Characterization of miRNA-mediated response to multiple stress conditions in *Cucumis melo*

Luis Manuel Cervera Seco

Máster en Bioinformática y Biología Computacional

MÁSTERES
DE LA UAM
2019 - 2020

Facultad de Ciencias

UAM Universidad Autónoma de Madrid

Universidad Autónoma de Madrid
Escuela Politécnica Superior

Máster en Bioinformática y Biología Computacional

Trabajo Fin de Máster

Characterization of miRNA-mediated response to multiple stress conditions in *Cucumis melo*

Autor: Luis Manuel Cervera Seco
Tutor: Gustavo Germán Gómez
Ponente: Gonzalo Martínez Muñoz

Febrero 2020

CHARACTERIZATION OF MIRNA-MEDIATED RESPONSE TO MULTIPLE STRESS CONDITIONS IN *CUCUMIS MELO*

Autor: Luis Manuel Cervera Seco
Tutor: Gustavo Germán Gómez
Ponente: Gonzalo Martínez Muñoz

Instituto de Biología Integrativa y de Sistemas
Parc Científic
CSIC - Universitat de València
Febrero 2020

Summary

Resumen

Los microRNAs (miRNAs) han sido descritos como elementos clave que intervienen en la regulación de la respuesta a estrés en plantas. Así, el mecanismo de inducción de RNA de interferencia surge como una nueva tecnología con el potencial de aumentar la tolerancia a ciertas condiciones ambientales adversas. Sin embargo, solo existen unos pocos estudios que caracterizan detalladamente la red de interacciones de miRNAs, así como los efectos transcripcionales en cascada relacionados con la expresión diferencial de miRNAs ante situaciones múltiples de estrés. Por tanto, en este trabajo se describe una estrategia computacional basada en secuenciación de pequeños RNAs (sRNAs) y en cuantificación de su transcriptoma para analizar la acumulación, funcionalidad y organización estructural de los miRNAs de melón reactivos a condiciones de estrés bióticas (*Monosporascus cannonballus*) y abióticas (frío, sequía, salinidad y día corto), en combinaciones dobles y triples. Se identificaron 97 secuencias de miRNA diferencialmente expresadas, agrupadas en 23 familias. La mayor parte de ellos (alrededor del 75%) están infra-representados, excepto las familias de miRNAs 398 y 408, las cuales están sobre-expresadas. Además, los resultados muestran que las dianas de aquellos miRNAs infra-expresados participan en el proceso de respuesta a estrés de la planta, mientras que los miRNAs sobre-expresados controlan procesos como la señalización o el metabolismo del cobre. Como observación adicional, se constató la predominancia de frío como condición de estrés determinante de la respuesta global de la planta de melón, de tal manera que el nivel de expresión de un gen controlado por un miRNA sensible a cambios de temperatura se mantiene bajo la presencia combinada de frío y otras situaciones adversas. Por otro lado, los miRNAs relacionados con estrés se suelen clasificar en tres grupos según el rango de respuesta a estrés, a saber: amplio, intermedio y específico. Los miRNAs de respuesta amplia, que intervienen en la modulación de la expresión de factores de transcripción y en la señalización celular mediada por cobre, representan los nodos más centrales y altamente conectados con el resto de componentes en la red de respuesta a estrés. El presente análisis puso en relevancia que los miRNAs generalistas o de respuesta amplia juegan un papel muy destacable en la tolerancia frente a estreses complejos, mientras que los miRNAs específicos presentan un rango de acción mucho más reducido, normalmente ceñido a situaciones de estrés simples. En resumen, los resultados obtenidos en este trabajo de investigación corroboran la hipótesis de que las plantas utilizan regulaciones sistémicas mediadas por miRNAs como mecanismo crucial para responder de manera eficaz y efectiva frente a situaciones ambientales adversas. Asimismo, este análisis minucioso de expresión diferencial de miRNAs en respuesta a estreses combinados bióticos y abióticos, en un contexto de relevancia económica y agronómica de un cultivo como el melón, ofrece la posibilidad de transferir conocimiento en el campo de la mejora genética con el objetivo de producir nuevas variedades con características mejoradas, en este caso, la tolerancia a condiciones de estrés múltiples.

Palabras clave

Tolerancia a estrés, Cambio climático, Silenciamiento por ARN, Micro-ARN, ARN no codificantes.

Abstract

MicroRNAs (miRNAs) have been recently described to be key elements in regulation of the stress response in plants and their induced RNA-interference process emerges as an innovative technology for improving tolerance to adverse environments. However, there are only a few studies that comprehensively report the miRNA interaction network, and the consequent transcriptional effects related to miRNA differential expression under multiple stress conditions. Here, we describe a computational approach based on sRNA sequencing and transcript quantification data to analyze the accumulation, functionality and structural organization of melon miRNAs reactive to biotic (*Monosporascus cannonballus*) and abiotic (cold, drought, salinity, and short day) stress conditions, at double and triple combinations. We have identified 97 differentially expressed miRNA-related sequences, which are grouped into twenty-three miRNA-families. Most of them (around 75%) are downregulated, but miR398 and miR408 families (known as copper-related miRNAs) are generally strongly up-regulated. Furthermore, our results show that the miRNAs that are down-regulated upon stress predominantly have target genes that are known to participate in the stress response by the plant, whereas the miRNAs that are up-regulated control genes linked to signalling and copper-processing. An additional remark involves the predominance, in melon, of cold as the stress condition determining the global response of the plant, in such a way that the level of expression of a given gene controlled by a cold-responsive miRNA is maintained upon the combined presence of low temperatures and diverse stress conditions. On the plus side, stress-related miRNAs are proven to be categorized in three groups showing a broad-, intermediate- or narrow- response range. MicroRNAs reactive to a broad range of environmental cues (involved in the modulation of Transcription Factors and related to copper-mediated plant signalling) appear as central and highly connected components in the stress-response network. Other significant insight of the analysis was that broad-response miRNAs play indeed a critical role in the tolerance of complex stress conditions, while narrow-response miRNAs display a reduced-range response, mainly restricted to single stress situations. In conclusion, the results obtained in this work support the emerging idea that plants use miRNA-mediated regulations, as a pivotal mechanism to quickly respond in adverse environmental situations. This comprehensive analysis of miRNA expression in response to combined biotic and abiotic stresses in a agronomical relevant crop offers the possibility that, in the future, this knowledge may be transferred to breeding programs to obtain new melon varieties with enhanced tolerance to multiple stress conditions.

Key words

Stress tolerance, Climate change, RNA silencing, MicroRNAs, Non-coding RNAs.

Acknowledgements

I would like to thank, in first place, my supervisors at the laboratory María Carmen Marqués Romero and Gustavo Gómez since I do really appreciate their exceptional guidance, and my acknowledgement as well for Gonzalo Martínez Muñoz, for his help throughout the year with academic counselling and my professional future. I would also like to express my gratitude to my friend Paula Gomis Rosa, for the provided training in LaTeX formatting and processing. You have always been the best friend! And finally, I strongly want that this Master's Thesis will be the final closure of my academic life at university to start off with a job or an exciting PhD project somewhere around the world. Thanks to all people that have support me along these 6 years at university in València, Madrid and Nottingham.

Contents

List of Figures

List of Tables

1 Introduction

1.1 Plant stress

Plant development is highly dependent on interactions with the environment. If any of these conditions impose a negative repercussion on plant growth and development, it is hence described as a stress [61, 63]. The origins of the stress are really diverse, including pathogen attacks (biotic stress) and drastic climatic and environmental changes, namely extreme temperatures, light intensity fluctuations, drought or different degrees of salinity (abiotic stress conditions) [16, 33]. Since these impacts may unbalance the homeostatic state of the plant, and as a consequence of their sessile nature, plants have been impelled to develop highly sophisticated and robust molecular mechanisms and systemic pathways to counteract the adverse effects triggered by prejudicial environmental cues [30].

Occasionally, when those mechanisms are not sufficient to cope with the detrimental effects, the crop production is dramatically affected, leading to plant collapse. Hence, the subsequent adverse physiological alterations limit crop yield and lead to massive economic and agricultural losses, to the most extreme case [14, 33]. Nowadays, within the context of climate change and the rise of extreme climate phenomena, it is becoming dramatically critical to face this issue by identifying pivotal molecules in order for plants to better tolerate the increasingly frequent environmental oscillations. Additionally, the current frame of an exponentially growing human population across the globe oblige to increase massively plant productivity, and design novel strategies to ensure a sustainable crop production [14, 19, 33, 43].

1.2 Small RNAs and microRNAs in stress response

From the beginning of the 21st century, the advent of *Omic* Technologies and the improvement and development of high-throughput Next-Generation Sequencing (NGS) have allowed to elucidate that plants are required to reprogram their transcriptional activity to modulate the expression of thousands of genes, thus recovering that initial homeostatic state. Consequently, the recognition of altered gene expression has led to identify and characterize differentially expressed genes which may be pivotal for stress tolerance in plants [19, 48, 52]. Despite the lack of certainty involving the regulatory network governing the stress-response process, which still

remains to be completely unveiled [51], there are numerous evidences suggesting that one of the crucial regulatory elements playing a significant role in several species are small RNAs (sRNAs), a class of 20- to 25-nt long RNAs which do not codify for any protein, though exhibit important functions in plants at transcriptional or post-transcriptional levels, namely systemic and tissue development, phase transition, heterochromatin formation, and the non-negligible silencing of core genes related to biotic and abiotic stress tolerance [2, 13, 14].

Small RNAs are classified, in turn, into diverse classes according to their function and biogenesis process, including microRNAs (miRNAs), small interfering RNAs (siRNAs), PIWI-interacting RNA (only in animals), and transfer RNA-derived small RNAs as the most prominent groups [59]. There have been recent evidences pointing out that subgroups of siRNAs are differentially expressed in response to stress, especially in the case of trans-acting small intereference RNAs (ta-siRNAs) in melon [14, 17]. However, over the last years there have been an increasing collection of analysis indicating the putative role of miRNAs in stress regulation in plants, including species such as melon, sugarcane, arabidopsis, maize, wheat or rice [12, 33, 48, 51, 54, 55]. To be more specific, those stress-responsive miRNAs have been characterized and identified in plants subjected to bacterial, fungal and viral pathogenesis [33, 62], in addition to abiotic stresses, namely salinity, drought, low and high temperature and nutrient deprivation [51, 54, 55, 61, 63].

To be more precise, miRNAs are usually 20- to 22-nt long sRNAs that have been generated from a longer precursor of around 200-nt long. These small pieces of RNA target gene transcripts from different loci at different locations, in a similar way to other sRNAs, like ta-siRNAs. Hence, its main function is the suppression of gene expression, this is, gene silencing [2, 12, 55].

1.3 miRNA biogenesis

Regarding the miRNA biogenesis process in plants, also named microprocessor, several proteins are recruited sequentially for the maturation of the final miRNA, as depicted in Figure 1.1. Even though most of these proteins are required for the majority of the sRNA processing pathways, some particular types, e.g. DICER-LIKE 1 (DCL1), are specific of the miRNA biogenesis pathway [2]. Moreover, miRNA generation in plants differs from the respective equivalent process in animals, both in terms of biogenesis and silencing [1].

To give a more precise insight into the topic, the initial stage of the miRNA processing commences in the nucleus, and involves the transcription of the miRNA-encoding genes by RNA polymerase II (Pol II) to generate a primary transcript or precursor of the miRNA gene, which is termed primary miRNA (pri-miRNA), This precursor folds itself into a hairpin-like secondary structure, with a capped 5' end and a polyadenylated 3' end, as well as an upper and a lower stem, and the region containing the RNA duplex [53, 59]. The pri-miRNA is further processed thanks to the recognition of the two arms of the precursor by DICER-like RNase endoculeases. This subsequent step of the processing involves the formation of the precursor miRNA (pre-miRNA) as an intermediate stem-loop scaffold containing solely the miRNA/miRNA* duplex along with the loop [1, 53, 59].

In this context, several proteins are recruited to process the pri-miRNA, emphasising the role of the accesory proteins HYPONASTIC LEAVES 1 (HYL1) - a double-stranded RNA-binding protein - and SERRATE (SE) - a zinc-finger protein - , whose main function is the improvement of the specificity of DICER-LIKE 1 (DCL1) protein by increasing overall cleavage efficiency and accuracy [59]. Therefore, HYL1 and SE are the core of the microprocessor, and confer an additional constraint to ensure the robustness of the cleavage and the correct identity of the final miRNA duplex [1, 20, 53]. Indeed, the levels of accumulation, hence the stability, of the pri-miRNA mainly determines the prospective targeting activity of the mature miRNA. In turn,

the stability may fluctuate by the length of the precursor, the hairpin structure, the interaction with the accessory proteins, and the alternative processing of some stem-loops arising from long precursors, which may give rise to different miRNAs from one single precursor [8].

Upon the cleavage of the pri-miRNA by a DICER-LIKE (DCL) protein and its accessory proteins, the pre-miRNA is formed, which is subsequently processed by DCL1 endonuclease leading to the formation of a miRNA/miRNA* duplex, which is already around 21- to 22-nt long normally [2, 5, 53, 59]. However, since different plant species have different numbers of DCL proteins, it is accepted that other DCL proteins may play a pivotal role in the pri-miRNA and pre-miRNA processing [59]. Conversely, the DCL-generated duplex presents a 2-nt overhang at both ends, in addition to two hydroxyl groups at 3' end (2' OH and 3' OH). Only the second hydroxyl is essential for the subsequent methylation of the 3' ends, hence it is 2'O-methylated, by the RNA methyltransferase HUA1 ENHANCER (HEN1), which directly interacts with HYL1 and DCL1 [1, 5, 53]. This nascent duplex is thought to be exported into the cytoplasm by the protein HASTY (HST), the homologous protein to the animal Exportin 5 [1, 2, 59].

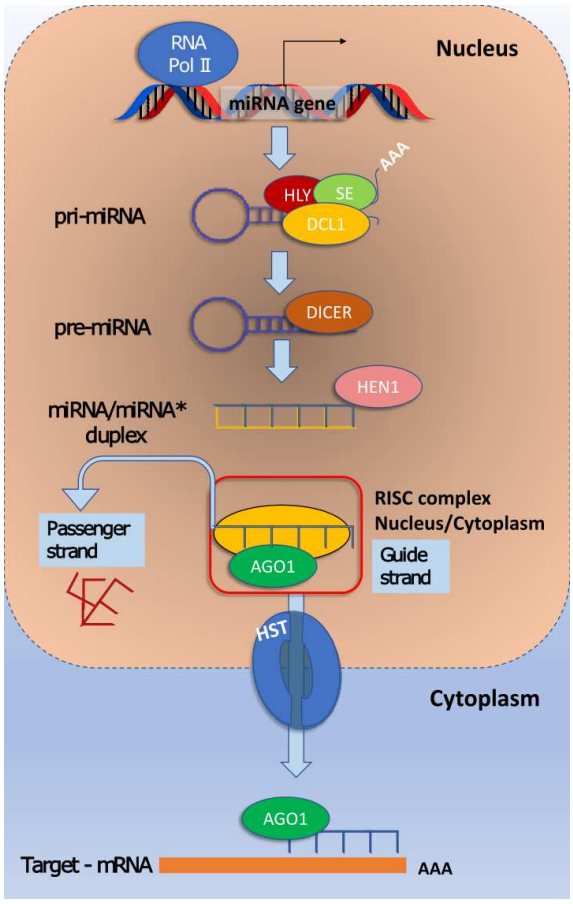


Figure 1.1: Schematical representation of miRNA biogenesis, and final targeting and degradation of mRNA target RNAs, according to Wang *et al* [59].

