159-loaded AGO1 does occur at the canonical site, which we defined here between positions 0 and +1 (Figure 2A). This was further

experimentally validated in the present study because MYB33 RISC 5'-cleavage fragments ending at position 0 accumulate in a genetic background abolishing uridylation (detailed later in Figure 6B).

To study MYB33 RISC 5'-cleavage fragments by 3' RACE-seq, we first analyzed the aerial part of 24-day-old plants grown *in vitro* corresponding to four biological replicates for WT and four biological replicates for the *heso1-1* mutant. We obtained a total of 29,689 reads for WT and 34,096 reads for *heso1-1* (Supplementary Table S2 and Supplementary Data Sheet S1). The WT data were first used to monitor the distribution of 3' extremities mapped in the sequence to which miR159 binds. The majority of reads (up to 85%) mapped at position 0 (Figure 2C). Therefore, we conclude that the 3' extremities of RISC-cleaved MYB33 are accurately mapped by 3' RACE-seq.

## Respective Contributions of HESO1 and URT1 in the Uridylation of MYB33 5'-Cleavage Fragments

To analyze untemplated nucleotides added after RISC-mediated cleavage of MYB33 mRNAs, the nucleotide extensions for reads that map to position 0 were analyzed first for the WT samples. Up



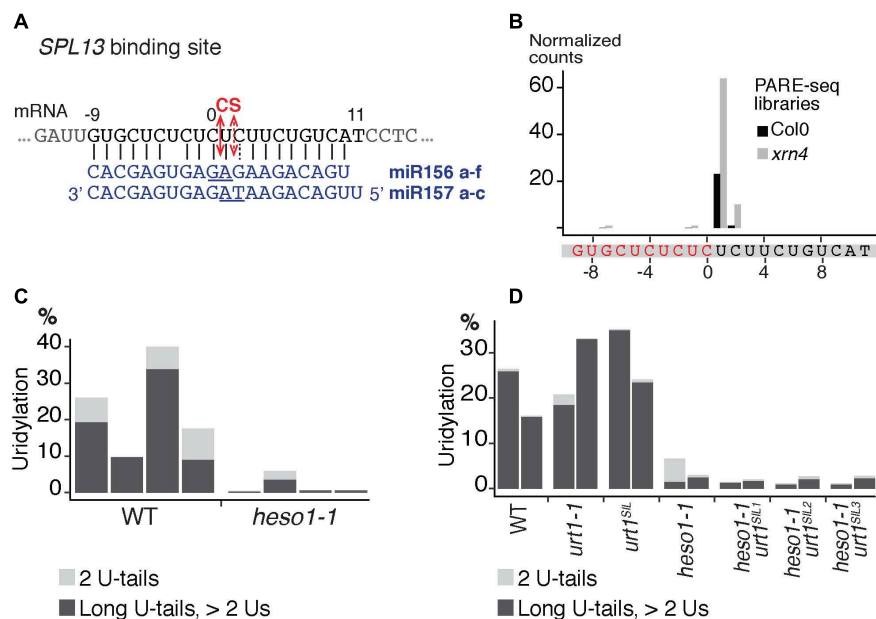
**FIGURE 6 |** Nibbled MYB33 RISC 5'-cleavage fragments accumulate in the absence of HESO1. Positions of 3' extremities of MYB33 RISC 5'-cleavage fragments mapped in a -10/0 window for (A) four biological replicates, WT and *heso1-1*, or (B) two biological replicates for WT, *urt1-1*, the *urt1<sup>SIL</sup>* line, *heso1-1*, and the three *heso1-1 urt1<sup>SIL</sup>* lines. Graphs are shown separately for each of the replicates in **Supplementary Figures S3, S4**.

to 98 % of MYB33 RISC 5'-cleavage fragments in WT are tailed by nucleotide extensions, which are predominantly composed of uridines (**Supplementary Table S3**). This result is in agreement with previous observations (Shen and Goodman, 2004; Ren et al., 2014; Zhang et al., 2017). Most U-rich tails were longer than 2 Us in the four WT biological replicates (**Figure 3A** and **Supplementary Table S3**). We then compared the impact of HESO1 on the uridylation MYB33 RISC 5'-cleavage fragments. A major decrease in uridylation was observed in *heso1-1* as compared with WT samples (**Figure 3B**). This observation confirmed that HESO1 is the main TUTase uridylating MYB33 5'-cleavage fragments, as shown here using four independent biological replicates in the Col-0 genetic background. In addition, and as previously observed (Ren et al., 2014), the size of U-tails detected in *heso1-1* was reduced as compared to WT, with mainly short U-tails (<2 Us) detected in *heso1-1* (**Figure 3**).

