Universidad de Oviedo

Departamento de Biología de Organismos y Sistemas

Programa de Doctorado: Biogeociencias

Identificación de nuevos biomarcadores de adaptación a estrés por altas temperaturas en pináceas mediante una aproximación de biología de sistemas.

Systems biology-driven identification of novel heat stress biomarkers in pine species.

TESIS DOCTORAL

Laura Lamelas Penas

Oviedo 2022

Universidad de Oviedo

Departamento de Biología de Organismos y Sistemas

Programa de Doctorado: Biogeociencias

lentificación de nuevos biomarcadores de adaptación strés por altas temperaturas en pináceas mediante un aproximación de biología de sistemas.

Systems biology-driven identification of novel heat stress biomarkers in pine species.

TESIS DOCTORAL

Laura Lamelas Penas

Directoras

María Jesús Cañal Villanueva

Mónica Meijón Vidal

Oviedo, Diciembre 2022

Universidad de Oviedo

RESUMEN DEL CONTENIDO DE TESIS DOCTORAL

1.- Título de la Tesis	
Español/Otro Idioma Identificación de nuevos biomarcadores de adaptación a estrés por altas temperaturas en pináceas mediante una aproximación de biología de sistemas:	Inglés: Systems biology-drived identification of novel heat stress biomarkers in pine species.

2.- Autor	
Nombre: Laura Lamelas Penas	
Programa de Doctorado: Biogeociencias	
Órgano responsable: Universidad de Oviedo	

RESUMEN (en español)

El cambio climático y el calentamiento global se han agudizado en las últimas décadas, ejerciendo una enorme presión ambiental sobre los organismos vivos. Por lo que el actual contexto de emergencia climática enfatiza la urgencia de descifrar y entender cómo las plantas perciben, responden, memorizan y se adaptan a eventos climáticos extremos como las olas de calor, que se encuentran entre los estreses más perjudiciales para las plantas y que además son extremadamente difíciles de controlar en condiciones de campo. A pesar de la relevancia de estos procesos de aclimatación y memoria, estos fenómenos están poco estudiados a nivel molecular, especialmente en especies no forestales modelo como *Pinus radiata*. *Pinus radiata* es una de las especies forestales más plantadas debido a su rápido crecimiento y calidad de la madera, además las poblaciones silvestres y manejadas de esta especie están distribuidas por todo el mundo.

Teniendo esto en cuenta, el objetivo principal de esta tesis es identificar los mecanismos de adquisición de memoria y señalización de respuesta a alta temperatura más relevantes en dicha especie empleando un enfoque de proteómica subcelular y un sistema experimental realista acorde con el incremento de las temperaturas estimado para los próximos años para proporcionar un conjunto de indicadores robustos útiles para la selección temprana de árboles y semillas termotolerantes.

Para ello, en primer lugar, se llevo a cabo la optimización de un protocolo de extracción de núcleos y cloroplastos, y se desarrolló una nueva herramienta bioinformática (pRocessomics), que permitiría analizar de un modo más exhaustivo y reproducible los datos generados en esta Tesis.

A continuación, para comprender mejor los mecanismos moleculares que conducen a la adquisición de la termotolerancia, así como a los procesos de establecimiento de la memoria, se simuló una ola de calor severa en condiciones controladas que permitió caracterizar los cambios que dan forma a la respuesta al estrés por calor. Estas alteraciones se monitorizaron a través de la determinación de biomarcadores fisiológicos, la cuantificación del proteoma subcelular de núcleos y cloroplastos, la inmunolocalización de metilación de citosina y los niveles de expresión génica de candidatos seleccionados, incluidos ARNm y microARN.

La comparativa del proteoma del cloroplasto entre dos poblaciones isogénicas (silvestre y manejada) reveló una memoria heredada, y permitió distinguir diferentes estrategias de aclimatación, demostrando que las condiciones de crecimiento de las plantas parentales tienen un efecto de primado transgeneracional relacionado con alteraciones en el Fotosistema II y proteínas osmoprotectoras y que, además, esta memoria transgeneracional ayudó a superar el estrés aplicado.

El análisis del proteoma nuclear (de la población no primada) durante el estrés y después de una fase de recuperación mostró el papel fundamental de los mecanismos epigenéticos (metilación del ADN, modificaciones de histonas y microARN) en la adquisición de la memoria (durante el estrés) así como en el mantenimiento de esta (después de la recuperación o fase latente). La inmunolocalización de 5 mC mostró una pérdida de metilación del ADN durante el estrés por alta temperatura, en paralelo con un descenso significativo en la abundancia de S-ADENOSYL METIONINE SYNTHASE, una proteína involucrada en el ciclo de metilación del ADN. Esta pérdida tanto en el nivel de metilación del ADN como en la abundancia de S-ADENOSILMETIONINA SINTASA se recuperó e incrementó su acumulación una vez las plantas estuvieron aclimatadas. Además, la histona H2A.X se identificó como variante termolábil e impulsora de la respuesta, lo que destaca el papel de la conformación de la cromatina durante el estrés y la recuperación.

La integración de ambos subproteomas permitió obtener una comprensión más profunda de la comunicación nuclear y del cloroplasto para coordinar la respuesta de calor celular, mostrando la dinámica de la señalización anterógrada y retrógrada bajo condiciones de estrés, y cómo proteínas relacionadas con microARN como ARGONAUTE1, responsable del silenciamiento génico postranscripcional, parecen dirigir la aclimatación a altas temperaturas en esta especie.

Finalmente, los niveles de expresión de los genes candidatos seleccionados se analizaron en semillas y plántulas (antes, durante y después de un tratamiento de estrés en los individuos primados y en los no primados). Esto permitió validar las hipótesis de los capítulos anteriores y diseñar un panel de biomarcadores de termotolerancia, donde miR160 y S-ADENOSILMETIONINA SINTASA parecen ser los candidatos más prometedores.

RESUMEN (en Inglés)

Climate change and global warming have worsen over the last decades, exerting a huge environmental pressure to living organisms. Hence, the current context of climate emergency emphasizes the urgency to decipher how plants sense, response, memorize and adapt to climate extreme events such as heat waves, which are among the most detrimental stresses for plants and extremely difficult to control in the field. Despite the relevance of acclimatization processes, these molecular phenomena are poorly studied, especially in non-model forest species as *Pinus radiata*. *Pinus radiata* is one of the most widely planted forest species due to its rapid grow and wood quality, and wild and managed populations of this species are distributed worldwide.

Then, the main objective of this thesis is to identify the most relevant high-temperature response signaling and memory acquisition mechanisms in *Pinus radiata* employing an integrative subcellular proteomics approach and using a realistic scenario of high temperature increase in order to provide a set of reliable and useful biomarkers for early selection of thermotolerant or primed trees and seeds.

To this end, a nuclei and chloroplast extraction method was optimized, and a bioinformatic tool (pRocessomics R package) was developed, both of them required in order to perform the subsequent analyses of this thesis.

To further understand the molecular mechanisms that lead to thermotolerance acquisition as well as memory establishment processes, in this thesis, a severe heat wave was mimicked in controlled conditions to characterize the changes that shape the response to heat stress. These alterations were tracked through physiological biomarkers determination, nuclei and chloroplast subcellular proteome quantification, cytosine methylation immunostaining and gene expression levels of selected candidates including mRNAs and microRNAs.

The chloroplast proteome comparative between two isogenic populations (wild and managed) revealed an inherited memory, and allowed to distinguish different acclimatization strategies, demonstrating that the growing conditions of the parental plants have a transgenerational priming effect related to alterations in Photosystem II and osmoprotective proteins which helped to overcome the applied stress.

Universidad de Oviedo

The nuclear proteome analysis of the non-primed population during the stress and after a recovery phase showed the pivotal role of epigenetics (DNA methylation, histone modifications and microRNAs) on the memory acquisition (during the stress) and maintenance (after recovery or latent phase) processes. 5mC immunostaining showed a DNA methylation loss during the applied heat stress in parallel with the depletion during the stress of S-ADENOSYL METHIONINE SYNTHASE, a protein involved in the DNA methylation cycle. Interestingly, these changes in DNA methylation level and S-ADENOSYLMETHIONINE SYNTHASE accumulation were the opposite after the plants were allowed to recover. Additionally, the thermosensitive H2A.X variant was identified as a heat response driver, highlighting the role of chromatin conformation during the stress and the recovery.

The integration of both subproteomes allowed to gain a deeper understanding of nuclear and chloroplast communication for coordinating cellular heat-response showing the dynamics of the anterograde and retrograde signaling under the stress, and how microRNA related proteins as ARGONAUTE1, responsible for post transcriptional gene silencing, were finely tuned during the acclimation to high-temperature.

Finally, the selected candidate gene expression levels were tracked in seeds and seedlings (prior, during and after a single or double heat stress treatment). This, allowed to validate the hypotheses from the previous chapters and the designing of a thermotolerance biomarker panel, where miR160 and S-ADENOSYLMETHIONINE SYNTHASE were the most promising candidates.

SR. PRESIDENTE DE LA COMISIÓN ACADÉMICA DEL PROGRAMA DE DOCTORADO EN OVIEDO, a 13 de Diciembre de 2022

Esta tesis ha sido realizada en

Universidad de Oviedo
Departamento de Biología de Organismos y Sistemas.
Área de Fisiología Vegetal.

Instituto Universitario de Biotecnología de Asturias
Unidad de Biotecnología de las Plantas.
Bases moleculares y epigenéticas del desarrollo en plantas (EPIOMIDES).

Leibniz-Institut für Gemüse- und Zierpflanzenbau (IGZ)
Functional Plant Biology Departament

Universidade de Aveiro
Centro de Estudos do Ambiente e do Mar

Financiación personal y de la investigación

Ministerio de Economía y Competitividad

Beca predoctoral BES-2017–082092

Proyecto AGL2014-54995-P: Caracterización epigenómica de marcadores de tolerancia adaptación a estrés UV y térmico en especies agroforestales. Aplicabilidad a la selección de genotipos de interés.

Proyecto AGL2016-77633-P: Selección de marcadores moleculares de estrés UV y térmico en pináceas. validación del efecto de memoria. epigenética.

Proyecto PID2019-107107GB-100: Estreses abióticos en pináceas; aproximación ómica integrativa y memoria epigenética

CHAPTER I. General introduction

1.1 Climate emergency

1.1.1 From The Greenhouse Effect to Climate Extremes

The Greenhouse Effect was firstly described as a natural driver of the Earth's surface temperature by Joseph Fourier in 1822. Fourier introduced the idea of the atmosphere as a strong component of the Earth's climate. Since the surface temperature of Earth could not be explained only by the received solar radiation, then the atmosphere must play a role as a heat-trapping gaseous layer (Fourier, 1822).

A few decades later, in the 1850s Eunice Newton Foote identified H_2O and CO_2 as the main molecules retaining the energy of the longer wavelengths and promoting temperature increase (Foote, 1856). Foote theorized this as a possible driver of Earth warming and provided the first experimental results on this topic. John Tyndall further explained this phenomenon from a physiochemical perspective in his studies on gases infrared absorption (Tyndall, 1861). Svante Arrhenius in 1896 created the first climate model, predicting that an one-fold increase in atmospheric CO_2 would elevate in 5-6 ºC the temperature of the Earth (Arrhenius, 1896). His findings still hold up currently.

However, the scientific concern about changes in the climate started only in the early 1960s, with several studies that monitored CO_2 seasonal concentrations in different locations worldwide (Keeling, 1960). The potential severity of those variations in CO_2 remained however controversial.

The term global warming, which describes the increase in the Earth's surface average temperature due to greenhouse gas emissions, was coined in 1975 by Wallace Broecker (Broecker, 1975). Afterward, in the 1990s, when the climate models reached a reasonable accuracy and precision levels for short-term predictions, the seriousness of the global warming was finally acknowledged. Accordingly, climate research has experienced unprecedented growth over the past years, becoming a hot topic for scientists in a wide variety of disciplines, covering the causes and the consequences for the years to come.

Since then, much effort has been put into understanding the impact that +0.5 ºC (achieved in 1995) or +1.1 ºC (achieved in 2020) in annual global average temperatures may have on living organisms. Despite these increments may seem modest, temperatures keep increasing (Fig. 1.1) heavily and irreversibly impairing the complex climate equilibrium. This imbalance was named as climate change, which was defined as a long-term change in the average weather patterns that have come to define Earth's local, regional and global climates.

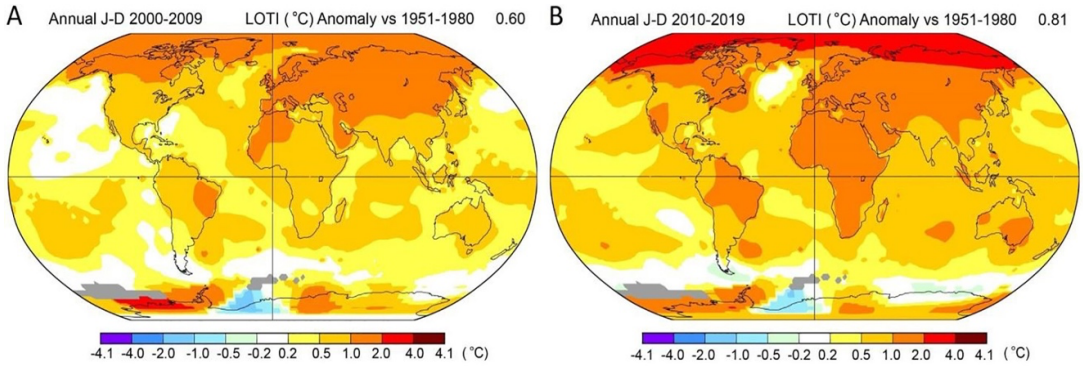


Fig. 1.1 Trends in global temperature change over the last two decades relative to 1951–1980. Map of the annual mean temperature change (°C) during (A) 2000–2009 and (B) 2010–2019. The data for land surface air temperature are from GISS (Goddard Institute for Space Studies) analysis based on global historical climatology network v4, and the data of sea surface temperature are from extended reconstructed sea surface temperature v5. The number at the top right-hand corner of the map plot is an estimate in Celsius degrees of the global mean of the calculated area. Gray areas signify missing data. Ocean data are not used over land nor within 100 km of a reporting land station. The maps were made using the website of GISS Surface Temperature Analysis (https://data.giss.nasa.gov/gistemp/maps/index.html).

Climate change causes and intensifies the so-called weather extremes or weather extreme events (Stott, 2016). These are weather phenomena at the extremes of the historical distribution and are rare for a particular place and/or time. Weather extremes are considered unequivocal signs of the unfolding climate emergency we are currently facing. All of them have shown to have a deep impact on living organisms and their habitat which is predicted to continue to worsen in the decades to come (Coumou & Robinson, 2013; Abatzoglou & Barbero, 2014).

1.1.2 Heat waves in the context of climate change

Such meteorological incidents include thunderstorms, snowstorms, ice storms, blizzards, flooding, hurricanes, megadroughts, and heat waves. These cause severe impairments to natural habitats and magnify the climate equilibrium disruption as well as the frequency of occurrence of other events (Stott, 2016). Therefore, weather extremes

seem to be deeply interconnected. In this context, heat waves affect climate change more negatively than other events, since they have a positive feedback effect on temperature and hence on global warming. Heat waves have therefore been identified as a main climate change indicators.