After being methylated, a member of the ARGONAUTE family, ARGONAUTE 1 (AGO1) mainly since its the more relevant effector protein in plants, is recruited along with different secondary components to form the RNA-Induced Silencing Complex (RISC). For a long time, it was believed that RISC was constituted in the cytoplasm after the exportation of the pre-miRNA into this space. Nonetheless, recent evidences indicate that RISC is assembled mainly in the nucleus and AGO1 and the rest of the components are present in the nucleus for this purpose [7]. At the same time, it has been recognized the possibility that some miRNA are exported into the cytosol in the duplex conformation, just after the DCL1 cleavage, whereas RISC is assembled in the cytoplasm [1, 7, 53, 59]. Others have even described that both the DCL-mediated cleavage of pre-miRNAs and the methylation may occur in the cytosol [5].

What is evident is that the passenger strand of the miRNA/miRNA* duplex is selectively degraded normally, and only the guide strand, the mature single-stranded miRNA is loaded and assembled into RISC, which modulates the accumulation of the target RNA [1, 2, 5, 53, 59, 61]. This regulation is carried out via complementarity of the miRNA aligning to the target. The miRNA-mediated modulation of the gene expression acts at two different scopes: either by cleavage and/or by translational inhibition [1, 2, 5, 9, 53]. Additionally, it has been described that plant miRNAs usually degrade or inhibit their targets by cleaving specifically between the 10th and 11th position from the 5' end [35].

1.4 The role of miRNA-mediated networks in stress regulation

MiRNAs are grouped into families according to function and sequence similarity. Members of the same families are proposed to diverge between them due to genome duplication events, which may also explain, along with natural selection, the emergence of new miRNA families, whose origin is a common ancestor [11]. From the numerous characterized families, just a few of them are conserved across species. Thus, it has been hypothesized that miRNA genes emerged really early during the evolution, since most of them are species-specific [9, 35, 48]. Recently, more evidences have been found highlighting that miRNAs are crucial in a wide range of processes, from plant development and growth to signal transduction, ta-siRNA generation, and, as many other sRNAs, the coordinated regulation of stress response under abiotic and biotic stress conditions. [9, 54, 61]. Briefly, the modulation of the miRNA biogenesis pathway is indispensable for stress tolerance and plant development [2, 12, 53, 61].

Furthermore, since the majority of miRNAs, especially the members of the same family, derive from a common ancestor, the stress-response network of many species share pivotal, central genes that bear the control over a broad range of adverse environmental factors. In contrast, other genes play specialised roles in the response to a few specific stress conditions, displaying a narrow-range response [61]. Therefore, general or conserved mechanisms governing the miRNA-mediated response to stress are expected to be observed within families of conserved miRNAs, showing an alteration of expression in the same direction, either up- or down-regulated [48].

1.5 Interest in agronomic species - melon

Besides the comprehensive analysis of model plant species, including *Arabidopsis thaliana*, in order to gain knowledge and unveil the molecular basis of the miRNA networks, it is vital to detail the particular regulatory mechanisms in species with agronomic interest, since the stress response might not be exactly equivalent among them, and results may be applied ultimately to increase crop productivity and protection to relevant plagues and climate changing conditions [61, 63]. Moreover, agronomic species have an economic importance in Spain, hence the deep understanding of the miRNA network might be transferred to breeding programs easier and faster.

Our group have directed the focus of the research on melon (*Cucumis melo*), because of the economic relevance of this horticultural crop in Spain and across the world. In terms of economic indicators, a small overview of provisional data from international and national institutions provide a good evidence of the trade involving melon production. According to the United Nations, melon exportations from Spanish State encompass a volume of more than one million tonnes, with a trade value of 873,000 dollars. Germany, France and the United Kingdom are the main importers of Spanish melons [18]. Regarding governmental data sources, the Spanish Ministry of Agriculture, Fishing and Food (MAPA in Spanish) states that the total annual melon yield is around 664,353 tonnes, cultivated in an arable area of 19,025 Ha [41]. Hence, statistics manifest an obvious predominant status of Spain in the leadership of global melon trade, and the most important across Europe [23].

From the point of view of genomics, melon is a dicotyledonous, diploid plant species whose genome is arranged into 12 chromosomes (2n = 24). However, despite the agronomic interest and both the genetic and morphological diversity of this member of the *Cucurbitaceae* family, there were a few studies able to characterize the entire melon genome. Since the advent of novel NGS technologies, EST collections, and TILLING strategies, international initiatives led by the International Cucurbit Genome Initiative (ICuGI) and the Spanish consortium Melonomics

have granted the public access to melon genome sequences, transcriptome and other annotation information updated in different databases [15, 40, 46]. The creation of such platforms ensures the development of genomic studies and the identification of active regulatory small RNAs involved in stress tolerance, including miRNAs, e.g. one of the first projects of Gonzalez-Ibeas *et al* [23, 51, 54].

In addition to the mentioned advances in sequencing methods and databases, diverse *in silico* technologies have been devised to exploit the information quality obtained from sequencing files. In order to achieve this purpose, some issues are needed to be efficiently addressed, starting with the sequence quality. It is widely recognized that sequencers are biased for a specific type of error through the sequencing process depending on the employed NGS method. In order to correct this error-prone sequencing, several tools are available, with an emphasis on FastQC, which filters sequences by quality scores stored in the FastQ files, and eliminates those poor-quality reads, e.g. normally the 3' ends [10]. Other relevant software to tackle this problematic issue is Trimmomatic or CutAdapt, which trims sequences derived from the sequencing process, for example, adapters, from paired-end reads, so as to obtain the real read, without any accessory sequence [6, 42].

Once the raw sequences have been correctly filtered, the characterization of the miRNA molecules can be performed, e.g. via differential gene expression. For this purpose there are many packages available in many programming languages, although the most utilised are hosted in the Bioconductor platform, and are written in R, especially EdgeR and DeSeq2 are widely used for the identification of differentially expressed sRNAs. Finally, for an adequate representation of a biological network, Cytoscape might be a straightforward and user-friendly programme of choice, since it favours the interpretation of results and interconnections between members of a determined pathway or route [50]. In short, all of these technologies favour the design of experimental studies employing melon as the research species, with significant practical applications for Systems Biology and Climate Change.

All in all, in spite of the studies identifying stress-responsive miRNAs, the basis on how miRNA biogenesis pathway is affected remains obscure yet. Besides, they are lacking complete experimental assays measuring differential miRNA expression over different exposure times at different stress conditions, with identical plant developmental stages. Further to that, neither the effect of applying multiple stress situations simultaneously nor the direction of the alteration of the stress response upon this exposure have not been determined. In this sense, more efforts are required to overcome this lack of information about the functioning of miRNA-mediated stress response and to determine completely which miRNA genes are differentially expressed both qualitatively and quantitatively, whose function will be consequently critical in facing environmental cues [33, 54].

2

Objective

All things stated, the principal purpose of this project is the *in silico* identification of the stress-responsive miRNAs differentially expressed under complex stress situations, combining abiotic and biotic cues. Additionally, we aim at the characterisation of the differential alterations of miRNA expression levels, as well as the comprehensive descripiton of the inferred miRNA-governed regulatory network in the process of response to multiple stress conditions. Thus, it is desirable to establish effective comparisons between the individual melon responses to one specific stress condition and the simultaneous combination of those simple detrimental exposures upon a multiple stress subjection. In order to achieve that, the considered miRNAs were first identified from the sequenced sRNA libraries, computed by our research group, emerged from a specific variety of melon plants. Ultimately, the gain of theoretical and basic knowledge arisen from this work would entail a new scope of action to address a cutting-edge research in the field of plant adaption to climate change. Hence, it would be potentially applied to breeding programmes, in such a way that tolerant melon plants might be engineered to face harmful, extreme environmental conditions, in the current context of climate changes, hence avoiding dramatic crop losses [61].

Briefly, in order to accomplish the global aim, sequential steps were designed up to the point of unveiling the full miRNA-regulated newtork in *C. melo*.

1. Filtration and pre-processing of sRNA sequences.
2. Detail of the sRNA distribution of the sequenced libraries by stress and size.
3. Identification of miRNAs among all sRNAs.
4. Determination of miRNA differential expression and visualization.
5. Characterization of miRNA accumulation profiles using all sRNA counts across stress libraries.
6. Study of alteration effects upon stress combinations on miRNA expression.

3

Materials and methods

3.1 Biological assays prior to the current Bioinformatic analysis

Although the pipeline of the project was devised to be thoroughly *in silico* and computer-based, there was a biological assay upon which the sRNA transcriptomic profile was determined, by means of sRNA sequencing, hence having the FASTQ files from all sRNA libraries as the starting point. The previous specific biologic experiment, developed at the laboratory of Dr. Gustavo Gómez, in València, involved the exposure of several melon plants (cv. "Piel de Sapo") to multiple stress conditions, both biotic and abiotic. Regarding the biotic stress, *C. melo* plants were infected with *Monosporascus cannonballus* (MON), a soil-borne fungal pathogen representing the major plague in melon, typically in arid or semi-arid areas, which may cause root rot, defoliation, canopy collapse and total necrosis [45]. On the other side, the selected abiotic stress situations were cold temperatures (C), this is 20 ^{o}C; high salinity (SA); drought (D) and short-day (SD), this is 8-hour light period along with 16-hour darkness. There were other biotic stresses included in the original analysis, but those were dealt with in a previous paper of our group, focused on the miRNA-mediated network in response to simple stress situations in *C. melo* [48].

Apart from the five conditions, there was a control group, not exposed to any stress, in order to be compared to the altered samples. Moreover, the multiple stress combinations were the following: cold with drought (C.D), cold with salinity (C.SA), cold with salinity and short-day conditions (C.SA.SD), cold with short-day (C.SD), drought with *M. cannonballus* (D.MON), and drought with salinity (D.SA). It is important to mention that, although the pipeline was addressed to investigate the potential role of miRNAs expressed under multiple damaging environmental phenomena, the considered stress libraries for the study included both simple and combined stress situations, leading to a total number of 12 libraries. Two additional single conditions were included, namely biological infections by *Hop Stunt Viroid* (HSVd), a long non-coding RNA that triggers a sRNA-mediated response upon a wide range of cucurbit hosts are challenged [37], and *Agrobacterium tumefaciens* (A), a typical plant-invader bacteria. As for the the original biological assay [14, 48], after the exposure to the stress, leaf tissue was sampled at 11 days post-treatment. The extractions from each plant were by triplicate, except for the control samples, which were extracted by quadruplicate.

3.2 Construction of the sRNA libraries

Regarding the sequencing protocol, once the RNA extractions were prepared from leaf samples, the solutions were enriched in sRNAs using an isolation kit from Stratagene, which was a constraint to obtain pure sRNA libraries. The sequencing procedure of the 14 sRNA libraries was carried out by Sistemas Genómicos (https://www.sistemasgenomicos.com), and included a reverse transcription of the enriched sRNA samples to cDNA libraries, by ligating adapters to both 5' and 3' ends, according to Illumina's recommendations. Later on, clusters were generated from the equimolar pool of the cDNA libraries, then sequenced by paired-end sequencing (100 x 1) in a HiSeq 2000 sequencer from Illumina.

Subsequently, all adapters at 5' and 3' as well as low quality reads from the FASTQ files were trimmed off using the Cutadapt tool version 1.10 implemented in Python. The trimmed sequences constituted the final version of the constructed sRNA libraries, stored in the form of FASTA files. Thus, this is indeed the initial point for establishing the analytic pipeline in order to identify and characterize the differentially expressed miRNAs in response to multiple stresses. The sRNA libraries contained RNAs ranging from 20 to 25 nt in length, which were assigned an own code for easing their identification: SNCMelX, being the X a numerical code specific to the unique sequence of sRNA. For more details of the previous informatic processing and the melon plants development, refer to the work describing the identification of ta-siRNA molecules in melon from our laboratory [14].

3.3 Filtration and pre-processing of sRNA libraries

The next step was the pre-processing of the sRNA libraries to eliminate the sequence artefacts and obtain only the real sRNAs. First of all, the reads corresponding to ribosomal RNA (rRNA), transfer RNA (tRNA), small nucleolar RNA (snoRNA) and small nuclear RNA (snRNA) were filtered out from the libraries using Rfam database (http://rfam.xfam.org/), since they are putative sequences that cannot play a role in stress regulation, i.e. they are not regulatory sRNAs, but other deceptive small sequences [29]. Thus, all reads aligning to any sequence deposited in the Rfam database were discarded.

After that, the counts associated to each unique sRNA read in the Rfam-filtered FASTA files were normalised using the Reads per Million (RPM) scaling factor. In this way, the factor was calculated by dividing the the total number of reads by 1,000,000, hence one million factor. Then, read counts were divided by this value, normalising the sequencing depth. Finally, a table was created for each stress situation, composed of read counts from both specific stress replicates and the non-treated samples, to have more readily available data in less files.

3.4 Detail of the sRNA distribution of the sequenced libraries by stress and size

Upon the normalisation of the read copy number, sRNA sequences were grouped according sRNA length, ranging from 20 to 25 nt, and the 14 possible libraries. All counts were collapsed and the mean values along with the standard deviation computed from the deviations of the respective replicates were used to represent the distribution of the cumulated counts to check the validity and robustness of the sequencing process, since it was expected that 24-nt long sRNAs were predominant. From this point of the pipeline on, the simple stress conditions which were not involved in the multiple combinations, i.e. *Hop Stunt Viroid* and *A. tumefaciens*, were not

taken into account and they were discarded from the analysis, leaving 12 sRNA libraries to be characterised.

Furthermore, a further characterization of the sRNA libraries via the study of the correlation between biological replicates and the stress conditions was performed using a principal component analysis (PCA). The PCA plot was computed via the "prcomp" function with scaling factor of the "stats" library implemented in R, along with the "plot3d" function of the "rgl" R package [34]. The statistical significance of differences in distances between libraries represented as data clusters was assessed via the Mann-Whitney-Wilcoxon test, using the "wilcox.test" function. This test requires the calculation of the Euclidean distances between members of the same cluster and among one cluster to the others, i.e. testing that replicates are closer to themselves than members of any other group. Moreover, to complement the analysis, ellipses showing a 95% confidence region were drawn, which depicts the area where a real data point of the cluster, i.e. a sRNA library, should probably appear.

