Characterisation of differentially expressed trans-acting siRNA (ta-siRNAs) facing stress conditions in melon

Luis Manuel Cervera Seco

Bachelor's degree Final Project

Supervisor: María Carmen Marqués Romero

Supervisor: Gustavo Germán Gómez

Academic Tutor: María Carmen Gisbert Doménech

Universitat Politècnica de València

Escuela Técnica Superior de Ingeniería Agronómica y del Medio Natural (ETSIAMN)

Academic year 2017-2018

BSc in Biotechnology

València, June 2018

Running title

Characterisation of differentially expressed trans-acting siRNA (ta-siRNAs) facing stress conditions in melon.

Abstract

Trans-acting siRNAs (ta-siRNAs) are a class of small regulatory RNAs (20 to 24 nts) described in plants, which are produced from specific genome regions called TAS genes. Transcribed fragments derived from these genes are recognised and processed by a miRNA and an RNA dependent RNA polymerase 6 (RDR6) in order to generate a dsRNA, in turn processed by DCL4 into 21-nt phased fragments. One of these pieces, once loaded into AGO, constitutes a mature ta-siRNA able to regulate many target genes *in trans*, at post-transcriptional level. To identify and characterise potential *Cucumis melo* (melon) ta-siRNAs implicated in stress-responsive processes, sRNA libraries were obtained from melon plants exposed to seven different stress situations both biotic and abiotic. Via bioinformatic approaches, a total of 242 candidate TAS genes were predicted. Two of the most promising candidates (TAS 915 and TAS 917) were further validated following computational algorithms and biological assays. Their sequences showed a high homology with the TAS3 family of other plant species, whose most relevant targets are Auxin Response Factors, transcriptional regulators involved in development and stress responses. Besides, trans-acting activity of these TAS genes was confirmed via identification and validation of ta-siRNA targets, their correlation to TAS transcript accumulation and degradome assays. Our results represent the first identification and functional validation of ta-siRNAs in *C. melo* and may provide novel insights into the role of this particular family of small non-coding RNAs in melon stress responses.

Resumen

Los trans-acting siRNAs (ta-siRNAs) son una clase de pequeños RNAs (20 a 24 nts) reguladores descritos en plantas. Se producen a partir de regiones específicas del genoma llamadas genes TAS. Los transcritos derivados de estos genes TAS son reconocidos y procesados por un miRNA y por una RNA polimerasa RNA dependiente 6 (RDR6) para generar un RNA de doble cadena, que es procesado por DCL4 en fragmentos en fase de 21 nts, llamados ta-siRNAs. Los ta-siRNAs funcionales se cargan en AGO y regulan a nivel post-transcripcional diversos genes *en trans*. Para identificar y caracterizar potenciales ta-siRNAs de *Cucumis melo* (melón) implicados en el proceso de respuesta a estrés, se trabajó con librerías de sRNAs obtenidas a partir de plantas de melón expuestas a 7 situaciones diferentes de estrés biótico y abiótico. Mediante una aproximación bioinformática, se predijeron 242 genes TAS candidatos. Dos de los más prometedores (TAS 915 y TAS 917) fueron posteriormente validados mediante ensayos biológicos y algoritmos computacionales. Sus secuencias presentaban homología con genes de la familia TAS3 de otras especies vegetales, cuyas dianas más relevantes son los Factores de Respuesta a Auxinas, que son reguladores a nivel transcripcional involucrados en procesos del desarrollo y respuesta a estrés. Además, la actividad de estos genes TAS fue confirmada a través de la identificación y validación de los genes diana, su correlación con la acumulación de los transcritos TAS y análisis del degradoma. Los resultados obtenidos suponen la primera identificación de ta-siRNAs funcionales en *C. melo*, y ofrecen una nueva perspectiva en cuanto a mecanismos de respuesta a estrés en melón regulados por estos pequeños RNAs no codificantes.

Keywords

Stress-responsive sRNA, TAS genes, ta-siRNA, *Cucumis melo*, RNA silencing

Palabras clave

Respuesta a estrés mediada por sRNAs, genes TAS, ta-siRNA, *Cucumis melo*, silenciamiento por RNA

Project information

Author: Luis Manuel Cervera Seco
Academic Tutor: María Carmen Gisbert Doménech
Supervisor: María Carmen Marqués Romero
Supervisor: Gustavo Germán Gómez
Place and date: València, June 2018

Acknowledgements

I would like to thank, in first place, Alejandro Sanz Carbonell for the magnificent aid in design of scripts and training in computer programming. Also, Joan Márquez Molins, for his invaluable support in vector designing and manipulation of plant materials, as well as for his revision of the final document. I also really appreciate the exceptional guidance provided by my supervisors María Carmen Marqués Romero and Gustavo Germán Gómez throughout the entire year, and my acknowledgement as well for Carmina Gisbert Doménech, for her help in formal corrections of the project document. I would also like to express my gratitude to my friend Paula Gomis Rosa, for the provided training in LaTeX formatting and processing. The synthetic ta-siRNA and artificial miRNA backbones utilised for the *in vivo* validation assay were supplied by Alberto Carbonell's laboratory at the Instituto de Biología Molecular y Celular de Plantas (IBMCP), based in València, Spain.

Contents

List of Figures

List of Tables

Abbreviations

A ⟶ *Agrobacterium tumefaciens*
ABA ⟶ Abscisic acid
AGO ⟶ Argonaute Protein
amiRNA ⟶ Artificial micro RNA
AREB ⟶ ABA responsive element binding
ARF ⟶ Auxin Response Factor
C ⟶ Cold
cDNA ⟶ Complementary DNA
ChFP ⟶ Cherry fluorescent protein
CMV ⟶ *Cucumber mosaic virus*
D ⟶ Drought
DCL ⟶ Dicer-like protein
DCL4 ⟶ Dicer-like 4 protein
DNA ⟶ Deoxyribonucleic acid
DRB4 ⟶ Double-stranded RNA binding factor 4
DREB ⟶ Dehydration responsive element binding
dsRNA ⟶ Double-stranded RNA
EST ⟶ Expressed sequence tag
hc-siRNA ⟶ Heterochromatic small interference RNA
HEN1 ⟶ HUA ENHANCER 1
HSVd ⟶ *Hop stunt Viroid*
ICuGI ⟶ International Cucurbit Genome Initiative
LFC ⟶ Log fold change
MAP ⟶ Mitogen-activated protein
MES ⟶ 2-(N-morpholino)ethanesulfonic acid
miRNA ⟶ micro RNA
mRNA ⟶ Messenger RNA
MNSV ⟶ *Melon necrotic spot virus*
MON ⟶ *Monosporascus cannonballus*
NCBI ⟶ National Center of Biotechnology
ncRNA ⟶ Non-coding RNA
nt(s) ⟶ Nucleotide(s)
OD600 ⟶ Optical density at 600 nm
PCR ⟶ Polymerase chain reaction
Pol II ⟶ RNA polymerase II
PPR ⟶ Pentatricopeptide-repeat containing protein
PR ⟶ Pathogenesis-related

phasiRNA ⟶ Phased small interference RNA
qPCR ⟶ Quantitative polymerase chain reaction
qRT-PCR ⟶ Quantitative reverse transcription polymerase chain reaction
RDR6 ⟶ RNA-dependent RNA polymerase 6
RISC ⟶ RNA-induced silencing complex
RNA ⟶ Ribonucleic acid
RPM ⟶ Reads per million
RT-PCR ⟶ Reverse transcription polymerase chain reaction
SA ⟶ Salinity
SD ⟶ Short day
SGS3 ⟶ Suppressor of gene silencing 3
siRNA ⟶ Small interference RNA
sRNA ⟶ Small RNA
syn-tasiRNA ⟶ Synthetic trans-acting small interference RNA
ta-siRNA ⟶ Trans-acting small interference RNA
T0 ⟶ Time at which the treatment to melon plant is just applied, initial time
T1 ⟶ Time 1; 2 days post-treatment
T2 ⟶ Time 2; 4 days post-treatment
T3 ⟶ Time 3; 11 days post-treatment
T_m ⟶ Melting temperature
UPE ⟶ Unpaired energy
WMV ⟶ *Watermelon mosaic virus*

1 Introduction

1.1 Plant stress

Plants are constantly exposed to diverse types of abiotic stresses, including extreme temperatures, light intensity fluctuations, drought or different degrees of salinity. As well as environmental factors, plants face insect and pathogen attack which may unbalance their homeostasis (Kinoshita and Seki, 2014). As opposed to most of animals able to flee from the unfavourable environmental conditions, plants are sessile organisms so that they have been impelled to develop a vast range of signalling pathways and metabolic routes to counteract the detrimental effects arisen when challenged by these stresses and return to the initial homeostatic state (Zhu, 2016). Nowadays, the need for modifying molecular mechanisms or designing novel strategies to adapt rapidly to those changes is becoming crucial in order to avoid massive crop losses along with a reduction of productivity, especially in the context of a growing world population and global warming, which may possibly imply doubling plant productivity by mid-century (Qin *et al.*, 2011; Kumar, 2014).

Plant hormones are responsible for the modulation of stress responses. Among them, abscisic acid (ABA) is considered a key regulator of abiotic stress, mainly due to the action of AREB proteins, the transcriptional activation of gene expression mediated by DREB transcription factors and the activation of MAP kinases cascades. Additionally, other responses to unfavourable environmental conditions are ABA-mediated, namely, osmolyte production -which include glycine betaine, proline or polyalcohols-, and reduction of transpiration rate by stomata closure. In addition, ABA-mediated regulation plays significant roles in developmental processes such as seed dormancy, seed germination and lateral root growth as well (Qin *et al.*, 2011; Zhu, 2016).

Regarding biotic stress, a complex interplay between salicylic acid-, jasmonic acid- and ethylene-mediated responses modulates the defence against plant pathogens via the activation of specific genes, such as pathogenesis-related (PR) genes, among others. Briefly, several signalling and hormone pathways seem to be involved in regulation of stresses, notwithstanding comprehensive knowledge embracing the complex network of signalling hormones and pathways still needs to be fully unveiled (Derksen *et al.*, 2013; Zhang *et al.*, 2013).

1.2 Small RNAs in stress response

Over the last years, other molecular entities have grasped researchers' attention. These are able to confer resistance or tolerance to distinct kinds of stress, by controlling metabolic routes involved in fine tuning of stress response. Among these key genetic regulators are specific RNA molecules described as small RNAs (sRNAs) which are 20- to 24-nt long RNAs processed from

longer precursors (around 200-nt long), that repress target gene expression, usually at post-transcriptional level. This group has been proven to be involved in developmental processes in plants, phase transition, heterochromatin formation, changes in gene expression, and more remarkably, abiotic and biotic tolerance via either post-transcriptional or transcriptional endogenous gene silencing (Carthew and Sontheimer, 2009; Arikit *et al.*, 2013). Plant small RNAs are further classified into two main classes: microRNAs (miRNAs) and small interference RNAs (siRNAs), the latter of which is subdivided into phased siRNAs (phasiRNAs), including trans-acting siRNAs (Tasi-RNAs, Ta-siRNAs, ta-siRNAs or tasiRNAs), and heterochromatic siRNAs (hc-siRNAs) among others (Tang *et al.*, 2012; Arikit *et al.*, 2013). The majority of siRNAs target the same region which originated them, except for ta-siRNAs that, in a similar way to miRNAs, target transcripts (mRNAs) from different loci; in other words, they act *in trans* (Tang *et al.*, 2012).

As for the context of interest, ta-siRNAs are small regulatory pieces of RNA, normally 21-nt long, produced from the miRNA-guided phased cleavage of a non-coding transcriptional products of specific genome regions called TAS genes (Tang *et al.*, 2012). The characteristic term "phased" refers to the precise sRNA generation in tandem triggered by the initial miRNA-guided cleavage and production of 21-nt phases by DCL proteins starting from the miRNA cut site at the TAS transcript (Fei *et al.*, 2013).

1.3 ta-siRNA biogenesis

The biogenesis of ta-siRNAs involves the recruitment of various critical proteins, intervening sequentially: RNA polymerase II (Pol II), Argonaute proteins 1 and 7 (AGO1/7), RNA-dependent RNA polymerase 6 (RDR6), Suppressor of gene silencing 3 (SGS3), and DCL4 along with Double-stranded RNA binding factor 4 (DRB4), as it is depicted in Figure 1. It should be noted that the entire set of proteins implied in ta-siRNA biogenesis are shared with the rest of sRNA processing routes, even though different variants of AGO, DCL, RNA polymerases and RNA-dependent RNA polymerases act to produce those sRNAs (Fei *et al.*, 2013).

After transcription of TAS genes by Pol II, TAS transcripts are recognised and cleaved by an AGO protein, usually AGO1, based on their homology with a miRNA loaded onto it, which directs the reaction (Arikit *et al.*, 2013; Fei *et al.*, 2013; Tang *et al.*, 2012). Instead of AGO1, AGO7 may be the responsible for the cleavage of the transcript in the presence of miR390, which slices specifically TAS3 loci-derived mRNAs, or transcripts with two target sites for the miRNA (Montgomery *et al.*, 2008; Tang *et al.*, 2012). In addition, AGO7 usually acts at the 3' end and processing continues upstream from the site. In any case, this step give rise to an unprotected 200-nt long ncRNA approximately, which is subsequently converted into a double-stranded RNA (dsRNA) upon recruitment of the RDR6 (Arikit *et al.*, 2013; Fei *et al.*, 2013).

Furthermore, RNA polymerase may count on the assistance of SGS3 for the polymerisation of the miRNA-cleaved product. The role of SGS3 was not completely elucidated until the study *in vitro* developed by Yoshikawa *et al.* (2013) stated the function of SGS3 as to stabilise the miR173-mediated cleavage of the 3' end of mRNA, preventing from degradation and gene silencing, after recruitment by AGO1 upon binding to TAS transcript. In a further stage, dsRNA is cleaved and processed by DCL4, a member of the DCL family, within RNase III endonucleases group, that are known to be one of the key components in sRNAs biogenesis (Arikit *et al.*, 2013; Fei *et al.*, 2013). Its cutting activity initiated from the cleavage of the miRNA cut site produces regular 21-bp long ta-siRNA duplexes, with the characteristic mentioned phased pattern (Li *et al.*, 2012; Wu *et al.*, 2017). Additionally, DCL4 requires the presence of DRB4 for a proper cleavage of the long dsRNA in three ways, namely, interacting with DCL4, increasing its activity and binding to dsRNA (Fukudome *et al.*, 2011).

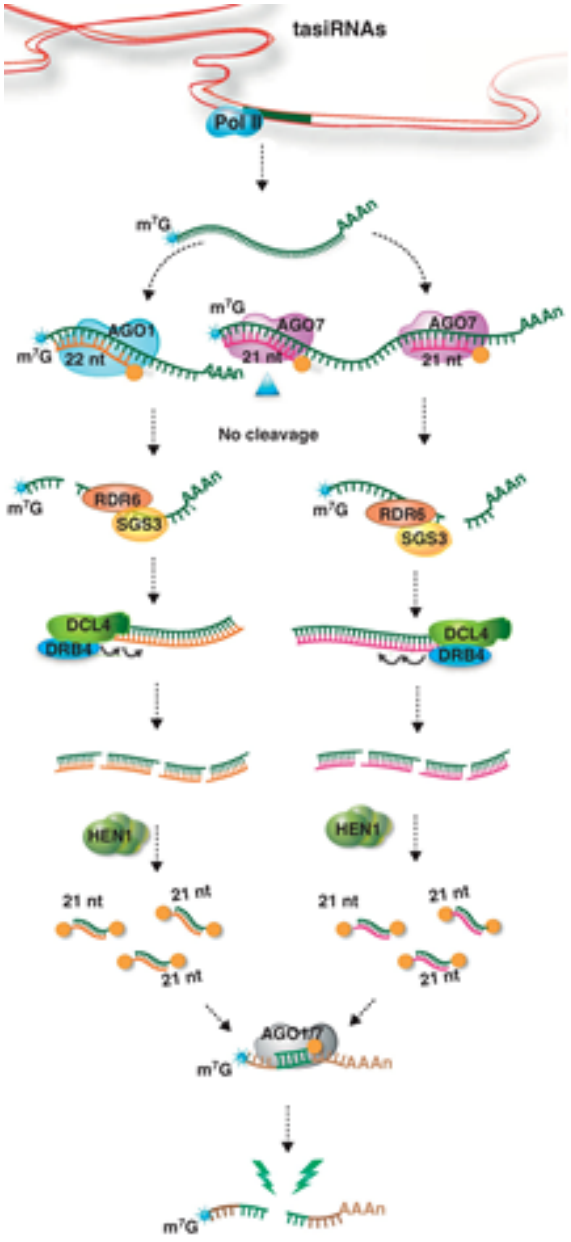


Figure 1: ***Schematical representation of ta-siRNA biogenesis, and final targeting and degradation of mRNA target genes.*** *Modified image taken from Arikit et al. (2013).*

It has been described the participation of an RNA methyltransferase in the pathway as well, HUA ENHANCER 1 (HEN1), whose function has been proven to be the methylation of 3' ends of DCL4-cleavage products in order to protect them from enzymatical modifications, such as 3'-uridylation, related to sRNA degradation (Zhai *et al.*, 2013). To complete the ta-siRNA biogenesis, the duplex ta-siRNA is disassembled, the passenger strand is degraded, and the functional ta-siRNA strand is incorporated into AGO1 or AGO7, depending on the transcribed TAS gene and the miRNA, which are the major components of the RNA-induced silencing complex (RISC). The formation of this complex leads to the

post-transcriptional silencing of the target genes in a homology-dependent manner between the ta-siRNA and its respective target gene (Li *et al.*, 2012; Arikit *et al.*, 2013; Fei *et al.*, 2013).