The most commonly used meteorological definition for a heat wave refers to three or more consecutive days where the maximum temperature is over the 90th percentile for a particular location at a particular time (Perkins & Alexander, 2013; Breshears *et al.*, 2021). The frequency of very hot days (those exceeding the 99th percentile of daily maximum temperature) has more than tripled during the past century (Scherrer *et al.*, 2016). In addition to the predictions of the increasing mean global temperatures (EEA European Environmental Agency, 2015), global climate models project an increase in the frequency, intensity and duration of heat waves in the future (Perkins-Kirkpatrick & Gibson, 2017; Guerreiro *et al.*, 2018), as shown in Fig. 1.2. This is coupled to increases in absolute record temperatures (Abatzoglou & Barbero, 2014). The land area affected by heat waves is expected to quadruple by 2040 (Coumou & Robinson, 2013).

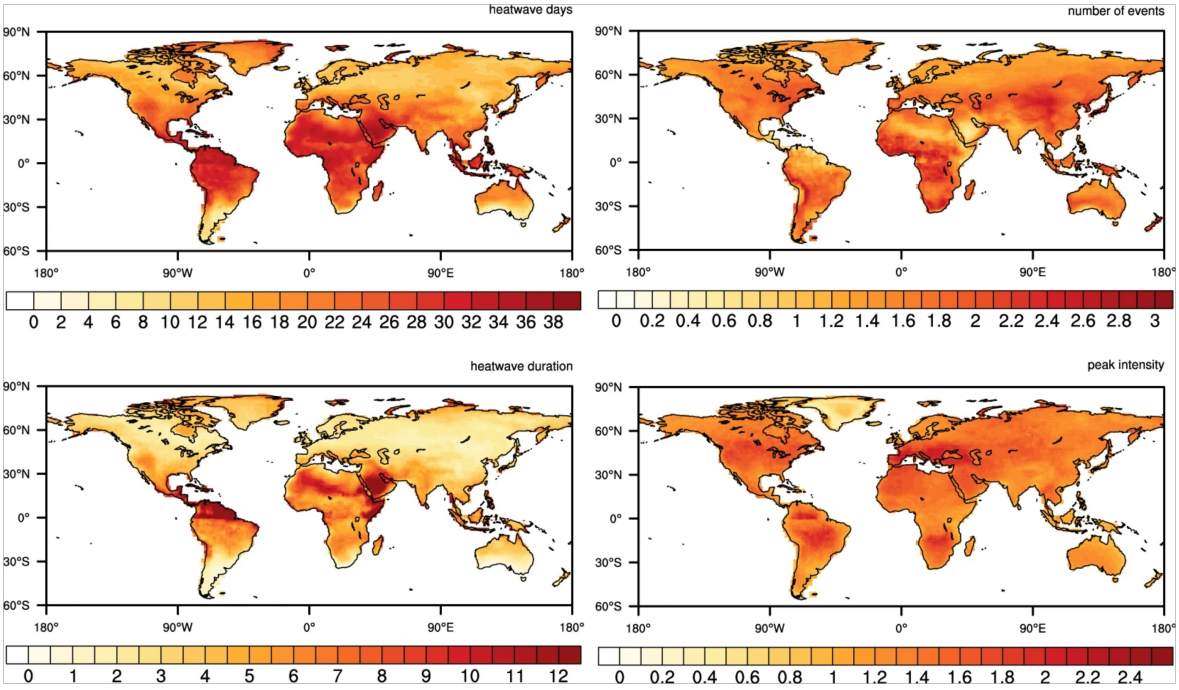


Fig. 1.2 Median regression coefficients estimated from the CMIP5 (Coupled Model Intercomparison Project Phase 5) model ensemble between global warming (ºC) and seasonal heat wave days, number of heat wave events, event duration, and peak heatwave intensity. Adapted from (Perkins-Kirkpatrick & Gibson, 2017)

1.2 Effects of heat waves on plant species

Despite heat waves have a short duration relative to plants' life cycle, their impact is devastating on plantations and wild populations growth and even survival. Heat stress is considered one of the most detrimental stresses. This stress is challenging to control in the field, causing billions of euros in annual losses in crop and forestry industries, as well as incalculable damage to ecosystems.

However, plants show a differential sensitivity to high temperature depending on the severity, duration, and developmental timing of the stress. Heat stress effects usually include photosynthetic impairment, protein unfolding, loss of cell wall and membrane integrity, and oxidative damage by the accumulation of toxic metabolites, such as reactive oxygen species (ROS).

1.2.1 Heat stress sensing and signalling

Due to their sessile nature, plants rely on a continuous monitoring and sensing of their environment that allows them to adjust accordingly and cope with stress. Thus, for all environmental responses, sensing is the primary step during which a sensor directly decodes a stimulus into cellular signaling by altering its own structure and/or activity or its interaction with other molecular components. This then prompts downstream responses that allow acclimatization and therefore survival.

Unlike other organisms as bacteria, there is not a unique thermosensor identified for plants as a master regulator and initiator of the response (Verslues *et al.*, 2022). Instead, a wide set of thermosensors have been identified covering chromatin structure (H2A variants) (Kumar & Wigge, 2010), cell wall and cell membrane (Ca^{2+} signals and fatty acids), ROS (Muench *et al.*, 2016), phytochromes (PHYTOCHROME INTERACTING FACTOR 4, PIF4) (Delker *et al.*, 2022) and GENOMES UNCOUPLED protein family (GUN) (Mochizuki *et al.*, 2001), as well as conformational changes in RNA and alternative splicing in proteins (Chung *et al.*, 2020). The multiple sensing and signaling cascades that plants have evolved, most of them redundant or shared among different biotic and abiotic stresses, are a clear sign of the relevance and complexity of this primary stage.

Chromatin conformational changes

Chromatin structure and dynamics regulate gene expression by controlling the accessibility of genomic DNA sequence by the general transcriptional machinery. Nucleosomes are the basic unit of chromatin, consisting of histone octamers (H2A, H2B, H3, and H4 – two molecules of each). In addition, a linker histone H1, seals the entry and exit of the nucleosomal DNA resulting in a more compact structure (Hergeth & Schneider, 2015). Apart from H4, all histone protein families (H2A, H2B, and H3) are characterized by the presence of histone variants (Bönisch & Hake, 2012). Among these, the histone variants H2A.Z and H2A.X are involved in several biological processes, such as DNA repair, transcriptional control and regulation of centromeric heterochromatic (Giaimo *et al.*, 2019). It has been proposed that the chromatin state also influences the expression of temperature-induced genes; and H2A.Z-containing nucleosomes, instead of the canonical H2A, is involved in temperature sensing in Arabidopsis (Kumar & Wigge, 2010).

Membrane and redox changes to heat stress

Heat stress causes an increased fluidity of the membrane, which leads to activation of lipid-based signaling cascades and to an increased Ca^{2+} influx and cytoskeletal reorganization. Signaling interaction between these routes leads to the production of osmolytes and antioxidants in response to heat stress in plants (Bita & Gerats, 2013). Ca^{2+} and ROS are involved in a signaling pathway that connects heat stress sensors and transcriptional regulators. Ca^{2+} enters the cytosol in response to plasma membrane fluidization, activating RESPIRATORY BURST OXIDASE HOMOLOGS (RBOH) and triggering ROS production (Saidi *et al.*, 2009). The increased Ca^{2+} cellular levels causes the stimulation of the calmodulin-dependent glutamate decarboxylase activity and γ-4-aminobutyric acid (GABA) synthesis (Kinnersley & Turano, 2000). Furthermore, studies revealed Ca^{2+} and ROS are the initial, indispensable factors that evoke heat stress response (Liu *et al.*, 2005; Volkov *et al.*, 2006), as schematized in Fig. 1.3.

Intracellular communication and signaling

After one or several thermosensors detect the stimulus, the alarm signal must be triggered in the cytoplasm and among the cellular organelles for developing the posterior response. In view of the presence of genes encoding organellar proteins in different cellular compartments of the plant cell, intracellular communication is required for the

regulation and coordination of the physiological processes that will lead to acclimatization. Anterograde signals originating from the nucleus and retrograde signals emerging from the chloroplast orchestrate this intracellular coordination (Woodson & Chory, 2008).

On top of that, recent studies highlight the chloroplast as a decision-making organelle raising further attention to understanding the role of retrograde communication over the recent years (Zhao *et al.*, 2020). By definition, retrograde signaling is a communication pathway whereby the transcriptional activities in the nucleus are regulated by signals derived from plastids and mitochondria (Woodson & Chory, 2008). The main functions of retrograde signaling are the developmental control of organelle biogenesis and the operational control to adjust and acclimate to environmental cues (Pogson *et al.*, 2008).

Retrograde signaling

One of the most well-studied examples of coordination between the nucleus and the chloroplast is the regulation of Photosynthesis Associated Nuclear Genes (PhANGs). Given that different subunits of the complexes required for photosynthesis are encoded in separate compartments, gene expression in chloroplast and nuclear genomes require the existence of sophisticated regulatory mechanisms that ensure adequate synthesis of proteins functioning in photosynthetic complexes. Thus, the tightly coordinated gene expression in both nucleus and chloroplast is required for the correct stoichiometric subunit composition of these complexes (Schlicke *et al.*, 2014; Tadini *et al.*, 2020).

From a historic perspective, the basis for postulating a plastid signal was firstly reported with the characterization of two barley chloroplast ribosome-deficient mutants, whose defects in plastid functions resulted in the downregulation of nuclear-encoded plastid proteins and albino phenotypes (Bradbeer *et al.*, 1979). Since this revolutionary discovery, studies have been focusing on the function of retrograde signaling in plastid development by coordinating chlorophyll biosynthesis with the expression of nuclear genes that encode plastid-localized chlorophyll-binding proteins in different plant species using inhibitors of plastid protein synthesis (Oelmüller & Mohr, 1986; Susek & Chory, 1992; Susek *et al.*, 1993).

A thoroughly described gene family related to plastid-nucleus communication was then characterized and named GUN. The *gun* mutants in which the communication

between the chloroplast and the nucleus is disrupted were very helpful in deciphering retrograde signaling cascades (Woodson & Chory, 2008; Barajas-López *et al.*, 2013). The mutant *gun1* was able to express genes related to photosynthesis despite showing defective chloroplast physiology or inhibited biogenesis. According to restrictions in specific steps in chlorophyll biosynthesis, identification of the *gun2*, *gun3*, *gun4*, and *gun5* mutants provided evidence that the accumulation of the chlorophyll intermediate Mg-protoporphyrin IX (Mg-Proto IX) is involved in the initiation of retrograde signaling (Koussevitzky *et al.*, 2007).

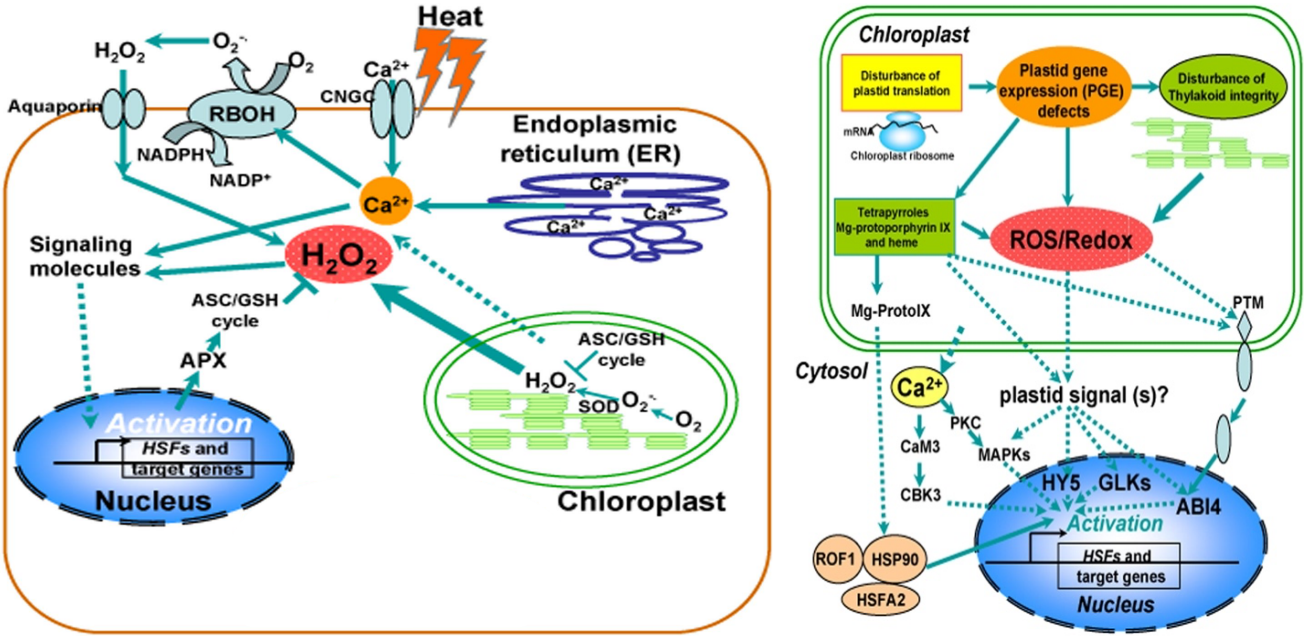


Figure 1.3. Membrane Ca2+ and ROS signaling in response to heat stress. Schematic representation of the major generation sites of ROS and transient calcium increase from different intracellular sources. The chloroplast is a major producer of ROS under heat stress contributing to sensing and signaling processes and contains a large array of ROS-scavenging mechanisms. Hydrogen peroxide (H_2O_2) and Ca^{2+} serve as messengers involved in heat-responsive activation of genes, such as heat shock transcription factors (HSFs), heat shock proteins (HSPs), and ascorbate peroxidase (APX). Under heat stress, several redox enzymes and metabolites, such as superoxide dismutase (SOD) and the ascorbate–glutathione (ASC–GSH) cycle are involved in the maintenance of ROS homeostasis. Heat stress is also sensed via an increased membrane fluidity and/or via a consequent increase in levels of Ca^{2+} controlled by a Ca^{2+} permeable channel (CNGC). Ca^{2+} influx activates downstream responses leading to the increase of ROS. Additionally, HS induces disturbance on the photosynthetic machinery and defects in plastid gene expression, again leading to generation of ROS and accumulation of tetrapyrroles such as Mg-ProtoIX and heme disrupting chlorophyll biosynthesis. The resulting ROS and disturbance of redox homeostasis in chloroplasts serve as retrograde signals to activate downstream signal cascades in the nucleus. Ca^{2+}, sequestered from chloroplasts under heat stress, via calmodulin activates calmodulin-binding proteins pathway. The stressed chloroplasts may sequester Mg-ProtoIX, that binds to HSP90, which can form a complex with a peptidyl prolyl cis/trans isomerase. The resulting complex mobilizes into the nucleus with the help of the heat stress transcription factor HSFA2, which allow the transcription of target genes such as HSPs required for establishing cellular heat tolerance. Adapted from (Sun & Guo, 2016)

The role of the chloroplast as a heat stress sensor

Plants have evolved complex signaling networks to sense and respond to environmental cues. In particular, chloroplasts act as specific sensors of intra- and extracellular stimuli and integrate a multitude of intracellular signals and pathways to sustain homeostasis at the cellular and organism level. In respect to chloroplast-nuclear signaling in response to environmental stimuli, intensive studies have been focusing on the initiation of signaling cascades in the chloroplast and transcriptional changes in the nucleus. In the past few years, a number of retrograde signaling pathways were identified and/or proposed as stress-specific organelles-to-nucleus retrograde signaling cascades.