3.5 Alignment of mature miRNAs to hairpin sequences

Once sRNA libraries have been characterised, we proceeded to the identification of both *Viridiplantae* and *C. melo* miRNAs among all sRNA libraries. In this step mature miRNAs sequences were downloaded from the primary database of described miRNAs: miRBase (release 22, March 2018), which is the main repository for described and identified miRNAs for plants and animals [31]. The miRBase mature miRNAs were subsequently aligned to the precursor miRNA sequences, otherwise hairpin precursors, which were downloaded as well from miRBase repository. The alignment was carried out by means of "bowtie" command line (version 1). Within the "bowtie" command, no mismatches were allowed with the option "-v 0" and all possible 100%-similarity pairwise alignments ("option -a").

Furthermore, not only mature melon miRNA were mapped, but also *Viridiplantae* miRNAs to the same melon hairpin precursors. In this case, the alignment was performed via "bowtie" as well, selecting the restrictive options. After that, the region of the precursor from which the mature miRNA is generated, i.e. it is selectively accumulated, was determined. In plants, miRNA phases may originate either from the upper stem of the hairpin, also called 5' or 5P region, or from the complementary lower stem, or 3'/3P region. As for the classification as 5P or 3P, miRNAs were assigned a label based on the fraction of the length of the precursor that they have aligned to, i.e. from half of the length of the precursor it was classified as 5P, and conversely, 3P. The miRNA identifiers were consequently updated by adding a "5P" or "3P" label. With regard to these calculations, they were implemented using "pandas" module in Python 3.7.

3.6 Determination of miRNA differential expression

With these new labels reflecting the precursor region, the determination of the differentially expressed multiple stress-responsive miRNAs was carried out. For this purpose, a differential expression analysis was performed using the sRNA reads in the Rfam-filtered libraries obtained in the first filtering step as an input. Three different statistical methods were employed to evaluate the sRNA differential expression, namely edgeR [38, 44], NOISeq [56] and DESeq2 [58], which are deposited as R packages. Since the differential expression data were referred to sRNA sequences, miRNAs were required to be recognised among the sRNA reads, and get the expression data related to them as well. As a result of the *in silico* computation, a master table

reflecting the differential expression log2FC values, along with statistical parameters, such as FDR, p-values, mean counts, standard deviations and so on, for each of the sRNA reads of the analysis. Moreover, other sets of master table, slightly smaller in size, stored the normalized counts of sRNA reads per library and stress, compared to the control replicates. This count table was used in a posterior section of the pipeline.

Upon processing the miRBase miRNA sequences in the most convenient format, the sRNA reads were matched to these mature miRNAs by sequence using the "match" function in the "grr" R package. Therefore, filtered sRNAs matching miRBase melon miRNAs, and those reads determined to be fully homologous to previously described *Viridiplantae* miRNAs in miRBase, were identified as stress-responsive melon miRNAs [48]. In addition, putative differentially expressed miRNAs were filtered according to two criteria in order to provide robustness and additional levels of confidence of the significance of the alteration of the expression levels. These criteria included a False Discovery Rate (FDR) value lower than 0.05 for the three differential expression methods used: edgeR, DeSeq2 and NOISeq (in this case, the probabilistic statistic threshold is called prob-N, working as an inverse p-value, i.e. significant results are imposed to be higher than 0.95 instead), and a second filter comprising the acceptance of miRNAs of 21 and 22 nt in length. After the selection of the most relevant miRNAs, the miRNA log2FC values were represented graphically in a heatmap, assigning the red colour to the down-regulated miRNAs and the blue to the up-regulated ones. Only the edgeR data was employed so that heatmap visualisation became more straightforward. To be more precise, the R package of choice for the heatmap was "morpheus" from the Broad Institute, that allows an interactive display of the diagram (https://github.com/cmap/morpheus.R).

Heatmaps by miRNA family were created from the usage of another R package: "ggplot2", useful for graphical representation of data sets. Regarding this package, the function "ggplot", that requires a reshaped object by means of "melt", deposited in the "reshape" R package, and the "aes" function from "ggplot2" to construct aesthetic mappings, providing the axis and parameters necessary for plotting. The combination of these parameters resulted in the production of a coloured squared, heatmap-like image, with identical colours and conditions as the general heatmap. Apart from heatmaps, boxplots were also generated, via the use of "Boxplot" function devised as a part of the "car" library in R, using as input the differential expression values grouped by miRNA family.

3.7 Characterization of miRNA accumulation profiles using all sRNA counts across stress libraries

In relation to the miRNA alignment to precursors, an additional bowtie alignment was conducted so as to map the entire set of sRNA reads from libraries to the melon miRNA precursors, whose sequences were obtained from miRBase. The complete list of sRNA unique reads (available in the format of a FASTA file) were extracted from the master table with differential expression values obtained in the previous section 3.6, by the application of the "unique" R command on the sRNA ID column, and numerically sorting them by SNCMel code. The information regarding ID, full sequence and stress library where they were present was retrieved from this master table for further downstream analysis. Once the FASTA file was compiled, the "bowtie" aligner (v1) was executed for this mapping, with no mismatches allowed, having melon hairpins from miRBase as the reference indices.

From this point on, the code of the pipeline was written in Python 3.7 with the aid of "pandas" module to create the basic support data frames. The SAM files were exported to csv and loaded using the "read_csv" function. Afterwards, the reads with a flag, indicating that

the sequence aligned to the reference was its reverse complementary, were reverse transcribed to get back the original sequence. Then, the compiled tables of read counts from stress libraries along with non-treated samples obtained in section 3.6 were imported as well. Here, the "merge" command from "pandas" core was vital to collapse all tables into a global one via a common key, i.e. the sRNA sequence, leading to a data frame dimensionally greater, with the 12 stress libraries (and 3 replicates each) as columns, in addition to the mean counts of each library and each of the replicates.

Besides, the differential-expression master table was imported as well, for miRNAs to be filtered based on log2FC, i.e. values $\leq$ -1 or $\geq$ 1, FDR of edgeR and DeSeq2 $\leq$ 0.05, and the p-value probability of NOISeq (prob.N) $\geq$ 0.95, as consensus requirements for truly miRNA candidates to be differentially expressed. Next, this table was reshaped to a data frame rendering the unique miRNAs as rows and the stress libraries as columns. To this data frame, by the pandas "merge" command, the loaded information from the SAM files was collapsed, in such a way that only the miRNAs aligning to the listed precursors were retained. What is more, the start and end positions of each miRNA within the precursor sequence was annotated as well. This was fused again to the tables of read counts.

The cumulated sum of all read counts at each position of the precursor, allowed by accessing to the start and end position data in the data frame, entailed the graphical representation of the sRNA accumulation onto the miRNA precursors. Specifically, reads coming from the different libraries were superposed on the same plot, and differentially coloured to be distinguished. The graphical description was carried out by the "matplotlib.pyplot" Python module. Likewise, control versus stress plots were produced as well, i.e. 13 plots per precursor since there are 13 stress libraries, so that changes in expression levels are better recognised.

3.8 Class clustering and miRNA-mediated network construction

As for the characterization of the miRNA-mediated network, miRNAs have already been clusterised in our laboratory in three classes according to the response range, namely broad, intermediate and narrow [48]. Following this line of research, further justification of this classification was provided by the elaboration of a binary table confronting presence ("1") and absence ("0") of expression in simple stress conditions, double or triple combinations. In other words, it depicts whether a miRNA is responsive to certain combinations of stress or not, either increasing or decreasing its expression. This presence/absence table was composed of four columns, expressing the miRNA name, and the appearance of that specific miRNA in the three groups of stress libraries (simple conditions, double and triple combinations). There was an additional column added afterwards which applied the sum of all values horizontally, for each row to render an approximate value related to connectivity. Using the "hclust" function from the R package "stats", a hierarchical clustering was performed and two cluster dendrograms (one vertical and one circular) were rendered, to identify the possible three response-range classes. The dendrogram was computed with the "ward.D2" method, making use of Ward linkage, whereas the test supporting the statistical significance of this clustering computation between classes was the Kruskal-Wallis test, via the "kruskal.test" implemented in R. A prior normality test was also applied via the "shapiro.test" R function.

The description of the miRNA-mediated network was complemented by the representation of a relationship diagram with connections among stress-responsive miRNAs and the stress conditions clustered by the number of combined adverse situations the plant has been exposed to, i.e simple (5 libraries), double (5 libraries) and triple (1 library). The assembly of the structure of the network was conducted in Cytoscape version 3.7.1. [50]. For the construction of

the network, an empty graph was created, introducing the use of the Python library "networkx", which allows the easy manipulation and assembly of graphs and networks [21]. The function "Graph" initiated the "networkx" graph.

Furthermore, the presence/absence binary table was interrogated, firstly, to create the nodes of the graph, importing the names of the miRNA reads from the first column and three extra nodes: "simple", "double" and "triple" with regard to the stress libraries. Then, edges were added via the use of filters and subsets of the binary data frame, where the existence of "1" in a cell meant the linkage of that miRNA with simple, double or triple stress conditions. In the end, the function "write.gml" within "networkx" exported the graph into a readable format for Cytoscape to read it and process it. To finally assess the relevance and influence of a determined miRNA or miRNAs within a group of neighbours, the betweenness centrality for nodes was calculated by means of the "betweenness_centrality" function from "networkx". The statistical validity between classes was evaluated again with the non-parametric "Mann-Whitney U" test for pairs of classes, and with the "Kruskal-Wallis" test for the three distributions, thanks to the "scipy.stats" Python library. Bar plots were produced to represent mean values of average centrality and betweenness centrality for each miRNA class, along with the standard deviation and the calculated p-values.

3.9 Study of miRNA agonism

As for the evaluation of the behaviour of miRNA expression levels upon the exposure to a combination of stresses, for each miRNA it was assessed whether the specific combination led to an additive, antagonistic or synergistic effect on this expression. For this purpose, mean read counts from each library were extracted from the counts table, along with the associated standard deviation, in order to compute the epistasis as the substraction of the sum of the control and the combined stress and the sum of the individual stresses. In addition to epistasis, the standard error of the epistasis was also calculated to compute a standardised epistasis value. The fitting of this ratio to a normalised distribution allowed the assessment of the degree of deviation from the expected additive effect [57]. Moreover, the distribution of these significant cases were analysed, although regarding the non-orthogonal basis of the experiment only comparisons of cold versus three stresses (salinity, short-day and drought) and drought with other two (*M. cannonballus* and salinity) were possible. Hence, the assessment of the combination effects was only measured as pairwise analyses between two stress conditions. These statistic calculations were performed using the statistical software SPSS [25].

For further support on the development of the computational methods, the custom scripts employed for obtaining the results of this project were uploaded to the following url address hosted in Github (`https://github.com/luicerse/TFM_Luis_Cervera`).

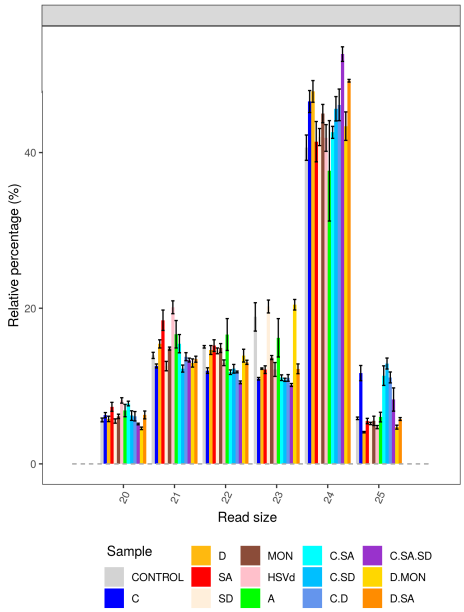
4 Results

4.1 Analysis of sRNA libraries

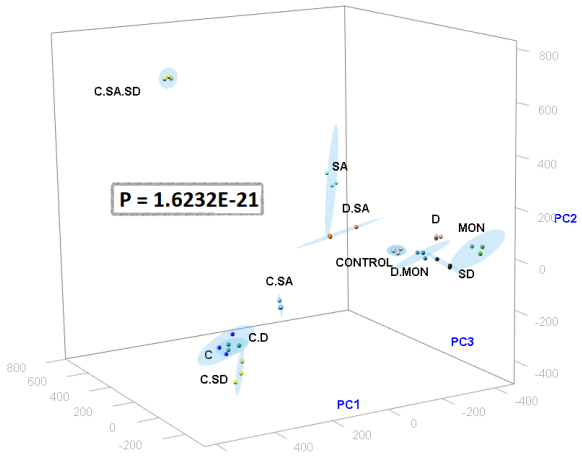
The identification of miRNA molecules was performed from the construction of sRNA libraries. After 11 days from the initial exposure to the seven simple stress conditions (A, C, D, HSVd, MON, SA, SD), the five double combinations (C.D, C.SA, C.SD, D.MON, D.SA) and the triple stress (C.SA.SD), a high-throughput sequencing was carried out from the RNA extraction samples. Once the FASTQ files were processed according to quality, the sequences were filtered by length (only the range between 20- and 25-nt long was kept) in addition to the elimination of reads mapping the sequences deposited in the Rfam data base, i.e. tRNAs, rRNAs, snoRNAs and snRNAs. These non-relevant non-coding RNAs represented between 2% and 5% of the total sRNA reads sequenced in the libraries. After these filters, the final sRNA libraries were composed of a total of 153,179,830 reads, from 799,492 unique sRNA sequences, composing all the replicates of the 14 conditions (42 libraries in total).

Upon the normalization of read counts by means of the Reads per Million method in order to counteract the library size effect, the relative abundance of sRNA sequences was exhibited as a bar plot based on sRNA length and library sample, as depicted in Figure 4.1a. First, the representation of the sRNA data set, there is an evident enrichment of 24-nt long sRNAs (40% of total reads), followed by similar accumulation levels of 21-, 22- and 23-nt long sRNAs, each of which constitute around a 15% of the relative accumulation of sequences. Conversely, the least abundant sRNA classes in length are the 20-nt and 25-nt long sRNA groups. More precisely, attending to the particular profiles within each class, no significant differences exist between the stress libraries and among control samples and stress conditions, hence they are equivalent in terms of read abundance and size.

Moreover, the correlation between the sRNA expression profiles in Figure 4.1a and the diverse stress exposures and replicates is evaluated and exhibited in Figure 4.1b with the Principal Component Analysis. The PCA assessment along with the Euclidean distances computation reveals that the replicates within the same treatment are more similar among them than with respect to the members of different conditions, since they are clustered into the same group, with a strong statistical significance, i.e. a p-value of 1.6×10^{-3}. The confidence provided by the statistical test ensures a consistent reproducibility of the experiment. As the Figure 4.1b



(a) Bar plot depicting the accumulation of the sRNA clean reads over the different libraries, ranging from 20- to 25-nt long. Control and treated samples are represented with different colours.