1.4 TAS genes and target genes

These proteins intervening in ta-siRNA biogenesis were first discovered in *Arabidopsis thaliana* mutants by screening, since it is the model plant par excellence in plant molecular biology. For instance, four DCL proteins and 10 different AGO proteins have been identified in this species. At the same time, discovery of TAS genes has been performed in *A. thaliana*, concretely, four TAS gene families (TAS1, TAS2, TAS3 and TAS4), comprising eight loci, have been identified (Arikit *et al.*, 2013; Fei et al., 2013). With respect to miRNA recognition, miR173 targets TAS1 and TAS2 families; TAS3 family is cleaved by miR390; and TAS4 is respectively recognised by miR828 (Fei *et al.*, 2013; Zhang, 2015). As a remarkable feature, TAS3 is the unique TAS gene family containing two target sites for the miR390 (two-hit locus) and there are TAS3 homologs found in several plant species, for example, maize or grapevine (Zhang *et al.*, 2012; Fei *et al.*, 2013; Gupta *et al.*, 2017). What is more relevant, Auxin Response Factor (ARF) coding genes are frequently regulated by TAS3 loci in every plant species in which they have been identified (Tang *et al.*, 2012; Arikit *et al.*, 2013). Thus, ta-siRNAs, to the extent that they may affect ARF gene expression, seem to control developmental processes, identity of lateral organs or root development, among others (Tang *et al.*, 2012; Fei *et al.*, 2013). ARF regulation, mediated by ta-siRNAs, is conserved with Arabidopsis in many organisms, since this regulatory pathway is present in rice, maize and other economically significant crops (Arikit *et al.*, 2013).

On the other hand, the other set of TAS genes are currently less characterised and most of their putative targets have not been described thoroughly, as they have only been predicted (Fei *et al.*, 2013; Singh *et al.*, 2016). However, there are some studies clarifying the role of some TAS genes, e.g. the identification of a TAS4 gene in tomato, as well as the role of ta-siRNAs regarding mitochondrial activity control facing oxygen deficit in Arabidopsis, or also pentatricopeptide genes and MYB factors as targets (Howell *et al.*, 2007; Moldovan *et al.*, 2010; Zhang, 2015; Singh *et al.*, 2016). As a general view, ta-siRNAs regulate expression of genes involved in plant development with partial or total sequence homology, at post-transcriptional level.

1.5 Interest in agronomic species - melon

Besides Arabidopsis, it is necessary to describe ta-siRNA regulatory pathways in vegetal species with agronomic interest to achieve a deep understanding of molecular mechanisms and bear damaging environmental conditions reducing the yield of these crops. In other words, not only investigation of model species to apply this knowledge should be performed, but also research in agronomic species, since results may not be comparable between them, in addition to the fact they do not have the same economic relevance.

In this project we have decided to focus on melon (*Cucumis melo*), due to its strong presence in Spanish market and across the world, as an economically-relevant horticultural species. In terms of economy, every year encompasses an annual production around one million tonnes in Spain. As a matter of fact, provisional data from the Spanish Ministry of Agriculture, Fishing, Food and Environment (MAPAMA in Spanish) about melon yield in the country reveals that there has been a total production of 649,767 tonnes, cultivated in an extensive area of 20,686 Ha (MAPAMA, 2016). More recently, as stated by the United Nations, Spanish exportations have a trade value of more than 300 million euros and a volume of 441,293 tonnes. The most relevant importers of Spanish melons were by this order: France, with 120,115 tonnes; Germany, with 108,334 tonnes and the United Kingdom, with 57,359 tonnes (COMTRADE, 2017). This fact confers Spain the status of one of the leaders in exportation across the world, being the leader in Europe (Gonzalez-Ibeas *et al.*, 2011).

Thus, melon is a potential species to thoroughly investigate genetic mechanisms in fruit and flower development due to the vast molecular and morphological diversity occurred after different domestication events across the world in history. However, not sufficient information regarding genome sequence was available until the development of some EST collections and cDNA libraries, especially thanks to the International Cucurbit Genome Initiative (ICuGI), and the Spanish consortium Melonomics, which provides one of the best published melon genome databases. The genome of this diploid, dicotyledonous species is composed of 12 chromosomes (2n = 24) (Gonzalez-Ibeas *et al.*, 2011). This melon genome sequencing implies the creation of a technological platform that will contribute to genome-based studies and to the strengthening of Spanish trade of melon internationally (Melonomics, 2018).

Despite its agronomic relevance across the globe and the progress made in terms of genome information, the small RNA molecules of the melon transcriptome have not been properly analysed. What is more, any ta-siRNA has ever been identified in melon, which gives us a broad field to investigate. It is important to acknowledge the achieved advances in model species such as *A. thaliana* and other agriculturally-relevant crops including tomato, grape, rice or maize (Zhang *et al.*, 2012; Arikit *et al.*, 2013; Singh *et al.*, 2016; Gupta *et al.*, 2017). In a broad perspective, all of them have proven the functionality of some TAS genes, especially TAS3 family, and their activity on a few target genes that might be involved in abiotic and biotic stress tolerance. By applying a similar workflow starting from melon plants exposed to several stress conditions, we hope to report the identification of the first functional TAS genes in melon, whilst focusing on a plant species with huge impact in the international market. In this way, manipulation of the potential ta-siRNA routes will give beneficial characteristics to new cultivars of melon to adapt more rapidly to the medium-term expected environmental changes.

1.6 ta-siRNA and target prediction via bioinformatic tools

Remarkably, for an easy identificacion and prediction of sRNA, and ta-siRNA in particular, a vast range of computational algorithms and bioinformatic tools developed by consortiums, universities and research groups are available to the scientific community. One of the most recently improved software package is the UEA sRNA Workbench, which includes a number of tools to analyse, predict and detect sRNA sequences, e.g. novel miRNAs, based on the development of bioinformatic algorithms (Stocks *et al.*, 2018). Among these tools, there is a ta-siRNA prediction tool which allows the user to identify phased 21-nt long sRNAs, characteristics of TAS genes or ta-siRNA loci. The trans-acting RNA prediction tool searches for phase accumulation, size and orientation of sRNA molecules that may correspond to the typical pattern of TAS gene loci. The only requirement is the uploading of an sRNA dataset from a plant and a reference genome to anneal with. By using or changing, if desired so, the default parameters, potential ta-siRNA loci may be detected (UEA, 2018).

Heavily important as well is the psRNATarget online tool, which was developed in order to predict the targets of characterised miRNAs available in existing libraries, because of its potential role in detection of miRNAs triggering the processing of transcribed TAS gene. This bioinformatic tool developed by Dai and Zhao (2011) considers not only the degree of complementarity between the miRNA and the target, in this case the TAS-derived mRNA, but also evaluates the accessibility to the binding site, by means of the calculation of the unpaired energy (UPE), this is the amount of energy required to modify the mRNA secondary structure for the miRNA to access that mentioned site. This factor has been critically proven to be crucial for correct prediction of miRNA targets (Marín and Vaníek, 2011). The user is able to submit own sRNA libraries, target candidates or both. The uploaded library or list is subsequently compared with deposited genome databases from different species or lists of identified miRNAs in several species (Dai and Zhao, 2011; Gupta *et al.*, 2017).

Many other predictive softwares are being developed by other research groups, for example PhaseTank to identify clusters of phasiRNA and ta-siRNAs and regulatory cascades depending on them, as well as development of several algorithms based on Perl (Chen *et al.*, 2007; Guo *et al.*, 2015).

2 Objective

All things considered, the aim of this final degree project is the identification and characterisation of ta-siRNAs derived from transcription of predicted TAS genomic regions in *C. melo* via both bioinformatic tools and biochemical procedures, employing sRNA libraries and different nucleic acid amplification techniques. Predicted ta-siRNAs, processed from miRNA cuts, may be used to unveil the specific target genes of these sRNAs, whose function might be critical for the development of melon plants, flowering, fructification, and so on. In short, description of novel ta-siRNAs is destined for the discovery of the potential function that these molecules may have in terms of signalling pathway regulation, which in turn are involved in stress-response processes via differential regulation and processing under specific stress conditions of derived products from target genes, which are members of these determined pathways. The ultimate purpose would be the genetic improvement of melon varieties tolerant to stresses in order to avoid biomass losses due to the increase of harmful environmental conditions in the context of climate change and the emergence of new plant pests, in addition to outnumber the current melon production, in Spain and globally.

To accomplish this primary goal and make it feasible, exhaustive steps are sequentially undertaken, in a such a way that huge raw data from libraries of *C. melo* is analysed and used to obtain a validated ta-siRNA and a corresponding target. Thus, the following definite objectives are set, to be known:

1. Prediction and identification of functional ta-siRNA sequences via bioinformatic tools and molecular techniques, including DNA amplification and quantification of ta-siRNA molecules.

2. Identification of potential targets for predicted melon ta-siRNAs.

3. Validation of ta-siRNA/target interactions by identification (via degradome assays) of residues of tasi-processed transcripts.

4. Further validation of functional activity of predicted ta-siRNA/targets interrelations by quantification of transcript accumulation.

5. Functional validation of targets by performance of an artificial *in vivo* assay in *Nicotiana benthamiana* consisting in transient co-expression of a Cherry fluorescent protein (ChFP) fused at the 3' end to the target, and a ta-siRNA construct in the form of either an artificial miRNA (amiRNA) or a synthetic ta-siRNA (syn-tasiRNA).

3 Materials and methods

3.1 Previous work of the group to construct sRNA libraries

In order to establish the initial framework to work with, the present project is based on the development of diverse sRNA libraries which were constructed previously, at our laboratory of Dr. Gustavo Gómez in Valencia, Spain (unpublished data). The developmental process started from melon seeds (*C. melo* cv. "Piel de Sapo") which were sown and allowed to grow in the greenhouse for 11 days, under standard conditions. All individuals were split into groups destined to grow under diverse abiotic and biotic stress conditions. The applied abiotic stresses were: cold (C), drought (D), high salinity (SA) and short-day photoperiod (SD). As for the biotic stress conditions, they included: infection with the major fungal pathogen in melon, *Monosporascus cannonballus* (MON), which affects root development negatively, causing one of the main diseases in melon: root rot (Armengol *et al.*, 2011); the bacterium *Agrobacterium tumefaciens* (A), the main bacterial vector used in plant transformation events; and the viroid *Hop Stunt Viroid* (HSVd), first discovered in *Humulus lupulus* but also invading other plant species including melon (Kappagantu *et al.*, 2017). Those three pathogens were selected as biological models for their respective class. There was also a control group to compare with, composed of plants grown under adequate conditions, with no stress. Detailed growing conditions are explained in Supplementary Table S1, in the annexes section.

After exposure to the specific stress condition, leaf material was sampled from each of the replicates. Sampling took place firstly just at the moment when the stress was applied, which is set as the T0, as well as 2 days after exposition to stress, which is defined as T1. Subsequent samples were taken at T2 (4 days post-treatment application), and at T3 (11 days post-treatment). Grouping and number of replicates are depicted in the Supplementary Figure S1, in addition to the statistical correlation between those replicates, to confirm molecular differences in terms of sRNA differential accumulation. Upon sampling, total RNA was extracted, followed by an enrichment of sRNA (<200-nt long) by means of miRACLE^{TM} kit (miRNA isolation kit, Stratagene).

Production and sequencing of the libraries were carried out by Sistemas Genómicos. Ninety-six (three replicates by each treatment and exposition time, except for Agro and control, with two and four replicates respectively) cDNA libraries were obtained by following Illumina's recommendations. Briefly, 3' and 5' adaptors were sequentially ligated to the RNA prior to reverse transcription and cDNA generation. cDNAs were enriched by PCR to create the indexed double stranded cDNA library. Size selection was performed using 6 % polyacrylamide gel. The quantity of the libraries was determined by quantitative real-time PCR (qRT-PCR) in a LightCycler 480 (Roche). Prior to cluster generation in cbot (Illumina), an equimolar pooling of the libraries

was performed. The pool of the cDNA libraries was sequenced by paired-end sequencing (100 x 1) in a HiSeq 2000 (Illumina). Adaptors and low quality reads were trimmed by using the Cutadapt tool (v. 1.10) in Python. Furthermore, these sequences were assigned to a code, to be easily identified in our lab: SNCMelX, being X the numeric code given to each of the unique sRNAs. In the end, created libraries contained the entire set of 20-to-25-nt length sRNAs, whose accumulation profiles are depicted in Supplementary Figure S2.

3.2 Data filtration and prediction of TAS genes

The first step of the sequential workflow was the prediction of genomic regions coding for ta-siRNA sequences, the so-called TAS genes. For this purpose, a bioinformatic tool was required: the UEA small RNA workbench. Although the online application is available, TAS genes prediction was conducted by means of command-line interface using Rstudio. Default parameters were selected for the prediction. Those overlapping ta-siRNA loci in the same chromosome were fused together and, from that point on, referred to as the same TAS gene, so that initial and final positions of the genes were reset, obtaining longer genomic regions for the fused overlapping genes. Besides, a unique TAS ID was given to identify them easily in the libraries, with the code TAS.Cmel-####### (hereafter named TAS as abbreviation). Thus, redundant data was eliminated, remaining only unique TAS regions.

Subsequently, melon genome was downloaded from melonomics CM. 3.5. database (Melonomics, 2018) to be aligned with collapsed TAS loci, hence, assigning a specific sequence to the position-identified TAS genes, linked to their own library, sample and stress treatment they were exposed to. To further restrict the initial genomic regions, a BLAST was performed calling the Nuccore database NCBI tool, as a homology-identification study. For the results filtration, length sequence was delimited between 100 and 10,000, and key expressions were submitted for the search of only already described *viridiplantae* TAS genes, in the form of "tasX", being X the TAS gene family classified from 1 to 100, i.e. tas1, tas2, tas3... Apart from the expression "tas" other similar words were used to broaden the search and include differently-identified TAS loci, namely, "tasirna", "ta-sirna" or "Trans-acting".

In addition, genomic DNA coding for ribosomal subunits was also discarded by using the "Grepl" command, which filters out certain expressions, in this case related to ribosome, appearing in the description of the DNA pieces, i.e. rRNA, 28S, 5.8S, 18S, 5S, ribosome and ribosomal, with at least 50 nt with complete homology with ribosomal DNA. Same criteria, employing the "Grepl" command to discard false candidates, was applied to other non-relevant parts of the genome, such as whole chromosomes, Bacterial Artificial Chromosomes, genomic scaffolds, sequenced, unassigned sequenced fragments and complete genomes. Best homology result according to bit score was added to the description of predicted TAS loci. Last limiting factors for

extensive filtration were selection of those TAS regions coding for 5 different sRNAs in tandem, at least, and actively transcribed in one of the directions (positive or negative).

3.3 Annealing of melon genome with sRNA libraries

Aiming at complementing the TAS genes selection, graphics for each of the initially predicted TAS loci were plot using "Bowtie" and "Sam Tools" commands in Rstudio from alignment files at T1, T2 and T3, in such a way that forward and reverse strands from the genes were paired with the transcribed sRNAs found in the libraries, with the permission of one mismatch. Finally, genome coverage bed was tested, this is, the sRNA accumulation or count number in every sample per each nucleotide, so that a functional transcription profile was obtained. There was a final normalisation of data considering the library size to allow real comparison between libraries, and a transformation of count number to RPM, by means of "Scale" command.