The residual uridylation in *heso1-1* indicates the involvement of an alternative TUTase. A good candidate for this activity is URT1, the second TUTase that has been identified in Arabidopsis

(Sement et al., 2013). To date, the possible involvement of URT1 in the uridylation of 5' RISC-cleaved mRNAs, including MYB33, has been proposed but not tested experimentally. Testing this hypothesis requires the production of a *heso1 urt1* double mutant. To this end, we crossed the *heso1-1* and *urt1-1* single mutants. However, we failed to recover the expected double mutant in the F2 progeny. This failure is yet unexplained but we could obtain lines that were originally designed to overexpress an inactive version of URT1, but that in fact co-suppress the endogenous URT1 gene in the *heso1-1* background. We selected three *heso1-1* lines for which the endogenous URT1 was not detected anymore by western blot analysis, revealing a drastic downregulation of URT1 (**Figure 4A**). These lines, which have no particular phenotype when grown under optimal conditions, are called *heso1-1 urt1<sup>SIL1</sup>*, *heso1-1 urt1<sup>SIL2</sup>*, and *heso1-1 urt1<sup>SIL3</sup>* thereafter. The uridylation of MYB33 5'-cleavage fragments was down to background levels in both biological replicates for three *heso1-1 urt1<sup>SIL</sup>* lines as compared with the single *heso1-1* mutant (**Figure 4B**). Therefore, both HESO1 and URT1 participate in





**FIGURE 7 |** Respective contributions of HESO1 and URT1 in the uridylation of *SPL13* RISC 5'-cleavage fragments. **(A)** Schematic representation of the cleavage of *SPL13* mRNA by AGO1 loaded with miR156 or miR157. The cleavage sites (CS) are predicted between the positions 0 and +1 or between position +1 and +2 for miR156 and miR157, respectively, and are defined by the 10th and 11th nucleotides (underlined) of the miRNAs. **(B)** 5' end mapping using PARE-seq data (German et al., 2008) of RISC 3'-cleavage fragments in WT and *xrn4*. **(C,D)** Percentages of uridylated *SPL13* RISC 5'-cleavage fragments in **(C)** four biological replicates for WT and *heso1-1* or **(D)** two biological replicates for WT, *urt1-1*, the *urt1<sup>SIL</sup>* line, *heso1-1*, and the three *heso1-1 urt1<sup>SIL</sup>* lines (i.e., six *heso1-1 urt1<sup>SIL</sup>* samples). Percentages of long (> 2 Us) and 2 U-tails are indicated by dark gray and light gray, respectively.

uridylating *MYB33* 5'-cleavage fragments, albeit HESO1 is clearly the main TUTase involved in uridylating these fragments.

Of note, HESO1 and URT1 might have a distinct contribution in the uridylation of *MYB33* 5'-cleavage fragments. HESO1 can synthesize short and long U-extensions, but URT1 seems to add only one or two uridines (Figure 4B). Interestingly, a similar distinction was proposed for HESO1 and URT1 in uridylating small RNAs. URT1 was proposed to add a single uridine to small RNAs to favor the subsequent action of HESO1, which prefers 3' extremities ending with uridines (Tu et al., 2015; Yu et al., 2017). A comparable scenario could exist for RISC 5'-cleavage fragments although additional investigation is required to confirm this hypothesis. In any case, and as previously observed for small RNAs, uridine addition by URT1 to RISC 5'-cleavage fragments does not seem to be a prerequisite to the action of HESO1, at least for *MYB33* 5'-cleavage fragments.