Most of operational retrograde signaling pathways are thought to be triggered by ROS and photosynthesis redox imbalance during stress conditions and play an important role in the acclimation of plants (Pogson *et al.*, 2008; Galvez-Valdivieso & Mullineaux, 2010; Suzuki *et al.*, 2012). On the other hand, the redox-status of the chloroplasts, which correlates with the ROS accumulation caused by abiotic stresses, may be transmitted by monitoring the state of the plastoquinone, ascorbate, and glutathione pools (Suzuki *et al.*, 2012).

Among a variety of ROS-dependent retrograde signaling pathways, most of the studies focused on the singlet oxygen pathway, which is independent of Magnesium Protoporphyrin IX (Mg-ProtoIX) and GUN1-mediated signaling (Suzuki *et al.*, 2012). Unlike H_2O_2, $^1O_2$ is a highly reactive radical that is involved in signaling pathway leading to cell death or to acclimation (Wagner *et al.*, 2004).

The singlet oxygen signaling pathway has been extensively studied in Arabidopsis using the conditional fluorescent in blue light (*flu*) mutants, which accumulate protochlorophyllide, a potent photosensitizer that generates large amounts of $^1O_2$ (Meskauskiene *et al.*, 2001; Wagner *et al.*, 2004; Kim & Apel, 2013). Importantly, the accumulated $^1O_2$ in the *flu* mutant chloroplasts correlated with the induction of stress responses, including dramatic alterations in nuclear gene expression and higher expression of genes related to the biosynthesis of stress. Moreover, these $^1O_2$-induced changes were regulated by the chloroplast proteins EXECUTER1 (EX1) and EXECUTER 2 (EX2) (Wagner *et al.*, 2004; Kim & Apel, 2013; Zhang *et al.*, 2014)

1.2.2 Heat stress response

Once the signalling cascades have been triggered, plant cells are required to respond in order to mitigate the exerted damage and reach an acclimated homeostatic state. Plants respond to heat stress by altering their transcriptome, and subsequently their proteome and metabolome (Escandón *et al.*, 2022). These changes to cope with the stress are orchestrated from the nucleus, which is the main regulatory hub of the cell. In plants, approximately up the 25% of transcriptome is upregulated two-fold or more under heat stress. Most of these transcripts code for proteins involved in primary and secondary metabolisms, translation, regulation of gene expression and protein folding.

Heat shock proteins in the heat stress response

The most widely studied response to high temperature is the accumulation of heat shock proteins (HSPs). HSPs are molecular chaperones involved in the stabilization of proteins denatured by heat stress and the maintenance of accuracy in early protein folding (Baniwal *et al.*, 2004). HSPs are widely conserved in all living organisms. All HSPs are characterized by the presence of a carboxylic terminal called heat-shock domain (Vierling, 1991; Helm *et al.*, 1993) and are located ubiquitously within the cell. These are usually classified in subfamilies according to their molecular weight or size, which ranges from 10 to more than 100KDa, in HSP100, HSP70/60, HSP40/30 and small HSP or HSP20.

Under heat stress, HSPs are induced by misfolded proteins and participate in protecting thermolabile proteins or complexes (Vierling, 1991; Helm *et al.*, 1993). Furthermore, the HSP20 family acts as an essential factor in the early development of seedlings under heat stress and promotes basal thermotolerance (Clarke *et al.*, 2004). In addition to their role as chaperones, some HSP100 and HSP20 protect plant cells from heat-induced programmed cell death (Rikhvanov *et al.*, 2007).

At the molecular level, Heat Shock Factors (HSFs) are the transcriptional activators of the *HSPs* genes (Banti *et al.*, 2010). The number of HSFs in plants is far more numerous than in mammalian cells, and its number depend on the plant species. In Arabidopsis, 21 HSF members have been identified (Jacob *et al.*, 2017). Typically, plant HSF proteins share a well-conserved modular structure. Its N-terminal DNA binding domain of this transcription factor family is characterized by a central helix-turn-helix motif that specifically binds to the heat stress elements (HSEs) in the target promoters, and subsequently

activates the transcription of stress-inducible genes (Scharf *et al.*, 2012). In general terms, overexpression of plant *HSFs* can enhance plant thermotolerance, but the gene knockouts of individual HSFs tested so far have had little effect on plant survival to high temperature. This implies that the HSF network may be redundant, perhaps reflecting the importance of this high temperature response.

1.2.3 Consequences after the heat stress. Thermomemory

Owing to their sessile nature and the current context of climate change, plants frequently face a wide range of recurring stress events during their life cycle. To counter these episodes, plants have evolved sophisticated molecular information storage mechanisms which allow them to adapt to such adverse conditions (Bäurle, 2016; Oberkofler *et al.*, 2021).

Stief *et al.*, (2014) described molecular memory as a phenomenon in which a stimulus of limited duration is perceived, stored and later retrieved, as evidenced by a modified response. Several studies have elucidated a pivotal role of epigenetics in the memory acquisition processes. The epigenetic regulation of gene expression encompasses three main components: DNA methylation, histone modifications, and the expression of small RNAs as microRNAs (miRNAs). Histone modifications include methylation, acetylation, phosphorylation and ubiquitination (Lämke *et al.*, 2016). Among these modifications, histone H3 lysine 4 trimethylation (H3K4me3) is typically associated with active gene transcription. It was reported that sustained H3K4 methylation marks on heat-responsive loci, such as HSP 18.2, HSP21, and HSP22.0, in Arabidopsis serves as a memory signature and are associated with hyper-induction of these genes upon repeated HS treatments (Lämke *et al.*, 2016). However, the inherent short life-cycle of Arabidopsis prevents to study long-term memory (over months or even years).

On the other hand, but also related to epigenetics, several miRNA pathways as well as some of their components (e.g. ARGONAUTE1 (AGO1), DICER-LIKE1 (DCL1) and SUO) were recently shown to involve in HS memory in Arabidopsis (Guan *et al.*, 2013; Stief *et al.*, 2014). It was hypothesized that the miR156 may integrate developmental pathways with stress conditions (Stief *et al.*, 2014); while miR398 negatively regulates several ROS-scavenging enzymes as COPPER/ZINC SUPEROXIDE DISMUTASE (CSD1, CSD2) and *COPPER CHAPERONE FOR SUPEROXIDE DISMUTASE* (*CCS*) genes (Guan *et al.*, 2013). Also, Fang *et al.*, (2019) recently reported that the biogenesis

of miRNAs is positively regulated by tocopherols. Tocopherols are required for accumulation of 3′-phosphoadenosine-5′-phosphate to inhibit exoribonuclease (XRN)-mediated primary miRNA degradation (Fang *et al.*, 2019). These results indicate that a chloroplast to nucleus retrograde signaling favors miRNA biogenesis under heat stress and subsequently improves high temperature tolerance.

1.2.4 Cross-tolerance

Cross-tolerance, also referred to as cross-resistance or cross-protection (Pastori & Foyer, 2002), refers to the enhanced ability of a plant to tolerate multiple stresses. Plants growing in wild conditions are constantly exposed, either sequentially or simultaneously, to many abiotic and/or biotic stress factors. As a result, plants have evolved shared strategies to respond to ever-changing environmental conditions, enabling them to monitor their surroundings and trigger their metabolic systems to maintain homeostasis.

Currently, there are three main approaches to characterize cross-tolerance processes; these include the identification of the transcriptional overlap between stress responses which is based on the study of the shared responses to different stressors; the study of the priming of plant stress responses following exposure to a different type of stress or compound (induced cross-tolerance) which relies on exposing plants to a training stressor and then compare the performance under a different or testing stressor; and the description of the genetic overlap between resistant populations to different stresses (inherent cross-tolerance), based on identifying a shared genetic signatures as single nucleotide polymorphisms (SNPs) or quantitative trait locus (QTLs) that some populations exhibit coupled to a basal tolerance to several stresses.

Recently, priming and induced cross-stress tolerance have attracted considerable interest within the scientific community as a potential means of stress management and for the production of stress-resistant plants.

1.2.5 Adaptation and transgenerational memory

The modifications occurring during the memory acquisition process have the potential to be transmitted to the progeny of stressed plant through a phenomenon called transgenerational stress memory (Bilichak & Kovalchuk, 2016). This is thought to create (epi)genetic variation for natural selection to help adaptation to HS at species-wide level

(Stief *et al.*, 2014). This represents a key mechanism by which plants can adapt to harsh environments. The concept of a possible heritability of attained traits caused by environmental conditions dates back to Lamarck and has often been referred to as ‘soft inheritance’ (Jablonka & Lamb, 2008; Jablonka, 2017). Numerous studies have made an attempt to analyze the molecular mechanisms of the acquisition of transgenerational inheritance in plants as previously reviewed (Dickins & Rahman, 2012). Overall, it has been accepted that when phenotypic characters are transferred to the offspring without the modification of a gene sequence, thus the underlying mechanism is epigenetic. The response to stress involves immediate and delayed responses directed towards the survival of an individual plant. At the somatic level, plants often acquire a certain memory of stress exposures that prepares them to withstand further encounters with similar and dissimilar stresses more efficiently. Despite the great interest and potential that this phenomenon holds, there is still scarce information about how this memory can be transmitted from the parents to their progeny.

1.3 Omics sciences to study complex process of plant physiology

Systems Biology has emerged in the last years as a comprehensive set of tools which provides high-throughput biological information by integrating different cell organization levels, classically -omics datasets, providing the relationships and correlations among them. Besides, this kind of approaches also enable the possibility of tracking different cell compartments, toward modelling how and why cells are synchronized under different conditions. Contrary to targeted approaches, the untargeted analyses represent an unbiased approach that far from introducing noise or scatter the focus of the research can provide new candidates, pathways or counterintuitive mechanisms otherwise undetectable.

To perform this characterization several analytical methods mainly, next generation sequencing, mass spectrometry, and statistical and modelling tools must be used. Omics approaches are based on the overlap of different information layers and allows us to discover and unravel the complex mechanisms and their interplay in plant systems, as well as track the molecular variability among individuals and link it to phenotype differences (Fukushima *et al.*, 2014; Großkinsky *et al.*, 2018). The integration of these omics layers is nowadays one of the most powerful strategies to deeper understand and analyze a wide variety of biological processes from a holistic point of view (Mochida & Shinozaki, 2011).

1.3.1 Subcellular proteomics

Among omics techniques, proteomics has emerged as a powerful strategy to unravel the complex mechanisms that drive cell responses to a determined stress, allowing us to evaluate the result of the previous regulation layers such as transcriptional regulation. Proteomics relies on three basic technological cornerstones that include a method to extract complex protein or peptide mixtures, mass spectrometry (MS) to acquire the data necessary to identify individual proteins, and bioinformatics to analyze and assemble the MS data.

The complexity of the plant cell proteome, exhibiting thousands of proteins which abundance vary in several orders of magnitude, makes impossible to cover most of the plant proteins using standard shotgun-based approaches. Despite for a general proteome

description this complexity is not a big issue (current protocols and instrumentation allow the identification of several thousand proteins per injection), low or medium abundant proteins cannot be detected most of times, being necessary to fraction or perform targeted analyses in order to detect and quantify them. Among fractioning choices, cell fractioning in its different organelles is a good strategy not only for gaining a deeper coverage of the proteome, but also the basis for understanding organelle function, protein dynamics and trafficking within the cell, as nuclear and chloroplast communication.

Despite its many virtues, one of the limitations for proteomics studies in non-model species is the requirement of a reference protein database, since protein identification is done by comparing an *in vitro* digestion of a given proteome database to the sequenced peptides from the spectra. Thus, the number of identified proteins heavily relies on the previous available knowledge and represent a challenge difficult to tackle in non-model organisms which lack a reference genome. Fortunately, next generation sequencing is becoming more affordable and for challenging species that have megagenomes, *de novo* transcriptomics studies can also yield a reliable database, enhancing proteome characterization including protein identification and annotation.

1.3.2 Data analysis and integration

In omics science, each advance of high-throughput techniques is usually geared towards obtaining a greater amount of molecular data. The analysis of these data was often considered the bottleneck of omics studies (Bino *et al.*, 2004; Ritchie *et al.*, 2015), to overcome this issue, novel bioinformatic tools are constantly being developed or updated to fulfil ongoing demand, and therefore bioinformatics is an exponentially growing field which is required to develop coupled to omics techniques (Manzoni *et al.*, 2016; Rajasundaram & Selbig, 2016). However, despite the recent development of several tools, most of them are focused on sequence analysis and annotation pipelines, hindering a proper biostatistics analysis of omics datasets. Thus, there is still a lack of user friendly software to perform state-of-art analysis.

For proteomics datasets, data quantitative analysis usually starts with a preprocess step that allows to reduce noisy variables or leverage variability between samples due not to biological causes but to the fragile nature of mass spectrometry. Once this step is performed the analysis workflow usually includes sample clustering, to infer the effect of the applied treatment to the subject of study and variable clustering by quantification to

detect co-expression and clusters of proteins that behave similarly, or by the biological processes related to that proteins and functional enrichment analysis.

Integrative analyses are typically depicted as interaction networks which allow the establishment relationships based on clustering algorithms to find out co-expression patterns among the different studied omics levels (Moreno-Risueno *et al.*, 2010) or protein-protein interaction networks in the case of proteomics datasets. The latter can be based on correlations among the proteins, can include the already known biological interactions using databases, or both to enhance their functionality by the combination of omics derived mathematical correlations and the current interactome knowledge.

1.4 Approach and Objectives

1.4.1 *Pinus radiata* as a model species for memory acquisition process

Forests provide the world's population with many far-reaching benefits. Among them, wood supplies and socio-economic goods, the delivery of vital long-term environmental benefits such as clean air and water, stable soils, and biodiversity (FAO, 2015), and more relevantly, the promise of mid-term mitigation of increasing atmospheric CO_2 concentrations which has been recently considered as a key strategy in the COP26 (United Nations climate change conference, hosted in 2021) with the launch of the Forests and Climate Leaders' Partnership financially supported with 19.2 billion US$ to fight deforestation and climate change.

Hence, current models of vegetation dynamics predict profound landscape alterations affecting not only natural forests but also plantations due to an increased average temperature and extreme drought and/or heat periods. Besides reducing growth during these periods, stressful conditions will also have an effect over the following years. These changes in natural and managed forest landscapes will have a major impact over environment (conservation, CO_2 capture) and productivity, endangering the sustainability of forest use, in a period in which the required production and consumption of wood products and wood energy are expected to increase largely, following historical trends (Aspinwall *et al.*, 2019).

Among forests tress, *Pinus radiata* is currently the most widely planted pine species for forestry due to its fast growth, acceptable wood quality, and economically profitable production (Mead, 2013). Therefore, in order to reach future sustainability of forest ecosystems and supply the demand of wood products there is an essential need to further characterize the signaling, response and acclimatization of *Pinus radiata* to heat stress as well as explore molecular memory mechanisms, especially being high temperature one of the most detrimental stresses that limits the growth of temperate forest trees (Wahid *et al.*, 2007) being lethal sometimes and already producing forest die-back and decreasing distribution areas (Sire *et al.*, 2022).