3.4 Identification of TAS transcripts and target cleavages by miRNAs and ta-siRNAs

For the prediction of the candidate miRNA which sets the start site for TAS transcript processing, psRNATarget tool was employed. For this purpose, both the sequence of the five predicted TAS genes and a list of 101 differentially-expressed melon miRNAs described at our laboratory were submitted for the search. The sequences of TAS genes 915, 917, 408, 498 and 959 were downloaded from Nuccore NCBI database, specifying the species *C. melo* and the position on each chromosome where they have been predicted. Extra 200 bp region upstream and downstream of the TAS gene boundaries was downloaded as well, since the recognition site for the miRNA is expected to be identified within 100 bp upstream of the beginning of the TAS region. Furthermore, both the sense and antisense strands were submitted to psRNATarget in order to verify which is the transcribed strand from which the functional ta-siRNA phase is expressed. Scoring conditions were set as the default standards using the Schema V1 (2011), except for the expectation, setting the threshold value at 7.0, instead of the standard (3.0), in order to be less restrictive for candidate TAS prediction.

Applicability of psRNATarget encompassed prediction of targets degradation events by ta-siRNAs. The sequence of identified TAS 915 and TAS 917 was uploaded to the programme to be aligned with both ICuGI and Melonomics melon transcript databases, contained in psRNATarget, setting a maximum acceptable expectation for scheme V1 at 3.0 and V2 at 5.0 in order to limit the excessive release of target candidates. Emerged results were compiled into a unique file, ordered by expectation and classified by TAS gene.

3.5 PCR from genomic DNA

Total genomic DNA extraction was performed, recovering leaf material from different control melon plants, following the protocol developed by Dellaporta *et al.* (1983). After quantification of extracted DNA with NanoDrop® ND-1000 (Thermo Fisher Scientific), a PCR reaction was conducted to specifically amplify selected TAS genes. Primer design for each genomic region is detailed in Table 1. The entire set of primers was design to amplify the entire TAS including the miRNA recognition site. PCR was performed in a thermocycler LifeECO TC-96/G/H(b)C (Hangzhou Bioer Technology Co. Ltd.). Specifically, each PCR reaction was prepared for a final volume of 20 μL, composed of: 1 μL of genomic DNA, 5 μL of each primer (forward and reverse at 1 μM), 2 μL of buffer 10X (final concentration of 1X), 0.5 μL of dNTPs mix (1 μM), 0.5 μL of Platinum^{TM} Taq DNA Polymerase from Thermo Fisher Scientific (final concentration of 1 U/μL) and 6 μL of sterile water. Cycle conditions were set as follows: Denaturation at 95 °C for 20 secs, annealing at 60 °C for 30 secs, and extension at 72 °C for 30 secs; preceded by an initial step of 95 °C for 2 min and a final extension step at 72 °C for 5 min.

Table 1: ***Primer sequences employed in PCR from genomic DNA to amplify TAS genes, along with the melting temperature (T_m).***

TAS gene	Primer name	Primer sequence ($5' \rightarrow 3'$)	T_m (°C)
408	TAS408_F	AGGGCCTGGAGACAATACCTAC	58.1
	TAS408_R	CCCTGTCGCTCACTATACTTGTCG	58.8
498	TAS498_F	TCAAAGCAAGATAAGATTGTCTGCCATC	57.4
	TAS498_R	CAGACTAAAAAAGAACCCGAAGCCAT	57.5
915	TAS915_F	TCAAAGCAAGATAAGATTGTCTGCCATC	57.6
	TAS915_R	CAGACTAAAAAAGAACCCGAAGCCAT	57.2
917	TAS917_F	TCAAAGCAAGATAAGATTGTCTGCCATC	60.0
	TAS917_R	CAGACTAAAAAAGAACCCGAAGCCAT	58.3
959	TAS959_F	TCAAAGCAAGATAAGATTGTCTGCCATC	58.2
	TAS959_R	CAGACTAAAAAAGAACCCGAAGCCAT	57.9

For the visualisation of PCR results, an electrophoresis procedure was performed in a 2% agarose gel, mixing 10 μL of PCR product and 2 μL of loading buffer in each well, applying a voltage of 120 V for 25 min. The marker ladder used was "100 bp DNA ladder" (Thermo Fisher Scientific).

3.6 Sequencing analysis

Amplified DNA fragments corresponding to TAS genes were sent to the Sequencing Service in order to verify the identity of the sequence of each TAS locus, expected to be identical or quite similar to the one uploaded to NCBI database. Only 3 μL of PCR product were required as

well as 1.5 μL of water and 1 μL of ExoSAP-IT^{TM} PCR Product Cleanup Reagent from Thermo Fisher Scientific. The latter reagent is composed of exonucleases and dephosphorylases and it is added to the treatment to hydrolyse excess nucleotides and primers and prepare samples for sequencing, since the enzymes in the mix are inactivated. The reaction took 15 min at 37 °C and 15 min at 80 °C to complete.

3.7 RT-PCR assay

Total RNA was extracted as is described above, with the creation of sRNA libraries. A mix of samples from all stress conditions was used as the starting material for the initial reverse transcription (RT) reaction, which was performed by using the kit RevertAid First Strand cDNA Synthesis Kit (Thermo Fisher Scientific). Each PCR reaction for every TAS locus was prepared for a final volume of 20 μL, composed of: 3 μL of synthesised cDNA, 5 μL of each primer (forward and reverse at 1 μM), 2 μL of buffer Taq 10X (final concentration of 1X), 0.5 μL of dNTPs mix (1 μM), 0.5 μL of Platinum^{TM} Taq DNA Polymerase from Thermo Fisher Scientific (1 U/μL) and 6 μL of sterile water. The temperature cycles are equivalent to those of PCR from genomic DNA assay. BLAST tool from NCBI was also used to anneal each TAS actual sequence with melon genome, so as to check specificity of DNA and cDNA amplification.

3.8 qRT-PCR and degradome analysis for TAS transcripts and targets

Real time quantitative PCR was performed to quantify accumulation of transcripts related to TAS genes 915 and 917, and their corresponding targets. The first step was the synthesis of cDNA from total extracted RNA following the same protocol as RT-PCR, upon the construction of sRNA libraries. In this case, a mixture of all replicates exposed to each stress condition was prepared, to perform the reverse transcription reaction per each stress, so that control and treated samples were quantified independently at every time (T1, T2 and T3). The qRT-PCR analysis was based on extractions where sRNAs, and to the extent ta-siRNAs, were differentially accumulated in such a way that differences in transcript accumulation levels may be correlated to target differential concentration, with reasonably opposite behaviour.

Target genes investigated in this study were identified by means of bioinformatic algorithms analysing the degradome libraries, which were constructed following the protocol described in Zhai *et al.* (2014) with minor modifications. Briefly, a 5' RNA oligonucleotide adaptor was ligated to poly(A+) RNA molecules purified from total RNA, and the product was subsequently reversed transcribed using a dT primer. After cDNA amplification, 3' adaptors were further ligated to cDNA molecules, and specific size fragments were selected to fulfil Illumina's recommendations for sequencing.

For this assay, reactions were performed by triplicate in a thermocycler ABI 7500 Fast-Real Time qPCR, from Applied Biosystems. Moreover, the enzymatic mix PyroTaq EvaGreen qPCR Mix Plus (ROX) from Cultek Molecular Bioline was used, noting that this particular pyroTaq DNA polymerase requires a hot start, i.e. an incubation for 15 min at 95 °C for activation. Hence, standard conditions for SYBR Green as dye were set, other than 5 min longer initial denaturation. On the plus side, primers for TAS 915 transcripts amplification were TAS915_F, listed in Table 1, and TAS915_R1, with sequence CAACTCTTCAAAAATCCCATCACCATCG, and a melting temperature (T_m) of 58.2 °C. In addition, for TAS 917 transcripts, TAS917_F (see Table 1) and TAS917_R2 primers (sequence: GAGAAATAGAAACGGAAATAAGGAAG and a T_m of 52 °C) were synthesised. Primers detailed in Table 2 were designed for targets amplification, aiming specifically at the 21-nt long ta-siRNA recognition site. All resulting amplicons were from 120- to 180-bp long. Each component for qPCR was calculated for a final volume of 10 μL, namely: 2 μL of synthesised cDNA (using a 1:40 dilution from the original solution), 0.6 μL of each primer (forward and reverse at 1 μM), 2 μL of pyroTaq mix (1X), and 4.8 μL of miliQ water.

Table 2: ***List of primers utilised for qRT-PCR assays concerning ta-siRNA targets.*** *For ARF3 variants, there was a consensus sequence flanking miRNA cleavage site able to be used for primer design. For ARF3 and ARF4, two pairs were designed, one for each target site.*

TAS gene	Target	Primer name	Primer sequence ($5' \rightarrow 3'$)	T_m (°C)
915	Glucosyltransferase	GCT_F	CAGAATGGGGAATCGGAATGG	55.8
		GCT_R	CCGCATCGGCCATCAATTTC	57.2
	Pentatricopeptide	PTR_F	TTGGTGAGTGAATTAGGACTTGAG	54.6
		PTR_R	GGCTTATATGGCCAACCAAAAAC	55.0
917	ARF3	A3_F	CCCACTTATCAACTCCGAAACTTC	55.6
		A3_R	CCACGGCCCTACCATAATAGG	57.0
		A3_F1	CAGAATTGGGCTCTCTTCCAC	55.6
		A3_R1	ATTGTCACCGTCAGGGGGAG	59.3
	ARF4	A4_F	CTCTTCTTTCACAGGCTTTATGG	53.7
		A4_R	GAACTTTCCCCAAACACCTAAAC	54.3
		A4_F1	GGGGCGGTTTTATGGACTTTG	56.7
		A4_R1	TGCAGAAGATCGCACATCAAAC	56.0

The qPCR reaction efficiency was determined by the standard curve method, for which 1:10 serial dilutions of cDNA mix were prepared from treated and non-treated samples. Relative gene expression levels were calculated by means of the $\Delta\Delta$ CT comparative method, this is, they were normalised to the profile of profilin mRNA (NM_001297545.1), which is considered as a constitutive expression gene, as well as internally normalised to the non-treated samples. Statistical significance of results was evaluated by means of a paired T-test.

qRT-PCR data of ta-siRNAs was further supported by a differential expression analysis. Using the sRNA libraries massive sequencing files, the Rstudio command "EdgeR" was applied to obtain the logarithmic fold change (in base 2) of specific ta-siRNA sequences by normalising sample expression levels of every particular ta-siRNA to the equivalent level in control plants.

3.9 *In vivo* validation assay

To finally validate the functional activity of TAS 917 and the identity of ARFs as its targets, an *in vivo* transient expression assay was devised, to overexpress artificially in *N. benthamiana* the proper cassettes containing the potential TAS 917 functional phase and the target recognition site fused to the 3' end of the Cherry fluorescent protein.

3.9.1 Binary plasmids

First, the required overexpressing plasmids, suitable for agroinfiltration, were designed using Snapgene. Predicted TAS 917 21-nt functional phase was cloned into synthetic backbones, one as amiRNA, and another as syn-tasiRNA, which were provided by Alberto Carbonell's lab at IBMCP, València. The former permits the processing as a TAS, and thus the generation of the correspondent ta-siRNAs when co-expressed with *A. thaliana* miR173. These three constructs were cloned separately into the binary plasmid pCMD32. Moreover, the predicted target site (21-nt long) was fused to the 3' end of the sequence of ChFP and cloned into pMOG 800 under the control of the constitutive promoter CaMV 35S.

3.9.2 Plant materials and agroinfiltration

Subsequently, *N. benthamiana* leaves of 3-4 weeks old plants were agroinfiltrated with the pertinent cultures of *A. tumefaciens* strain C58C1, previously transformed with the desired construct. The overnight grown bacterial culture was diluted in infiltration buffer (MES 0.1 M, $MgCl_2$ 1 M) up to an optical density at 600 nm (OD600) of 0.2 for the ChFP carrying the target sequence, and 0.5 for the amiRNA, syn-tasiRNA and miR173 constructs, and injected on the abaxial side of the leaves using a 1 mL needle-less syringe. Agroinfiltrated plants were grown in a phytotron and analysed two days post-agroinfiltration. The phytotron growing conditions were as follow: photoperiod of 16 hours under visible light (wavelength between 400-700 nm) with an irradiance of 65-85 $\mu mol \cdot m^{-2}s^1$ and 8 hours of darkness; temperature cycles of 25 °C (light) and 18 °C (darkness) and relative humidity of 60-65 % (light) and 95-100 % (darkness).

3.9.3 Confocal microscopy

ChFP fluorescence was observed using inverted Zeiss LSM 780 confocal microscope and ZEN software (Carl Zeiss). Leaf dishes were cut and mounted in water. Then, images were acquired using an objective plan-apochromat 40x/1.4 Oil DIC M27 (0.5 cm diameter).

4 Results and Discussion

4.1 Prediction and selection of candidate TAS genes

The extensive task of discovering novel ta-siRNAs in melon relied on the previous construction of sRNA libraries in order to extract from them the potential sRNA to be considered as plausible ta-siRNA. Each library was classified according to the time at which RNA extraction was performed (T0, T1, T2, T3) and the stress the plants were exposed to, namely: cold (C), drought (D), salinity (SA), short day (SD), *M. cannonballus* (MON), *Hop Stunt Viroid* (HSVd) and *A. tumefaciens* (A), in addition to non-treated melon plants. Once sRNA libraries became available for the present project, the workflow to be followed was designed to ultimately identify functional ta-siRNAs as well as their targets, as shown in Figure 2.

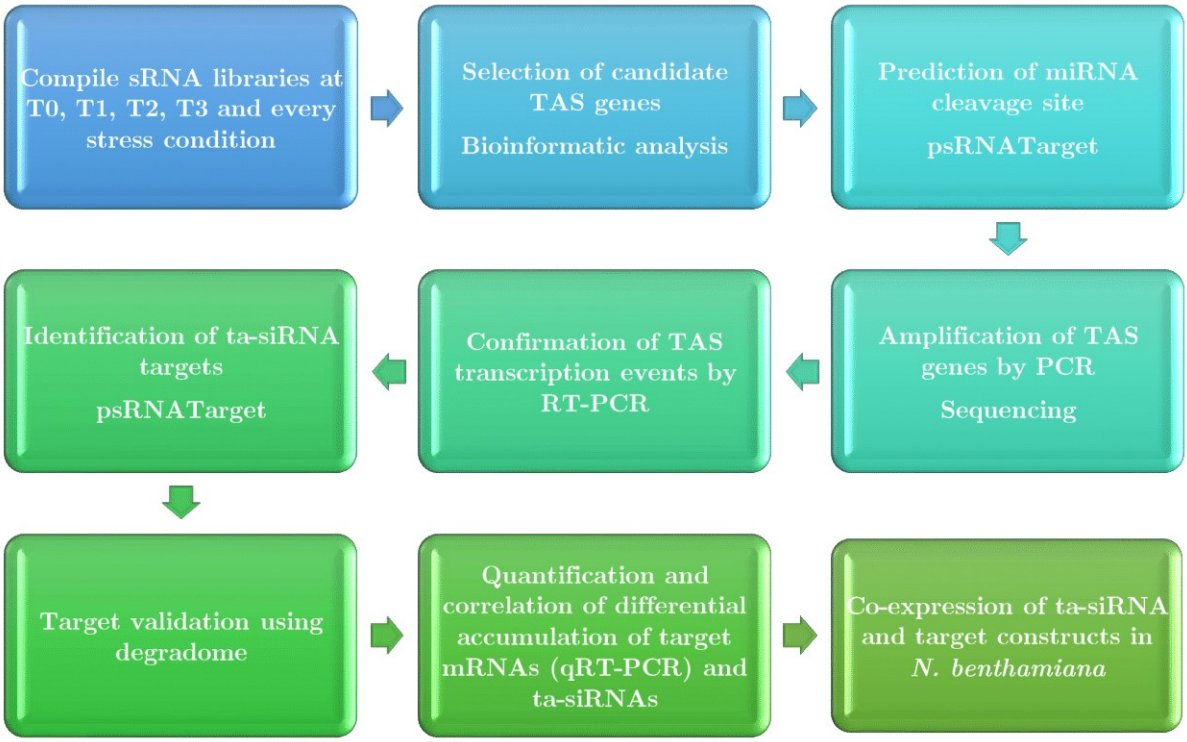


Figure 2: ***Flow diagram of the followed protocol for a positive identification and validation of critical ta-siRNAs and their corresponding targets.***

Upon compilation of every sRNA library, there was a subsequent prediction of candidate TAS genes by means of the ta-siRNA predictive algorithm of the software provided by UEA small RNA workbench. More than 15,000 candidate TAS genes were obtained from sRNAs contained in libraries. Then, collapsing and data filtration was performed to reduce the data set, foster selection of adequate TAS candidates, and avoid discrepancies in TAS gene prediction since transcribed subregions differing in length corresponding to the same TAS region exist in all 96

libraries, with differences in abundance. A table was generated containing all 15,000 TAS genes, associated to their position in the corresponding chromosome (250 to 300 bp length each, on average), not to mention the grouping of non-classified TAS genes into an artificial chromosome, i.e. chromosome 0.