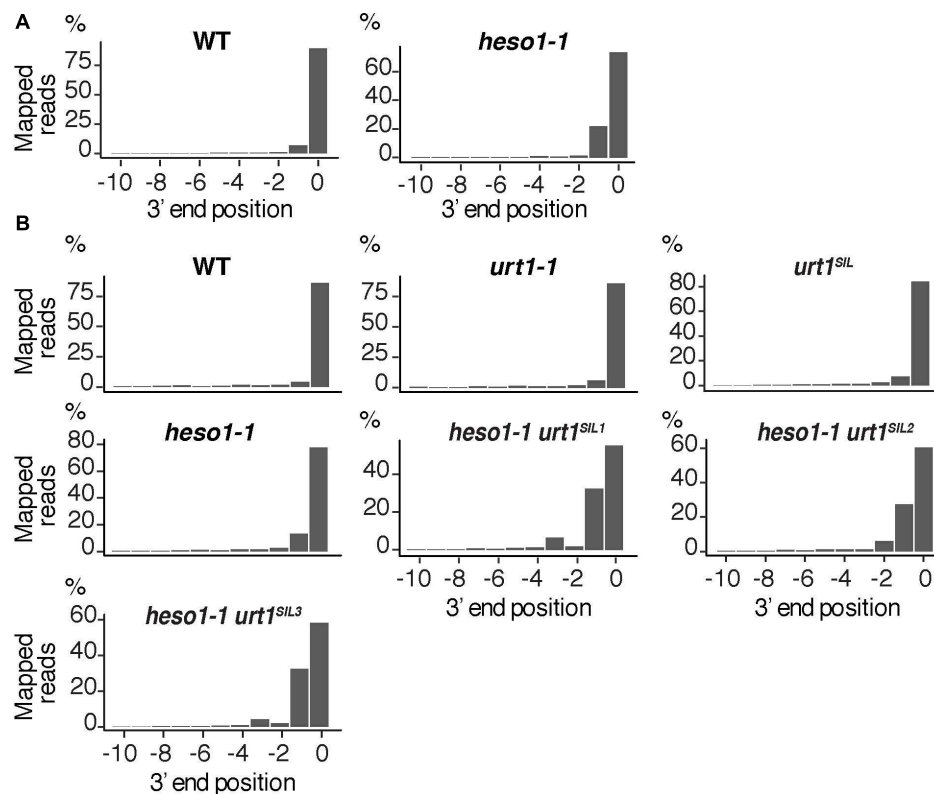
## Respective Contribution of HESO1 and URT1 in the Accumulation of *MYB33* 5'-Cleavage Fragments

To further check the predominant role of HESO1 in the metabolism of *MYB33* 5'-cleavage fragments, we analyzed their accumulation by northern blot analysis and phosphorimager quantification (Figure 5A). The accumulation of *MYB33* RISC 5'-cleavage fragments in each sample was calculated relative to its full-length mRNA and each ratio was normalized to the ratio obtained for the WT control for each of the two

replicates. As previously observed (Ren et al., 2014), *MYB33* 5'-cleavage fragments accumulated to higher levels in *heso1-1* with respect to WT (Figure 5B), although for unknown reasons the accumulation seemed variable in both replicates. Yet, our northern analysis confirmed that uridylation by HESO1 likely destabilizes *MYB33* 5'-cleavage fragments. The single *urt1* mutation seemed to have no major effect on this accumulation. Furthermore, *MYB33* 5'-cleavage fragments accumulated to similar levels in the *heso1-1 urt1<sup>SIL</sup>* lines as compared to the single *heso1-1* mutant (Figure 5). In other words, there was no additive effect of the lack of URT1 and HESO1, and this observation points to HESO1 as the main TUTase controlling the accumulation of *MYB33* 5'-cleavage fragments. Of note, miR159 accumulated to similar levels when HESO1 is absent, ruling out a higher rate of production of *MYB33* 5'-cleavage fragments in *heso1-1* mutants (Figure 5C). Altogether, the 3' RACE-seq and northern analyses indicate that HESO1 is the main TUTase modifying *MYB33* 5'-cleavage fragments. Although URT1 could add short uridine extensions to *MYB33* 5'-cleavage fragments, it does not appear to be a limiting factor neither in the uridylation nor in the destabilization of this fragment produced by RISC cleavage.