(b) Principal Component Analysis of the sRNA libraries obtained from Control and plants exposed to double and triple combinations of stress. Light blue ellipses represent the 95%-confidence region where library replicates should be found.

Figure 4.1: Combined diagram depicting the distribution of the sRNA reads (a) and the PCA representation of the sRNA libraries (b).

shows, the combined-stress libraries are observed to be distributed in an intermediate position with respect to their corresponding simple stress treatments, especially in the case of the double combinations involving cold treatment, since they are clustered together very closely to the cold replicates, to the left region of the PC1. At the same time, not unexpectedly, the triple stress library (C.SA.SD) shows the most divergent sRNA accumulation profile, as the interaction of three treatments simultaneously modulates heavily the sRNA expression. As for the cumulated variance, the six first principal components explain more than 55% of the total variability.

4.2 Determination of 5P/3P reads

Since the confidence in the robustness of the sRNA libraries was confirmed, then known miRNA sequences deposited in miRBase were imported so that mature miRNA reads may be identified from the entire set of sRNAs. However, previously the miRNA sequences were aligned to the melon hairpin precursors from miRBase in order to determine the pre-miRNA region where the mature miRNA is successfully processed. For the mapping of miRNAs, Bowtie version 1 is more suitable, as it has been optimised for the overlapping of short reads aligning to large sets of genome sequences, like melon, hence faster and more sensitive than Bowtie2 [28]. Furthermore, not only mature melon miRNA were mapped, but also *Viridiplantae* miRNAs to the same melon hairpin precursors, in order to pinpoint putative miRNAs that have not been characterised yet in melon, but may originate from any *C. melo* miRNA precursor.

The processing of the SAM files originated from the alignment revealed such results in Table 4.1 that, according to the miRNA position respective to the hairpin, 401 mature miRNAs positively matched at least one region of one melon pre-miRNA or precursor, 144 from which were

described as melon miRNAs, and the rest (257) belonging to the *Viridiplantae* plant families, i.e. green plants. Interestingly, whereas melon miRNAs were predominantly identified as 5P (arising from the 5' proximal region of the hairpin), *Viridiplantae* miRNAs have been preponderantly assigned as 3P (constituted from the complementary sequence of the precursor, at the 3' end). However, some of the stress-responsive miRNAs in the libraries were not identified either as 5P or as 3P due to the lack of the corresponding precursor within the miRBase-described hairpins, or the failure to align with no mismatches to the precursor.

Table 4.1: Table summarizing the assignment of 5P and 3P regions to the mature melon and *Viridiplantae* miRNAs.

-	***C. melo***	***Viridiplantae***	**Total**
mature miRNAs	144	257	401
5P	**81**	114	195
3P	63	**143**	206

4.3 miRNAs are generally down-regulated

Once the description of the sRNA libraries was completely characterized, the alteration effects on sRNA transcription were assessed by statistically comparing control and stress-exposed samples by three differential expression methods, namely edgeR, DeSeq2 and NOISeq, in order to give a major robustness to our detection system. From the 141,838 unique sequences determined to be stress-responsive sRNAs according to the three analytical systems, miRNA reads were identified by aligning them to the sequences deposited in miRBase database. After the application of filters based on statistical significance parameters explained in the methods section 3, the truly differentially expressed miRNAs were exhibited in a heatmap, as shown in Figure 4.2. As it can be observed, 97 different miRNA sequences were identified, which belong to 23 unique miRNA families. The general tendency for a stress-responsive miRNA is to be down-regulated in response to any adverse condition, in a proportion close to 75% of the cases. However, there is a cluster of a few miRNA families (miR398, miR408, miR397 and miR162), which appear to be up-regulated upon the application of any stress as compared to the control plants. Surprisingly, miR398 and miR408 families show consistently extremely higher levels of accumulation, with log2FC values between 5 and a maximum of 9.82 under combined stress conditions involving cold and/or drought. On the other side, there is also a wide scope of accumulation levels of down-regulated reads, especially relevant for miR156, miR166, miR167, miR169, miR171, miR396 families, and so on, reaching log2FC values up to -8 at most. Additionally, as mentioned in the previous section 4.2, some differentially expressed miRNAs in the heatmap of Figure 4.2 have not been identified as 5P or 3P yet, hence no "5P" or "3P" label has been assigned to them, since the alignment to any of the precursors was not successful.

4.4 Cold predominance disguises the effect of other stress conditions

Another characteristic pattern of the differential expression analysis is that melon miRNAs identified as responsive to multiple stress conditions (double and triple combinations) show almost no differential expression in plants exposed to high salinity (SA), *M. cannonballus* infection (MON) and short-day (SD) conditions, either up- or down-regulation. The only simple stress situations in which stress-responsive miRNAs apparently modify their expression significantly are

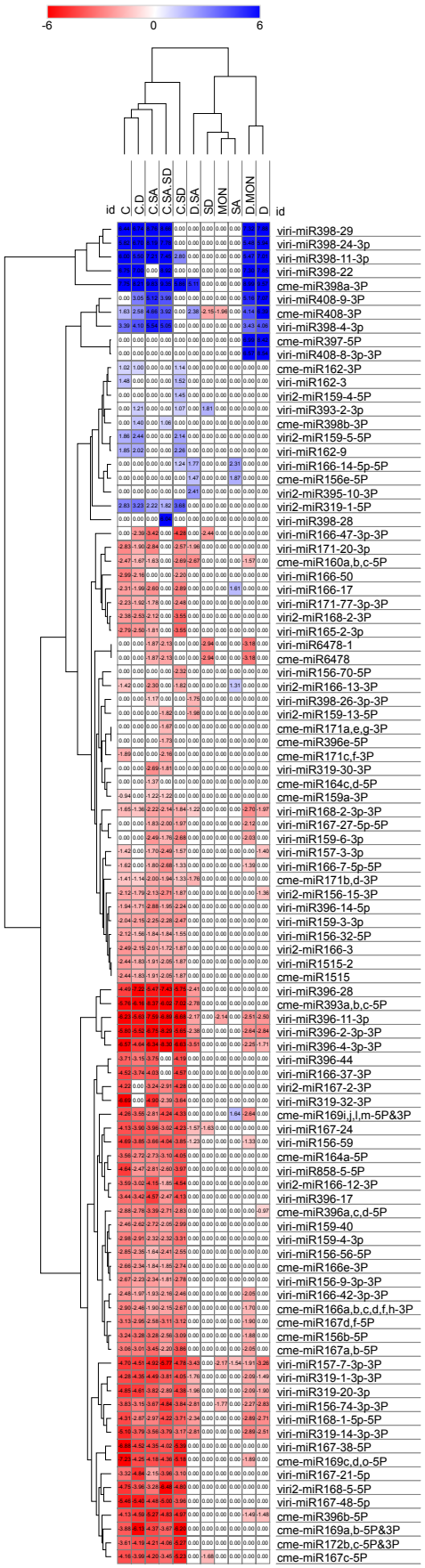


Figure 4.2: Heatmap representation of the differentially expressed sRNA filtered by length and FDR values based on the three differential expression methods. Down-regulated miRNAs appear in red, whereas up-regulated are gradually labelled in blue.

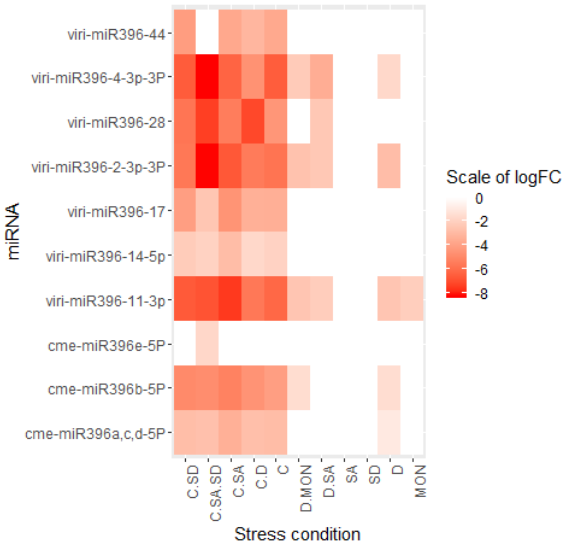
cold treatment and drought. Indeed, it is especially remarkable the predominant effect of cold temperatures, since the direction of the regulation determined by this stress is maintained both qualitatively and quantitatively in combined multiple stress treatments, in addition to the fact that cold is the environmental cue that induces more changes in miRNA expression levels (80% of the sequences) than any other. This is depicted in Figure 4.2 for the core of the down-regulated miRNA mature reads (significant in C, C.SA, C.SA.SD, C.SD and C.D libraries), as well as for the up-regulated miR398 and miR408 families. In this sense, drought stress also operates in a similar way, regarding the double combination of drought and *M. cannonballus* exposures (D.MON), which presents the same level of alteration (similar log2FC) than drought libraries themselves. This fact is mainly noticed for up-regulated miRNA families (miR397, miR398, miR408) and some down-regulated sequences. However, upon the simultaneous interaction of cold and drought, cold prevalence veils any further modulation by drought treatment.

To evaluate the similar behaviour of miRNA members of the same families in terms of differential expression under determined stress conditions, heatmaps and boxplots were plotted by family, as shown in Figure 4.3 for three specific miRNA families among the 23 in total. The selected plots for this section include a representative pattern for significantly down-regulated miRNA families as compared to controls in Figures 4.3a and 4.3d for miR396, and the miR398 family in Figures 4.3b and 4.3e, which present a high diversity of mature miRNA, differentially up-regulated in response to most of stress conditions. Generally, across the 23 different families, the majority of their members follow a shared direction or tendency in terms of differential expression. In other words, there are only a few stress-responsive miRNA whose differential accumulation sense is opposed to the rest of the cluster members, as observed in miR159, miR166 or miR393. Furthermore, in the miR319 family plots in Figures 4.3c and 4.3f, the mature miRNA viri2-miR319-1-5P is up-regulated under cold or combinations of cold and other adverse situations and identified as an outlier for some libraries upon the representation of the Boxplot (Figure 4.3f), whereas the rest of the members, which have been classified as "3P" melon miRNAs, are down-regulated. This may indicate a functional alternative processing and accumulation of 5P and 3P reads, as already manifest in one of our previous research experiments [12].

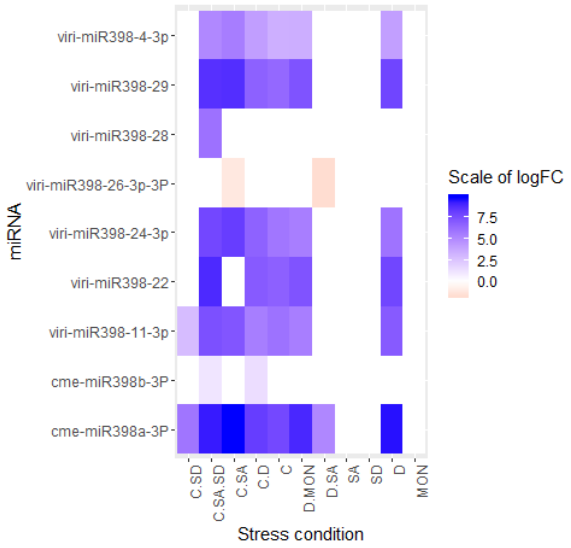
4.5 Small RNAs accumulation profiles onto *C. melo* miRNA precursors

Since in the miRBase database there are only deposited and characterised known mature and hairpin miRNAs, we addressed the issue of the identification of putative miRNAs among all the sRNA molecules, which may originate as a result of any form of alternative processing of existing miRNA precursors or pre-miRNAs. After the alignment of all differentially expressed sRNA reads to the miRNA precursors with Bowtie, the read counts associated to mapped sequences to observe the miRNA accumulation profiles. Thus, these profiles computed in Python are shown in Figure 4.4. From the initial 141,838 stress-responsive (differentially expressed) sRNAs identified in the constructed master table with differential expression data, only 40,117 passed the probabilistic and log2FC filters. In addition, from the 799,492 unique sRNA reads, only 538 aligned to the miRNA precursors, and after collapsing the data from the 40,117 differentially expressed sRNAs, 381 sRNAs remained out of 538. The majority of the sRNAs (40.7%) are 21-nt long, the expected length of a mature miRNA, followed by 20-nt long (26.6%) and 22-nt long (20.7%). Moreover, the final 43 identified pre-miRNAs are grouped into 16 families.

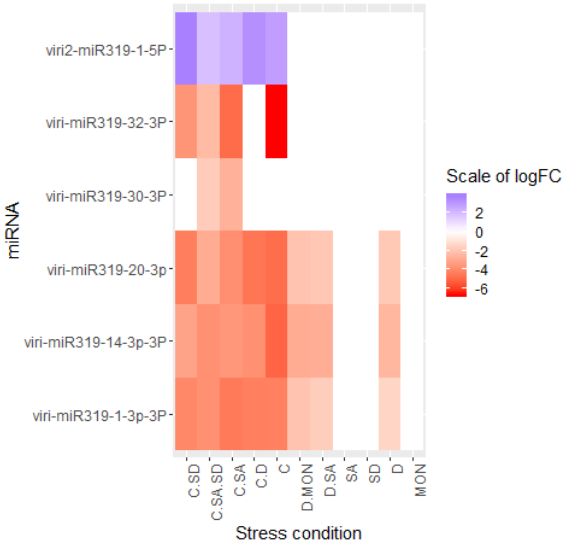
Here, we present the results just for a representative group of miRNA stem-loop hairpins. The usual profile for the accumulation of putative miRNAs, originated from a true hairpin is the one displayed by miR156d in Figure 4.4a and miR166b in Figure 4.4b, as well as the majority of the characterised precursors (more than 75% of them). Specifically, the characteristic profile