The next step in TAS gene prediction was the annealing with melon genome. From the initial 15,000 TAS genes, only 1,411 were successfully associated to a sequence in the melon genome. Furthermore, BLAST resulted in 194 positively identified regions as candidate TAS genes, based on homology with TAS orthologs. Nonetheless, homology is not the most significant criterion to select a predicted TAS gene as functional in *C. melo*, since they may not be conserved along evolution, and not being actual orthologous genes, or the coding sequence may have mutated, and therefore, the target. Additionally, most of deposited TAS genes in databases like NCBI were just predicted and not validated, which make the result of homology studies more unreliable to solely select candidate TAS genes based on homology (Fei *et al.*, 2013). Among the 1,411 TAS genes, those regions containing descriptions such as Ribosomal DNA, Bacterial Artificial Chromosomes, genomic scaffolds and whole chromosomes were discarded due to their irrelevant value to the project, because none of these regions are assumed as able to express any ta-siRNA.

After this extra selection criterion, only 965 TAS genes were still considered for the analysis. Applying the rest of filtration criteria, this are, consideration of TAS regions expressing 5 different sRNAs in tandem, at least, and actively transcribed in one of the directions (positive or negative), 242 putative TAS genes were selected, each of which allows the expression of differentially accumulated sRNAs, with variable number of counts.

Among 242 candidate TAS genes, five TAS were finally selected to undergo the entire validation procedure, according to specific parameters defined from sRNA libraries data. The five final candidate TAS genes as well as analysed parameters on which they were selected, are represented in Table 3, with every TAS gene around 250- to 330-bp long. Ordered by relevance, those parameters were: (1) good differential accumulation of 21-nt ta-siRNAs based on the alignment graphics; (2) “count sequence” parameter, or number of sRNA copies per sRNA molecule, i.e. all sequences divided by the number of unique sRNAs; (3) “different sequence" parameter, i.e. the average sequence number per stress, which is related to the amount of different transcribed sRNAs differing in sequence; (4) “phase sequence” parameter, i.e. the sum of all sRNA found in phase, (5) length of alignment, and (6) percentage of homology with acknowledged TAS regions.

It is not unusual or unreasonable that one TAS locus registered a higher value for “different sequence” than “count sequence”, because of a single TAS gene may express many different ta-siRNAs being expressed under any stress condition presumably, so that each of them is not really abundant in every stressed plant, but the sum of all ta-siRNAs implies that “different sequence”

value is quite elevated. On the other hand, "count sequence" might be high when there is a massive accumulation of actively transcribed ta-siRNAs, depending on the type of stress, which is the case for the TAS 915 with a punctuation of 405.34 for this latter parameter.

According to selection criteria, the most probable candidates to be TAS genes seem to be both TAS 915 and TAS 917, which are predicted to belong to TAS3 family, in concordance to the homology with two organisms, *Solanum demissum* Lindl. and *Oryza sativa* L., respectively. Particularly, both have the best accumulation profiles among all the 242 candidate TAS genes, especially TAS 915, whose potential functional phase reaches an accumulation level up to 4,000 RPM is transcribed from the sense strand.

Table 3: ***Selection of 5 predicted TAS genes from the initial pool of 242 filtered loci.*** *Factors in bold are the key parameters why a specific TAS was selected. Explained data accounts for "count sequence", "different sequence" and "phase sequence" parameters, the chromosome where the TAS is located (chr), and the best homology result considering both high length and low number of mismatches or gaps (TAS homologs). TAS3 homologs are the best candidates to be validated as TAS genes.*

TAS gene (TAS.Cmel)	Chr	Count sequence	Different sequence	Phase sequence	Homology		TAS homolog
					Length	%	
0408	5	**23.46**	**26.50**	7.00	18	88.9	*A. thaliana* TAS1
0498	6	2.33	17.00	6.00	53	75.5	*A. thaliana* TAS1
0915	11	**405.34**	**40.57**	**8.79**	**101**	73.3	*S. demissum* TAS3
0917	11	18.16	**32.69**	**8.71**	44	97.7	*O. sativa* TAS3
0959	11	4.50	3.00	3.00	47	84.5	*A. thaliana* TAS2

These graphics obtained from genome alignment with sRNAs contained in libraries are depicted in Figure 3. Moreover, for both TAS there are two cleavage sites recognised by miRNA, and thus, two active phases since there are two peaks in 915 and 917, as seen in Figure 3. Differential accumulation depends on the cleavage efficiency of miRNA. In the case of 915, cold is the most prominent stress causing an upregulation of the TAS gene, of its antisense strand, whereas TAS 917 is mostly upregulated under short-day conditions, for both strands. Observing differential accumulation profile from T1 to T3 (not shown here), an evolutionary pattern is noticed, starting from basal transcription in control conditions, resulting in the maximum level of TAS transcription at T3, demonstrating that the longer the application of stress, the more regulatory effect is obtained in TAS transcription, either up or down regulation. Apart from differential accumulation and "count sequence", significant "different sequence" parameters were predicted for both TAS 915 and 917, meaning that, there are 40.57 and 32.69 distinct sequences on average, per stress for 915 and 917, respectively. Those parameters were relevant for the selection of these predicted TAS as well as the number of phased RNAs (9 per TAS gene roughly).

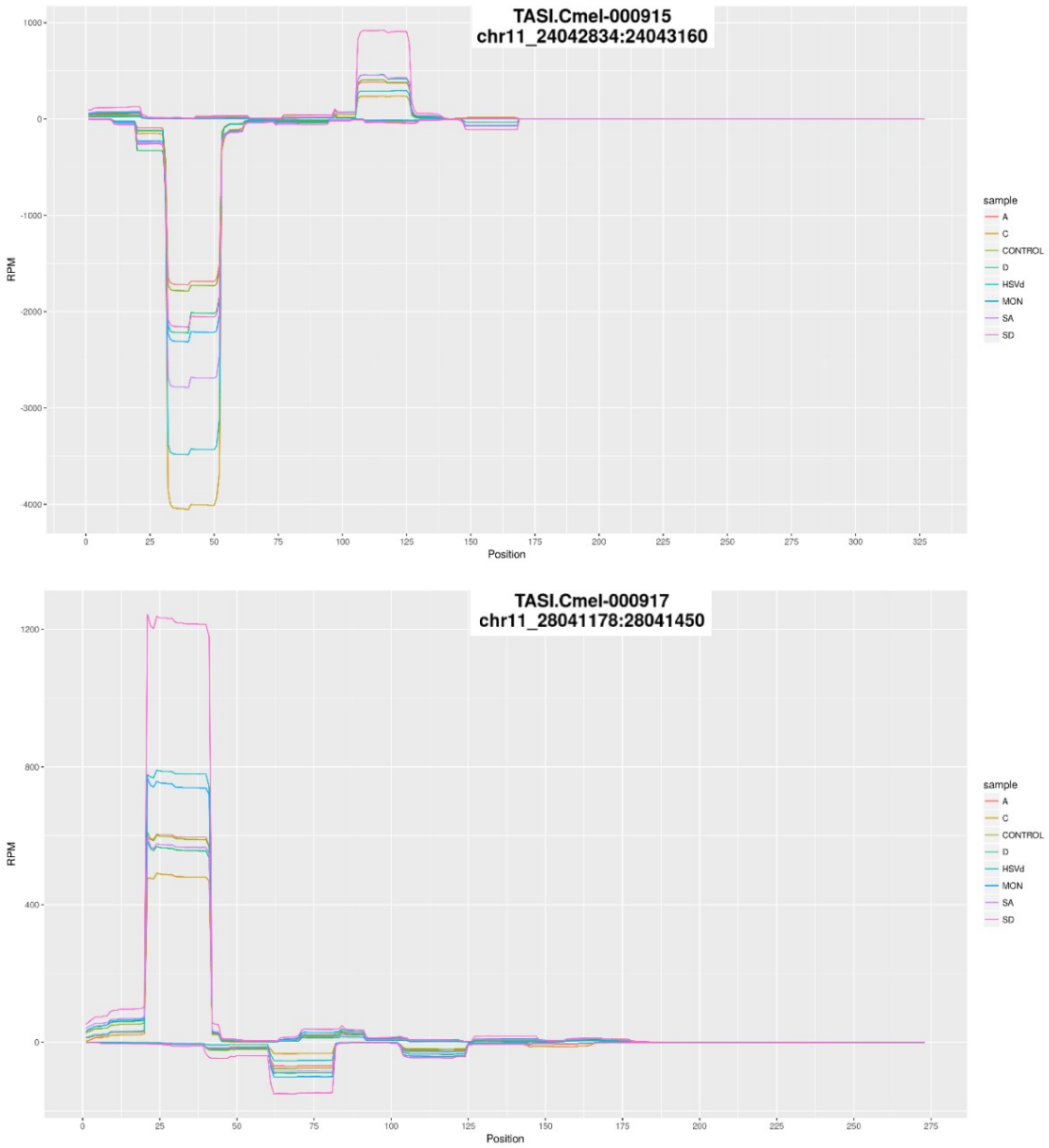


Figure 3: ***Differential accumulation profiles of transcribed TAS genes 915 and 917 at T3.*** *Upper graph depicts the transcription profile of TAS 915, whereas the bottom one, the TAS 917 accumulation profile. Positive side of the Y axis corresponds to the forward, or sense strand of the TAS gene, whereas the negative side, to the reverse strand or antisense. Peaks represent the most accumulated phase, which might be the functional phase loaded onto AGO, presumably. Slight variations in phase accumulations are due to frame change of DCL4. Legend: A - A.tumefaciens, C - cold, D - drought, HSVd - Hop Stunt Viroid, MON - M. cannonballus, SA - salinity, SD - short day.*

Next chosen TAS regions were 408 and 498, due to their relevant "different sequence" punctuations: 26.50 for TAS 408 and 17.00 for TAS 498, respectively, and their good accumulation profile at T3, so that we may characterise two possible members of the TAS1 family, according to the homology result. On the other hand, the choice of TAS 959 serves as a check point to validate the usefulness of our criteria-method procedure for TAS selection, since 959 presents worse results in every parameter compared to the others. In other words, TAS 959 may be used as a negative control.

4.2 miRNA 390 intervenes in processing of TAS3 transcripts in melon

By means of psRNATarget software we were able to identify the miRNAs involved in TAS transcript downstream processing of the five selected TAS genes. Complete set of data regarding psRNATarget results is explained in Table 4. We consider miRNA cleavage prediction to be reliable for an expectation lower than 5.0. In any case, two cleavage sites are noticed, one of them with a lower expectation, i.e. a higher efficacy of cleavage. This may occur due to the fact that only one miRNA-guided cleavage event is necessary for DCL4 to process some TAS transcript into phased, mature sRNAs. Thus, for TAS 408 and 959, the cleavage site with lower expectation is the one where the specific miRNA hybridises, whereas the other site is less probable to be recognised by any miRNA. Despite the requirement of only one recognition site, in the case of TAS3 family, they are described to contain two cleavage sites, recognised by the same miRNA, usually miR390 (Fei *et al.*, 2013), which is evident at the sight of Table 4, showing excellent fitting for TAS 915 and 917 for miR390 hybridisation, with expectations lower than 3.0 and 4.0 for the prominent site, respectively.

Table 4: ***Summary of psRNATarget results regarding miRNA targeting different TAS transcripts.*** *The number of cleavage sites is referred to as the number of sites at the TAS gene recognised by one single miRNA.*

TAS gene (TAS.Cmel)	Predicted miRNA	Expectation	Strand orientation	Number of cleavage sites
0408	miR396e/miR156j	5.0/6.0	Antisense	1
0498	miR393a/b	4.5/6.5	Sense	2
0915	**miR390**	**2.0/2.5**	Antisense	2
0917	**miR390**	**3.5/4.5**	Sense	2
0959	miR397b/miR156g	6.0/6.5	Antisense	1

More specifically, TAS3 genes are characterised to be processed from the 3' end, so that miR390 recognises more frequently the second site, at 3' end, which is observable in Table 4 as well, because the lower expectation for TAS 915 (2.0) correspond to that second site at 3'

boundaries of TAS gene (Fei *et al.*, 2013; Montgomery *et al.*, 2008). Although the expectation of the cut site at 3' for TAS 917 is not the lowest, 4.5 is still a good expectation result. In concert with prediction analysis, TAS 959 is the only TAS gene with both expectations for the two miRNA sites higher than 5.0, hence it fails to provide acceptable miRNA-cleavage predicted sites. All in all, TAS 959 might have emerged as a false TAS gene. To fully prove this hypothesis, a RT-PCR was performed to check the accumulation of TAS 959 transcripts.

4.3 Transcription events confirm the functionality of TAS 915 and TAS 917

Genomic PCR of the five TAS genes permitted to verify the existence of those genomic regions, because expected-length DNA was amplified as a result. Thus, designed primers for TAS 915 were able to amplify a fragment of 421 bp, as calculated; TAS 917 reaction resulted in a 502 bp band; a fragment of 833 bp was obtained for TAS 408; a band of 500 bp was correspondent to TAS 498 and TAS 959 reaction gave rise to a band of 511-bp long. Gel images depicting band fragments are represented in Supplementary Figure S3, in the annexes. Validation of PCR results was given by sequencing analysis, since the sequence of amplified fragments corresponding to TAS genes were exactly equivalent to the ones predicted and downloaded from genome database. Furthermore, in view of RT-PCR involving amplification of transcribed genes, we were able to finally discard the TAS 959 for the downstream analysis, since it was demonstrated it was not behaving as an actual TAS gene, considering the lack of transcription activity, in spite of repeated RT-PCR with stringent conditions and with different sets of primers. The RT-PCRs were performed utilising a mix of RNA from all stress conditions in different proportions. Therefore, since TAS 959 was not validated as a functional TAS gene, we may assert that this putative TAS region is not a real TAS gene, as expected. Data from scoring parameters in Tables 3 and 4 also indicate the adequate consideration of TAS 959 as a negative control for TAS validation via bioinformatic analysis. Electrophoresis of RT-PCR products is also depicted in Supplementary Figure S3.

Despite of being transcribed, according to RT-PCR results, TAS 408 and 498 were also rejected from the analysis since they present 99% homology with other genome regions, so that we cannot confirm the results from RT-PCR are valid since transcribed RNA may have been transcribed from the predicted TAS or from other regions. At this point, only TAS 915 and TAS 917 were still candidates to be functional TAS genes, which is confirmed by good scoring points in every molecular and bioinformatic evidences.

4.4 Glucosyltransferases and pentatricopeptides as potential targets for melon TAS 915

From numerous targets predicted by psRNATarget, only four of them were selected according to degradome analysis. Degradome is constructed basically, as commented in Materials and Methods section, adding adaptors to mRNA molecules which are potentially targeted, in this case

by ta-siRNA, followed by high-throughput sequencing of the amplified fragments. This analytical method highlights the existence of some targets which present specific accumulation profiles of sequences, indicating that there is an elevated cleavage activity by ta-siRNA. Nonetheless, candidate targets are discarded occasionally by other predictive software, such as psRNATarget, since they are considered less likely to be targets, and thus scored with a high expectation. Hence, degradome analysis may be one of the most reliable evidences to identify real ta-siRNA targets, despite offering an unfavourable expectation result (Li *et al.*, 2012). Particularly, abundance of cleaved TAS 915 target sequences verified by degradome annealing is depicted in Figure 4. The identified targets are 7-deoxyloganetin glucosyltransferase (LOC103502665, NCBI) and a pentatricopeptide-repeat containing protein (LOC103492176, NCBI), both with an expectation of 4.5. On the one hand, abundance of cleaved glucosyltransferase is predominant in high salinity conditions (SA), reaching an accumulation of 20 RPMs in T3 samples, whereas under cold stress, the pentatricopeptide-containing protein seems to be actively downregulated with an abundance of 2 RPMs, much lower than the maximum of glucosyltransferase, though.