## mRNA 5' Fragments Are Nibbled at RISC Cleavage Site in the Absence of Uridylation

The 3' truncation up to several hundreds of nucleotides upstream of the RISC cleavage site was previously observed for *MYB33*



**FIGURE 8 |** Nibbled *SPL13* RISC 5'-cleavage fragments accumulate in the absence of uridylation. Positions of 3' extremities of *SPL13* RISC 5'-cleavage fragments mapped in a -10/0 window for (A) four biological replicates, WT and *heso1-1*, or (B) two biological replicates for WT, *urt1-1*, the *urt1<sup>SIL</sup>* line, *heso1-1*, and the three *heso1-1 urt1<sup>SIL</sup>* lines. Graphs are shown separately for each of the replicates in **Supplementary Figures S5, S6**.

5'-cleavage fragments in the *heso1-2* mutant (Ren et al., 2014). We took advantage of the depth of the 3' RACE-seq procedure to analyze at high resolution the 3' extremities of *MYB33* 5'-cleavage fragments in the vicinity of the cleavage site. Although the vast majority of extremities in the four WT biological replicates mapped at position 0, different patterns were observed for *heso1-1*. The patterns were not completely identical in the four biological replicates, but they all revealed the same trend: the 3' extremities were spread over positions from -10 to 0 (Figure 6 and Supplementary Figure S3). This observation reveals that the *MYB33* 5'-cleavage fragments that accumulate in the absence of HESO1 are nibbled at close proximity to the cleavage site. This nibbling shortens *MYB33* 5'-cleavage fragments by up to 8–9 nucleotides (Figure 6A and Supplementary Figure S3). Such a nibbling was not observed in the single *urt1-1* mutant (Figure 6B and Supplementary Figure S4) but it was consistently observed in *heso1-1* and not aggravated in *heso1-1 urt1<sup>SIL</sup>* mutants (Figure 6B and Supplementary Figure S4). Therefore, the nibbling is solely attributed to the absence of HESO1, but not of URT1, in the case of *MYB33* 5'-cleavage fragments.

We then analyzed the respective contribution of HESO1 and URT1 in uridylating the 5' fragments produced by RISC cleavage of *Squamosa promoter-binding-like protein 13* (*SPL13*) mRNAs that are targets of miR156 and miR157 (Figure 7A). PARE-seq data identify a major and a minor 5' extremity for the 3'

fragments produced by RISC cleavage (Figure 7B). Therefore, it is possible that in addition to the major cleavage site denoted 0 in Figure 7A, a minor site exists at position +1. This minor site at +1 presumably results from the action of miR157 (Figure 7A; He et al., 2018). Because nucleotide +1 is a U, it is not possible to determine in the 3' RACE-seq data whether this U is encoded or added post-transcriptionally. To eliminate this uncertainty that could affect the proportion of uridylated versus non-uridylated fragments, we considered only tails of at least two nucleotides. Of note, not considering the 1 U extensions may lead to the underestimation of the action of URT1 and/or HESO1 in adding 1 U. The overall level of uridylation of *SPL13* 5'-cleavage fragments was lower than for *MYB33*, with percentage of uridylation below 40% and an increased variability between replicates (Figures 7C,D and Supplementary Table S4). Yet, a similar pattern was observed for both targets: uridylation of RISC 5'-cleavage fragments is mostly reduced in the absence of HESO1 and close to background levels in *heso1-1 urt1<sup>SIL</sup>* lines (Figures 7C,D). Interestingly, the nibbling of RISC 5'-cleavage fragments was increased in the six replicates of *heso1-1 urt1<sup>SIL</sup>* although to a lesser extent than the one observed for *MYB33* (Figure 8). This observation confirms the accumulation of RISC 5'-cleavage fragments that are nibbled close to the cleavage site in case of defective uridylation. The greater accumulation of nibbled fragments in absence of HESO1 and URT1 suggests

that in the case of *SPL13* 5'-cleavage fragments, the absence of uridylation *per se* is responsible for this accumulation.