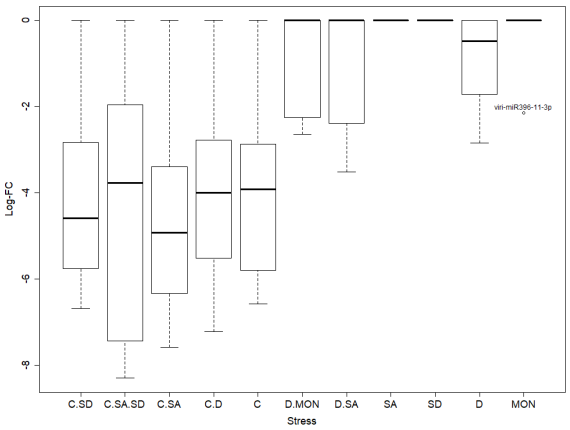
(a) Heatmap of miR396 family



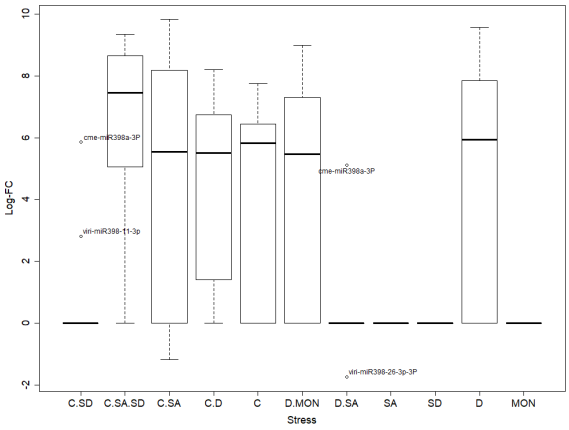
(b) Heatmap of miR398 family



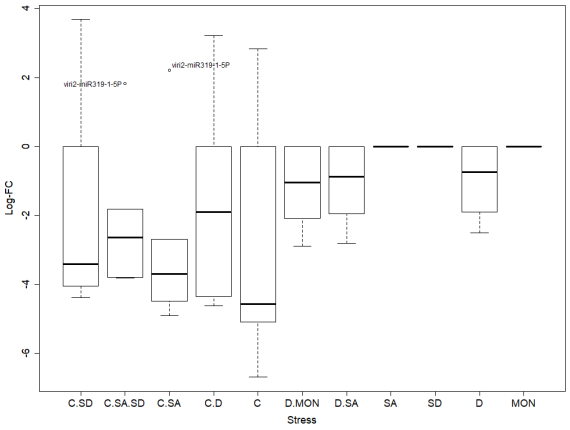
(c) Heatmap of miR319 family



(d) Boxplot of miR396 family

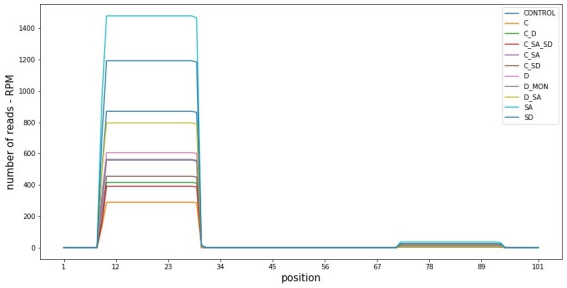


(e) Boxplot of miR398 family

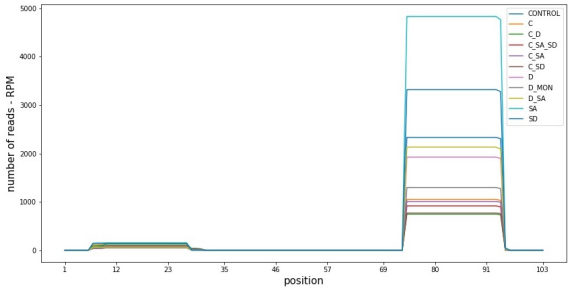


(f) Boxplot of miR319 family

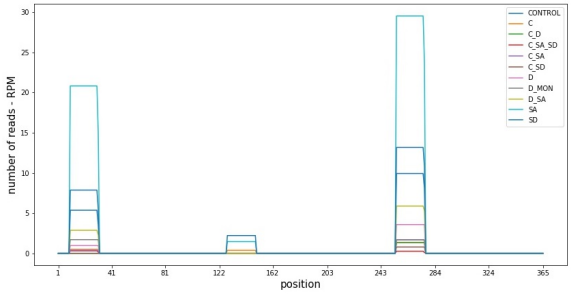
Figure 4.3: Representation of miRNA differential expression collapsed by the 23 families with heatmaps and boxplots.



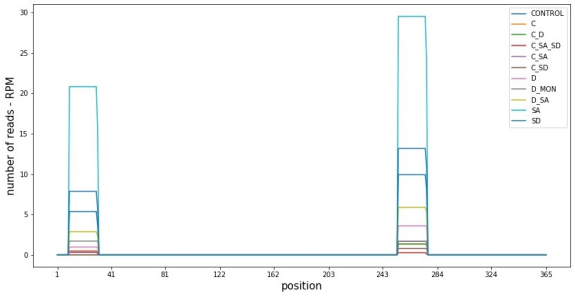
(a) sRNA accumulation profile onto the miR156d precursor.



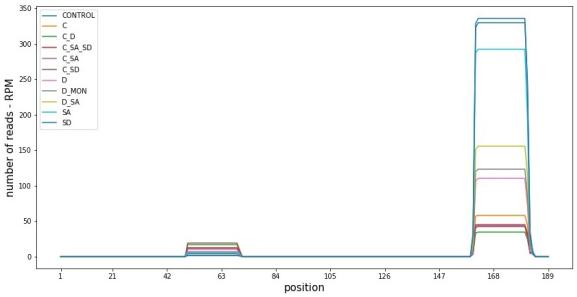
(b) sRNA accumulation profile onto the miR166b precursor.



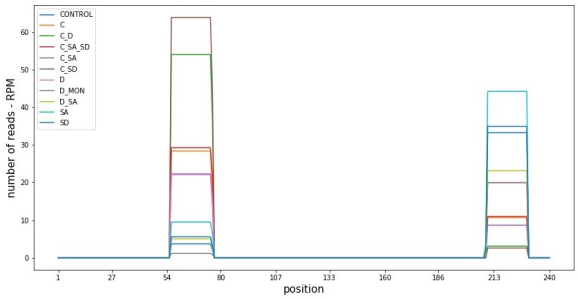
(c) sRNA accumulation profile onto the miR169a precursor.



(d) sRNA accumulation profile onto the miR169b precursor.



(e) sRNA accumulation profile onto the miR319a precursor.

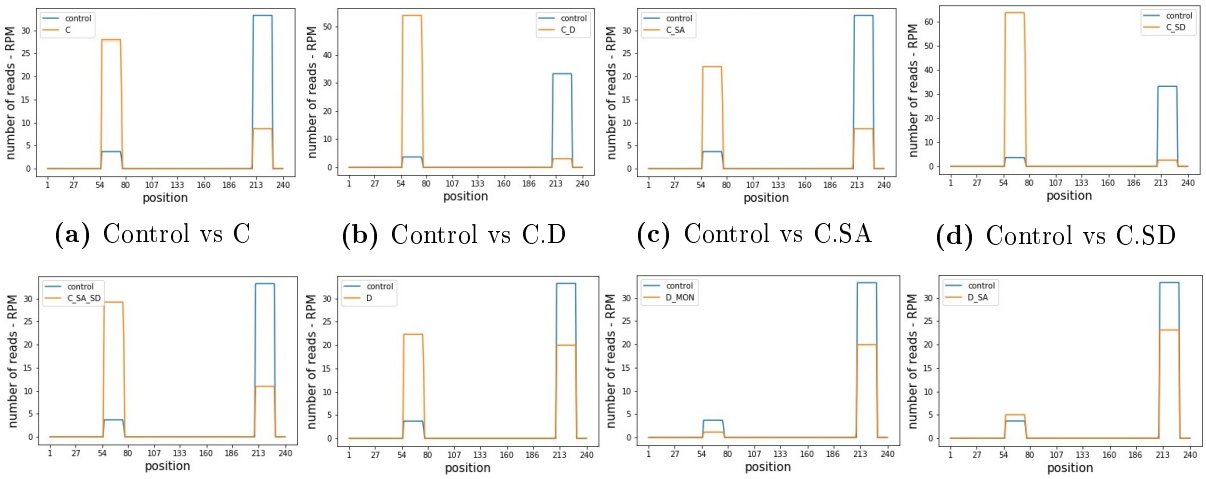


(f) sRNA accumulation profile onto the miR319c precursor.

Figure 4.4: Representation of sRNA accumulation profiles on four characteristic precursor families: 156,166, 169 and 319.

entails the massive transcription of 21-nt long reads from just one region of the precursor, which may be located at the 5P strand, like the miR156d hairpin, or at the 3P region, similarly to miR166b. However, there are some interesting situations which demand a careful attention to investigate them. On the one hand, the miR319 hairpin family is a good representative case, since the putative mature miRNAs emerging from the 3P region in the miR319a precursor (Figure 4.4e) are widely more abundant than the ones transcribed from the 5P region, as the normal situation. Conversely, miR319c precursor is processed in such a way that sRNA which are potential miRNAs accumulates almost equally from the two candidate regions of the precursors, as seen in Figure 4.4f. Furthermore, a differential accumulation profile dependent on the type of stress exposure is observed as well, i.e. cold-related stress combinations favour the accumulation of mature reads from the 5P strand, whereas other stress situations cause the accumulation of 3P-derived putative miRNAs. This alternative processing is shown in Figure 4.5.

Finally, the accumulation profile of the miR169 precursor family is even more complex, as an additional third peak also appears between the other two main miRNA-generative candidate



(a) Control vs C **(b)** Control vs C.D **(c)** Control vs C.SA **(d)** Control vs C.SD

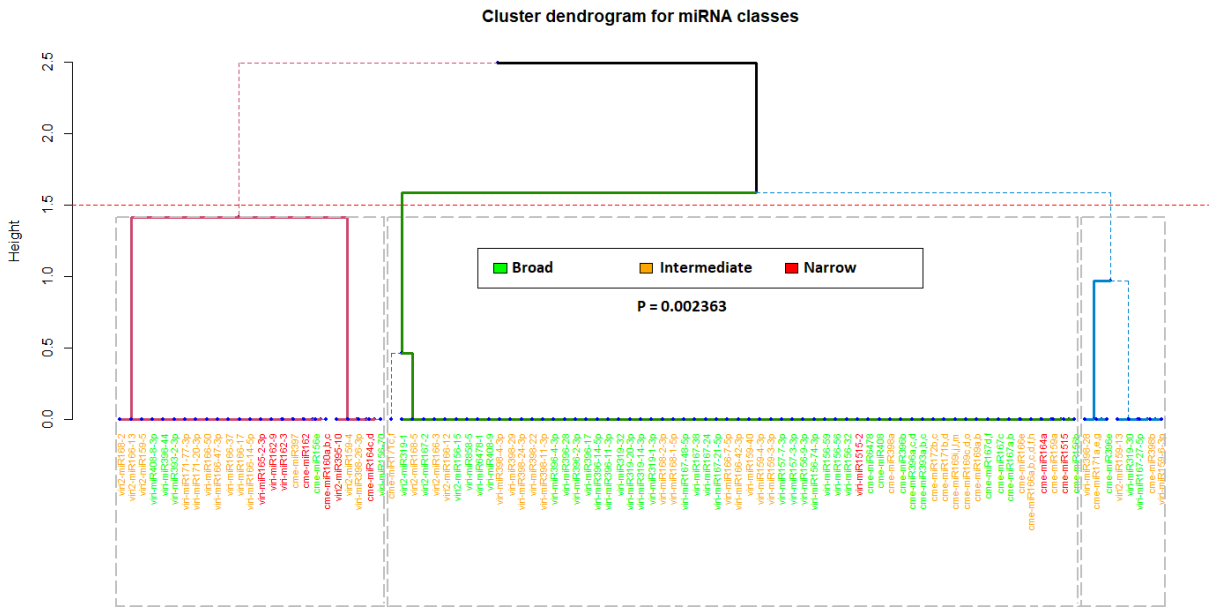
(e) Control vs C.SA.SD **(f)** Control vs C.D **(g)** Control vs D.MON **(h)** Control vs D.SA

Figure 4.5: sRNA accumulation profiles of control vs single stress samples for miR319c precursor.

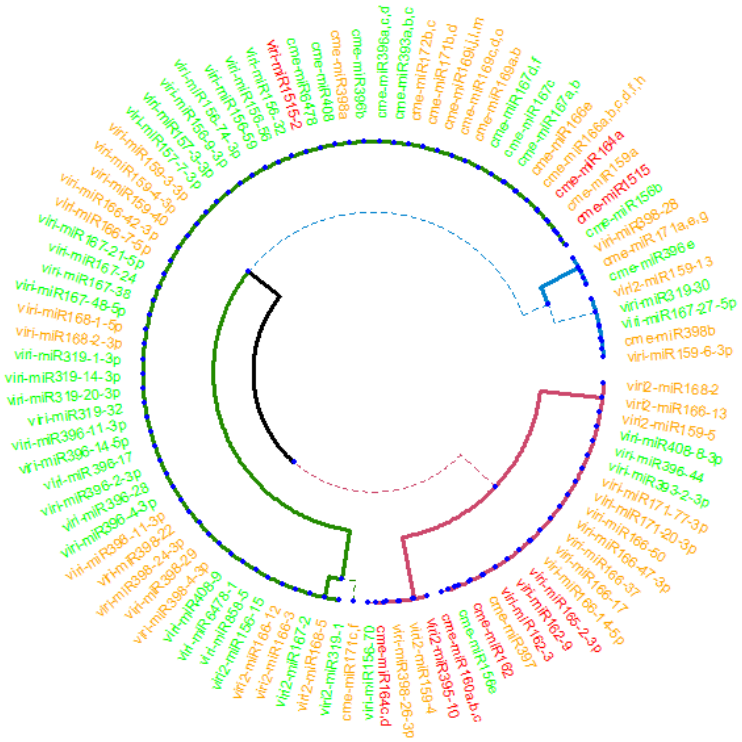
regions only in one of the precursors: miR169a (Figure 4.4c), though it is not noticeable in the other members of the family (Figure 4.4d). Besides, as opposed to the stress-dependent alternative processing displayed by miR319c hairpin, for the miR169 cluster the stress condition does not determine the selective accumulation of reads generated from one region, otherwise it is the level of expression of both regions which is dependent on stress. In other words, independently of the adverse situation, the 3P end always present a slightly more significant reads accumulation, and the difference between the 5P and 3P regions is maintained over all the libraries. Thus, high salinity (SA) libraries present the highest accumulation of putative stress-responsive miRNAs for the miR169 precursors, followed by control and short-day samples.

4.6 miRNA-mediated network is organised according to their stress-responsiveness range

The next step of the characterisation of the stress-responsive miRNAs was the construction of the miRNA-mediated network and the computation of some coefficients that may describe properly its nature. For this purpose, the elaboration of a presence/absence binary data frame was vital, to determine whether a miRNA was responsive to any stress conditions at three levels of exposure: simple, double and triple stress combinations. From this table, a clustering analysis was performed in order to try to classify them as the miRNA clusters according to their stress-responsiveness range proposed previously [48]: broad, intermediate and narrow response range. Therefore, in Figure 4.6, two visualisations of the Ward clustering are shown, a tree-shaped cluster dendrogram (Figure 4.6a) and a circularised dendrogram (Figure 4.6b). In both graphs, the 97 miRNAs have been clusterised into three clades (from left to right: red, green and blue), and miRNA classes have been associated to three different colours: green for broad miRNAs, orange for the intermediate ones, and red for narrow or specific miRNAs. As we may observe, 7 out of the 10 narrow-responsive miRNAs are classified within the first, red clade, along with the intermediate stress-responsive miRNAs. Although broad miRNAs are distributed more or less uniformly, they are the principal components of the green clade, in addition to their abundance in the blue clade along with the intermediate miRNAs in similar proportions. Hence, the dendrogram clustering enhance the comprehension of the classification of miRNAs based on stress-response range.