Glucosyltransferases are proteins that catalyse the glucose conjugation to metabolites in order to increase water solubility and stability. They are included within a big family of glycosyltransferase proteins, whose general function is to permit the transference of sugar moieties from one activated donor to an acceptor, allowing the formation a glycosidic bond. All proteins share a common structure in which their activity relies on, including a Rossman fold in N and C-terminal domains (Tang *et al.*, 2011; NCBI, 2018). Moreover, pentatricopeptide repeat proteins (PPRs) are composed of 35-amino acid motifs, known to interact with certain RNA molecules, acting as a scaffold for the union, via recognition of one nucleotide and binding to one repeat. A specific amino acid code has been deciphered in order to design new PPRs and to fully understand the molecular function of already characterised proteins (Miranda *et al.*, 2018).

For a further validation of targets, apart from degradome, differential analysis expression of TAS 915 and accumulation profiles of verified TAS 915 targets coming from qRT-PCR were compared to try to correlate the values to obtain a proportional linear relationship between TAS expression and target degradation under different stresses. As observed in Figure 5, the log fold change in base 2 (LFC) was calculated for every functional phase expressed in every sample, which may correspond to the peaks of accumulation profiles depicted in Figure 3. A negative correlation between the functional phase of TAS 915 differential expression and target transcripts accumulation is expected to be shown in Figure 5. For the sense strand the functional predicted phase was: TTTGAACAATTTTCGTGGACA, as for the antisense strand: CTTGGGCTCTTCTAATCATTT, which is indeed the most accumulated by far, and thus, the more relevant plausibly. ta-siRNA sequence in red in degradome for TAS 915 is the functional phase from antisense strand as well.

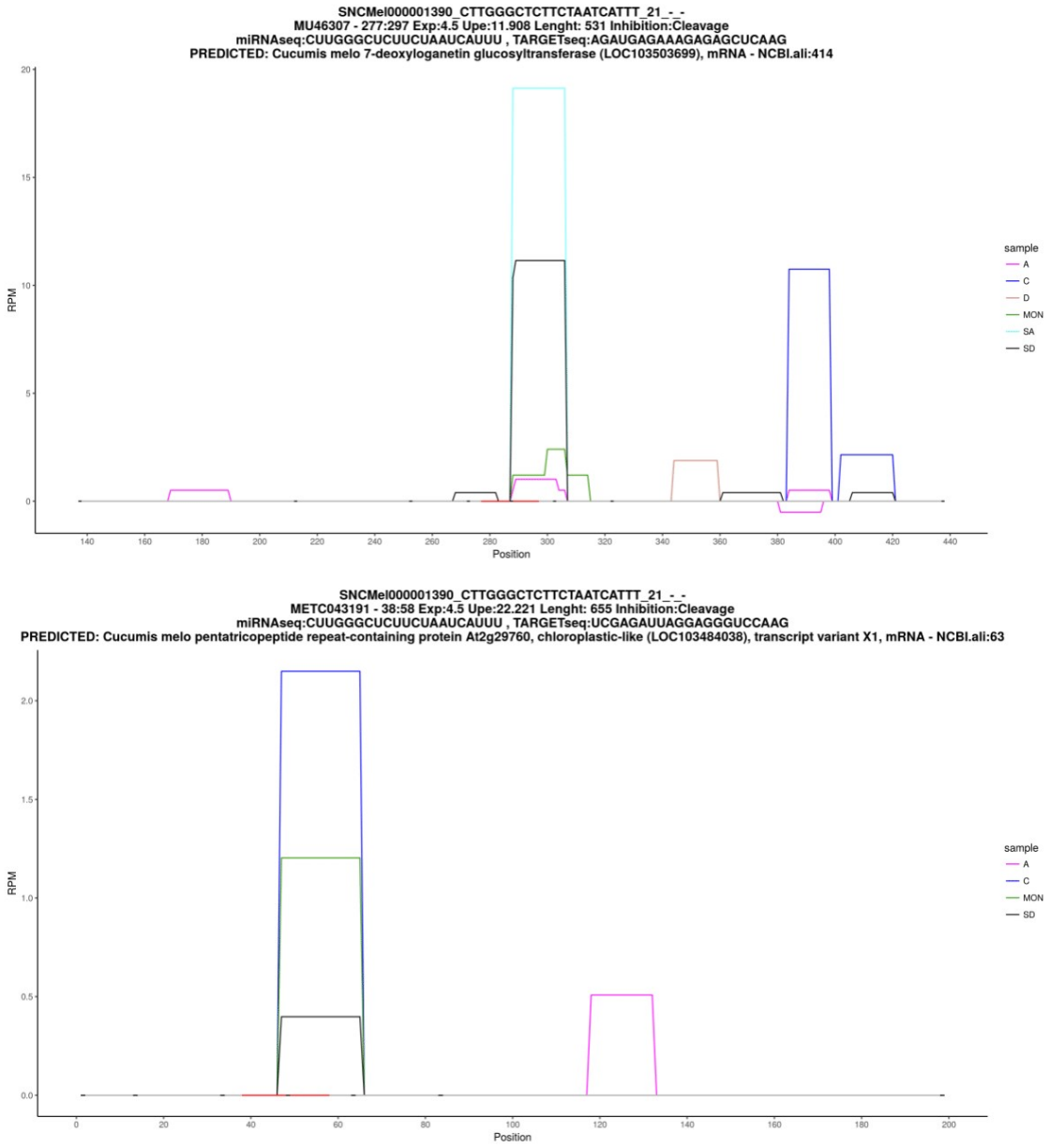


Figure 4: ***Degradome profile of TAS 915 targets glucosyltransferase and pentatricopeptide (PPR), at T3.*** *Glucosyltransferase degradome is shown in the upper image, whereas the bottom one depicts the PPR degradome. Only 150 bp up and downstream from the ta-siRNA recognition site of the target mRNA are represented on the X axis. Sequence of functional ta-siRNA derived from TAS 915 is highlighted in red, as well as the cleavage site in the middle as a black point. Abundance of sequences is expressed in RPMs. Legend: A - A.tumefaciens, C - cold, D - drought, MON - M. cannonballus, SA - salinity, SD - short day.*

TAS 915		
Samples	**Sense strand LFC**	**Antisense strand LFC**
Control	1.00	1.00
C - Cold	**0.29**	**2.48**
D - Drought	0.82	1.91
SA - Salinity	0.83	1.46
SD - Short Day	1.90	1.76
MON - *M. cannonballus*	0.82	1.78
HSVd - *Hop Stunt Viroid*	**0.29**	**2.06**
A - *A. tumefaciens*	0.82	**2.13**

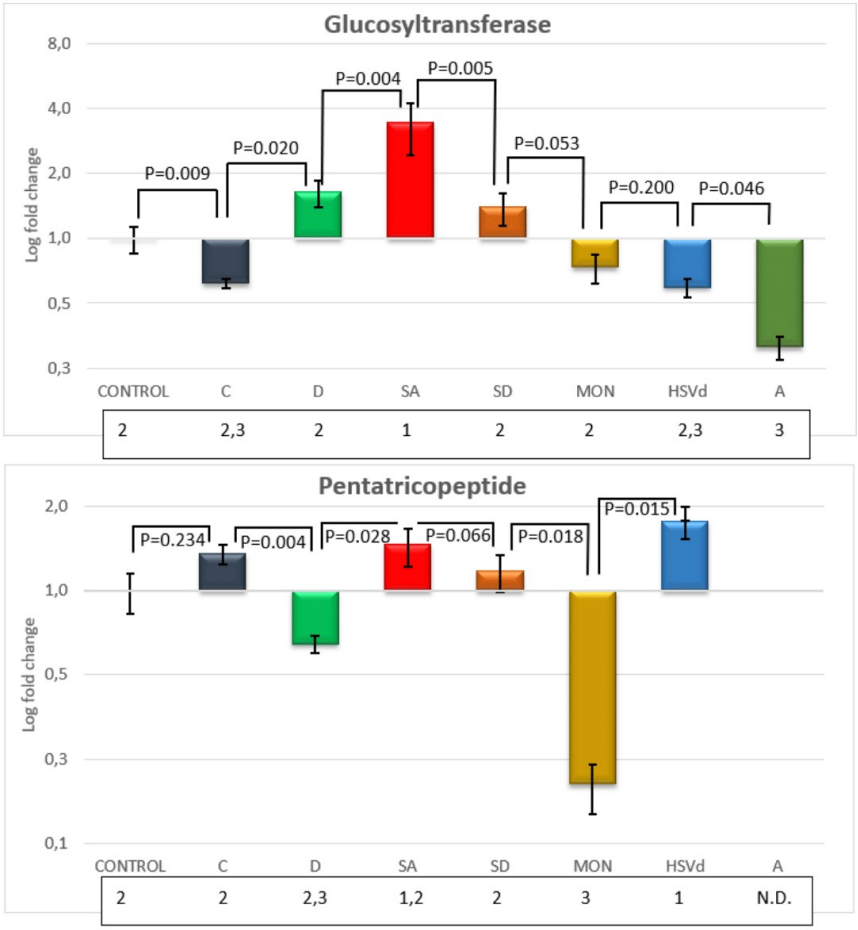


Figure 5: ***Relative TAS 915 transcripts accumulation in terms of log fold change (LFC) according to sRNA libraries, and its corresponding targets glucosyltransferase and pentatricopetide (PPR) by qRT-PCR at T3.*** *Significant LFC factors for TAS transcripts accumulation (upper table) are highlighted in red for downregulation (LFC < 0.5) and in green for upregulation (LFC > 2.0). The accumulation of PPR transcripts in sample A was not determined (N.D.) due to negligible concentration. Sample classification into statistically differentiated groups is shown below X-axis legend, assigning the number 1 to the sample with the highest expression level, and subsequent distinct groups in descendent order. Statistical significance was given by the paired T-test between every pair of samples. The p-value of T-tests of only adjacent samples is depicted for each target.*

With the control sample as the reference, the transcript emerged from antisense strand of TAS 915 is always upregulated under any stress condition, according to Figure 5. Cold condition is the stress in which the RNA phase is accumulated the most (LFC = 2.48), followed by *A. tumefaciens* sample (LFC = 2.13) and HSVd (LFC = 2.06). We expected that accumulation of target genes transcripts may be negatively correlated to the antisense phase, since it is the most abundant, and not to the sense phase. This hypothesis fits with the degradome of PPR in Figure 4, which is more degraded in cold conditions, coincident with the upregulation of TAS 915 antisense strand.

M. cannonballus and HSVd samples present upregulation of that functional phase which also implies a logical increase of degraded pentatricopeptide target (LFC = 1.78, and 1.76 respectively). Glucosyltransferase, however, does not behave similarly, since the degradation pattern does not correspond to differential expression analysis. The only possible correlated result might involve SD samples if the functional phase affecting the target was the one transcribed from the sense strand with an LFC of 1.90 and 10 RPMs of accumulated, degraded glucosyltransferase.

The correlation with qRT-PCR amplification cycles was even more challenging. There was an easily observed behaviour involving C, HSVd and A samples, because they appear to be the most downregulated in qRT-PCR for glucosyltransferase, so that they were grouped into the third statistical class of the Figure 5. Precisely, under these three stress conditions, the differential expression assay shows the highest upregulation level of TAS 915 antisense strand, as commented above. Opposite to what was expected, under high salinity (SA) the accumulation of glucosyltransferase transcripts appears to rise (qRT-PCR) along with an increase in degradation. In this case, there might be other regulative ta-siRNAs or intermediate molecules that interfere with the overall, final accumulation level. Nevertheless, qRT-PCR should be considered until a certain extent, because we believe degradome to be more reliable since data is coming from rounds of high-throughput sequencing. On the contrary, since the qRT-PCR was performed with a sparse number of replicates (2 or 3) and samples, hence it is more likely to obtain a p-value greater than 0.05, as well as the unspecific amplifications that may occur. Also, there might be other events on the cell which are affecting the overall accumulation of transcripts, independently of ta-siRNA cut efficiency, such as mediation of intermediate messengers, or regulation by secondary pathways. Hence, the validation of ta-siRNA targets via degradome could prevail the qRT-PCR results, if not coincident in some cases.

Further support to downregulation of glucosyltransferase-like proteins upon infection with virus or bacteria, in a similar way to *Hop Stunt Viroid* and *A. tumefaciens,* is provided by the work of Gómez-Aix *et al.* (2016). In this project, the *cytokinin-O-glucosyltransferase2* gene was detected to be completely deregulated in the *Cucumber mosaic virus* (CMV), *Watermelon mosaic virus* (WMV) and *Melon necrotic spot virus* (MNSV) treated melon plants. In other

plants, like tomato, other glucosyltransferases collaborate in the transport and synthesis of other cytokines, e.g. the zeatin O-glucosyltransferase-like (Ye *et al.*, 2018).

With regard to pentatricopeptide repeats, there have been evidences proving that PPRs are intermediates in ta-siRNA generative pathways, hence, they are involved in ta-siRNA production and could be considered secondary TAS genes (Chen *et al.*, 2007). In a general way, TAS1 and TAS2 families may derive from PPR genes, which are targeted by miR173 in Arabidopsis, and whose 3'-cleaved products are PPRs and act as ta-siRNAs. It has also been proposed that they have a dual role in protein production and triggering of target degradation via siRNA generation, in such a way that pentatricopeptides might be cleaved via miRNA recognition and be the targets of primary ta-siRNAs. This allows the formation of new ta-siRNAs from PPR transcripts, which could be related to our pentatricopeptide TAS 915 target (Chen *et al.*, 2007; Guillaumot *et al.*, 2017). Not only PPR genes are firm candidates to be TAS genes, but they also intervene in general RNA editing and RNA processing. For instance, there are PPR proteins catalysing the deamination of cytosines into uracil in mitochondria and chloroplasts, e.g. protein partners of plastidial CLB19 PPR editing factor, and many other gene expression processes in those organelles at the post-transcriptional level, such as translational activation, intron splicing, and so on (Howell *et al.*, 2007; Colcombet *et al.*, 2013; Guillaumot *et al.*, 2017; Miranda *et al.*, 2018).

4.5 Auxin Response Factors as potential targets for melon TAS 917

Similarly to 915 targets, full validation of TAS 917 targets was accomplished by means of degradome analysis, qRT-PCR comparative assay and correlation with differential accumulation levels of TAS transcripts, related to TAS gene expression. Results of alignment of ta-siRNA phases from TAS 917 with predicted target genes, i.e. degradome obtention, are shown in Figure 6. Here, two acknowledged ta-siRNA targets were verified, namely, Auxin Response Factor 3 (LOC103484487, NCBI) and Auxin Response Factor 4 (LOC103501935, NCBI), both with an acceptable expectation of 2.5, provided by psRNATarget tool. Both ARF targets gave rise to excellent degradome profiles. In the case of ARF3, degradation frequency is increased in drought conditions (RPMs = 2.0), and to a lesser extent in *M. cannonballus*-treated samples and short-day conditions. On the contrary, ARF4 is only targeted under cold stress, with a perfect alignment of functional phase of ta-siRNA with target recognition site.

The most common Auxin Response Factors include ARF2, ARF3 and ARF4, all of them being upregulated in the presence of auxins. These hormones play a critical role in morphophysiological plant processes, especially development of lateral roots, other lateral organs and shoots (Montgomery *et al.*, 2008; Marin *et al.*, 2010). Establishment of leaf polarity, or anther development in cold stress in wheat are other processes in which ARFs are involved in (Montgomery *et al.*,

2008; Tang *et al.*, 2012).