Two, non-mutually exclusive, interpretations can explain the accumulation of nibbled RISC 5'-cleavage fragments in the absence of uridylation. First, uridylation of the nibbled fragments could trigger their degradation. Their fast turn-over would explain that they are not detected in WT plants. However, those fragments would accumulate in the absence of the TUTases. The second alternative possibility would be that in the presence of HESO1 and/or URT1, the 3' extremities are not accessible to the activity, presumably a 3'-5' exoribonucleolytic activity, that generates the nibbled RNA species. Such a possibility was previously evoked to explain the accumulation of truncated 5'-cleavage fragments in the *heso1-2* mutant (Ren et al., 2014). Solving this question entails the identification of all ribonucleases involved in the metabolism of 5' RISC-cleaved transcripts.

## CONCLUSION

Here, we report the respective contribution of HESO1 and URT1 in the metabolism of two 5' RISC-cleaved mRNAs. In addition, we show the applicability of 3' RACE-seq to map the 3' ends of 5' RISC-cleaved transcripts and to identify untemplated nucleotides added at these 3' ends. The depth of 3' RACE-seq will be useful for both qualitative and quantitative comparisons across different targets, tissues, conditions or genotypes. For instance, different RISC 5'-cleavage fragments could be investigated to identify both common and specific behaviors of these RNA fragments produced by post-transcriptional gene silencing. Also, the full machinery involved in the degradation of RISC 5'-cleavage fragments needs to be characterized. This is an on-going process with the recent identification of RICE exoribonucleases (Zhang et al., 2017) or the recent description that components of the NSD pathway and the Ski complex, a major co-factor of the cytosolic RNA exosome, are involved in the degradation of RISC 5'-cleavage fragments (Branscheid et al., 2015; Szádeczky-Kardoss et al., 2018). Yet the direct involvement of the RNA exosome in the clearance of RISC 5'-cleavage fragments remains to be demonstrated in Arabidopsis. The impact of SUPPRESSOR OF VARICOSE (SOV), whose ortholog is called Dis3L2 in non-plant eukaryotes, on the degradation of RISC 5'-cleavage fragments could also be investigated. Dis3L2 is a 3'-5' exoribonuclease belonging to the RNase II family and whose activity is stimulated by uridylation in fission yeast, fruit fly or human cells (De Almeida et al., 2018). Whether SOV participates in the clearance of uridylated 5' fragments of RISC-cleaved transcripts could be reliably addressed using 3' RACE-seq and by comparing Col-0 and *Ler* accessions, because a point mutation affects SOV activity in Col-0 (Zhang et al., 2010). All these examples illustrate that a large number of samples must be analyzed with sufficient depth and replicates to draw reliable conclusions. The 3' RACE-seq method adapted to the analysis of RISC 5'-cleavage fragments will contribute to fully characterize the tailing and nibbling events linked to the metabolism of these fragments and to address the respective roles

of distinct factors of the RNA degradation machinery in this process.

## AUTHOR CONTRIBUTIONS

DG and HZ conceived and designed the study, wrote the paper, and acquired funding. HZ, A-CJ, and HS performed the experiments. HZ performed the bioinformatics analysis. A-CJ and HS edited the manuscript. HZ and HS prepared the illustrations.

## FUNDING

This work was supported by the Centre National de la Recherche Scientifique (CNRS, France) and the Agence Nationale de la Recherche (ANR, France) as part of the Programme d'Investissements d'Avenir in the frame of the LabEx NetRNA (ANR-2010-LABX-36) to DG and in the frame of the IdEx Unistra to HZ.

## ACKNOWLEDGMENTS

The authors are grateful to Camille Noblet for technical help.

## SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fpls.2018.01438/full#supplementary-material>

**FIGURE S1 related to Figure 4** | Uncropped images of the western blot analysis and the membrane stained with Coomassie blue shown in **Figure 4**.

**FIGURE S2 related to Figure 5** | Uncropped images of the northern blot analysis and the membrane stained with methylene blue for **(A)** MYB33 5' fragment analysis and for **(B)** miR159 analysis.