(a) Three-clade dendrogram representing miRNA stress-responsiveness classes. The shown p-value comes out from the Kruskal-Wallis test performed over the three different groups.

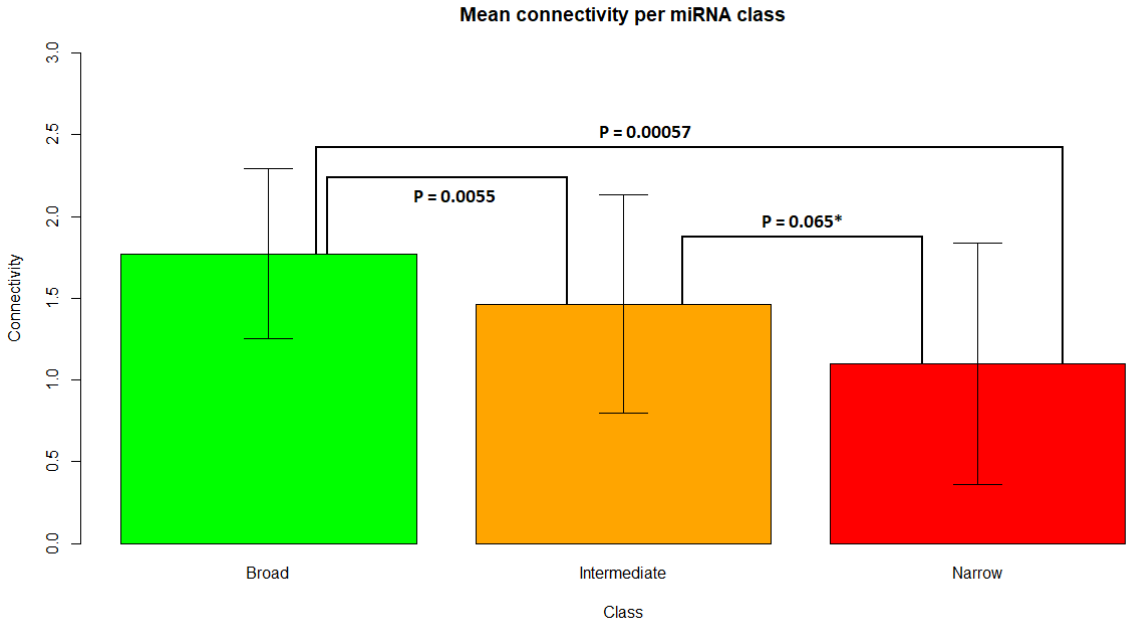


(b) Circular dendrogram showing the same classification of miRNAs into three clades.

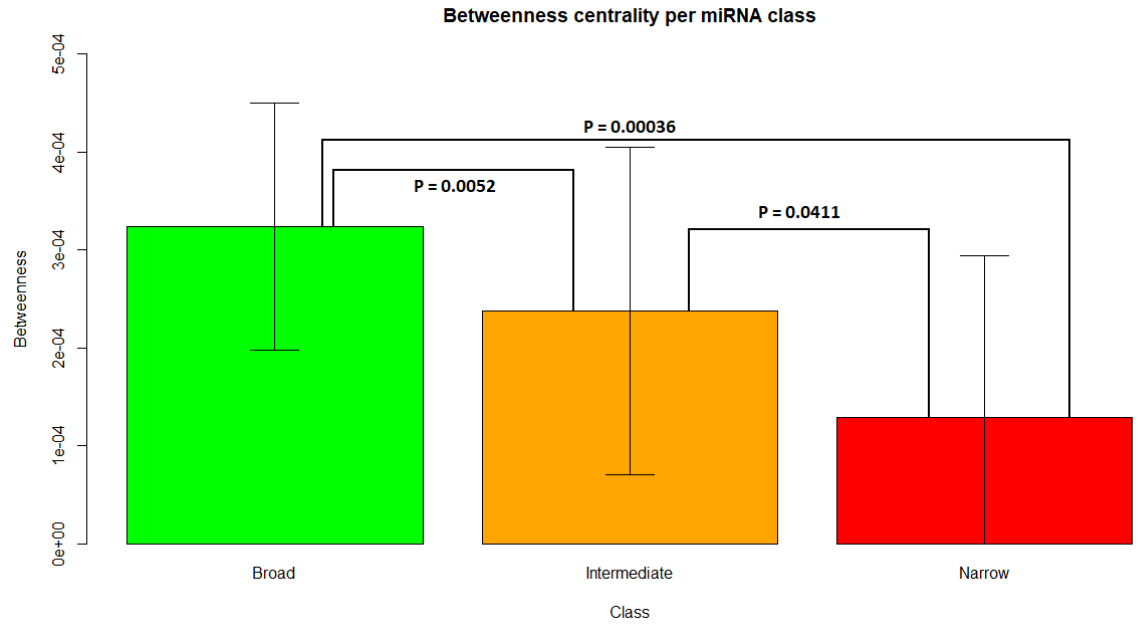
Figure 4.6: miRNA network organised as clusters according to the stress-responsiveness range of miRNA sequences.

To support the evidences for the significant differentiation of miRNAs into classes, we computed two measures of centrality or connectivity degree per each class. The first calculation, called "mean connectivity", involved the mean sum value of the binary table, being 2 the maximum score if the determined miRNA is differentially expressed at any simple, double and triple combinations. Secondly, values of betweenness centrality of each of the nodes (miRNAs) were

also computed as an additional measure of node connectivity. For their visualisation, bar plots were produced for the "mean connectivity" in Figure 4.7a and for the "betweenness centrality" parameter in Figure 4.7b. The p-values calculated with the non-parametric Mann-Whitney-Wilcoxon test in pairs shown the statistical significance in the difference between the mean values of two non-paired classes, i.e. broad vs. intermediate, intermediate vs. narrow and broad



(a) Mean connectivity values for the three miRNA response-range classes.



(b) Betweenness centrality values for the three miRNA response-range classes.

Figure 4.7: Barplots representing the miRNA connectivity of broad, intermediate and narrow classes for both mean connectivity (4.7a) and betweenness centrality (4.7b). The higher these values, the more connected is the miRNA node. Broad miRNAs are coloured in green, intermediate in orange, and narrow miRNAs are represented by the red columns.

vs. narrow.

Since ANOVA model was not possible to be applied due to the lack of normal distribution of the libraries, an essential assumption for ANOVA, Kruskal-Wallis is used instead, which is the non-parametric version of the ANOVA. As depicted in Figure 4.7, both measures of connectivity differs between classes, having broad-response miRNAs as the most connected components (1.77 for mean connectivity out of a maximum of 2), followed by the intermediate (1.46) and narrow miRNAs (1.10). What is more relevant is that all differences in terms of mean connectivity and betweenness centrality, except for the pair "intermediate-narrow" for the mean connectivity, with a p-value higher than 0.05 (0.065). As for the classes assignation, 44 miRNAs are considered as general or broad, 43 are labelled as intermediate and 10 of them are narrow or specific. Hence, a higher sample of narrow miRNAs might be necessary to better assess the significance in terms of differences in connectivity.

In the end, the final representation of the full miRNA-mediated network consisted in a graph composed of the 97 miRNAs as nodes, grouped by the three categories (green for broad, orange for intermediate and red for specific miRNAs), as shown in Figure 4.8. The 97 miRNA nodes are connected to other 3 nodes representing the five simple stress conditions ("Simple stress" node), five double combinations ("Double stress" node) and the triple combination of cold, salinity and short-day ("Triple stress" node) depending on their differential expression in at least one of the referred libraries per group. Thus, in this Figure 4.8, it is evident the lack of connectivity of narrow-response miRNA nodes to the triple stress, whereas almost all the 10 miRNA nodes are linked by edges to double and simple stress nodes, i.e. specific miRNAs are usually differentially expressed under specific, simple stress situations. On the contrary, intermediate- and broad-range miRNAs are involved more uniformly in the three types of multiple stress conditions, as they show a higher connectivity and betweenness centrality.

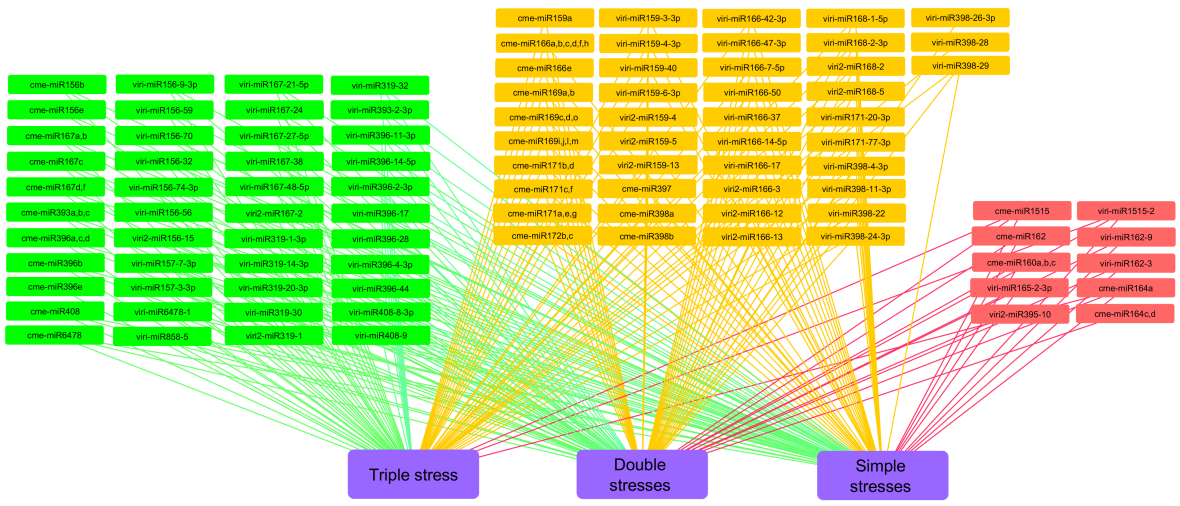
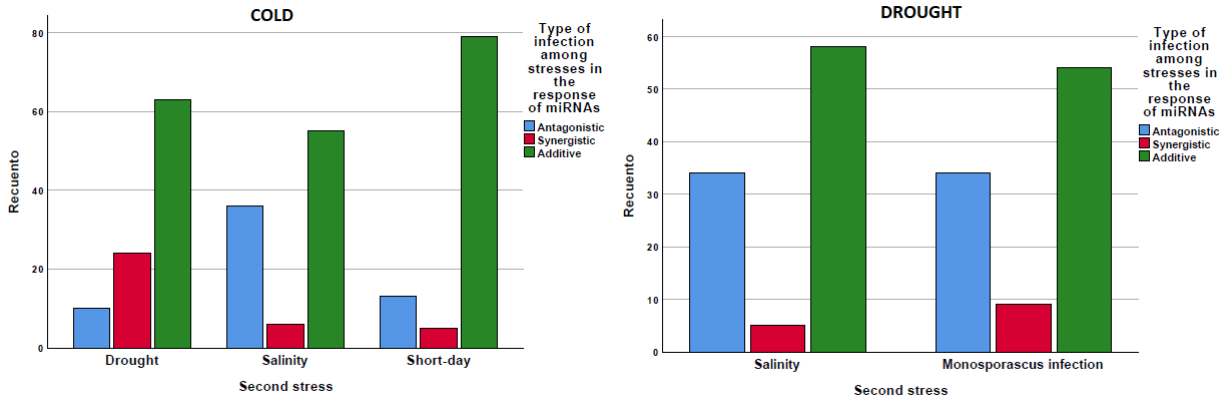


Figure 4.8: Representation of the miRNA network showing the specific weight and relevance of the three response-range classes within the multiple stress situations (simple, double and triple stresses).

4.7 The additive effect is predominant on miRNA differential expression upon multiple stress conditions

To study the alteration effect on the miRNA expression triggered by the combination of multiple stresses, we performed a pairwise analysis of the superposition between each pair of stresses, in which the magnitude and direction of the alteration is measured by the normalised epistasia calculation. The difference between the combined control reads and the multiple stress as compared to the sum of the two individual stresses determines the sign of the epistatic interaction, hence the deviation from the expected additive effect. A negative epistasia means an antagonistic interaction among two stress conditions (the combined effects are lower than the sum of the individual stresses), whereas a positive one is related to synergistic effects (the combined effects are higher than the sum of the individual stresses).

As for the first set of paired comparisons in Figure 4.9, cold treatment with respect to a second stress, as well as the effect of drought along with another stress both display the same profile, this is, around a 55% or over of additive effects on miRNA expression, followed by an antagonistic effect (40% approximately), and the synergistic being the least common. Upon the application of salinity (SA) and *Monosporascus* infection (MON) as the second stress this proportions are maintained both for cold (Figure 4.9a) and drought (Figure 4.9b) as the main stresses, respectively. However, for cold treatment there is an unbalance upon the combination with short-day conditions, since almost 80% of the cases of the alteration of miRNA expression present an additive effect of both conditions. More importantly, in combination with drought as a second stress there is a drop in the number of antagonistic relationships to 10 cases, and an increase in synergistic effects. Remarkably, most of these 10 miRNAs displaying an antagonistic combination are members of the miR398 and miR408 families, belonging to the cluster of up-regulated miRNAs in Figure 4.2. Conversely, the enrichment in synergistic effects is explained by miRNA sequences, the majority of which belong to the down-regulated families, e.g. miR396, miR167 or miR171.



(a) Stress interaction effects with cold as the main adverse condition.

(b) Stress interaction effects with drought as the main adverse condition.

Figure 4.9: Graphical representation in bar plots rendered by the SPSS program, accounting for the interaction effects affecting stress-responsive miRNAs under cold (Figure 4.9a) and drought (Figure 4.9b) multiple stress conditions.

5

Discussion

The advent of sequencing technologies that revealed the existence of the regulatory sRNAs has led to the identification and characterisation of the first miRNAs in plants and their associated functionality [9, 11, 23, 61]. The recent collection of high-throughput data has reassert the regulatory role of miRNAs as key elements governing a wide variety of global mechanisms controlling plant development and response to environmental fluctuations that alter the homeostatic state of the plant. Hence, miRNAs provide an extensive regulatory layer to counteract the adverse cascade effects triggered by stress situations by the modulation of the expression of crucial stress-tolerant genes [2, 33, 54]. However, although the mode of action of stress-responsive miRNAs under specific prejudicial situations has been practically unveiled, there is an important lack of understanding in the miRNA modulation upon the interaction of multiple stress conditions, which remains obscure.

To address this unanswered issue, we questioned the sRNA libraries in order to identify miRNA reads involved in the stress response, hence giving a profound insight into multiple stress interactions. As a first step, the highest proportion of the characterised sRNAs resulting from a general description of the sRNA libraries were 24-nt long. This indication has been previously suggested in transcriptomic pipelines analysing *C. melo* and other *Cucurbitaceae*, in which the predominance of 24-nt sRNAs was manifest [23, 24, 26]. It is worth noticing that miRNA expected length is 21 nt, according to the biogenesis process [1, 2, 59], therefore our focus was based on the identification of 21-nt long sRNA for the differential expression analysis (and 22-nt long to recover possible erroneously extended miRNA phases in the DCL-cleavage).