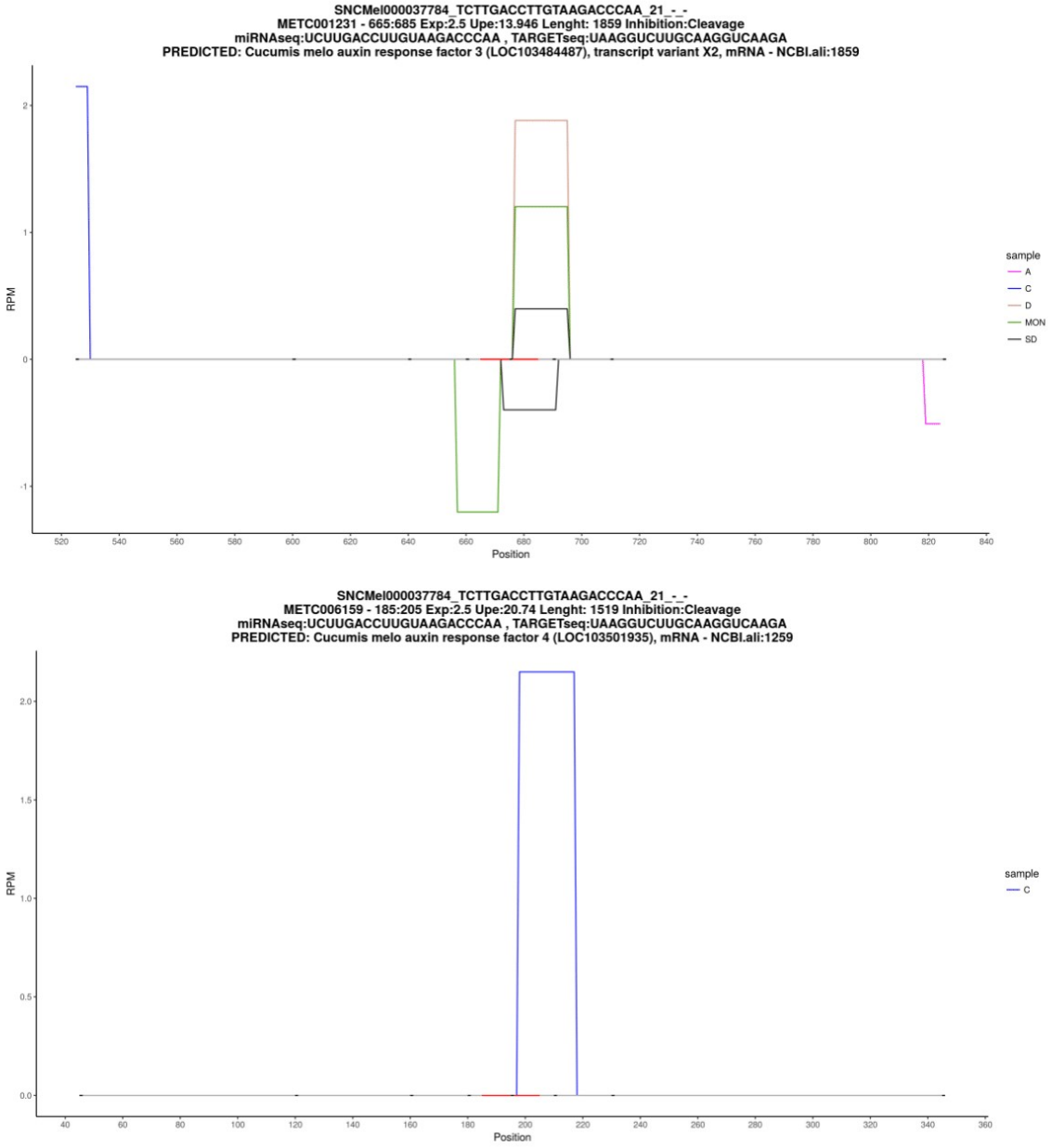


Figure 6: ***Degradome profile of TAS 917 targets Auxin Response Factor 3 and Auxin Response Factor 4 at T3.*** *ARF3 degradome is shown in the upper image, whereas the bottom one depicts the ARF4 degradome. Only 150 bp up and downstream from the ta-siRNA recognition site of the target mRNA are represented on the X axis. Sequence of functional ta-siRNA derived from TAS 917 is highlighted in red, as well as the cleavage site in the middle as a black point. Abundance of sequences is expressed in RPMs. Legend: A - A.tumefaciens, C - cold, D - drought, MON - M. cannonballus, SD - short day.*

Auxin treatment also induce brassinosteroid expression mediated by ARFs in Arabidopsis and rice (Sakamoto and Fujioka, 2013). In relationship to our case study, ta-siRNAs arising from TAS3 genes, more importantly identified in Arabidopsis, negatively regulate the activity of ARFs (Fei *et al.*, 2013). At structural level, ARF3 and ARF4 consist of different amino acid sequences, as a consequence of the different coding sequence for the totality of the gene length. However, the ta-siRNA recognition and cleavage site are 100% conserved, being the only equivalent part between these proteins. The degree of homology was evaluated with BLAST tool from NCBI. Moreover, none of the recognition sites is corresponding to any miRNA sequence at miRbase database, so that this site is unique for ta-siRNAs.

Applying the same procedure as for TAS 915, log fold change (base 2) was calculated in Figure 7 for every functional phase expressed in every sample, corresponding to functional phases of peaks in Figure 3 for TAS 917. A negative correlation between the predicted functional phase of TAS 917 differential expression and target transcripts accumulation is expected to be shown in Figure 7. The analysed ta-siRNA phases were: TCTTGACCTTGTAAGACCCAA for the sense strand, and CATCCATTTGATTTCATACTC for the antisense one. Since the sense strand, according to Figure 3 is the utmost accumulated, it is the more probable to be functional and actively cleaving target transcripts, hence, it was the preferred for degradome alignment and performance of differential expression assays.

Focusing on LFCs of sense strands (Figure 7, upper subfigure), plants submitted to severe cold temperatures show an intense drop in TAS 917 transcription rate, extrapolating from low accumulation profiles, as proven by the extremely low LFC of 0.05, which in turn is correlated to the abundance of ARF3 and ARF4 in assayed samples by qRT-PCR under cold conditions. Therefore, C sample is assigned to the first statistical group solely in both ARF3 and ARF4 (LFC equal to 9.23 and 14.30, respectively). Although qRT-PCR results offer information regarding amplification of the first cleavage site, other reactions were performed to amplify the second site, close to 3' ends of ARFs, and similar tendencies were obtained approximately. The significance of downregulation of TAS 917 in cold was more evident in ARF4, which was the target gene with the best degradome, not surprisingly, depicting a strong ta-siRNA regulation upon exposure to low temperatures. Looking at the rest of samples, no evident correlation was discovered between differential expression of TAS 917 and the accumulation of neither the sense phase nor the antisense, although the major ARFs upregulation event is found in plants grown under drought conditions. The general tendency, though, is that under stress conditions the accumulation of ARF3 and ARF4 is greater than control plants, which does not correlate with upregulation of TAS 917 under any stress, except for HSVd and A samples. This fact might be due to translation inhibition by TAS transcripts and mRNA accumulation. In addition, there might be other regulatory pathways affecting the overall ta-siRNA 917 interference.

TAS 917		
Sample	**Sense strand LFC**	**Antisense strand LFC**
Control	1.00	1.00
C - Cold	**0.05**	0.85
D - Drought	1.39	0.86
SA - Salinity	1.40	0.76
SD - Short Day	1.93	**2.76**
MON – *M. cannonballus*	1.33	1.27
HSVd - *Hop Stunt Viroid*	0.60	**0.50**
A – *A. tumefaciens*	0.85	1.43

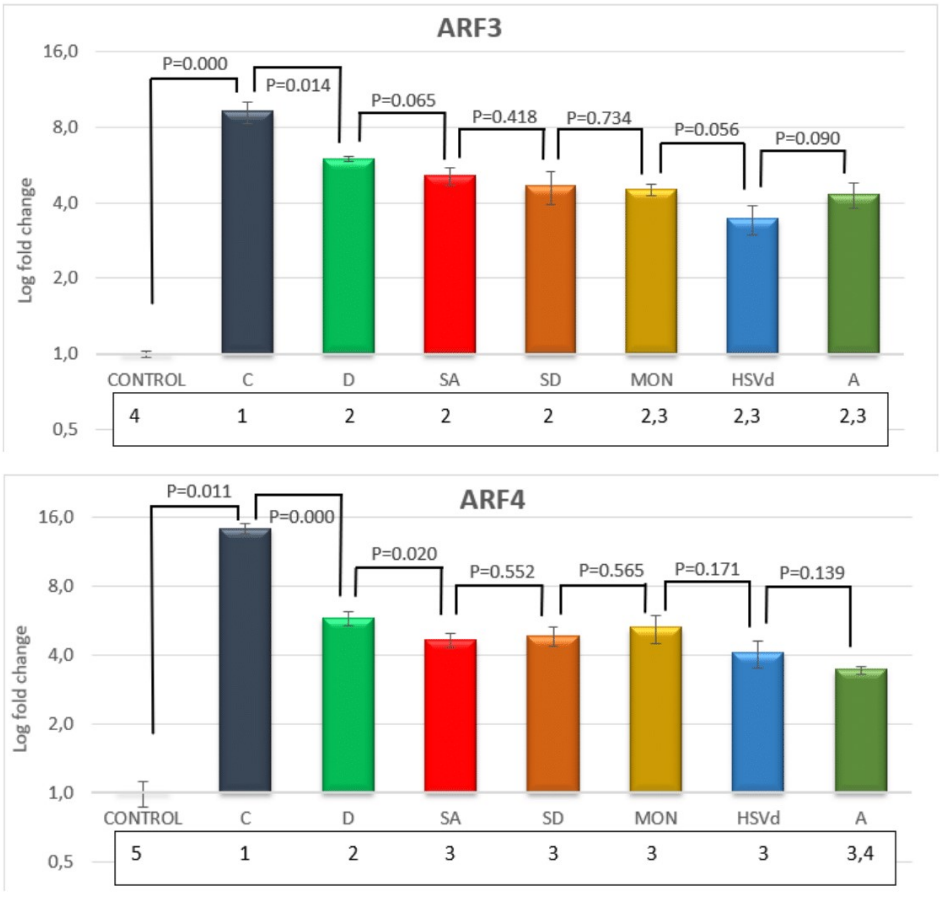


Figure 7: ***Relative TAS 917 transcripts accumulation in terms of log fold change (LFC) according to sRNA libraries, and its corresponding targets ARF3 and ARF4 by qRT-PCR of the first recognition sequence at T3.*** *A negative correlation between TAS differential expression and target transcripts accumulation is expected. Significant LFC factors for TAS transcripts accumulation (upper table) are highlighted in red for downregulation (LFC < 0.5) and in green for upregulation (LFC > 2.0). Sample classification into statistically differentiated groups is shown below X-axis legend, assigning the number 1 to the sample with the highest expression level, and subsequent distinct groups in descendent order. Statistical significance was given by the non-paired T-test between every pair of samples. The p-value of T-tests of adjacent samples is depicted for each target.*

TAS 917 was predicted as a member of the TAS3 gene family according to homology analysis. Since TAS 915 was also predicted as a TAS3 gene, we searched for any ARF homologous sequence within the TAS gene to be in phase. Strikingly, we found that by manual phasing in TAS 915, there was a functional phase targeting ARFs, corresponding to the sixth phase from the miRNA site at the 3' end. References in sRNA libraries accounting for the presence of this phase reveal the poor accumulation of the new phase (less than 4 counts at the T3, control library), even though the expectation to cleave ARF3 and ARF4 is outstanding (1.0).

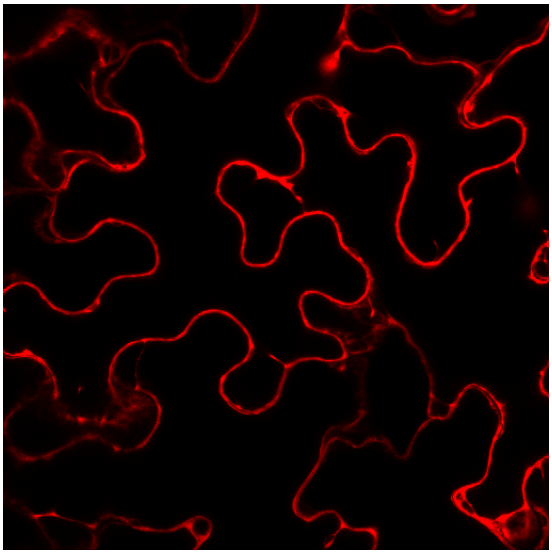
On the other side, two variations of the newly identified phase in TAS 915 were found in TAS 917, differing in number of thymines at the 5' end of the phase. One of them was not found in libraries, whereas the other also improves the cleavage expectation to 0.5 of both ARFs and is abundantly accumulated in libraries. Anyway, it was not predicted as functional by all bioinformatic tools employed. All in all, this implies the followed downstream procedure to validate ta-siRNAs was adequate.

As prospective perspectives, despite the positive identification of functional TAS genes in melon, further research should be performed to find a balance between bioinformatic prediction and biological validation procedures, since both assays may give rise to misleading results, hence, huge efforts should be made to confirm TAS activity using both methods, and to clarify other identified phases, which are not validated yet.

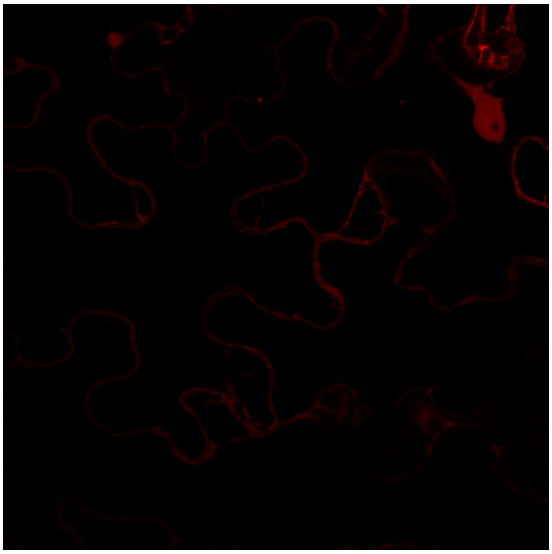
4.6 *In vivo* validation assay confirms ARFs as TAS 917 targets

Functional activity of TAS 917 was validated by an *in vivo* assay which consisted in transiently overexpressing both the functional phase from TAS 917 and the predicted target recognition site fused to the 3' end of a fluorescent protein, in this case ChFP. This assay was performed in *N. benthamiana*, due to the impossibility of detecting expression of the fluorescent protein in *C. melo* leaves. Thus, it is a heterologous assay in which a decrease in fluorescence can be related to the activity of the overexpressed TAS, corroborating its functionality. Overexpression of TAS 917 involves engineering an artificial microRNA (amiRNA) and a synthetic trans-acting small interfering RNA (syn-tasiRNA), which might silence a transcript carrying the target recognition site, if loaded onto AGO (Carbonell and Daròs, 2017). Images taken at the confocal microscope are depicted in Figure 8.

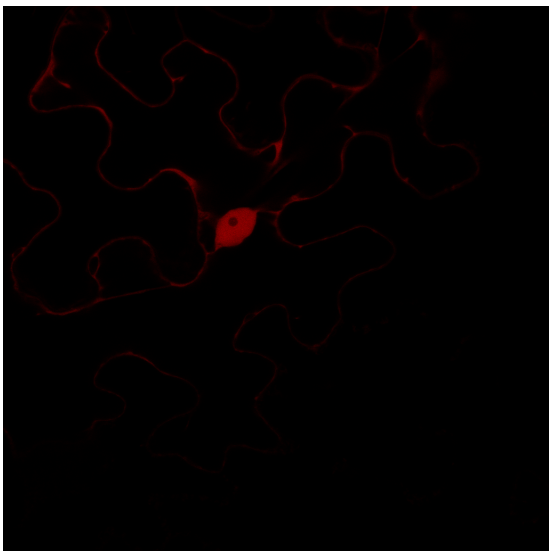
In order to compare properly the decrease in fluorescence levels, one of *N. benthamiana* leaves was agroinfiltrated with a solution containing only the ChFP fused at the 3' end to the target site (subfigure A from Figure 8). According to this image, the *in vivo* experiment was correctly designed and the plant was highly expressing the ChFP-fused target construct, since strong red



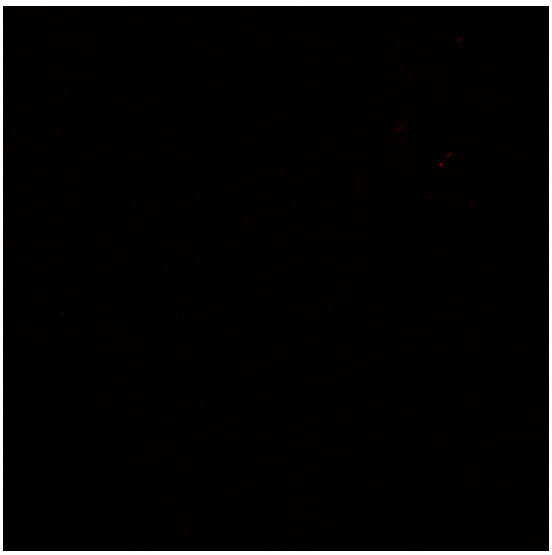
(a) *Control: Expression of ChFP-fused ARF target site.*



(b) *Co-expression of ChFP-fused ARF target site and amiRNA from TAS 917.*



(c) *Co-expression of ChFP-fused ARF target site and syn-tasiRNA from TAS 917.*



(d) *Expression of amiRNA from TAS 917 one day prior to agroinfiltration of ChFP-fused ARF target site.*

Figure 8: ***Display of the ChFP fluorescence levels under confocal microscope emitted by the ChFP fused to the ARF target site sequence.*** *A significant reduction in ChFP fluorescence is expected between the control (subfigure A) and the co-expression of TAS and target constructs in N. benthamiana, due to the catalytic activity of ta-siRNA functional phase, in the form of amiRNA (subfigure B) or syn-tasiRNA (subfigure C). A more abrupt fluorescence decay might be expected in subfigure D, because of the early transient expression of the amiRNA construct, one day prior to agroinfiltration of ChFP-fused target construct.*

fluorescence is observed across the tissue and cells, including cytoplasm and nuclei. Hence, the control allow us to compare the levels of fluorescence of the rest of the constructs.