1.4.2 Experimental system

The rapid climate change framework underlines the urgency of generating applicable knowledge that is global in scope, portraying vulnerability of organisms and ecosystems to heat wave events. To gain an integrative understanding of the magnitude, scope and vulnerability of key plants globally to heat waves, it has been suggested that a battery of trials should be conducted using controlled (e.g. growth chamber), standardized single heat wave events (Breshears *et al.*, 2021).

In this thesis it has been settled a common heat wave treatment for all the performed experiments, consisting of at least 5 days with a maximum temperature of 45 ºC maintained over 6 hours a day, as a heat treatment severe enough to enable the characterization of the heat stress response.

It is of vital importance to establish an experimental system that takes into account both future climatic conditions (such as heat waves) and the most relevant players in the triggering of the response and in the establishment of the possible epigenetic memory consequence of one or several exposures to thermal stress in a long life cycle species as radiata pine.

To further study these mechanisms, chloroplast and nuclear untargeted proteomics are thought to be useful approaches since the central importance of chloroplasts as heat signal transduction and the relevance of nuclei in the response orchestration and the memory acquisition processes. Additionally, due to the multiple components and types of molecular memory (as described above), it is of great importance focus the analysis in a *Pinus radiata* population without a pre-existing priming component to heat or other stresses, that may introduce noise and produce unreproducible results in non-primed progenies

1.4.3 Aim and partial objectives

In conclusion, the main objective of this thesis is to **identify the most relevant high-temperature response signaling and memory acquisition mechanisms in *Pinus radiata* employing an integrative subcellular proteomics approach and using a realistic scenario of high temperature increase in order to provide a set of reliable**

and useful biomarkers for early selection of thermotolerant or primed trees and seeds. To accomplish this main goal, the following set of partial objectives was defined:

1 Development of an universal protocol for (prote)-omics data analysis and integration (Chapter 2).

2 Characterization of the chloroplast proteome in two wild isogenic populations grown in different environments in optimal conditions and under a controlled heat wave. Exploration of transgenerational memory through chloroplast proteomics. (Chapter 3).

3 Characterization of the nuclear proteome in response to heat stress, before, during, and after the heat stress exposure in non-primmed *P. radiata* population. (Chapter 4)

4 Integration and evaluation of the synchronization of nuclear and chloroplast proteomes in response to heat stress prior and during a controlled heat wave. (Chapter 5)

5 Identification of a panel of thermotolerance and thermomemory biomarkers in seeds and seedlings in response to high temperature response in *Pinus radiata* (Chapter 6).

1.5 References

Abatzoglou JT, Barbero R. **2014**. Observed and projected changes in absolute temperature records across the contiguous United States. *Geophysical Research Letters* **41**: 6501–6508.

Arrhenius S. **1896**. XXXI. On the influence of carbonic acid in the air upon the temperature of the ground. *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science* **41**: 237–276.

Aspinwall MJ, Pfautsch S, Tjoelker MG, Vårhammar A, Possell M, Drake JE, Reich PB, Tissue DT, Atkin OK, Rymer PD, *et al.* 2019. Range size and growth temperature influence Eucalyptus species responses to an experimental heatwave. *Global Change Biology* **25**: 1665–1684.

Baniwal SK, Bharti K, Chan KY, Fauth M, Ganguli A, Kotak S, Mishra SK, Nover L, Port M, Scharf KD, *et al.* 2004. Heat stress response in plants: A complex game with chaperones and more than twenty heat stress transcription factors. *Journal of Biosciences* **29**: 471–487.

Banti V, Mafessoni F, Loreti E, Alpi A, Perata P. **2010**. The Heat-Inducible Transcription Factor HsfA2 Enhances Anoxia Tolerance in Arabidopsis. *Plant Physiology* **152**: 1471–1483.

Barajas-López J de D, Blanco NE, Strand Å. **2013**. Plastid-to-nucleus communication, signals controlling the running of the plant cell. *Biochimica et Biophysica Acta (BBA) - Molecular Cell Research* **1833**: 425–437.

Bäurle I. **2016**. Plant Heat Adaptation: priming in response to heat stress. *F1000Research* **5**: 694.

Bilichak A, Kovalchuk I. **2016**. Transgenerational response to stress in plants and its application for breeding. *Journal of Experimental Botany* **67**: 2081–2092.

Bino RJ, Hall RD, Fiehn O, Kopka J, Saito K, Draper J, Nikolau BJ, Mendes P, Roessner-Tunali U, Beale MH, *et al.* 2004. Potential of metabolomics as a functional genomics tool. *Trends in Plant Science* **9**: 418–425.

Bita CE, Gerats T. **2013**. Plant tolerance to high temperature in a changing environment : scientific fundamentals and production of heat stress-tolerant crops. *Frontiers in Plant Science* **4**: 1–18.

Bönisch C, Hake SB. **2012**. Histone H2A variants in nucleosomes and chromatin: more or less stable? *Nucleic acids research* **40**: 10719–10741.

Bradbeer JWi, Atkinson YE, Börner T, Hagemann R. **1979**. Cytoplasmic synthesis of plastid polypeptides may be controlled by plastid-synthesised RNA. *Nature* **279**: 816–817.

Breshears DD, Fontaine JB, Ruthrof KX, Field JP, Feng X, Burger JR, Law DJ, Kala J, Hardy GESJ. **2021**. Underappreciated plant vulnerabilities to heat waves. *New Phytologist* **231**: 32–39.

Broecker WS. **1975**. Climatic Change: Are We on the Brink of a Pronounced Global Warming? *Science* **189**: 460–463.

Chung BYW, Balcerowicz M, Di Antonio M, Jaeger KE, Geng F, Franaszek K, Marriott P, Brierley I, Firth AE, Wigge PA. **2020**. An RNA thermoswitch regulates daytime growth in Arabidopsis. *Nature Plants* **6**: 522–532.

Clarke SM, Mur LAJ, Wood JE, Scott IM. **2004**. Salicylic acid dependent signaling promotes basal thermotolerance but is not essential for acquired thermotolerance in Arabidopsis thaliana. *The Plant journal : for cell and molecular biology* **38**: 432–447.

Coumou D, Robinson A. **2013**. Historic and future increase in the global land area affected by monthly heat extremes. *Environmental Research Letters* **8**.

Delker C, Quint M, Wigge PA. **2022**. Recent advances in understanding thermomorphogenesis signaling. *Current Opinion in Plant Biology* **68**: 102231.

Dickins TE, Rahman Q. **2012**. The extended evolutionary synthesis and the role of soft inheritance in evolution. *Proceedings of the Royal Society B: Biological Sciences* **279**: 2913–2921.

EEA European Environmental Agency. **2015**. Global and European temperatures.

Escandón M, Valledor L, Lamelas L, Álvarez JM, Cañal MJ, Meijón M. **2022**. Multiomics analyses reveal the central role of nucleolus and nucleoid machinery during heat stress acclimation in *Pinus radiata*; *bioRxiv*: 2022.07.08.499117.

Fang X, Zhao G, Zhang S, Li Y, Gu H, Li Y, Zhao Q, Qi Y. **2019**. Chloroplast-to-Nucleus Signaling Regulates MicroRNA Biogenesis in Arabidopsis. *Developmental Cell* **48**: 371-382.e4.

Foote E. **1856**. Circumstances Affecting Heat Suns-Rays. *American Journal of Science and Arts* **22**: 382–383.

Fourier J. **1822**. Memoir sur la theorie analitique du le chaleur.

Fukushima A, Kanaya S, Nishida K. **2014**. Integrated network analysis and effective tools in plant systems biology. *Frontiers in Plant Science* **5**: 1–9.

Galvez-Valdivieso G, Mullineaux PM. **2010**. The role of reactive oxygen species in signalling from chloroplasts to the nucleus. *Physiologia plantarum* **138**: 430–439.

Giaimo BD, Ferrante F, Herchenröther A, Hake SB, Borggrefe T. **2019**. The histone variant H2A.Z in gene regulation. *Epigenetics & Chromatin* **12**: 37.

Großkinsky DK, Syaifullah SJ, Roitsch T. **2018**. Integration of multi-omics techniques and physiological phenotyping within a holistic phenomics approach to study senescence in model and crop plants. *Journal of Experimental Botany* **69**: 825–844.

Guan Q, Lu X, Zeng H, Zhang Y, Zhu J. **2013**. Heat stress induction of miR398 triggers a regulatory loop that is critical for thermotolerance in Arabidopsis. *The Plant Journal* **74**: 840–851.

Guerreiro SB, Dawson RJ, Kilsby C, Lewis E, Ford A. **2018**. Future heat-waves, droughts and floods in 571 European cities. *Environmental Research Letters* **13**.

Helm KW, LaFayette PR, Nagao RT, Key JL, Vierling E. **1993**. Localization of small heat shock proteins to the higher plant endomembrane system. *Molecular and cellular biology* **13**: 238–247.

Hergeth SP, Schneider R. **2015**. The H1 linker histones: multifunctional proteins beyond the nucleosomal core particle. *EMBO reports* **16**: 1439–1453.

Jablonka E. **2017**. The evolutionary implications of epigenetic inheritance. *Interface focus* **7**: 20160135.

Jablonka E, Lamb MJ. **2008**. Soft inheritance: Challenging the modern synthesis. *Genetics and Molecular Biology* **31**: 389–395.

Jacob P, Hirt H, Bendahmane A. **2017**. The heat-shock protein/chaperone network and multiple stress resistance. *Plant Biotechnology Journal* **15**: 405–414.

Keeling CD. **1960**. The Concentration and Isotopic Abundances of Carbon Dioxide in the Atmosphere. *Tellus* **12**: 200–203.

Kim C, Apel K. **2013**. Singlet oxygen-mediated signaling in plants: Moving from flu to wild type reveals an increasing complexity. *Photosynthesis research* **116**.

Kinnersley AM, Turano FJ. **2000**. Gamma Aminobutyric Acid (GABA) and Plant Responses to Stress. *Critical Reviews in Plant Sciences* **19**: 479–509.

Koussevitzky S, Nott A, Mockler TC, Hong F, Sachetto-Martins G, Surpin M, Lim J, Mittler R, Chory J. **2007**. Signals from Chloroplasts Converge to Regulate Nuclear Gene Expression. *Science* **316**: 715–719.

Kumar SV, Wigge PA. **2010**. H2A.Z-Containing Nucleosomes Mediate the Thermosensory Response in Arabidopsis. *Cell* **140**: 136–147.

Lämke J, Brzezinka K, Altmann S, Bäurle I. **2016**. A hit-and-run heat shock factor governs sustained histone methylation and transcriptional stress memory. *The EMBO Journal* **35**: 162–175.

Liu H, Sun D, Zhou R. **2005**. Ca2+ and AtCaM3 are involved in the expression of heat shock protein gene in Arabidopsis. *Plant, Cell & Environment* **28**: 1276–1284.

Manzoni C, Kia DA, Vandrovcova J, Hardy J, Wood NW, Lewis PA, Ferrari R. **2016**. Genome, transcriptome and proteome: the rise of omics data and their integration in biomedical sciences. *Briefings In Bioinformatics* **19**: 286–302.

Mead DJ. **2013**. *Sustainable management of Pinus radiata plantations*.

Meskauskiene R, Nater M, Goslings D, Kessler F, op den Camp R, Apel K. **2001**. FLU: A negative regulator of chlorophyll biosynthesis in Arabidopsis thaliana. *Proceedings of the National Academy of Sciences* **98**: 12826–12831.

Mochida K, Shinozaki K. **2011**. Advances in omics and bioinformatics tools for systems analyses of plant functions. *Plant and Cell Physiology* **52**: 2017–2038.

Mochizuki N, Brusslan JA, Larkin R, Nagatani A, Chory J. **2001**. Arabidopsis genomes uncoupled 5 (GUN5) mutant reveals the involvement of Mg-chelatase H subunit in plastid-to-nucleus signal transduction. *Proceedings of the National Academy of Sciences* **98**: 2053–2058.

Moreno-Risueno MA, Busch W, Benfey PN. **2010**. Omics meet networks - using systems approaches to infer regulatory networks in plants. *Current Opinion in Plant Biology* **13**: 126–131.

Muench M, Hsin CH, Ferber E, Berger S, Mueller MJ. **2016**. Reactive electrophilic oxylipins trigger a heat stress-like response through HSFA1 transcription factors. *Journal of Experimental Botany* **67**: 6139–6148.

Oberkofler V, Pratx L, Bäurle I. **2021**. Epigenetic regulation of abiotic stress memory: maintaining the good things while they last. *Current Opinion in Plant Biology* **61**: 102007.

Oelmüller R, Mohr H. **1986**. Photooxidative destruction of chloroplasts and its consequences for expression of nuclear genes. *Planta* **167**: 106–113.

Pastori GM, Foyer CH. **2002**. Common Components, Networks, and Pathways of Cross-Tolerance to Stress. The Central Role of “Redox” and Abscisic Acid-Mediated Controls. *Plant Physiology* **129**: 460–468.

Perkins-Kirkpatrick SE, Gibson PB. **2017**. Changes in regional heatwave characteristics as a function of increasing global temperature. *Scientific Reports* **7**: 12256.

Perkins SE, Alexander L V. **2013**. On the Measurement of Heat Waves. *Journal of Climate* **26**: 4500–4517.

Pogson BJ, Woo NS, Förster B, Small ID. **2008**. Plastid signalling to the nucleus and beyond. *Trends in Plant Science* **13**: 602–609.

Rajasundaram D, Selbig J. **2016**. More effort - more results: Recent advances in integrative ‘omics’ data analysis. *Current Opinion in Plant Biology* **30**: 57–61.

Rikhvanov EG, Romanova N, Chernoff YO. **2007**. Chaperone Effects on Prion and Nonprion Aggregates. *Prion* **1**: 217–222.

Ritchie MD, Holzinger ER, Li R, Pendergrass SA, Kim D. **2015**. Methods of integrating data to uncover genotype-phenotype interactions. *Nature Reviews Genetics* **16**: 85–97.

Saidi Y, Finka A, Muriset M, Bromberg Z, Weiss YG, Maathuis FJM, Goloubinoff P. **2009**. The heat shock response in moss plants is regulated by specific calcium-permeable channels in the plasma membrane. *The Plant cell* **21**: 2829–2843.

Scharf K-D, Berberich T, Ebersberger I, Nover L. **2012**. The plant heat stress transcription factor (Hsf) family: Structure, function and evolution. *Biochimica et Biophysica Acta (BBA) - Gene Regulatory Mechanisms* **1819**: 104–119.

Scherrer SC, Fischer EM, Posselt R, Liniger MA, Croci-Maspoli M, Knutti R. **2016**. Emerging trends in heavy precipitation and hot temperature extremes in Switzerland. *Journal of Geophysical Research: Atmospheres* **121**: 2626–2637.

Schlicke H, Hartwig AS, Firtzlaff V, Richter AS, Glässer C, Maier K, Finkemeier I, Grimm B. **2014**. Induced deactivation of genes encoding chlorophyll biosynthesis enzymes disentangles tetrapyrrole-mediated retrograde signaling. *Molecular Plant* **7**: 1211–1227.