**FIGURE S3 related to Figure 6** | Nibbled MYB33 RISC 5'-cleavage fragments accumulate in the absence of HESO1. Positions of 3' extremities of MYB33 RISC 5'-cleavage fragments mapped in a -10/0 window for four biological replicates in WT and *heso1-1*. Graphs are shown separately for each of the four replicates.

**FIGURE S4 related to Figure 6** | Nibbled MYB33 RISC 5'-cleavage fragments accumulate in the absence of HESO1. Positions of 3' extremities of MYB33 RISC 5'-cleavage fragments mapped in a -10/0 window for two biological replicates for WT, *urt1-1*, the *urt1*<sup>SIL</sup> line, *heso1-1*, and the three *heso1-1 urt1*<sup>SIL</sup> lines. Graphs are shown separately for each of the two replicates.

**FIGURE S5 related to Figure 8** | Positions of 3' extremities of *SPL13* 5'-cleavage fragments mapped in a -10/0 window for four biological replicates in WT and *heso1-1*. Graphs are shown separately for each of the four replicates.

**FIGURE S6 related to Figure 8** | Nibbled *SPL13* RISC 5'-cleavage fragments accumulate in the absence of HESO1 and URT1. Positions of 3' extremities of *SPL13* RISC 5'-cleavage fragments mapped in a -10/0 window for two biological replicates for WT, *urt1-1*, the *urt1*<sup>SIL</sup> line, *heso1-1*, and the three *heso1-1 urt1*<sup>SIL</sup> lines. Graphs are shown separately for each of the two replicates.

**TABLE S1** | List of primers used in this study.

**TABLE S2** | Summary of the number of reads analyzed at each step of the data processing for each 3' RACE-seq library.

**TABLE S3** | Exhaustive list of extensions found by 3' RACE-seq for *MYB33* 5'-cleavage fragments in WT for four biological replicates from dataset #1.

**TABLE S4** | Exhaustive list of extensions found by 3' RACE-seq for *SPL13* 5'-cleavage fragments in WT for four biological replicates from dataset #1.