After the statistical confirmation of the robustness of the replicates conforming the sRNA libraries by the calculation of Euclidean distance between and among samples in PCA, we analyzed the processing of primary and precursor miRNAs (pri-miRNA and pre-miRNA) and the consequent stabilization of the miRNA/miRNA* duplex at the 5' arm, and the equivalent duplex at the 3' region, henceforth named 5P and 3P regions of the pre-miRNA. From the 401 known miRNAs identified in the libraries, the fact that the more stable melon miRNA sequences are mainly derived from the 5P region, whereas mature miRNAs described in other *Viridiplantae* organisms are potentially expressed from the 3P end of the precursor, may indicate a differential processing dependent on some specific molecular events. For instance, coupling between pre-miRNA and intron splicing has been mentioned to have a significant effect on the preferential accumulation of one of the miRNA duplexes, as the presence of introns at the 3P region and

the location of the splice site at the 5' end favour the selective loading of 3P-derived miRNAs into ARGONAUTE proteins, hence an increase in abundance [1]. Also, the partial cleavage of the 5' arm of the stem-loop may affect [12]. Although the processing of most pre-miRNAs results in the preferential accumulation of one of the strands from the miRNA/miRNA* duplex, usually named as "mature miRNA" or "functional miRNA", there are evidences that under certain circumstances its complementary strand is predominant, calling into question the nomenclature used to date. Therefore, our assignment of aligned *Viridiplantae* and melon miRNA sequences as mature melon miRNAs represents the first attempt to rename miRNAs according to the 5P and 3P regions of hairpin precursors in melon, as being already performed for other species in miRBase [32].

Furthermore, the performance of the differential expression analysis was carried out with the renamed miRNA reads. The use of three different differential expression algorithms (edgeR, DESeq2 and NOISeq) provided an additional level of robustness for the conclusions of the work. Thus, our results indicate that only a small group of known miRNA sequences are characterised as truly stress-responsive, concretely 97 miRNAs are differentially expressed according to the applied filters, which are grouped into 23 families, as shown in the general heatmap (Figure 4.2). Indeed, the highly restrictive filters are employed in view of the increasing amount of data generated by high-throughput technologies, hence there is a need for minimizing the false positive events (in other words, the False Positive Rate) [4]. Specifically, the identified stress-responsive melon miRNAs have already been reported as reactive to both abiotic and biotic stress conditions in other plant species, e.g. miR393, miR167 and miR397 in rice; miR160 and miR164 in maize; miR395 in wheat, miR164 and miR396d in barley; as well as others in soybean, cowpea and *Arabidopsis thaliana* are some representative examples [2, 3, 19, 61].

Most of the stress-reactive miRNAs are down-regulated (around 80% of them), whereas specific miRNA families (miR398, miR397 and miR408) displays significant up-regulated expression levels. The functional and biological implications of the direction of these alteration of miRNA accumulation must follow a general theory of gene regulation, which implies that genes involved in common functions critical for plant development such as plant growth, or hormone regulation are supposed to be positively regulated or activated in a natural state, since their action is pivotal for the plant. However, other genes associated to low-demanded functions are usually negatively regulated or repressed, e.g. stress-responsive genes [49]. Since miRNAs are naturally intended for target repression, it is logical that, upon a prejudicial situation, miRNA levels descend, so that target transcripts accumulate, and repressed genes involved in stress response are activated, although other molecular mechanisms might explain the sense of the interplay between miRNAs and targets.

Therefore, those targets activating crucial genes for stress regulation are normally transcription factors. In this sense, the majority of down-regulated miRNAs are targeting transcription factors, which are related to plant development. For instance, miR156 family interacts with SPL protein and miR172 with APETALA2, hence regulating adult transition from the juvenile phase, in addition to miR319 interaction with TCP and miR172-APETALA2 controlling the flowering process. Other miRNA families repressing transcription factors under normal conditions are miR396, or miR164 [47]. Conversely, other miRNAs seems to be activated under certain stress conditions inducing signal transduction, e.g. miR169 and miR319 as well, which may explain the identification of some up-regulated members of this family in the master heatmap in Figure 4.2. Notably, the miRNA phase generated from the 5P strand from another miRNA duplex, which is up-regulated as depicted in Figure 4.3c, indicates the existence of an alternative processing of the pre-miR319, already described in a previous experiment of our group [12]. Although the functional implications have not been elucidated, there are enough evidences to consider that cold-related stress situations trigger the misprocessing of the 5' arm, leading to the down-regulation of 3P miRNA reads and the accumulation of the 5P miRNA in melon

(Figure 4.5). Moreover, up-regulated miR397, miR398 and miR408 families have been clearly identified as regulatory miRNAs affecting copper and cadmium processing in *Brassica napus* and other species [22]. Nevertheless, it is more logical for these miRNAs to play a general role on regulation of altered ionic homeostatic state, rather than cope with a low presence of copper ions as a consequence of the effects of cold- and drought-responsive pathways.

Precisely, in regard to cold stress, low temperatures are observed to be the predominant stress causing a massive down-regulation or up-regulation of the implicated miRNA genes, which represents a total of 90% of the overall miRNA sequences. Furthermore, miRNAs involved in cold-complex stress conditions(C.SA, C.SD, C.SA.SD, C.D) are differentially expressed in the same direction as the samples exposed to individual cold stress, this is, cold is the stress condition determining the global response of the plant, in such a way that the level of expression of a given gene controlled by a cold-responsive miRNA is maintained upon the combined presence of low temperatures and diverse stress conditions. The scope of action underlying the exceptional subjection to cold conditions is so extensive that multiple pathways are affected, including ROS signalling, trigger of calcium influx, activation of kinase cascades, DNA methylation, chromatin organization and histone modifications. The systemic response of the plant is hence so complex that it may mask the effective response to other abiotic stress conditions, when they are applied in combination with low temperatures [39]. Additionally, complex stresses with drought as the dominant condition are observed as well (for D.SA and D.MON libraries), although the miRNA accumulation levels are shortly affected by drought exposure upon the combination with cold stress, as we have noticed. Moreover, as shown in Figure 4.2 and 4.1b, the combination of salinity with drought does not entail a similar stress response as compared with D and D.MON libraries, since salinity may trigger different molecular pathways in melon. In short, even though previous experiments have described the detrimental changes for plants upon cold and other abiotic alterations, further research is required in order to provide a profound insight into the predominance of particular stress conditions [33, 51, 52].

Regarding sRNA alignment to miRNA precursors, to discover non-described miRNAs in any database but potentially generated and present in our sRNA libraries, a similar pattern to the alternative processing of miR319c was noticed for sRNA accumulation profiles onto melon miRNA precursors. Generally, it was certainly believed that only one of the strands of the miRNA/miRNA* duplex was accumulated and fully functional as it was solely incorporated to an AGO protein, hence loaded onto the RISC complex. Consequently, the expected uniform accumulation profile is the one shown for miR156d and miR166b in Figures 4.4a and 4.4b, with a single region giving rise to the mature guide miRNA strand, while the other, the passenger, is degraded [27, 60]. However, many studies have provided strong evidences that, occasionally, at specific times and specific locations, the stability of the passenger strand is improved due to changes in miRNA precursor folding and structure, in such a way that the accumulation magnitude of the passenger is even as abundant as the guide miRNA. This situation is observed in the case of miR319c in Figure 4.3c and 4.4f, as well as for many other stem-loop melon hairpins [36, 60].

These results are explained by the renamed nomenclature of mature miRNAs, substituting the miRNA/miRNA* terminology referred to the fact that this passenger read was always degraded, as we have justified previously [32]. According to the recent discoveries, the two strands from the duplex were renamed as 5P and 3P, attending to their origin region within the precursor. Nevertheless, either 5P or 3P might become the functional strand, or even accumulate both under some specific stress conditions, as we discussed. In addition, there is a particular accumulation profile, for the miR169a precursor (Figure 4.4c), in which a small set and low abundance putative miRNAs are originated from a third region of the precursor, which does not appear in other precursors of the family. It may occur that miR169 hairpin is a dimerised pre-miRNA, and the intermediate phase is a combination of the 3' and the 5' regions of two

precursors.

Besides, another vital point for our analysis was the confirmation of our hypothesis that miRNAs are classified according to the range of the stress response into broad, intermediate and narrow miRNAs. In our group, we reported in previous works a custom assignment of miRNA sequences attending to the number of stresses in which they were implicated, this means, miRNAs present in 1 or 2 stress conditions were categorised as narrow response-range miRNAs, the intermediate miRNAs were present in 3 to 4 cues, and if they are identified in 5 or more stress libraries, they are considered as broad miRNAs [48]. Here, our aim was to provide further evidences of this clustering, by confirming that there are significant differences in terms of presence of mature miRNAs playing a critical role in the regulation of simple, double and triple stress combinations. The final hypothesis implies that broad miRNAs will be implicated in the control of a more global response to diverse natures of stress, hence vital for the regulation of complex stress situations (double and triple), whereas narrow-range stress-responsive miRNAs may be only necessary for the regulation of very specific processes upon single stress conditions. The most complex situations require the action of key central miRNAs, governing the central core of plant developmental processes, which ensure survival and growth.

The first set of results regarding this classification using cluster dendrograms offered an unclear division of miRNA class members, although most of the narrow miRNAs seem to be clustered in a single cluster in Figure 4.6. Then, to obtain more relevant evidences, two parameters were calculated: the mean connectivity and the betweenness centrality of the members of the network. The statistically significant differences between grouped miRNAs reaffirms our established theory, this time from the point of view of multiple stress conditions, in which broad miRNAs are proved to be pivotal elements for an efficient response to combined stress, whereas narrow miRNAs play more specific roles. However, there was just a non-significant result for the computation of mean connectivity, involving the comparison between intermediate and narrow classes (p-value of 0.065). We could believe that there is no certainty about the differential behaviour or presence of these two groups of miRNAs, but attending to the validity of the Mann-Whitney test for betweenness centrality between those classes and to the small size of the narrow miRNA samples (n=10), we may consider that more members belonging to the narrow class are required to find a significant difference.

Finally, the brief overview about the interaction between different stress conditions, and the deviation from the expected additive effect provided additional evidences for the full characterisation of the miRNA network. Here, the relevance of cold stress was demonstrated again, since the differential expression levels of miRNAs fluctuated more with cold as the primary stress, in combination to another adverse situation. Interestingly, cold with short-day conditions leads for a variation of the accumulation of most miRNAs simply summing the individual differential expression values, whereas drought as a second stress operates inversely. Indeed, the synergistic miRNA accumulation for down-regulated genes often occurs under cold and drought samples, as well as the observation of the opposite behaviour of miR398 and miR408 as separated clusters (showing antagonistic effects under these conditions). This fact clearly distinguishes the downstream targets and functionality of miRNAs, either up-regulated (as miR398 and miR408 targets are usually copper and/or cadmium regulatory elements) or down-regulated (having transcription factors as the main targets). Anyway, this differential relationship is not observed for other multiple combinations. One plausible reason may be related to the strong influence of cold and drought stresses, as we have mentioned previously in this section. Although we may observe synergistic and antagonistic effects on miRNA expression upon multiple stress conditions, cold effects are so dramatic that these interaction effects are veiled by the initial differential response of the plant to low temperatures. In short, more research is required to elucidate the full scope of the interactions of cold and drought stresses on certain groups of miRNAs and decipher the patterns of miRNA differential expression upon multiple stress conditions.

6

Conclusion and prospective work

In conclusion, the precise notions provided by these analyses manifest the pivotal role of miRNA-mediated response to stress as one of the main mechanisms for melon plants to cope with the exposure to simultaneous adverse situations [33, 61, 63]. We have elucidated the components of this miRNA network, first by successfully identifying the region from which miRNAs are generated in melon, which is the first research work on this topic. Additionally, the general down-regulation events affecting miRNAs, especially under cold stress may indicate a potential background to investigate thoroughly the mechanistic action of the stress-response in melon. Furthermore, the lack of fully deciphered interaction effects between multiple stress conditions may represent another field of study to identify subsequent patterns of miRNA differential expression grouped by miRNA families or, even, miRNA responsiveness levels based on their significance, connectivity and centrality within the miRNA-mediated response network. More importantly, the comprehensive study of this network within the framework of an agronomically-relevant species like melon may entail the transference of this knowledge to breeding programmes in order to use miRNAs to develop new melon cultivars or varieties with the enhanced characteristic of tolerance to abiotic and biotic stress conditions [61].

Abbreviations

- **A**: *Agrobacterium tumefaciens* infection
- **AGO**: ARGONAUTE protein
- **AGO1**: ARGONAUTE 1 protein
- **ANOVA**: Analysis of Variance
- **C**: Cold
- **C.D**: Cold with drought
- **C.SA**: Cold with salinity
- **C.SA.SD**: Cold with salinity and short-day conditions
- **C.SD**: Cold with short-day conditions
- **D**: Drought
- **D.MON**: Drought with *Monosporascus cannonballus* infection
- **D.SA**: Drought with salinity
- **DCL**: DICER-LIKE protein
- **DCL1**: DICER-LIKE 1 protein
- **FDR**: False Discovery Rate
- **Ha**: Hectares
- **HEN1**: HUA1 ENHANCER protein
- **HST**: HASTY protein
- **HSVd**: *Hop Stunt Viroid* infection
- **HYL1**: HYPONASTIC LEAVES 1 protein
- **log2FC**: Log2-Fold Change
- **miRNA**: MicroRNA
- **MON**: *Monosporascus cannonballus* infection
- **NGS**: Next-Generation Sequencing
- **nt**: Nucleotide(s)
- **PC**: Principal Component

- **PCA**: Principal Component Analysis
- **Pol II**: RNA polymerase II
- **pre-miRNA**: Precursor miRNA
- **pri-miRNA**: Primary miRNA
- **RISC**: RNA-Induced Silencing Complex
- **RNA**: Ribonucleic Acid
- **rRNA**: Ribosomal RNA
- **RPM**: Reads per Million
- **ROS**: Reactive Oxygen Species
- **SA**: Salinity
- **SD**: Short-day conditions
- **SE**: SERRATE protein
- **siRNA**: Small interfering RNA
- **snoRNA**: Small nucleolar RNA
- **snRNA**: Small nuclear RNA
- **sRNA**: Small RNA
- **ta-siRNA**: Trans-acting small interfering RNA
- **tRNA**: Transfer RNA

Bibliography

[1] Achkar, N. P., Cambiagno, D. A., and Manavella, P. A. (2016). miRNA Biogenesis: A Dynamic Pathway. *Trends Plant Sci*, 21(12):1034–1044.

[2] 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(2):170–179.