Sire L, Yáñez PS, Wang C, Bézier A, Courtial B, Cours J, Fontaneto D, Larrieu L, Bouget C, Thorn S, *et al.* 2022. Climate-induced forest dieback drives compositional changes in insect communities that are more pronounced for rare species. *Communications Biology* **5**: 57.

Stief A, Altmann S, Hoffmann K, Pant BD, Scheible W-R, Bäurle I. **2014**. Arabidopsis miR156 Regulates Tolerance to Recurring Environmental Stress through SPL Transcription Factors . *The Plant Cell* **26**: 1792–1807.

Stott P. **2016**. How climate change affects extreme weather events. *Science* **352**: 1517–1518.

Sun A-Z, Guo F-Q. **2016**. Chloroplast Retrograde Regulation of Heat Stress Responses in Plants . *Frontiers in Plant Science* **7**.

Susek RE, Ausubel FM, Chory J. **1993**. Signal transduction mutants of Arabidopsis uncouple nuclear CAB and RBCS gene expression from chloroplast development. *Cell* **74**: 787–799.

Susek RE, Chory J. **1992**. A Tale of Two Genomes: Role of a Chloroplast Signal in Coordinating Nuclear and Plastid Genome Expression. *Functional Plant Biology* **19**: 387–399.

Suzuki N, Koussevitzky S, Mittler RON, Miller GAD. **2012**. ROS and redox signalling in the response of plants to abiotic stress. *Plant, Cell & Environment* **35**: 259–270.

Tadini L, Jeran N, Peracchio C, Masiero S, Colombo M, Pesaresi P. **2020**. The

plastid transcription machinery and its coordination with the expression of nuclear genome: Plastid-encoded polymerase, nuclear-encoded polymerase and the genomes uncoupled 1-mediated retrograde communication. *Philosophical Transactions of the Royal Society B: Biological Sciences* **375**.

Tyndall J. **1861**. XXXVI. On the absorption and radiation of heat by gases and vapours, and on the physical connexion of radiation, absorption, and conduction.—The bakerian lecture. *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science* **22**: 273–285.

Verslues PE, Bailey-Serres J, Brodersen C, Buckley TN, Conti L, Christmann A, Dinneny JR, Grill E, Hayes S, Heckman RW, *et al.* 2022. Burning questions for a warming and changing world: 15 unknowns in plant abiotic stress. *The Plant Cell*: koac263.

Vierling E. **1991**. The Roles of Heat Shock Proteins in Plants. *Annual Review of Plant Physiology and Plant Molecular Biology* **42**: 579–620.

Volkov RA, Panchuk II, Mullineaux PM, Schoffl F. **2006**. Heat stress-induced H(2)O (2) is required for effective expression of heat shock genes in Arabidopsis. *Plant molecular biology* **61**: 733–746.

Wagner D, Przybyla D, op den Camp R, Kim C, Landgraf F, Lee KP, Würsch M, Laloi C, Nater M, Hideg E, *et al.* 2004. The Genetic Basis of Singlet Oxygen□Induced Stress Responses of Arabidopsis thaliana. *Science* **306**: 1183–1185.

Wahid A, Gelani S, Ashraf M, Foolad MR. **2007**. Heat tolerance in plants: An overview. *Environmental and Experimental Botany* **61**: 199–223.

Woodson JD, Chory J. **2008**. Coordination of gene expression between organellar and nuclear genomes. *Nature reviews. Genetics* **9**: 383–395.

Zhang S, Apel K, Kim C. **2014**. Singlet oxygen-mediated and EXECUTER-dependent signalling and acclimation of Arabidopsis thaliana exposed to light stress. *Philosophical transactions of the Royal Society of London. Series B, Biological sciences* **369**: 20130227.

Zhao X, Huang J, Chory J. **2020**. Unraveling the Linkage between Retrograde Signaling and RNA Metabolism in Plants. *Trends in Plant Science* **25**: 141–147.

CHAPTER II. pRocessomics: a new Omics data analysis R package

2.1 Abstract

Data analyses and their interpretation usually represent the ultimate bottleneck in omics science, not due to the lack of tools aimed to perform these tasks but the incompatibilities among them and their complexity. pRocessomics is an open source R package designed for facilitating a complete state-of-the-art Omics data analysis in a comprehensive way. This package enables a wide set of analytics, including single and multi-level omics data analysis and integration, and produces high quality images and tables for publication. In addition to native functions for performing data analysis, building plots and exporting the results, this package contains wizard functions to help non-expert users. This feature makes pRocessomics an user-friendly tool by explaining the main statistical aspects required to select the best parameters or type of tests for each dataset as a step-by-step guide. pRocessomics benefits of having an all-in-one package for increasing the reproducibility of the statistical protocols and exploratory analyses. This package has been successfully applied to answer a broad range of biological questions and employed in the rest of the chapters of this thesis.

2.2 Introduction

Omics datasets comprise quantitative information of thousands of features, leading to massive datasets especially when they are combined for their integration. This fact, also known as dimensionality curse, frequently represents a challenge for non-expert data analysts; since contrary to omics data, classic statistics tests and analyses are designed for data with a larger number of observations than features. However, the high dimensionality of the data is the main reason that enables exhaustive and unbiased analysis, that are able to detect previously ignored candidates and raise unbiased hypotheses.

Then, in order to handle and interpret these complex datasets, it is mandatory to develop a strong statistical background and a very specific set of skills related to analytical and statistical competences. This heavily increases analyses time and, in many cases, represents a burden for data interpretation. On top of that, to adequately explore omics data several methodologies must be applied in order to preprocess, analyze, integrate and visualize data (Fukushima *et al.*, 2009; Subramanian *et al.*, 2020). All of them are necessary to identify the most important variables in a biological system, their accumulation patterns, whether those variables are clustered or which features describe the sample clustering and ultimately, to harness the data to the maximum and reach reliable conclusions.

Addressing these queries, which are common for most of the research areas, is mandatory to raise robust data-driven hypothesis. Currently, answering them requires advanced data science knowledge, the use of several tools, and in most cases the development of new code. These problems come with the associated time investment, all far beyond of what most bench scientists are willing to commit. Besides, as omics data usually share most of the characteristics across different biological systems and experiments, there is a need to adopt an analytical workflow that enables the reproducibility of these analyses.

The integration of multiple omics datasets represents a statistical challenge due to the limited number of individuals, the high number of variables and the heterogeneity of the datasets to integrate (Tini *et al.*, 2019; Subramanian *et al.*, 2020). Despite several tools already exist to meet these challenges, as Perseus (Tyanova *et al.*, 2016) , MetaboAnalyst (Chong *et al.*, 2018), PaintOmics (Hernández-de-Diego *et al.*, 2018), instantClue (Nolte *et*

al., 2018), and BioMex (Taverna *et al.*, 2020), demonstrating the need for these software; most of them lack some key aspects such as a thoroughly preprocess step or a complete workflow focused either only in one aspect of the analysis or in the omics integration stage.

While many available tools lack advanced functionalities for integrating different datasets such as Multiple Co-Inertia Analysis (MCIA), which outperformed other integration tools (Meng *et al.*, 2014; Tini *et al.*, 2019) or Block Discriminant Analysis (BDA) (Singh *et al.*, 2019), which allow to build multi-and intra-omics level interaction networks and to identify the most relevant features of each omics layer; those including this state-of-art analyses, are usually difficult to install or run, lack an implementation for bifactorial experimental designs, or require the tuning of too much parameters. Additionally, none of them provide built-in tools to guide the users step by step during analysis.

To fill this gap, pRocessomics is presented in this Chapter as a package for performing data preprocess, univariate and multivariate integrative omics analyses. Within pRocessomics a non-specialist in data science will find wizards for performing complete and exhaustive analyses in an easy way. On the other hand, specialist will find a wide set of functions that can be called for performing exploratory analyses over different omics layers with only a single line of code. In both cases, these analyses will yield ready-to-publish tables and figures to fully describe the data by following a straightforward but versatile workflow.

2.3 Implementation and design

pRocessomics has been designed in a modular manner to offer the greatest possible versability, so the users can decide which modules to use with their data (Fig 2.1). This modular design also applies within each analysis or visualization.

Each analysis module consist of at least two functions, one for the qualitative or quantitative analysis itself and the another for the plotting step, these functions are denoted in all cases with the suffix "_analysis" and "_plot" respectively (Table 2.1).

Additionally, all analyses devoted functions share among them, arguments related to the structure of the data; similarly, some plotting functions may share the same type of representation, accordingly the parameters for these functions are always identical, creating a proprietary grammar that enhances the usability of the whole package and decrease the learning time. In this direction, pRocessomics also allows for the definition of a color theme which will be used for all the suitable representations.

Table 2.1 List of pRocessomics functions

	Analysis	**Visualization**	**Wizard**
Import			importannotation importfromexcel
Preprocess	preprocess_omic_list		preprocess_wizard
Normalization	transformdata		transformation_wizard
Filtering	featureselection		featureselection_wizard
Qualitative Analysis	Venn_analysis	Venn_plot	Venn_diagram_wizard
Biological clustering	mapman_group	mapman_plot	mapman_wizard
Univariate Analysis	univariate	univariate_plot	univariate_wizard
Multivariate Analysis	pca_analysis ica_analysis da_analysis kmeans_analysis	pca_plot ica_plot da_plot kmeans_wizard	pca_analysis_wizard ica_analysis_wizard da_analysis_wizard kmeans_plot
Integrative Analysis	spls_analysis mcia_analysis bda_analysis	spls_plot mcia_plot bda_plot	spls_analysis_wizard mcia_analysis_wizard bda_analysis_wizard
Theme	set_custom_palettes		select_palette
Export	export_table export_plot		

In case the user is unsure about the meaning of the parameters or how a function works step by step, pRocessomics includes wizard functions as console interactive dialogues. These explain each step and argument of the function in use and ask for the desired values to call the main function. This results in an identical functionalities and outcome independently of the use of wizards or classic functions. Thus, non-expert users will be able to get the same results regardless of their previous R or statistical knowledge. The implementation of this dialogue-based approach improves the usability of the package and represents a unique feature in the currently available tools for omics analysis.

2.3.1 Module description

Data import

Dataset import in R can lead to issues, especially when collating text labels (as sample short descriptions or annotations) and numeric data in the same object. These mix-ups are difficult to spot and may lead to incorrect results or errors when running the functions in the next steps. Moreover, non-expert users may not be familiar with data structures in R programming. Hence, data import for both quantitative and variable description (annotations) is available in pRocessomics with the wizards "importfromexcel()" and "importannotation()" respectively, which store the information in a proper format for further steps. Data import module is not mandatory, but strongly recommended to use to ensure a proper run of posterior analyses.

Data preprocess

Before starting the analyses, it is essential to manage initial data heterogeneity (Fukushima *et al.*, 2009). Raw data in this kind of analysis often comes from three main platforms: microarray, Next Generation Sequencing (NGS) and Mass Spectrometry (MS). Thus, datasets must be preprocessed according to their origin (Kim & Tagkopoulos, 2018).

The most common pretreatments are missing values imputation, filtration, normalization and transformation (van den Berg *et al.*, 2006; Fukushima *et al.*, 2009; Gardinassi *et al.*, 2017). To ease data imputation there are several algorithms already available in R as K-nearest neighbor (KNN) imputation (Hastie et al., 2017) and Random Forest (RF), a machine learning based algorithm to assess '*not available*' values. After missing values imputation (Liaw and Wiener, 2002), the next proposed step is to remove

those variables that are nearly empty, employing a consistency criterion by the application of a user-defined threshold. Once missing values have been imputed and empty variables filtered, the last step is to balance sample total quantity, i.e. abundance balancing. This operation can be expressed as a fraction or times one, considering all the samples within the dataset contained the same amount (sample-centric approach) or considering that all the samples for each treatment or biological condition contained an equivalent amount.

Data transformation and feature selection

The functions of this module are intended to provide a quick way for normalizing data in order to meet the requirements for the use of a parametric statistic test. This normalization is often achieved by transforming the data with logarithm, square root, sin, among other functions. In addition, in this module it is possible to select features (variables) based on their variability in order to drop from the posterior analyses those variables that show the greatest erraticism. To this end inter-quartile range (IQR), coefficient of variation (CV), and p or q-value are accessible.

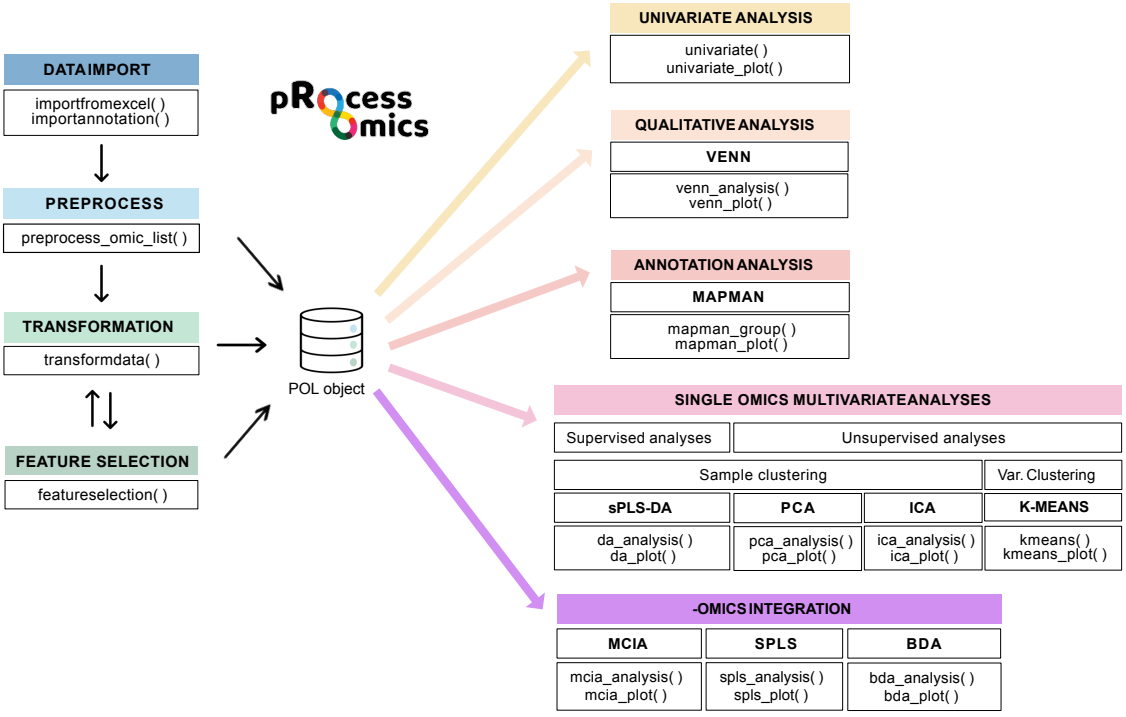


Fig 2.1 pRocessomics consist of several modules: (0) data import, (1) data preprocess, (2) data transformation and feature selection, (3) univariate tests including parametric and non-parametric tests, (4) Qualitative and annotation analyses, (5) Single omics multivariate analyses. and (6) multiple omics integrative analyses.