## REFERENCES

- Arribas-Hernández, L., Kielpinski, L. J., and Brodersen, P. (2016). mRNA decay of most Arabidopsis miRNA targets requires slicer activity of AGO1. *Plant Physiol.* 171, 2620–2632. doi: 10.1104/pp.16.00231
- Bartel, D. P. (2018). Metazoan microRNAs. *Cell* 173, 20–51. doi: 10.1016/j.cell.2018.03.006
- Borges, F., and Martienssen, R. A. (2015). The expanding world of small RNAs in plants. *Nat. Rev. Mol. Cell Biol.* 16, 727–741. doi: 10.1038/nrm4085
- Branscheid, A., Marchais, A., Schott, G., Lange, H., Gagliardi, D., Andersen, S. U., et al. (2015). SKI2 mediates degradation of RISC 5'-cleavage fragments and prevents secondary siRNA production from miRNA targets in Arabidopsis. *Nucleic Acids Res.* 43, 10975–10988. doi: 10.1093/nar/gkv1014
- Brodersen, P., Sakvarelidze-Achard, L., Bruun-Rasmussen, M., Dunoyer, P., Yamamoto, Y. Y., Sieburth, L., et al. (2008). Widespread translational inhibition by plant miRNAs and siRNAs. *Science* 320, 1185–1190. doi: 10.1126/science.1159151
- Chang, H., Lim, J., Ha, M., and Kim, V. N. (2014). TAIL-seq: genome-wide determination of poly(A) tail length and 3' end modifications. *Mol. Cell* 53, 1044–1052. doi: 10.1016/j.molcel.2014.02.007
- Chen, X. (2004). A MicroRNA as a translational repressor of APETALA2 in Arabidopsis flower development. *Science* 303, 2022–2025. doi: 10.1126/science.1088060
- De Almeida, C., Scheer, H., Zuber, H., and Gagliardi, D. (2018). RNA uridylation: a key post-transcriptional modification shaping the coding and non-coding transcriptome. *WIREs RNA* 9:e1440. doi: 10.1002/wrna.1440
- German, M. A., Pillay, M., Jeong, D. -H., Hetawal, A., Luo, S., Janardhanan, P., et al. (2008). Global identification of microRNA-target RNA pairs by parallel analysis of RNA ends. *Nat. Biotechnol.* 26, 941–946. doi: 10.1038/nbt1417
- He, J., Xu, M., Willmann, M. R., McCormick, K., Hu, T., Yang, L., et al. (2018). Threshold-dependent repression of SPL gene expression by miR156/miR157 controls vegetative phase change in *Arabidopsis thaliana*. *PLoS Genet.* 14:e1007337. doi: 10.1371/journal.pgen.1007337
- Iwakawa, H., and Tomari, Y. (2015). The functions of MicroRNAs: mRNA decay and translational repression. *Trends Cell Biol.* 25, 651–665. doi: 10.1016/j.tcb.2015.07.011
- Kwak, J. E., and Wickens, M. (2007). A family of poly(U) polymerases. *RNA* 13, 860–867. doi: 10.1261/rna.514007.
- Li, S., Liu, L., Zhuang, X., Yu, Y., Liu, X., Cui, X., et al. (2013). MicroRNAs inhibit the translation of target mRNAs on the endoplasmic reticulum in Arabidopsis. *Cell* 153, 562–574. doi: 10.1016/j.cell.2013.04.005
- Martínez de Alba, A. E., Elvira-Matlot, E., and Vaucheret, H. (2013). Gene silencing in plants: a diversity of pathways. *Biochim. Biophys. Acta* 1829, 1300–1308. doi: 10.1016/j.bbagr.2013.10.005
- Munoz-Tello, P., Rajappa, L., Coquille, S., and Thore, S. (2015). Polyuridylation in eukaryotes: a 3'-end modification regulating RNA Life. *Biomed. Res. Int.* 2015:968127. doi: 10.1155/2015/968127
- Orban, T. I., and Izaurralde, E. (2005). Decay of mRNAs targeted by RISC requires XRN1, the ski complex, and the exosome. *RNA* 11, 459–469. doi: 10.1261/rna.7231505
- Reis, R. S., Hart-Smith, G., Eamens, A. L., Wilkins, M. R., and Waterhouse, P. M. (2015). Gene regulation by translational inhibition is determined by dicer partnering proteins. *Nat. Plants* 1:14027. doi: 10.1038/nplants.2014.27
- Ren, G., Chen, X., and Yu, B. (2012). Uridylation of miRNAs by hen1 suppressor1 in Arabidopsis. *Curr. Biol.* 22, 695–700. doi: 10.1016/j.cub.2012.02.052
- Ren, G., Xie, M., Zhang, S., Vinovskis, C., Chen, X., and Yu, B. (2014). Methylation protects microRNAs from an AGO1-associated activity that uridylates 5' RNA fragments generated by AGO1 cleavage. *Proc. Natl. Acad. Sci. U.S.A.* 111, 6365–6370. doi: 10.1073/pnas.1405083111
- Scheer, H., Zuber, H., De Almeida, C., and Gagliardi, D. (2016). Uridylation earmarks mRNAs for degradation and more. *Trends Genet.* 32, 607–619. doi: 10.1016/j.tig.2016.08.003