[3] Axtell, M. J. and Bartel, D. P. (2005). Antiquity of MicroRNAs and Their Targets in Land Plants. *Plant Cell*, 17(6):1658–1673.

[4] Axtell, M. J. and Meyers, B. C. (2018). Revisiting Criteria for Plant MicroRNA Annotation in the Era of Big Data. *Plant Cell*, 30(2):272–284.

[5] Basso, M. F., Ferreira, P. C. G., Kobayashi, A. K., Harmon, F. G., Nepomuceno, A. L., Molinari, H. B. C., and Grossi-de Sa, M. F. (2019). MicroRNAs and new biotechnological tools for its modulation and improving stress tolerance in plants. *Plant Biotechnol J*, 17(8):1482–1500.

[6] Bolger, A., Lohse, M., and Usadel, B. (2014). Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics*, 30(15):2114–2120.

[7] Bologna, N. G., Iselin, R., Abriata, L. A., Sarazin, A., Pumplin, N., Jay, F., Grentzinger, T., Peraro, M. D., and Voinnet, O. (2018). Nucleo-cytosolic Shuttling of ARGONAUTE1 Prompts a Revised Model of the Plant MicroRNA Pathway. *Mol Cell*, 69(4):709–719.

[8] Bologna, N. G., Schapire, A. L., Zhai, J., Chorostecki, U., Boisbouvier, J., Meyers, B. C., and Palatnik, J. F. (2013). Multiple RNA recognition patterns during microRNA biogenesis in plants. *Genome Res*, 23(10):1675–1689.

[9] Borges, F. and Martienssen, R. A. (2015). The expanding world of small RNAs in plants. *Nat Rev Mol Cell Biol*, 16(12):727–741.

[10] Brown, J., Pirrung, M., and McCue, L. A. (2017). FQC Dashboard: integrates FastQC results into a web-based, interactive, and extensible FASTQ quality control tool. *Bioinformatics*, 33(19):3137–3139.

[11] Budak, H. and Akpinar, B. A. (2018). Plant miRNAs: Biogenesis, Organization and Origins. *Funct Integr Genomics*, 15(5):523–531.

[12] Bustamante, A., Marques, M. C., Sanz-Carbonell, A., Mulet, J. M., and Gomez, G. (2018). Alternative processing of its precursor is related to miR319 decreasing in melon plants exposed to cold. *Sci Rep*, 19(8):15538.

[13] Carthew, R. W. and Sontheimer, E. J. (2009). Origins and Mechanisms of miRNAs and siRNAs. *Cell*, 136(4):642–655.

[14] Cervera-Seco, L., Marques, M. C., Sanz-Carbonell, A., Marquez-Molins, J., Carbonell, A., Daròs, J.-A., and Gomez, G. (2019). Identification and Characterization of Stress-Responsive TAS3-Derived TasiRNAs in Melon. *Plant Cell Physiol*, 60(11):2382–2393.

[15] Clepet, C., Joobeur, T., Zheng, Y., Jublot, D., Huang, M., Truniger, V., Boualem, A., Hernandez-Gonzalez, M. E., Dolcet-Sanjuan, R., Portnoy, V., Mascarell-Creus, A., Caño-Delgado, A. I., Katzir, N., Bendahmane, A., Giovannoni, J. J., Aranda, M. A., Garcia-Mas, J., and Fei, Z. (2011). Analysis of expressed sequence tags generated from full-length enriched cDNA libraries of melon. *BMC Genomics*, 12(1):252.

[16] Courbier, S. and Pierik, R. (2019). Canopy light quality modulates stress responses in plants. *Plant Cell Physiol*, 22(1):441–452.

[17] de Felippes F. F., Marchais, A., Sarazin, A., Oberlin, S., and O., V. (2017). A single miR390 targeting event is sufficient for triggering TAS3-tasiRNA biogenesis in Arabidopsis. *Nucleic Acids Res*, 45(9):5539–5554.

[18] Department of Economic and Social Affairs - United Nations (2018). UN COMTRADE database. http://comtrade.un.org/data/. Accessed on 30-12-2019.

[19] Djami-Tchatchou, A., Sanan-Mishra, N., Ntushelo, K., and Dubery, I. (2017). Functional roles of microRNAs in agronomically important plants—potential as targets for crop improvement and protection. *Front Plant Sci*, 8:378.

[20] Dong, Z., Han, M. H., and Fedoroff, N. (2008). The RNA-binding proteins HYL1 and SE promote accurate in vitro processing of pri-miRNA by DCL1. *Proc Natl Acad Sci USA*, 105(29):9970–9975.

[21] Elliott, A., Leicht, E., Whitmore, A., Reinert, G., and Reed-Tsochas, F. (2018). A non-parametric significance test for sampled networks. *Bioinformatics*, 34(1):64–71.

[22] Fu, Y., Mason, A. S., Zhang, Y., Lin, B., Xiao, M., Fu, D., and Yu, H. (2019). MicroRNA-mRNA expression profiles and their potential role in cadmium stress response in *Brassica napus*. *BMC Plant Biol*, 19:570.

[23] 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.

[24] Herranz, M. C., Navarro, J. A., Sommen, E., and Pallas, V. (2015). Comparative analysis among the small RNA populations of source, sink and conductive tissues in two different plant-virus pathosystems. *BMC Genomics*, 16(1):117.

[25] IBM Corporation (Version 26.0, 2019). *IBM SPSS Statistics*. Armonk, NY, USA.

[26] Jagadeeswaran, G., Nimmakayala, P., Zheng, Y., Gowdu, K., Reddy, U. K., and Sunkar, R. (2012). Characterization of the small RNA component of leaves and fruits from four different cucurbit species. *BMC Genomics*, 13:329.

[27] Jin, H. Y., Gonzalez-Martin, A., Miletic, A. V., Lai, M., Knight, S., Sabouri-Ghomi, M., Head, S. R., Macauley, M. S., Rickert, R. C., and Xiao, C. (2015). Transfection of microrna mimics should be used with caution. *Front Genet*, 6:340.

[28] Johns Hopkins University (2019). Bowtie 2 Fast and sensitive read alignment. http://bowtie-bio.sourceforge.net/bowtie2/manual.shtml#how-is-bowtie-2-different-from-bowtie-1/. Accessed on 06-01-2020.

[29] Kalvari, I., Argasinska, J., Quinones-Olvera, N., Nawrocki, E. P., Rivas, E., Eddy, S. R., Bateman, A., Finn, R. D., and Petrov, A. I. (2018). Rfam 13.0: shifting to a genome-centric resource for non-coding RNA families. *Nucleic Acids Res*, 46(D1):D335–D342.

[30] Kinoshita, T. and Seki, M. (2014). Epigenetic memory for stress response and adaptation in plants. *Plant Cell Physiol*, 55(11):1859–1863.

[31] Kozomara, A., Birgaoanu, M., and Griffiths-Jones, S. (2019). miRBase: from microRNA sequences to function. *Nucleic Acids Res*, 47(D1):D155–D16.

[32] Kozomara, A. and Griffiths-Jones, S. (2014). miRBase: annotating high confidence microRNAs using deep sequencing data. *Nucleic Acids Res*, 42(D1):D68–D73.

[33] Kumar, R. (2014). Role of microRNAs in biotic and abiotic stress responses in crop plants. *Appl Biochem Biotechnol*, 174(1):93–115.

[34] Lakota, K., Thallinger, G. G., Sodin-Semr, S., Rozman, B., Ambrozic, A., Tomsic, M., Praprotnik, S., Cucnik, S., Mrak-Poljsak, K., Ceribelli, A., Cavazzana, I., Franceschini, F., Vencovsky, J., Czirják, L., Varjú, C., Steiner, G., Aringer, M., Stamenkovic, B., Distler, O., Matucci-Cerinic, M., and Kveder, T. (2012). International cohort study of 73 anti-Ku-positive patients: association of p70/p80 anti-Ku antibodies with joint/bone features and differentiation of disease populations by using principal-components analysis. *Arthritis Res Ther*, 14(1):R2.

[35] Li, R., Chen, D., Wang, T., Wan, Y., Li, R., Fang, R., Wang, Y., Hu, F., Zhou, H., Li, L., and Zhao, W. (2017). High throughput deep degradome sequencing reveals microRNAs and their targets in response to drought stress in mulberry (Morus alba). *PLoS One*, 12(2):1–30.

[36] Liu, W., Meng, J., Cui1, J., and Luan, Y. (2017). Characterization and Function of MicroRNA*s in Plants. *Front Plant Sci*, 8:2200.

[37] Martinez, G., Castellano, M., Tortosa, M., Pallas, V., and Gomez, G. (2014). A pathogenic non-coding RNA induces changes in dynamic DNA methylation of ribosomal RNA genes in host plants. *Nucleic Acids Res*, 42(3):1553–1562.

[38] Maza, E. (2016). In Papyro Comparison of TMM (edgeR), RLE (DESeq2), and MRN Normalization Methods for a Simple Two-Conditions-Without-Replicates RNA-Seq Experimental Design. *Front Genet*, 7:164.

[39] Megha, S., Basu, U., and Kav, N. (2019). Regulation of low temperature stress in plants by microRNAs. *BMC Plant Biol*, 19:570.

[40] Melonomics (2019). Melon genome database 3.6.1. https://www.melonomics.net/melonomics.html#/. Accessed on 31-12-2019.

[41] Ministerio de Agricultura, Pesca y Alimentación - Gobierno de España (2018). Superficies y producciones anuales de cultivos. https://www.mapa.gob.es/es/estadistica/temas/estadisticas-agrarias/agricultura/superficies-producciones-anuales-cultivos/. Accessed on 30-12-2019.

[42] Park, C. H., Yeo, H. J., Park, Y. E., Baek, S.-A., Kim, J. K., and Park, S. U. (2019). Transcriptome Analysis and Metabolic Profiling of Lycoris Radiata. *Biology*, 8(3):63.

[43] Qin, F., Shinozaki, K., and Yamaguchi-Shinozaki, K. (2011). Achievements and challenges in understanding plant abiotic stress responses and tolerance. *Plant Cell Physiol*, 52(9):1569–1582.

[44] Robinson, M. D., McCarthy, D. J., and Smyth, G. K. (2010). edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics*, 26(1):139–140.

[45] Roig, C., Fita, A., Ríos, G., Hammond, J. P., Nuez, F., and Picó, B. (2012). Root transcriptional responses of two melon genotypes with contrasting resistance to *Monosporascus cannonballus* (Pollack *et* Uecker) infection. *BMC Genomics*, 13:601.

[46] Ríos, P., Argyris, J., Vegas, J., Leida, C., Kenigswald, M., Tzuri, G., Troadec, C., Bendahmane, A., Katzir, N., Picó, B., Monforte, A. J., and Garcia-Mas, J. (2017). ETHQV6.3 is involved in melon climacteric fruit ripening and is encoded by a NAC domain transcription factor. *Plant J*, 91(4):671–683.

[47] Samad, A. F. A., Sajad, M., Nazaruddin, N., Fauzi, I. A., Murad, A. M. A., Zainal, Z., and Ismail, I. (2017). MicroRNA and transcription factor: key players in plant regulatory network. *Front Plant Sci*, 8:565.

[48] Sanz-Carbonell, A., Marques, M. C., Bustamante, A., Fares, M. A., G., R., and Gomez, G. (2019). Inferring the regulatory network of the miRNA-mediated response to biotic and abiotic stress in melon. *BMC Plant Biol*, 19(1):78.

[49] Savageau, M. (1998). Demand theory of gene regulation. I. Quantitative development of the theory. *Genetics*, 149(4):1665–1676.

[50] Shannon, P., Markiel, A., Ozier, O., Baliga, N. S., Wang, J. T., Ramage, D., Amin, N., Schwikowski, B., and Ideker, T. (2003). Cytoscape: A Software Environment for Integrated Models of Biomolecular Interaction Networks. *Genome Res*, 13(11):2498–2504.

[51] Shriram, V., Kumar, V., Devarumath, R. M., Khare, T. S., and Wani, S. H. (2016). MicroRNAs as potential targets for abiotic stress tolerance in plants. *Front Plant Sci*, 7:817.

[52] Silva, R. G. d., Rosa-Santos, T. M., França, S. d. C., Kottapalli, P., Kottapalli, K. R., and Zingaretti, S. M. (2019). Microtranscriptome analysis of sugarcane cultivars in response to aluminum stress. *PLoS One*, 14(11):1–12.

[53] Stepien, A., Knop, K., Dolata, J., Taube, M., Bajczyk, M., Barciszewska-Pacak, M., Pacak, A., Jarmolowski, A., and Szweykowska-Kulinska, Z. (2017). Posttranscriptional coordination of splicing and miRNA biogenesis in plants. *WIREs RNA*, 8(3):e1403.

[54] Sunkar, R., Chinnusamy, V., Zhu, J., and Zhu, J. H. (2007). Small RNAs as big players in plant abiotic stress responses and nutrient deprivation. *Trends Plant Sci*, 12(7):301–309.

[55] 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(2):721–738.

[56] Tarazona, S., Furió-Tarí, P., Turrà, D., Pietro, A. D., Nueda, M. J., Ferrer, A., and Conesa, A. (2015). Data quality aware analysis of differential expression in RNA-seq with NOISeq R/Bioc package. *Nucleic Acids Res*, 43(21):e140.

[57] Trindade, S., Sousa, A., Xavier, K. B., Dionisio, F., Ferreira, M. G., and Gordo, I. (2009). Positive epistasis drives the acquisition of multidrug resistance. *PLoS Genetics*, 5(7):1–9.

[58] Varet, H., Brillet-Guéguen, L., Coppée, J.-Y., and Dillies, M.-A. (2016). SARTools: A DESeq2- and EdgeR-Based R Pipeline for Comprehensive Differential Analysis of RNA-Seq Data. *PLoS One*, 11(6):e0157022.

[59] Wang, J., Mei, J., and Ren, G. (2019). Plant microRNAs: Biogenesis, Homeostasis, and Degradation. *Front Plant Sci*, 10:360.

[60] Yaish, M. W., Sunkar, R., Zheng, Y., Ji, B., Al-Yahyai, R., and Farooq1, S. A. (2015). A genome-wide identification of the miRNAome in response to salinity stress in date palm (*Phoenix dactylifera* L.). *Front Plant Sci*, 6:946.

[61] Zhang, B. (2015). MicroRNA: A new target for improving plant tolerance to abiotic stress. *J Exp Bot*, 66(7):1749–1761.

[62] Zhang, T., Zhao, Y.-L., Zhao, J.-H., Wang, S., Jin, Y., Chen, Z.-Q., Fang, Y.-Y., Hua, C.-L., Ding, S.-W., and Guo, H.-S. (2016). Cotton Plants Export microRNAs to Inhibit Virulence Gene Expression in a Fungal Pathogen. *Nat Plants*, 2(10):16153.

[63] Zhu, J. (2016). Abiotic stress signaling and responses in plants. *Cell*, 167(2):313–324.