Univariate analysis

Univariate tests aim to compare the different features, one by one, across the applied treatments or the different experimental conditions within the data. Student t-test (parametric) or Mann Whitney U test (non-parametric) for testing two conditions; ANOVA (Analysis of Variance) and Kruskal Wallis as well as their respective post-hoc and False Discovery Rate (FDR) tests are currently implemented. Besides, a parametricity check is also offered in the wizard dialogue corresponding to this module to guide the user to properly choose the univariate test to apply.

Exploratory analyses: Qualitative and Biological Quantitative analyses

This module has been conceived for grouping, classifying, and visualizing data according to MapMan (Lohse *et al.*, 2014) or any other supplied custom annotation categories. Heatmaps and bar and circular plots on individual or treatment-grouped samples may be produced in this module. Additionally, it includes a qualitative test for evaluating the presence or absence of the variables across the treatments. Both functionalities may be performed over the unfiltered datasets, after data preprocessing and prior to feature selection.

Multivariate analyses

This module contains supervised or discriminant : sparse Partial Least Squares-Discriminant Analysis (sPLS-DA) (Lê Cao *et al.*, 2011) and unsupervised: Principal Component Analysis (PCA), Independent Component Analysis (ICA) and kmeans clustering analyses to cluster samples or variables as indicated in Fig. 2.1. kmeans clustering aims to detect patterns among the variables across the treatments and group them in different units (Likas *et al.*, 2003); while sPLS-DA (Lê Cao *et al.*, 2011), PCA, and ICA (Helwig, 2022) are intended to reduce dimensionality and classify the samples according to artificial components that are set of variables that gather the variance among the samples (Mevik & Wehrens, 2007). This module can be fed with preprocessed, transformed or filtered data and provide resulting analysis as excel tables and pdf vectorized plots.

Integrative analysis

The integration module consist of a set of functions for the integration of two or more omics layers, where one of them is used to predict the other in the case of sPLS (Lê Cao *et al.*, 2011), while MCIA (Meng *et al.*, 2014), and BDA (Singh *et al.*, 2019) for integrating more than two omics layers and reveal the relations among them, MCIA explores the cohesion of the different samples across the input omics layers while BDA can be used for biomarker identification and interaction identification. The outcome of these functions includes correlation networks, which can be exported in a cytoscape compatible format for further aesthetics editing and circus plots.

Tables and figures export

Finally, after each analysis is performed the resulting tables and plots can be exported to excel and pdf formats respectively using the functions “export_table()” and “export_plot()”.

2.3.2 Wizards

Wizard functionalities are designed to help non-expert users to carry out every step of the pRocessomics workflow, enabling a communication channel between the user and the package. By calling the wizards, without any arguments (e.g. preprocess_wizard()), a dialogue will prompt in the console.

Wizards search within the global environment for the data stored in the proper format for each analysis providing the list of elements suitable for each analysis. Or in the case of external data import, will prompt the operative system window interface to ease the localization and the acquisition of the data. This family of user-friendly functions guide the user by a dialogue in the console. These dialogues include the numeric analysis itself, the visualization options available, as well as the exporting functionalities. In addition, pRocessomics generates a plain text document (.txt) containing a log file with all the dialogue.

2.3.3 Code and wiki availability

pRocessomics, a comprehensive wiki, and fully detailed case-studies are available in the GitHub public repository https://github.com/Valledor/pRocessomics).

2.4 Discussion

Despite its broad applicability and growth over the last years, there still remain some entry barriers to omics sciences. Probably the most frequent, is the difficulty of analyzing the ever-growing generated datasets, which may represent a limiting obstacle for bench researchers.

Despite there are some available tools, contrary to pRocessomics most of them are missing an adequate data pretreatment and variable filtering workflows, have a more limited set of “common” analyses or lack the flexibility to meet the requirements of different types of datasets. Additionally, none of them incorporates a step by step dialogue to guide the users through the different analysis stages.

pRocessomics can be used for processing and integrating virtually any omics and phenotype datasets, including for example geoclimatic data, to link omics and environment or in general terms, any quantitative variables. It has been developed to facilitate scientists with the arduous task of processing omics datasets in a comprehensive fashion, particularly for user with a limited R background, through the incorporation of wizards. This family of functions create log files after each analysis facilitating the reproducibility, shareability, and reviewability of the analyses; and could be used as a didactical tool to introduce researchers to the field of omics data science and R language.

To date, pRocessomics package has been already employed in several research articles in different species and through transcriptomics, proteomics, metabolomics and, physiological data in response to biotic and abiotic stresses (Colina *et al.*, 2020b,a; López-Hidalgo *et al.*, 2021; Valledor *et al.*, 2021; Amaral *et al.*, 2021; García-Campa *et al.*, 2022; Escandón *et al.*, 2022) and in the following chapters of this thesis, demonstrating its capabilities.

2.5 References

Amaral J, Lamelas L, Valledor L, Castillejo MÁ, Alves A, Pinto G. **2021**. Comparative proteomics of Pinus–Fusarium circinatum interactions reveal metabolic clues to biotic stress resistance. *Physiologia Plantarum* **173**: 2142–2154.

van den Berg RA, Hoefsloot HCJ, Westerhuis JA, Smilde AK, van der Werf MJ. **2006**. Centering, scaling, and transformations: Improving the biological information content of metabolomics data. *BMC Genomics* **7**: 1–15.

Chong J, Soufan O, Li C, Caraus I, Li S, Bourque G, Wishart DS, Xia J. **2018**. MetaboAnalyst 4.0: towards more transparent and integrative metabolomics analysis. *Nucleic Acids Research* **46**: W486–W494.

Colina FJ, Carbó M, Cañal MJ, Valledor L. **2020a**. A complex metabolic rearrangement towards the accumulation of glycerol and sugars consequence of a proteome remodeling is required for the survival of *Chlamydomonas reinhardtii* growing under osmotic stress. *Environmental and Experimental Botany* **180**: 104261.

Colina F, Carbó M, Meijón M, Cañal MJ, Valledor L. **2020b**. Low UV-C stress modulates *Chlamydomonas reinhardtii* biomass composition and oxidative stress response through proteomic and metabolomic changes involving novel signalers and effectors. *Biotechnology for Biofuels* **13**: 110.

Dietrich S, Argelaguet R, Stegle O, Buettner F, Huber W, Marioni JC, Velten B, Arnol D, Zenz T. **2018**. Multi-Omics Factor Analysis—a framework for unsupervised integration of multi-omics data sets. *Molecular Systems Biology*.

Escandón M, Lamelas L, Roces V, Guerrero-Sanchez VM, Meijón M, Valledor L. **2020**. Protein Interaction Networks: Functional and Statistical Approaches. In: Methods in Molecular Biology. 21–56.

Escandón M, Valledor L, Lamelas L, Álvarez JM, Cañal MJ, Meijón M. **2022**. Multiomics analyses reveal the central role of nucleolus and nucleoid machinery during heat stress acclimation in *Pinus radiata*; *bioRxiv*: 2022.07.08.499117.

Fukushima A, Kanaya S, Nishida K. **2014**. Integrated network analysis and effective tools in plant systems biology. *Frontiers in Plant Science* **5**: 1–9.

Fukushima A, Kusano M, Redestig H, Arita M, Saito K. **2009**. Integrated omics approaches in plant systems biology. *Current Opinion in Chemical Biology* **13**: 532–538.

García-Campa L, Guerrero S, Lamelas L, Meijón M, Hasbún R, Cañal MJ, Valledor L. **2022**. Chloroplast proteomics reveals transgenerational cross-stress priming in *Pinus radiata*. *Environmental and Experimental Botany* **202**: 105009.

Gardinassi LG, Xia J, Safo SE, Li S. **2017**. Bioinformatics Tools for the Interpretation of Metabolomics Data. *Current Pharmacology Reports* **3**: 374–383.

Helwig NE. **2022**. ica: Independent Component Analysis.

Hernández-de-Diego R, Tarazona S, Martínez-Mira C, Balzano-Nogueira L, Furió-Tarí P, Pappas Jr GJ, Conesa A. **2018**. PaintOmics 3: a web resource for the pathway analysis and visualization of multi-omics data. *Nucleic Acids Research* **46**: W503–W509.

Kim M, Tagkopoulos I. **2018**. Data integration and predictive modeling methods for multi-omics datasets. *Molecular Omics*.

Lê Cao K-A, Boitard S, Besse P. **2011**. Sparse PLS discriminant analysis: biologically relevant feature selection and graphical displays for multiclass problems. *BMC Bioinformatics* **12**: 253.

Likas A, Vlassis N, J. Verbeek J. **2003**. The global k-means clustering algorithm. *Pattern Recognition* **36**: 451–461.

Lohse M, Nagel A, Herter T, May P, Schroda M, Zrenner R, Tohge T, Fernie AR, Stitt M, Usadel B. **2014**. Mercator: A fast and simple web server for genome scale functional annotation of plant sequence data. *Plant, Cell and Environment* **37**: 1250–1258.

López-Hidalgo C, Meijón M, Lamelas L, Valledor L. **2021**. The rainbow protocol: A sequential method for quantifying pigments, sugars, free amino acids, phenolics, flavonoids and MDA from a small amount of sample. *Plant, Cell & Environment* **n/a**.

Manzoni C, Kia DA, Vandrovcova J, Hardy J, Wood NW, Lewis PA, Ferrari R. **2016**. Genome, transcriptome and proteome: the rise of omics data and their integration in biomedical sciences. *Briefings In Bioinformatics* **19**: 286–302.

Meng C, Kuster B, Culhane AC, Gholami AM. **2014**. A multivariate approach to the integration of multi-omics datasets. *BMC Bioinformatics* **15**: 162.

Mevik B-H, Wehrens R. **2007**. The pls Package: Principal Component and Partial Least Squares Regression in R. *Journal of Statistical Software* **18**: 1–23.

Nolte H, MacVicar TD, Tellkamp F, Krüger M. **2018**. Instant Clue: A Software Suite for Interactive Data Visualization and Analysis. *Scientific Reports* **8**: 12648.

Singh A, Shannon CP, Gautier B, Rohart F, Vacher M, Tebbutt SJ, Lê Cao K-A. **2019**. DIABLO: an integrative approach for identifying key molecular drivers from multi-omics assays. *Bioinformatics* **35**: 3055–3062.

Subramanian I, Verma S, Kumar S, Jere A, Anamika K. **2020**. Multi-omics Data Integration, Interpretation, and Its Application. *Bioinformatics and Biology Insights* **14**: 1177932219899051.

Taverna F, Goveia J, Karakach TK, Khan S, Rohlenova K, Treps L, Subramanian A, Schoonjans L, Dewerchin M, Eelen G, *et al.* 2020. BIOMEX: an interactive workflow for (single cell) omics data interpretation and visualization. *Nucleic Acids Research* **48**: W385–W394.

Tini G, Marchetti L, Priami C, Scott-Boyer M-P. **2019**. Multi-omics integration—a comparison of unsupervised clustering methodologies. *Briefings in Bioinformatics* **20**: 1269–1279.

Tyanova S, Temu T, Sinitcyn P, Carlson A, Hein MY, Geiger T, Mann M, Cox J. **2016**. The Perseus computational platform for comprehensive analysis of (prote)omics data. *Nature Methods* **13**: 731–740.

Valledor L, Carbó M, Lamelas L, Escandón M, Colina FJ, Cañal MJ, Meijón M. **2018**. When the Tree Let Us See the Forest: Systems Biology and Natural Variation Studies in Forest Species. In: 353–375.

Valledor L, Guerrero S, García-Campa L, Meijón M. **2021**. Proteometabolomic characterization of apical bud maturation in *Pinus pinaster*. *Tree physiology* **41**: 508–521.

CHAPTER III. Comparative chloroplast proteome profiling to decipher transgenerational cross-tolerance to heat stress[1]

3.1 Abstract

How different stressors impact plant health and memory when they are imposed in different generations in wild ecosystems is still understudied. In this chapter, it is addressed whether different environments are able to shape heritable memory for the next generation. The performance of the seedlings belonging to two wild isogenic subpopulations of *Pinus radiata* (optimal fertirrigation vs lightly stressful soil conditions) was tested under heat stress through physiological profiling and comparative time-series chloroplast proteomics in the progenies belonging to both subpopulations. The obtained results showed differential responses between the progenies, evidencing a cross-stress transgenerational memory. Seedlings belonging to previously stressed subpopulation retained key proteins related to Photosystem II, chloroplast-to-nucleus signaling and osmoprotection which helped to overcome the applied heat stress. These finds not only delve into transgenerational cross-stress memory in trees, but also provide new information on how shared molecular mechanism enable differential responses in different climatic contexts.

[1] **Lamelas L, López-hidalgo C, Valledor L, Meijón M, Jesús M**. Like mother like son: transgenerational cross-tolerance from drought to heat stress is driven by retained osmoprotective related proteins and miR160. (Under review)

3.2 Introduction

The current context of climate change, with increasingly intense and frequent events, represents a massive threat for plant species (EEA European Environmental Agency, 2015; Lesk *et al.*, 2016; O'Neill *et al.*, 2017). Thus, the ecological consequences are expected to be out of proportion to the relatively short duration of these extreme weather events (Walter *et al.*, 2013). Fortunately, the exposure to one or repeated sub-lethal stress makes plants more tolerant to them. This phenomenon is known as stress memory or priming, and has been observed in various plants in response to different abiotic stresses (Bäurle, 2016; Crisp *et al.*, 2016; Wibowo *et al.*, 2016; Razi and Muneer, 2021; Roces *et al.*, 2022). Increasing our awareness on how plants overcome stress seems fundamental for understanding plant acclimation and, subsequent adaptation to changing and challenging environments. This knowledge has therefore outstanding potential for stabilizing ecosystems in times of intensifying climatic extreme events (Walter *et al.*, 2013).

This acquired stress memory may be heritable in nature and long-term transmitted leading to transgenerational tolerance in plants (Herman and Sultan, 2011; Wang *et al.*, 2016; Wibowo *et al.*, 2016). Hence, plants respond to environmental conditions not only by plastic changes to their own development and physiology, but also by altering the phenotypes expressed by their offspring. This inherited developmental plasticity may promote niche construction through adaptation to hostile environmental conditions, wherein the phenotypes of organisms are plastic and can better deal with their surrounding environment (Moczek *et al.*, 2015). It has been established that environmental challenges to a maternal plant can affect the quantity and composition of starch reserves, epigenetic status, microRNAs, mRNAs, proteins, hormones, and other primary and secondary metabolites packaged into seeds (Herman and Sultan, 2011) leading to altered seed provisioning (Zas *et al.*, 2013).

Evolutionary ecology studies have shown that, in some cases, these inherited effects can include specific growth adjustments that are functionally adaptive to the parental environmental conditions that induced them. These can range from contrasting states of controlled laboratory environments to the complex habitat variation encountered by natural plant populations (Herman and Sultan, 2011; Adrian-Kalchhauser *et al.*, 2020). The latter introduces the idea of stress cross-tolerance, which is based on the suitability of one stressor as a training stress for overcoming others. This occurs probably through

shared stress response mechanisms, such as HEAT SHOCK PROTEINS (HSPs) overaccumulation, which are widely triggered by a plethora of abiotic stresses (Timperio *et al.*, 2008), or more efficient signaling mechanisms involving mainly reactive oxygen species (ROS) (Dickinson *et al.*, 2018).