- Sement, F. M., Ferrier, E., Zuber, H., Merret, R., Alioua, M., Deragon, J. -M., et al. (2013). Uridylation prevents 3' trimming of oligoadenylated mRNAs. *Nucleic Acids Res.* 41, 7115–7127. doi: 10.1093/nar/gkt465
- Shen, B., and Goodman, H. M. (2004). Uridine addition after microRNA-directed cleavage. *Science* 306, 997–997. doi: 10.1126/science.1103521
- Sikorska, N., Zuber, H., Gobert, A., Lange, H., and Gagliardi, D. (2017). RNA degradation by the plant RNA exosome involves both phosphorolytic and hydrolytic activities. *Nat. Commun.* 8:2162. doi: 10.1038/s41467-017-02066-2
- Souret, F. F., Kastenmayer, J. P., and Green, P. J. (2004). AtXRN4 degrades mRNA in Arabidopsis and its substrates include selected miRNA targets. *Mol. Cell* 15, 173–183. doi: 10.1016/j.molcel.2004.06.006
- Szadeczyk-Kardoss, I., Csorba, T., Auber, A., Schamberger, A., Nyikó, T., Taller, J., et al. (2018). The nonstop decay and the RNA silencing systems operate cooperatively in plants. *Nucleic Acids Res.* 46, 4632–4648. doi: 10.1093/nar/gky279
- Tu, B., Liu, L., Xu, C., Zhai, J., Li, S., Lopez, M. A., et al. (2015). Distinct and cooperative activities of HESO1 and URT1 nucleotidyl transferases in microRNA turnover in Arabidopsis. *PLoS Genet.* 11:e1005119. doi: 10.1371/journal.pgen.1005119
- Vaucheret, H. (2008). Plant ARGONAUTES. *Trends Plant Sci.* 13, 350–358. doi: 10.1016/j.tplants.2008.04.007
- Wang, X., Zhang, S., Dou, Y., Zhang, C., Chen, X., Yu, B., et al. (2015). Synergistic and independent actions of multiple terminal nucleotidyl transferases in the 3' tailing of small RNAs in Arabidopsis. *PLoS Genet.* 11:e1005091. doi: 10.1371/journal.pgen.1005091
- Yang, L., Wu, G., and Poethig, R. S. (2012). Mutations in the GW-repeat protein SUO reveal a developmental function for microRNA-mediated translational repression in Arabidopsis. *Proc. Natl. Acad. Sci.* 109, 315–320. doi: 10.1073/pnas.1114673109
- Yang, Z., Ebright, Y. W., Yu, B., and Chen, X. (2006). HEN1 recognizes 21–24 nt small RNA duplexes and deposits a methyl group onto the 2' OH of the 3' terminal nucleotide. *Nucleic Acids Res.* 34, 667–675. doi: 10.1093/nar/gkj474
- Yu, B., Yang, Z., Li, J., Minakhina, S., Yang, M., Padgett, R. W., et al. (2005). Methylation as a crucial step in plant microRNA biogenesis. *Science* 307, 932–935. doi: 10.1126/science.1107130
- Yu, Y., Jia, T., and Chen, X. (2017). The “how” and “where” of plant microRNAs. *New Phytol.* 216, 1002–1017. doi: 10.1111/nph.14834
- Zhang, H., Xia, R., Meyers, B. C., and Walbot, V. (2015). Evolution, functions, and mysteries of plant ARGONAUTE proteins. *Curr. Opin. Plant Biol.* 27, 84–90. doi: 10.1016/j.pbi.2015.06.011
- Zhang, W., Murphy, C., and Sieburth, L. E. (2010). Conserved RNaseII domain protein functions in cytoplasmic mRNA decay and suppresses Arabidopsis decapping mutant phenotypes. *Proc. Natl. Acad. Sci. U.S.A.* 107, 15981–15985. doi: 10.1073/pnas.1007060107



- Zhang, Z., Hu, F., Sung, M. W., Shu, C., Castillo-González, C., Koiwa, H., et al. (2017). RISC-interacting clearing 3'-5' exoribonucleases (RICEs) degrade uridylated cleavage fragments to maintain functional RISC in *Arabidopsis thaliana*. *eLife* 6:e24466. doi: 10.7554/eLife.24466
- Zhao, Y., Yu, Y., Zhai, J., Ramachandran, V., Dinh, T. T., Meyers, B. C., et al. (2012). The Arabidopsis nucleotidyl transferase HESO1 uridylates unmethylated small RNAs to trigger their degradation. *Curr. Biol.* 22, 689–694. doi: 10.1016/j.cub.2012.02.051
- Zuber, H., Scheer, H., Ferrier, E., Sement, F. M., Mercier, P., Stupfler, B., et al. (2016). Uridylation and PABP cooperate to repair mRNA deadenylated ends in Arabidopsis. *Cell Rep.* 14, 2707–2717. doi: 10.1016/j.celrep.2016.02.060

**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 © 2018 Zuber, Scheer, Joly and Gagliardi. 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](http://www.frontiersin.org)

**Contact us:** [info@frontiersin.org](mailto: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