In regard to Figure 8 again, the co-expression of TAS and ChFP-fused constructs proved artificially the ocurrence of the cleavage of the ARF target site mediated by the ta-siRNA derived from TAS 917. A significant decrease in red fluorescence emitted by the ChFP is observed in subfigures B and C, indicating that ta-siRNA is actively degrading the target site sequence, and consequently, the ChFP fused to it. Still nuclei and nucleoli are distinguished though. Since non-relevant differences in fluorescence levels are noticed between subfigures B and C, it is concluded that the type of TAS construct, either amiRNA (subfigure B) or syn-tasiRNA (subfigure C), does not influence the cleavage efficiency, even though might be reasonable that the syn-tasiRNA could represent closely the natural biogenesis pathway of the actual ta-siRNA. To clarify the structure, the syn-tasiRNA construct consists in a series of ta-siRNA generative phases in tandem (from 5 to 6 phases), the first of which belongs to *A. thaliana* TAS1 family and it is recognised by *A. thaliana* miR173. Hence, this is the reason why the miR173 should be co-expressed with the syn-tasiRNA construct, to permit a correct processing of the syn-tasiRNA region. The second phase of the construct is the phase of interest whose potential functionality is to be demonstrated, in this case, the functional phase from TAS 917. On the contrary, amiRNA structure does not require of the miR173 to be processed, and is based on a central complementary double-stranded region, where the TAS phase is located, linked to a loop, so that it generates a miRNA-like structure, which is adequately processed in *N. benthamiana* (Carbonell and Daròs, 2017).

Moreover, there is a timing influence on efficiency of ta-siRNA-mediated cleavage of the target. The early expression of the TAS construct in *N. benthamiana*, one day prior to the infiltration of the cells containing the ChFP-fused target site construct allowed a higher expression of the first construct, this is the functional phase from TAS 917. Non-strikingly, as expected, the overexpressed TAS phase was able to cleave the practical totality of the ChFP-fused targets, leading to a massive diminution in fluorescence levels, with almost no detection of fluorescence, attending to subfigure D of Figure 8. This fact highlights the exceptional cleavage efficiency of ta-siRNA, since is able to degrade every synthesised ChFP, if fused to the ARF target site, with a time lapse of at least one day.

Ultimately, the *in vivo* assay validates ARF3 and ARF4 as TAS 917 targets, since the overexpression of the predicted functional ta-siRNA from TAS 917 implies the targeting and cleaving of ARF3 and ARF4, mediated by the recognition of the shared cleavage site by the functional ta-siRNA derived from TAS 917. Thus, in melon, an upregulation of TAS 917 under a particular stress condition leads to the degradation of ARF transcripts.

5 Conclusions

By means of bioinformatic predictive software, 242 candidate TAS genes, able to generate functional 21-nt long ta-siRNAs, were identified among the initial pool of sRNAs contained in the libraries. Five of them were selected to undergo the validation procedure, attending to sRNA accumulation profiles emerged from TAS regions, homology to other acknowledged TAS loci, and calculated parameters regarding accumulation of sRNAs transcribed from those TAS regions.

Only 2 candidates, TAS 915 and TAS 917, were especially relevant for biological validation of ta-siRNA activity, because of their outstanding scoring expectation for miR390 recognition, and accumulation profiles of ta-siRNA phases. These genes were associated to the TAS3 family. Moreover, the other 3 candidates were discarded according to amplification of TAS transcripts by RT-PCR, prediction of miRNA-mediated cleavage and homology sequence analysis.

Via degradome analysis, a glucosyltransferase and a PPR protein were predicted as potential targets for the functional ta-siRNA phase from the sense strand of TAS 915. In the case of TAS 917, the candidate identified targets were ARF3 and ARF4. By means of qRT-PCR and differential expression assays, relative accumulation of TAS transcripts under stress conditions were correlated to accumulation of those target candidates obtained by degradome analysis. For TAS 915, it was not observed any clear negative correlation between its targets and differential expression analysis, apart from the downregulation of glucosyltransferase under cold conditions and HSVd and *A. tumefaciens* infection. For the pentatricopeptide, the most clear evidence was the degradation of the target by ta-siRNA from TAS 915 under cold stress according to degradome and accumulation of TAS 915 transcripts.

However, regarding TAS 917, there were strong evidences for a negative correlation in cold stress for both ARF3 and ARF4, especially for ARF4 which was validated as a target by degradome and qRT-PCR correlation with TAS 917 transcripts differential accumulation. Furthermore, both ARF3 and ARF4 were upregulated under every stress conditions, in such a way that they behave similarly in terms of transcript accumulation. Apart from C samples, there was not observed any further clear negative correlation between targets and TAS transcripts accumulation under other stress condition, but a positive one. The *in vivo* validation assay in *N. benthamiana* allow us to definitely determine the functionality of TAS 917 and the identity of ARFs as TAS 917 targets.

In short, we can conclude that these are the first ta-siRNA targets ever identified in *C. melo*. This event may set the start point for a new field of potential regulation of key elements in stress-response processes to allow the genetic improvement of melon plants tolerant to stress conditions.

6 References

ARIKIT, S.; ZHAI, J. AND MEYERS, B.C. (2013). Biogenesis and function of rice small RNAs from non-coding RNA precursors. *Curr Opin Plant Biol* 16: 170–179.

ARMENGOL, J.; ALANIZ, S.; VICENT, A.; BELTRÁN, R.; ABAD-CAMPOS, P.; PÉREZ-SIRERA, A.; GARCÍA-JIMÉNEZ, J.; BEN SALEM, I.; SOULI, M. AND BOUGHALLEB, N. (2011) Effect of dsRNA on growth rate and reproductive potential of *Monosporascus cannonballus*. *Fungal Biol* 115: 236–244.

CARBONELL, A. AND DARÒS, J.A. (2017). Artificial microRNAs and synthetic *trans*-acting small interfering RNAs interfere with viroid infection. *Mol Plant Pathol* 18: 746-753.

CARTHEW, R.W. AND SONTHEIMER, E.J. (2009). Origins and Mechanisms of miRNAs and siRNAs. *Cell* 136: 642–655.

CHEN, H.M.; LI, Y.H. AND WU, S.H. (2007). Bioinformatic prediction and experimental validation of a microRNA-directed tandem trans-acting siRNA cascade in *Arabidopsis*. *Proc Natl Acad Sci* 104: 3318–3323.

COLCOMBET, J.; LOPEZ-OBANDO, M.; HEURTEVIN, L.; BERNARD, C.; MARTIN, K.; BERTHOMÉ, R. AND LURIN, C. (2013). Systematic study of subcellular localization of Arabidopsis PPR proteins confirms a massive targeting to organelles. *RNA Biol* 10: 1557–1575.

COMTRADE (2017). UN COMTRADE database. [online] Available at: `https://comtrade.un.org/data/` (Accessed 14 May 2018).

DAI, X. AND ZHAO, P.X. (2011). PsRNATarget: A plant small RNA target analysis server. *Nucleic Acids Res* 39: 155–159.

DELLAPORTA, S.L.; WOOD, J. AND HICKS, J.B. (1983). A plant DNA minipreparation: version II. *Plant Mol Biol Report* 1: 19–21.

DERKSEN, H.; RAMPITSCH, C. AND DAAYF, F. (2013). Signaling cross-talk in plant disease resistance. *Plant Sci* 207: 79–87.

FEI, Q.; XIA, R. AND MEYERS, B.C. (2013). Phased, Secondary, Small Interfering RNAs in Posttranscriptional Regulatory Networks. *Plant Cell* 25: 2400–2415.

FUKUDOME, A.; KANAVA, A.; EGAMI, M.; NAKAZAWA, Y.; HIRAGURI, A.; MORIYAMA, H. AND FUKUHARA, T. (2011). Specific requirement of DRB4, a dsRNA-binding protein, for the in vitro dsRNA-cleaving activity of *Arabidopsis* Dicer-like 4. *Rna* 17: 750–760.

GÓMEZ-AIX, C.; PASCUAL, L.; CAÑIZARES, J.; SÁNCHEZ-PINA, M.A. AND ARANDA, M.A. (2016). Transcriptomic profiling of *Melon necrotic spot virus*-infected melon plants revealed virus strain and plant cultivar-specific alterations. *BMC Genomics* 17: 429.

GONZALEZ-IBEAS, D.; BLANCA, J.; DONAIRE, L.; SALADIÉ, M.; MASCARELL-CREUS, A.; CANO-DELGADO, A.; GARCIA-MAS, J., LLAVE, C. AND ARANDA, M.A. (2011). Analysis of the melon (*Cucumis melo*) small RNAome by high-throughput pyrosequencing. *BMC Genomics* 12: 393.

GUILLAUMOT, D.; LOPEZ-OBANDO, M.; BAUDRY, K.; AVON, A.; RIGAILL, G.; FALCON DE LONGEVIALLE, A.; BROCHE, B.; TAKENAKA, M.; BERTHOMÉ, R.; DE JAEGER, G.; DELANNOV, E. AND LURIN, C. (2017). Two interacting PPR proteins are major Arabidopsis editing factors in plastid and mitochondria. *Proc Natl Acad Sci* 114: 8877–8882.

GUO, Q.; QU, X. AND JIN, W. (2015). PhaseTank: Genome-wide computational identification of phasiRNAs and their regulatory cascades. *Bioinformatics* 31: 284–286.

GUPTA, S.; KUMARI, M.; KUMAR, H. AND VARADWAJ, P.K. (2017). Genome-wide analysis of miRNAs and Tasi-RNAs in *Zea mays* in response to phosphate deficiency. *Funct Integr Genomics* 17: 335–351.

HOWELL, M.D.; FAHLGREN, N.; CHAPMAN, E.J.; CUMBIE, J.S.; SULLIVAN, C.M.; GIVAN, S.A.; KASSCHAU, K.D. AND CARRINGTON, J.C. (2007). Genome-Wide Analysis of the RNA-DEPENDENT RNA POLYMERASE6/DICER-LIKE4 Pathway in *Arabidopsis* Reveals Dependency on miRNA- and tasiRNA-Directed Targeting. *Plant Cell* 19: 926–942.

KAPPAGANTU, M.; VILLAMOR, D.E.V.; BULLOCK, J.M. AND EASTWELL, K.C. (2017). A rapid isothermal assay for the detection of Hop stunt viroid in hop plants (*Humulus lupulus*), and its application in disease surveys. *J Virol Methods* 245: 81–85.

KINOSHITA, T. AND SEKI, M. (2014). Epigenetic memory for stress response and adaptation in plants. *Plant Cell Physiol* 55: 1859–1863.

KUMAR, R. (2014). Role of microRNAs in biotic and abiotic stress responses in crop plants. *Appl Biochem Biotechnol* 174: 93–115.

LI, F.; ORBAN, R. AND BAKER, B. (2012). SoMART: A web server for plant miRNA, tasiRNA and target gene analysis. *Plant J* 70: 891–901.

MAPAMA (2016). Superficies y producciones anuales de cultivos. [online] Available at: `http://www.mapama.gob.es/es/estadistica/temas/estadisticas-agrarias/agricultura/superficies-producciones-anuales-cultivos/` (Accessed 13 May 2018).

MARIN, E.; JOUANNET, V.; HERZ, A.; LOKERSE, A.S.; WEIJERS, D.; VAUCHERET, H.; NUSSAUME, L.; CRESPI, M.D. AND MAIZEL, A. (2010). miR390, Arabidopsis TAS3 tasiRNAs, and Their AUXIN RESPONSE FACTOR Targets Define an Autoregulatory Network Quantitatively Regulating Lateral Root Growth. *Plant Cell* 22: 1104–1117.

MARÍN, R.M. AND VANÍEK, J. (2011). Efficient use of accessibility in microRNA target prediction. *Nucleic Acids Res* 39: 19–29.

MELONOMICS (2018). Melonomics. [online] Available at: `http://www.melonomics.net/melonomics.html#/` (Accessed 4 April 2018).

MIRANDA, R.G.; MCDERMOTT, J.J. AND BARKAN, A. (2018). RNA-binding specificity landscapes of designer pentatricopeptide repeat proteins elucidate principles of PPR–RNA interactions. *Nucleic Acids Res* 46: 2613–2623.

MOLDOVAN, D.; SPRIGGS, A.; YANG, J.; POGSON, B.J.; DENNIS, E.S. AND WILSON, I.W. (2010). Hypoxia-responsive microRNAs and *trans*-acting small interfering RNAs in *Arabidopsis*. *J Exp Bot* 61: 165–177.

MONTGOMERY, T.A.; HOWELL, M.D.; CUPERUS, J.T.; LI, D.; HANSEN, J.E.; ALEXANDER, A.L.; CHAPMAN, E.J.; FAHLGREN, N.; ALLEN, E. AND CARRINGTON, J.C. (2008). Specificity of ARGONAUTE7-miR390 Interaction and Dual Functionality in TAS3 *Trans*-Acting siRNA Formation. *Cell* 133: 128–141.

NCBI (2018). Conserved Protein Domain Family Glycosyltransferase_GTB_type. [online] Available at: `https://www.ncbi.nlm.nih.gov/Structure/cdd/cddsrv.cgi?uid=299143` (Accessed 17 May 2018).

QIN, F.; SHINOZAKI, K. AND YAMAGUCHI-SHINOZAKI, K. (2011). Achievements and challenges in understanding plant abiotic stress responses and tolerance. *Plant Cell Physiol* 52: 1569–1582.

SAKAMOTO, T. AND FUJIOKA, S. (2013). Auxins increase expression of the brassinosteroid receptor and brassinosteroid-responsive genes in *Arabidopsis*. *Plant Signal Behav* 8: e23509.

SINGH, A.; SARAF, S.; DASGUPTA, I. AND MUKHERJEE, S.K. (2016). Identification and validation of a virus-inducible ta-siRNA-generating TAS4 locus in tomato. *J Biosci* 41: 109–118.

STOCKS, M.B.; MOHORIANU, I.; BECKERS, M.; PAICU, C.; MOXON, S.; THODY, J.; DALMAY, T. AND MOULTON, V. (2018). The UEA sRNA Workbench (version 4.4): a comprehensive suite of tools for analyzing miRNAs and sRNAs. *Bioinformatics* bty338.

TANG, Q.; MA, X.; MO, C.; WILSON, I.W.; SONG, C.; ZHAO, H.; YANG, Y.; FU, W. AND QIU, D. (2011). An efficient approach to finding *Siraitia grosvenorii* triterpene biosynthetic genes by RNA-seq and digital gene expression analysis. *BMC Genomics* 12: 343.

TANG, Z.; ZHANG, L.; XU, C.; YUAN, S.; ZHANG, F.; ZHENG, Y. AND ZHAO, C. (2012). Uncovering Small RNA-Mediated Responses to Cold Stress in a Wheat Thermosensitive Genic Male-Sterile Line by Deep Sequencing. *Plant Physiol* 159: 721–738.

UEA (2018). TA-SI PREDICTION. [online] Available at: `http://srna-workbench.cmp.uea.ac.uk/ta-si-prediction-2/` (Accessed 15 May 2018).

WU, F.; CHEN, Y.; TIAN, X.; ZHU, X. AND JIN, W. (2017). Genome-wide identification and characterization of phased small interfering RNA genes in response to *Botrytis cinerea* infection in *Solanum lycopersicum*. *Sci Rep* 7: 3019.

YE, X.; FU, M.; LIU, Y.; AN, D.; ZHENG, X.; TAN, B.; LI, J.; CHENG, J.; WANG, W. AND FENG, J. (2018). Expression of grape ACS1 in tomato decreases ethylene and alters the balance between auxin and ethylene during shoot and root formation. *J Plant Physiol* 226: 154–162.