Increasing evidence underlines the relevance of the chloroplast in modulating the whole cellular redox metabolism (Locato *et al.*, 2018). The chloroplast is a key component in signaling and stress response in plants. Besides its primary role as a light-driven energy factory, the chloroplast perceives signals from the surroundings and transmits these signals to the nucleus. This signaling mechanism regulates alternative splicing (Petrillo *et al.*, 2014; Roces *et al.*, 2022), phytochrome interactions (Dickinson *et al.*, 2018) and, homeostasis under stress conditions (Chen *et al.*, 2010; Foyer *et al.*, 2014), allowing acclimation to ever-changing environmental cues. Chloroplast proteome analysis has been proven useful to explore abiotic stress response (Tamburino *et al.*, 2017; Watson *et al.*, 2018; García-Campa *et al.*, 2022).

As natural plant populations confront rapid environmental changes, researchers are focusing on immediate and transgenerational plasticity as potential sources of adaptive rescue (Beltrán *et al.*, 2018). However, little is known about whether a stress exposure suffered by the parent plants can trigger a better response to other stresses in their progeny, especially for long lived plants as trees in wild or field-managed environments. Thus, to explore the possible transgenerational cross-tolerance and adaptive molecular mechanisms along with population establishment and diversification, environmental variability has been incorporated in the experimental design of this work. Enabling a further characterization of this evo-devo processes at physiological and proteomics levels.

3.3 Materials and Methods

3.3.1 Plant material and stress treatment

Seeds of field-grown *Pinus radiata* were collected from two locations in the Biobío region in Chile: Escuadrón (E), Latitude 36° 56' 49.26" S, Longitude: 73° 8' 49.42" W; and Tranguilvoro (T), Latitude 37° 59' 35.04"S, Longitude 73º 21' 41.48"W. Both locations have a Mediterranean climate but were managed as follows: pines located in E were fertigated once a week using a drip system during the summer, while T pines suffered repetitive mild drought stress events (detailed information in Table S3.1). Originally, both subpopulations came from clone 0027. Ramets of 0027 clone were established by grafting in 1981 in both locations for wood production. Therefore, progenies will be considered isogenic, with the E progeny representing the experiment control line and the T progeny the previously stressed line.

Seeds belonging to both subpopulations were grown in 1 dm^3 pots in a climate chamber (Fitoclima 1200, Aralab) under the following day/night conditions: 16/8 h photoperiod (400 µmol m^{-2} s^{-1}), 25/15 °C and 50/60% relative humidity. Eight-month-old seedlings of the two progenies (height 25 ± 0.3 cm, showing no growth difference between them, t test, $P>0.05$) were used for the heat stress experiment.

The experiment was arranged in a completely randomized block design. The day before starting the experiment (T0), needles were sampled, and chlorophyll fluorescence measured in both progenies. The following day, the heat stress treatment began with an increasing temperature gradient from 15 °C to 45 °C, over a 5 h period, which was then maintained at 45 °C for 6 h. During the following 5 h, the temperature was returned to 15 ºC and maintained for 8 h, thus mimicking a day-night scenario. Heat-stressed plant material was sampled, and chlorophyll fluorescence measured at the end of the 6 h heat exposure on day 1 (T1), day 3 (T3), and day 5 (T5). All the samples and measurements (T0, T1, T3 and T5) were collected at the same time of day to avoid any circadian fluctuations as shown in Fig. 3.1. Plants of each progeny were divided into four pools, constituted of needles of three plants each. These pools were kept across the experiment and formed the four independent biological replicates analysed. Each needle sample was divided into three fractions, one was frozen in liquid nitrogen and stored at −80 ºC for further biochemical analysis, another was kept in ice for immediate fresh chloroplast extraction and the third was used for electrolyte leakage determination.

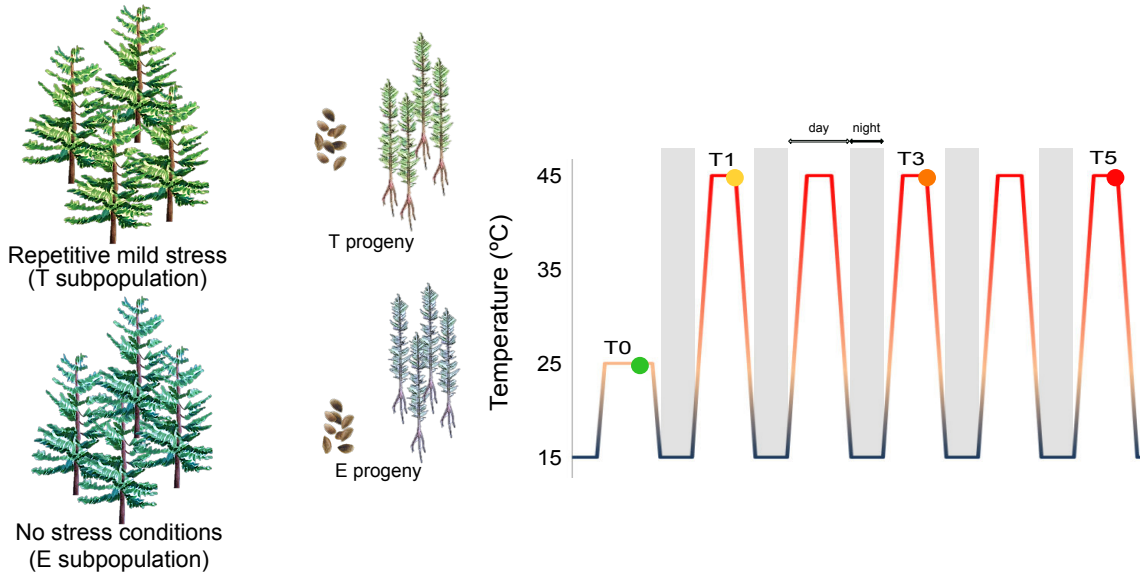


Fig 3.1 Experimental design. Seeds from two isogenic subpopulations with different hydric and nutritional environments were collected and grown under the same favorable conditions. Seedlings belonging to both progenies were exposed to a five-day lasting heat wave, with maximum temperature of 45ºC.

3.3.2 Physiological characterization

The percentage of electrolyte leakage (EL, %) was used to determine cell membrane damage in samples from phase I during the heat stressed treatments (T0, T1, T3, and T5), and previous experiment to establish the optimal temperature, according the following protocol. 70 mg of needles of each biological replicate were collected after at each heat exposure, cut in 1 cm long pieces, immediately immersed in sterile de-ionized water, and then incubated for 24 hours at room temperature under agitation at 30 rpm on a Shaker Ch-4103 (Infors HT) experimental conductivity (C_{exp}) was then measured. Maximum conductivity (C_m) was measured after autoclaving for 20 min at 1100 KPa and 121 °C (sensION +MM150 portable meter, Hach), and cooling at room temperature overnight under agitation. Electrolyte leakage (EL) were used to determine leaf membrane damage measurements (see above in plant material and experimental design). EL, was calculated using the equation $EL(\%) = [(C_{exp}-C_i)/(C_m-C_{ii})]$ x100, where C is water conductivity under control conditions (i), control conditions after 24h (ii).

Chlorophyll fluorescence measurements were taken just prior to sampling using a pulse-amplitude modulation fluorimeter (OS1-FL, Opti-Sciences, Hudson). Quantification of Chlorophyll a, Chlorophyll b, Carotenoids, Malondyaldehyde (MDA), Free Amino Acids (FAA), Total Soluble Sugars (TSS), Starch (STA), Total Flavonoids (TFL), and Total

Phenolics Compounds (TPC) were performed according to López-Hidalgo *et al.* (2021). This analysis was performed for all sampling points and both progenies starting from 10 mg of lyophilized needles.

Chloroplast isolation and protein extraction

Chloroplast enriched fraction for proteomics analyses was purified as described (Lamelas *et al.*, 2020). In brief, samples were homogenized, incubated in cellular lysis buffer, and then filtrated. Organelle enrichment was achieved with sucrose-Percoll discontinuous gradients for intact chloroplast isolation. Protein extraction was performed following a phenol-SDS protocol (Valledor and Weckwerth, 2014). As protein samples were dissolved with the detergent SDS, an in-gel digestion was performed using trypsin (Roche, cat. no. 03 708 969 001) according to the manufacturer's indications. Peptides were extracted and desalted as previously described (Valledor and Weckwerth, 2014). The peptides were analyzed using a 1D nano-flow LC coupled to an MS/MS Orbitrap Fusion spectrometer (ThermoFisher Scientific), using a 60 min gradient starting with 0.1% formic acid and with 80% acetonitrile as the mobile phase.

Protein identification, quantification, and *in silico* subcellular location

Protein identification and label free quantification were performed with Proteome Discoverer v2.2 (ThermoFischer). A combined database was compiled with three protein databases and used for protein identification, including the *Pinus taeda* genome v.1.01 (https://bioinformatics.psb.ugent.be/plaza/versions/gymno-plaza/), UniProt/SwissProt Viridiplantae, and *P. radiata* transcriptome (Escandón *et al.*, 2022).

Identified proteins were blasted using the following *in silico* localization tools: BUSCA (Savojardo *et al.*, 2018), Localizer (Sperschneider *et al.*, 2017), YLoc (Briesemeister *et al.*, 2010) and TargetP (v2.0) (Almagro Armenteros *et al.*, 2019). Then, proteins were annotated with sma3s (Casimiro-Soriguer *et al.*, 2017) and Mercator MapMan (Lohse *et al.*, 2014).

Finally, contamination sources were addressed by dropping for downstream analyses proteins with less than two matches for chloroplast location considering *in silico* localization tools or with no positive chloroplast location in the annotation according to Mercator Mapman or sma3s.

3.3.3 Statistical analysis

All statistical analyses were conducted in R (v 4.0.2) (R Core Team, 2020). Agricolae R package (http://CRAN.R-project.org/package=agricolae) was employed for biomarker significance testing according to de Mendiburu, (2017). Proteomics data were analyzed using pRocessomics R package (available at htttps://github.com/Valledor/pRocessomics) and described in Chapter 2, to perform data pre-processing, univariate (Venn) and multivariate (PCA and k-means) analyses. In brief, each proteomics dataset was pre-processed independently, keeping those proteins that were present in at least 15% of the samples or in all the replicates that constituted a treatment. Missing values were imputed using the Random Forest algorithm (Stekhoven and Bühlmann, 2012). After data pre-processing, univariate analyses were performed, and then both datasets were z-scaled for multivariate analysis.

3.4 Results

3.4.1 T progeny showed improved performance under heat stress associated to enhanced basal photosynthesis and altered starch and soluble sugars content

It was mimicked a five-day heat wave in a time-series experiment to elucidate whether there were differences in the response between the two plant isogenic progenies tested, whose direct ascendents grew in differential environmental conditions. The heat stress applied had a strong physiological impact over both groups of seedlings as shown in Fig. 3.2. Most of the biomarkers measured changed along with the applied stress ($P < 0.05$, HSD test), and more interestingly some of them (Fv/Fm, EL, TSS, and STA) exhibited differences prior to the stress exposure (T0; $P < 0.05$, t test). Additionally, the E progeny showed a wider dispersion in most of the quantifications.

Common to both progenies, photosynthesis, measured through Fv/Fm, suffered a decay since the first stress exposure. While chlorophyll a, and b, decreased continuously in E offspring plants, T descendant plants were able to re-raise their chlorophyll content values in the last day of stress. However, significant changes in carotenoids levels were not detected.

Regarding the carbohydrate metabolism, an accumulation of starch was observed in the last stress shock along with a moderate sugar content increase for E pines progeny. On its turn, the T progeny exhibited a higher amount in TSS and a more stable content of STA in all the tested experimental conditions. The secondary metabolism was characterized by TPC and TFL. Both progenies increased their needle TPC content. However, TFL lowered its abundance under stress in E seedlings. In contrast, T seedlings exhibited a minimum TFL content in T1 which returned to basal levels in the third and fifth day. FAA content drastically increased at T5 in the T group pointing to deep changes in protein metabolism dynamics after several days of stress. In summary, T progeny exhibited a maintained activity of the primary metabolism along the stress (TSS and STA) and enhanced secondary metabolism (TPC and TFL) at final stress point coupled to an increase in chlorophyll content. These differences can be attributed to a better performance towards the stress, indicating alternative acclimation processes and highlighting the role of chloroplasts in the response to severe heat stress.

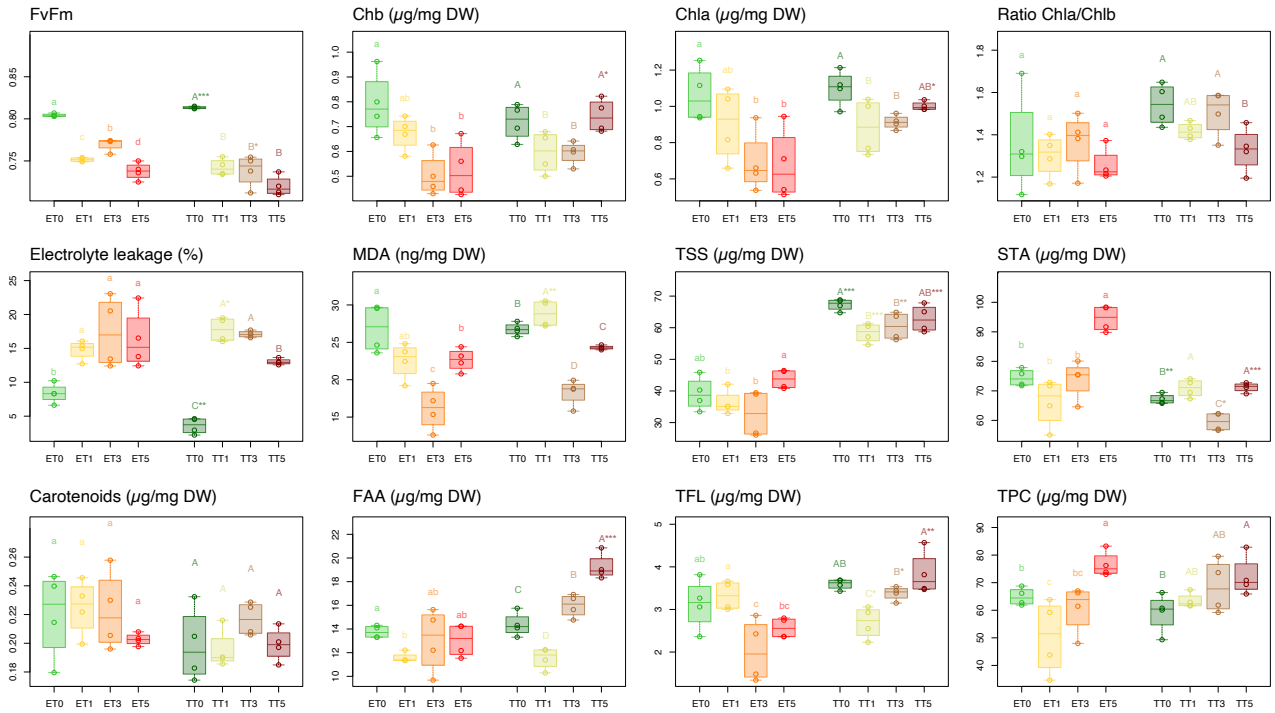


Fig 3.2. Physiological profile of the seedlings from E and T progenies including: Fv/Fm, Chlorophyll b, Chlorophyll a, ratio Chla/Chlb, Electrolyte leakage, malondialdehyde (MDA), total soluble sugars (TSS), starch (STA), Carotenoids, Free amino acids (FAA), total flavonoids (TFL) and total phenolic compounds (TPC) content in basal conditions (T0), after 1 day of stress (T1), 3 days of stress (T3) and 5 days of stress (T5). Different letters indicate significant differences according to HSD test (lower case for E progeny and upper case for T progeny; $p < 0.05$); asterisks indicate significant differences between the progenies for a specific sampling point. (* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$).Dots represent mean values of four independent biological replicas and three technical replicas each.