YOSHIKAWA, M.; IKI, T.; TSUTSUI, Y.; MIYASHITA, K.; POETHIG, R.S.; HABU, Y. AND ISHIKAWA, M. (2013). 3' fragment of miR173-programmed RISC-cleaved RNA is protected from degradation in a complex with RISC and SGS3. *Proc Natl Acad Sci* 110: 4117–4122.

ZHAI, J.; ARIKIT, S.; SIMON, S.A.; KINGHAM, B.F. AND MEYERS B.C. (2014). Rapid construction of parallel analysis of RNA end (PARE) libraries for Illumina sequencing. *Methods Vol* 67: 84-90.

ZHAI, J.; ZHAO, Y.; SIMON, S.A.; HUANG, S.; PETSCH, K.; ARIKIT, S.; PILLAY, M.; JI, L.; XIE, M.; CAO, X.; YU, B.; TIMMERMANS, M.; YANG, B.; CHEN, X. AND MEYERS, B.C. (2013). Plant MicroRNAs Display Differential 3' Truncation and Tailing Modifications That Are ARGONAUTE1 Dependent and Conserved Across Species. *Plant Cell* 25: 2417–2428.

ZHANG, B. (2015). MicroRNA: A new target for improving plant tolerance to abiotic stress. *J Exp Bot* 66: 1749–1761.

ZHANG, C.; LI, G.; WANG, J. AND FANG, J. (2012). Identification of *trans*-acting siRNAs and their regulatory cascades in grapevine. *Bioinformatics* 28: 2561–2568.

ZHANG, Y.; LUBBERSTEDT, T. AND XU, M. (2013). The Genetic and Molecular Basis of Plant Resistance to Pathogens. *J Genet Genomics* 40: 23–35.

ZHU, J.K. (2016). Abiotic Stress Signaling and Responses in Plants. *Cell* 167: 313–324.

7 Annexes

Table S1: ***Growing conditions of melon plants for libraries development.*** *Prior to stress exposure, seeds were sown and pre-germinated at 37 °C for 48 h, then germinated 24 h at 25 °C (16:8 photoperiod) and followed by a growth stage for 10 days at 28 °C/ 20 °C with 16 h light/8 h dark.*

Stress condition	Photoperiod (light/dark)	Temperature (light/dark)	Observations
C	16 h / 8 h	20 °C / 14 °C	Lower temperatures
D	16 h / 8 h	28 °C / 20 °C	No irrigation
SA	16 h / 8 h	28 °C / 20 °C	Irrigation solution with 200 mM LiCl
SD	8 h / 16 h	28 °C / 20 °C	Short day photoperiod
A	16 h / 8 h	28 °C / 20 °C	Agroinfiltration at 0.8 OD600, two cotyledons
HSVd	16 h / 8 h	28 °C / 20 °C	Two cotyledons infected with viroidal RNA
MON	16 h / 8 h	28 °C / 20 °C	50 mL solution with 1,000 CFU micellium from *M. cannonballus*
CONTROL	16 h / 8 h	28 °C / 20 °C	-

Time (dpi) / Stress condition	T1	T2	T3
CONTROL	CONTROL.1, CONTROL.2, CONTROL.3	CONTROL.1, CONTROL.2, CONTROL.3	CONTROL.2, CONTROL.3, CONTROL.4, CONTROL.5
A	A.1, A.2, A.3	A.1, A.2, A.3	A.2, A.3
C	C.1, C.2, C.3	C.1, C.2, C.3	C.1, C.2, C.3
D	D.1, D.2, D.3	D.1, D.2, D.3	D.1, D.2, D.3
HSVd	HSVd.1, HSVd.2, HSVd.3	HSVd.1, HSVd.2, HSVd.3	HSVd.1, HSVd.2, HSVd.3
MON	MON.1, MON.2, MON.3	MON.1, MON.2, MON.3	MON.1, MON.2, MON.3
SA	SA.1, SA.2, SA.3	SA.1, SA.2, SA.3	SA.1, SA.2, SA.3
SD	SD.1, SD.2, SD.3	SD.1, SD.2, SD.3	SD.1, SD.2, SD.3

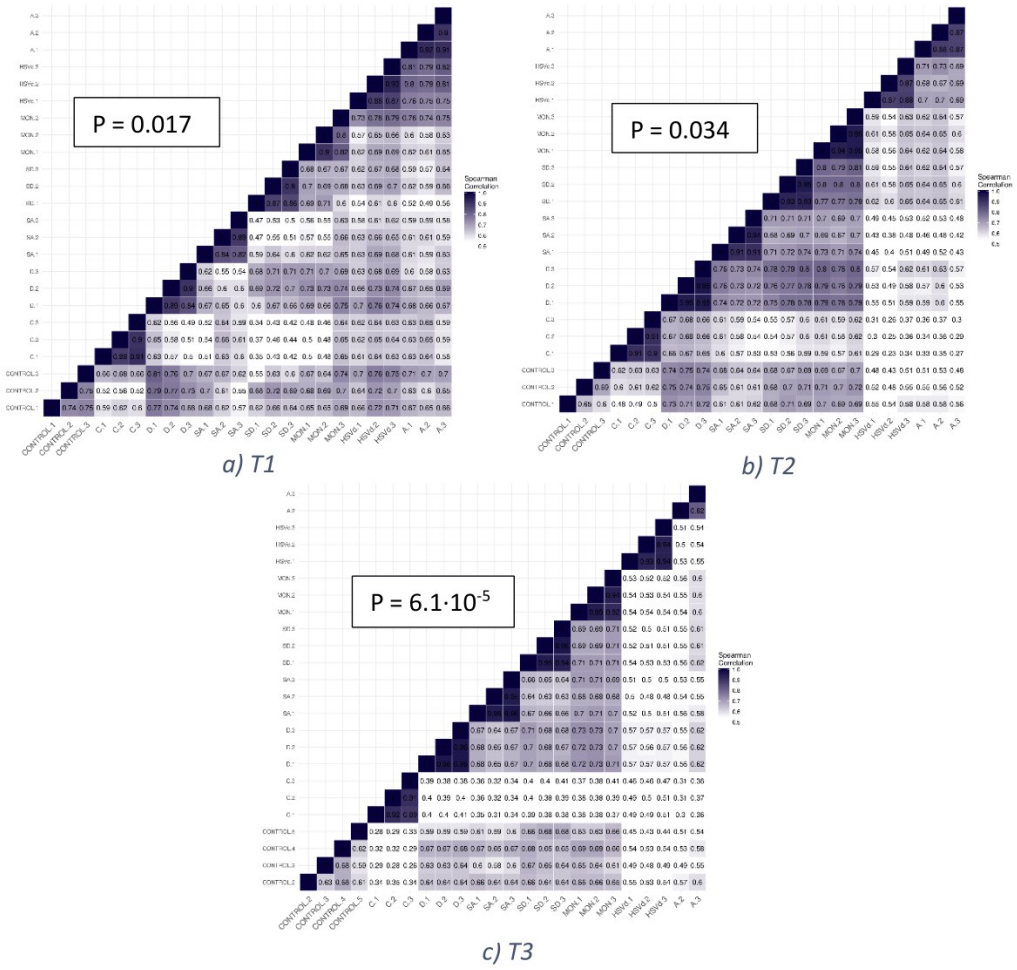


Figure S1: ***Replicates of samples from melon plants subjected to stress conditions and their corresponding statistical correlation between them.*** *Control plants and stress-treated plants are named in the upper table, whereas Spearman's correlation between all the replicates is depicted in image A (for T1), image B (for T2) and image C (for T3). High correlation is expected between replicates exposed to the same stress condition, i.e. the darker blue squares. P expresses the p-value of the null hypothesis, which is the probability of any of the samples to be statistically different.*

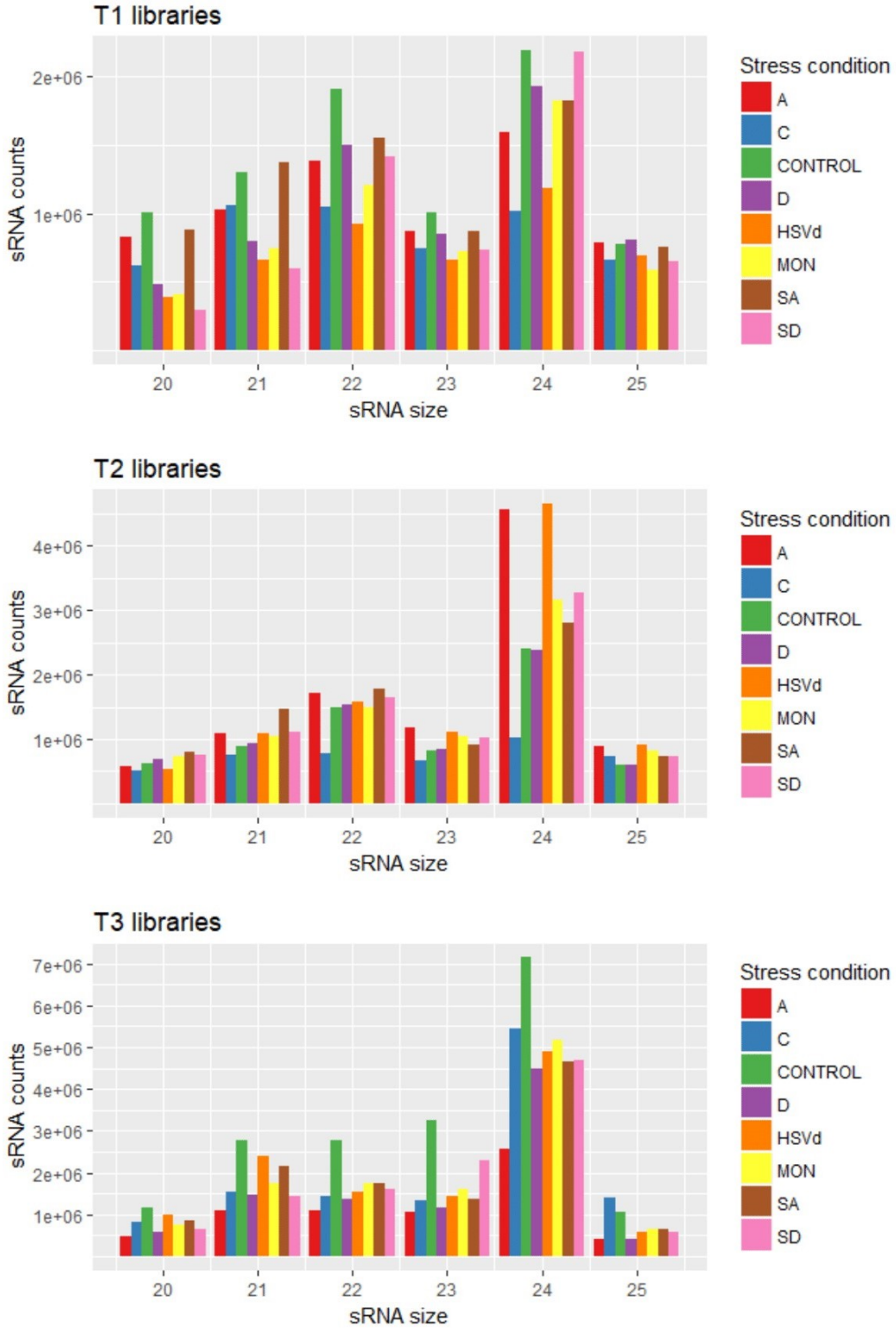


Figure S2: ***Accumulation profiles of unique sRNA counts identified in all the 96 constructed libraries, at the three different sampling times (T1, T2, T3), and classified by length for the control and exposed-to-stress samples.*** *These stress conditions are: A. tumefaciens (A), cold (C), drought (D), Hop Stunt Viroid (HSVd), M. cannonballus (MON), salinity (SA) and short day (SD). As shown in the picture, the number of sRNA counts increases, especially the 24-nt sRNAs, as the time passes since the initial application of the stress treatment.*

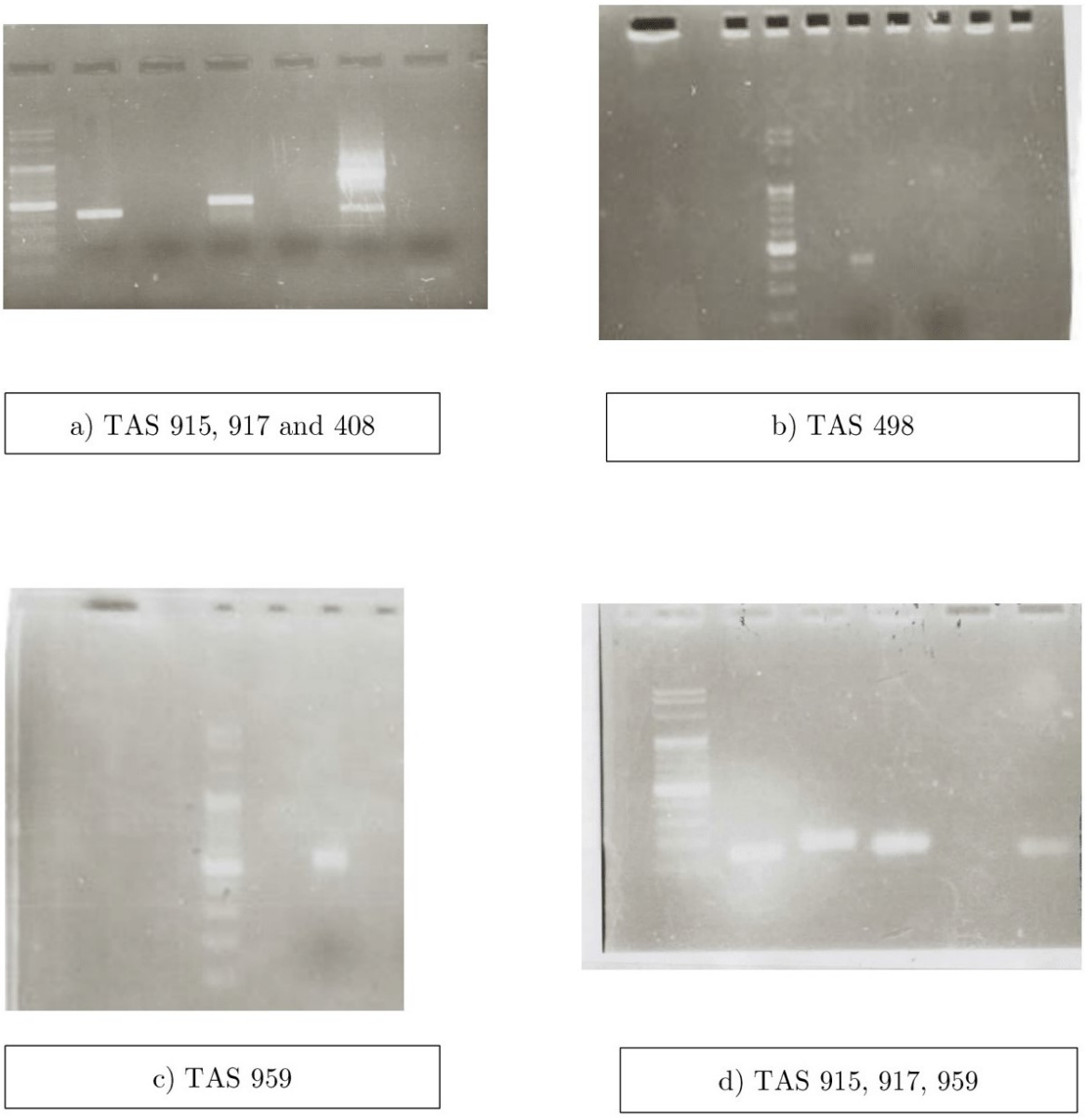


Figure S3: ***Agarose gel electrophoresis showing genomic amplification of TAS genes.*** *Concretely: in gel A, lane 1 corresponds to the marker of 1 kb, lane 2 to TAS 915, lane 4 to TAS 917, lane 6 to TAS 408, and lanes 3, 5 and 7 to negative controls of respective TAS 915, 917 and 408. For gel B, lane 3 depicts the mass ladder, lane 4 the negative control and lane 5 the amplified TAS 498. Same pattern is observed for gel C, but concerning lanes 2, 3 and 4. On the other hand, subfigure D depicts results of RT-PCR from RNA extraction. Here, ladder marker is in lane 1, amplified TAS 915 with two sets of primers correspond to lanes 2 and 3, amplified TAS 917 was placed in lane 4, and lane 5 shows no amplification of TAS 959. Lane 6 is considered to contain the positive control, a gene coding for auxin response factor 3 (ARF3), which is proven to be expressed. For A, B, C and D, the brightest band of molecular mass marker is 600-bp long, increasing or decreasing 100 bp each band on.*