3.4.2 Chloroplast proteome revealed differences in the basal photosynthetic machinery between the two progenies

A subcellular analysis was performed to determine whether the chloroplast proteome may shed light on the dissimilarities of heat stress response found at the physiological level between the two progenies. This approach allowed to identify and quantify 1847 proteins. Out of them, 1428 were found to be present in at least four samples out of 32. From these, 1243 proteins (Table S3.2) were confirmed to belong to the chloroplast by at least two *in silico* subcellular location tools or one annotation tool (Fig. S3.1a). These yielded an average ratio of 92.71% in abundance (Fig. S3.1b), validating the accuracy of the employed methodology (sample contamination distribution is shown in Fig. S3.1c).

A general depiction of the chloroplast proteome profile according to Mercator MapMan annotation principal category is provided in Fig. 3.3a. This analysis indicates high similarity between progenies, being noteworthy the peak of the multi-process regulation

cluster after three days in heat stress conditions and of protein homeostasis in T5 for both progenies. Additionally, they were found differences in the external stimuli response category, which reached its maximum in T5 for E progeny plants and earlier in T3 for T descendant plants, being maintained in T5. T offspring showed a milder decrease in hormonal action and secondary metabolism in T1 and T3 in comparation to E. However, prior to stress, clear differences were observed between both progenies for secondary metabolism, RNA biosynthesis, cell wall organization and redox homeostasis. A deeper look into the multi-process regulation (Fig. 3.3b), external stimuli response (Fig. 3.3c), and redox homeostasis (Fig. 3.3d) Mapmans bins highlighted the dynamics of terpenoids biosynthesis and morphogenesis related proteins in E progeny, the increased amount of drought related chloroplast proteins in T0 for T progeny and an altered tocopherol biosynthesis, which suggest variations in retrograde signaling between both subpopulations.

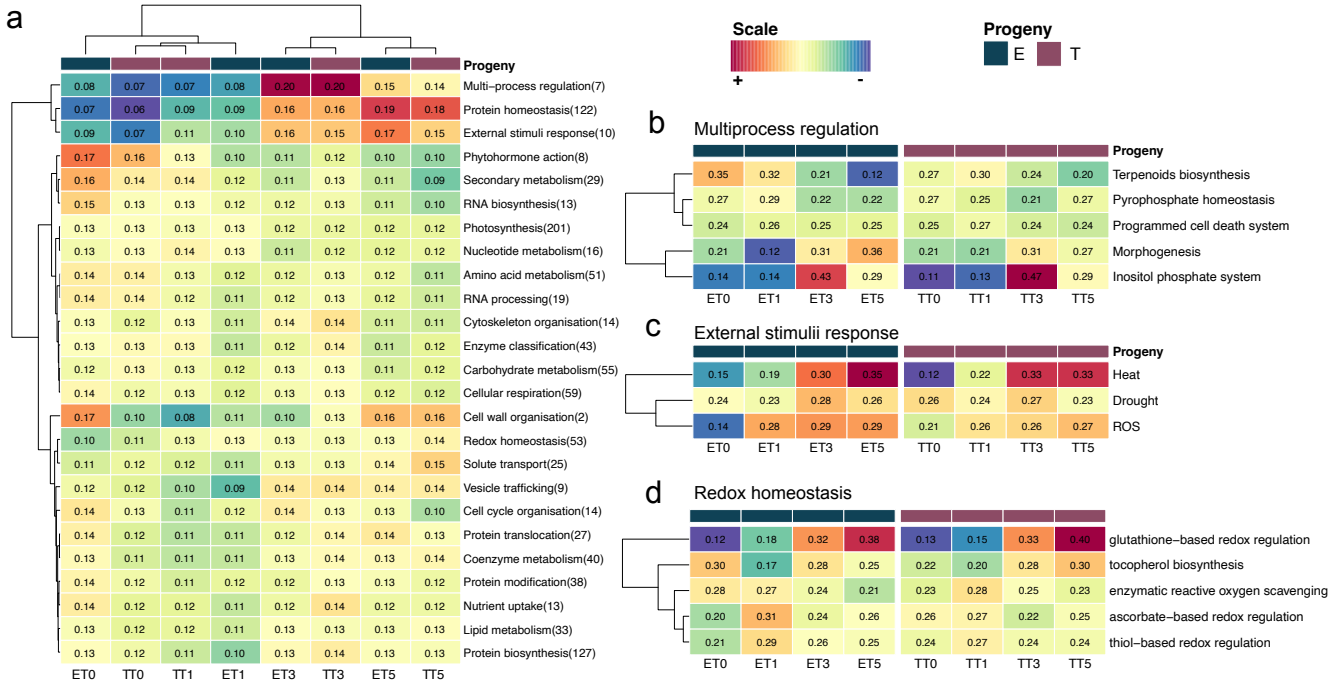


Fig 3.3 Heatmap-clustering analysis using MapMan categorization pathways in the chloroplast proteome of needles of *Pinus radiata* seedlings subjected to heat stress from both progenies. The numbers inside the cells indicate the scaled abundance according to each MapMan functional bin. Manhattan distance and Ward's aggregation method were used for hieratical clustering. The number of proteins included in each category are indicated for each Mapman bin for a) total chloroplast proteome, and sub-heatmaps of b) Multi-process regulation, c) External stimuli response and d) Redox homeostasis.

In order to further identify these differences, the chloroplast proteome for each sampling point by the environment of their ascendants was compared, as shown in Fig. 3.4. The most variable sampling point was T0 (Fig. 3.4a), with a total of 153 (32+15

qualitative and 25+81 quantitative) proteins being qualitative or significantly over- or under-accumulated; followed by 96 (22+16 qualitative and 25+33 quantitative) proteins in T1 (Fig. 3.4b); 46 (18+12 qualitative and 5+11 quantitative) proteins in T3 (Fig.3.4c); and 55 proteins in T5 (Fig. 3.4d) (17+9 qualitative and 15+12 quantitative). A closer look to the differentially accumulated proteins in T plants highlighted photosynthesis phosphorylation related processes in the chloroplast as key variation between the two subpopulations in T0. This increased basal concentration and variety of LHCII proteins in T descendant seedlings (Table S3.3, Supplementary File 3.1) may indicate an increased number of PSII complex in their chloroplasts, which was not coupled to an increased pigment basal concentration but to an enhanced FvFm (Fig. 3.2). Thus, suggesting a drought-induced basal protein fortification of PSII, which is known to be the most labile component of the photosynthetic machinery to several stressors.

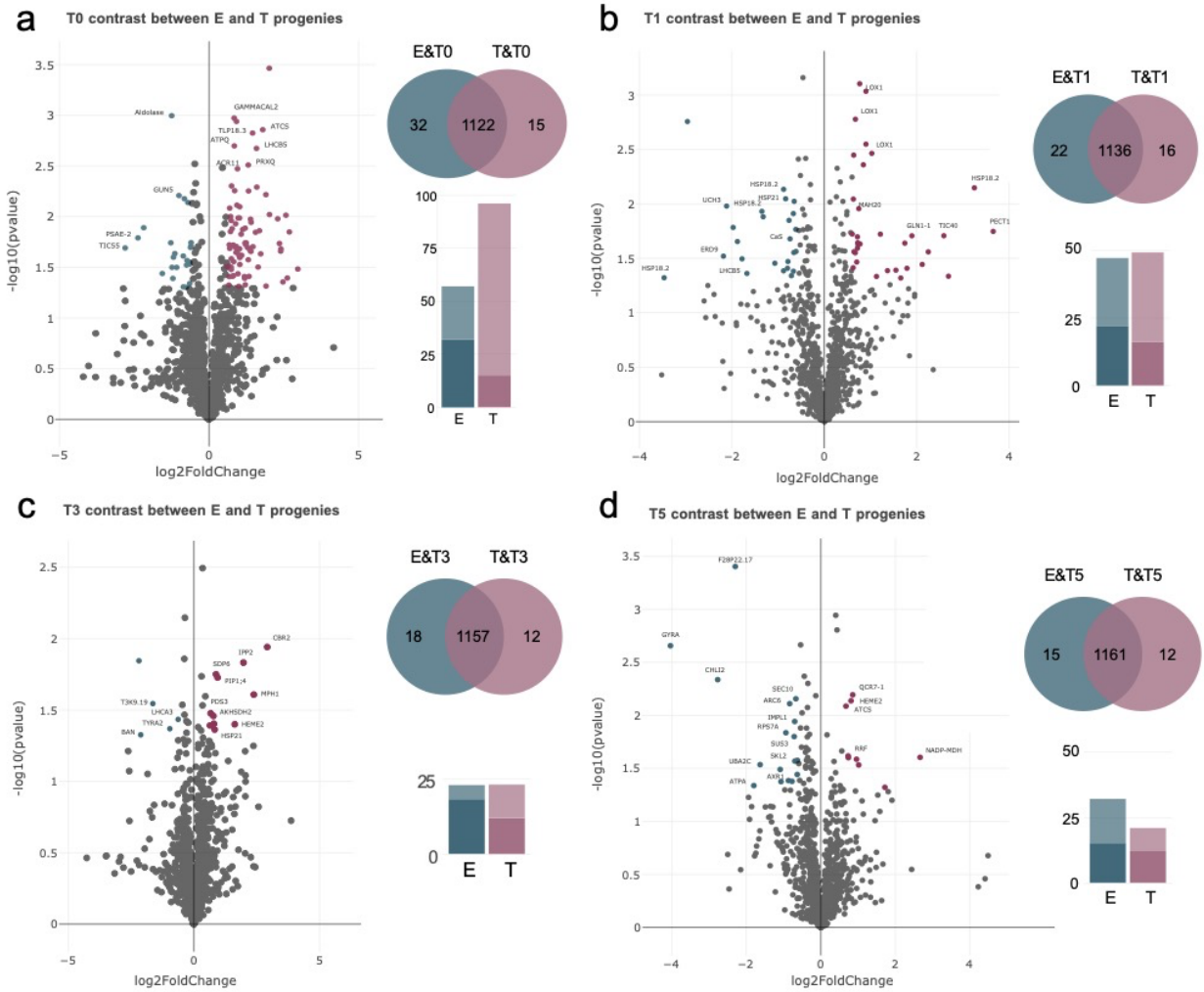


Fig. 3.4 Differential abundance and qualitative analysis comparing both progenies in a) T0, b) T1, c) T3, and d) T5. Volcano plots represent fold change and pvalue, Venn diagrams indicate qualitative chloroplast proteins and bar plots indicate the cumulative sums, where bold color indicates the number of qualitative proteins and light color the number of differentially over-accumulated proteins for each progeny and sampling point.

Interestingly, after the first stress shock non-primed plants showed an increased abundance of small HSPs (sHSPs) in T1, while the potentially primed plants showed an overaccumulation of fatty acids remodelers as lipoxygenases (LOX) protein family (Fig. 3.4b, Supplementary File 3.2, Table S3.3). The latter have been linked to drought and oxidative stress memory formation as dosage-dependent memory genes and a sign of adaptation to drought stress (Sofo et al., 2004; Zhou et al., 2009). Additionally, in T3 and T5, UROPORPHYRINOGEN DECARBOXYLASE (HEME2) (Fig 3.4c and d, Supplementary Files 3.3 and 3.4), a protein involved in the transcriptional regulation of tetrapyrrole biosynthesis (Kobayashi and Masuda, 2016), was found to be significantly higher concentrate in T progeny coupled to an increased chlorophyll content (Fig. 3.2).

3.4.3 Transgenerational cross-stress memory allowed for a wider variety of heat shock proteins after five days at high temperature

As shown in the PCA combining both progenies (Fig. S3.2, Table S3.4) and accordingly to Fig. 3.4, the greatest differences between the two progenies were detected at the initial time point. The first heat stress shock (T1) produced a drastic change in E offspring plants approaching them to T offspring T0 and T1 profiles, which showed a weak response to the first shock. This suggests that the latter group was better prepared for the high temperature. After this first-exposure divergence, both populations seemed to reach an equivalent proteomic profile at T3, evidenced by the samples overlap in Fig. S3.2. This profile was maintained for E subpopulation descendants at T5 while T subpopulation descendants kept evolving its chloroplast proteome. In addition, both progenies presented increased concentration of HSP at T5 as shown in PC1 positive loadings, representing a clear indicator of prolonged heat stress.

After the global analysis of the data, the specific candidates driving the heat stress response for each population can be found in the separate PCAs in Fig. 3.5, Table S3.6. Each subpopulation developed alternative response strategies and timing. As expected, and confirming the extent of the heat stimuli applied to the seedlings, HSPs predominate in both cases in T5 samples (Top positive loadings PC1, both progenies). However, T pines offspring displayed a wider variety of HSP and not only sHSP. It was reported that HSP90 and HSP60 were differentially expressed in distinct varieties according to their heat tolerance in other species (Inoue *et al.*, 2013; Yadav *et al.*, 2020). Moreover, Hsp90C and HSP70 are associated with protein import intermediates (Inoue *et al.*, 2013), through TIC-TOC (translocation at the inner/outer envelope of the chloroplast) components (Sung and

Guy, 2003), and Hsp60 with newly translated polypeptides folding, being Rubisco among its client proteins (Zhao and Liu, 2018).

In addition to the HSPs belonging to the Unfolding Protein Response (UPR) triggered in the chloroplasts in the longer heat stress time, (Fig. 3.5, PC2, top scoring loadings), the accumulation of SELENOPROTEIN O in T progeny at T3 provides some insight on the performance of light reactions. Since SELENOPROTEIN O interacts with Ferredoxin NADP Reductase (FNR) and Protein Proton Gradient Regulation, which are involved in electron transport from Photosystem I (PSI) and osmo-protection (Fichman *et al.*, 2018). Still regarding the PC2 in T progeny, and increased amount of photosynthesis-related proteins was found in T5: light harvesting complex proteins such as LHCB5, LHCA3, LHCB2 and LHC2; PSB 32, involved in the Photosystem II (PSII) repairment; the small subunit of RUBISCO; and PSAG and PSAH of Photosystem I (PSI) complex. These increments occurred parallel to the pigment content increase (Fig. 3.2) and suggest that after five days of heat stress a deep repairment of photosynthetic function took place exclusively in the offspring of stressed parental pines. Conversely, despite progeny E suffers considerable damage in photosynthesis components in T1, as indicated by Fv/Fm data (Fig. 3.2) and the loadings of PC2, did not seem to present an efficient repair system.

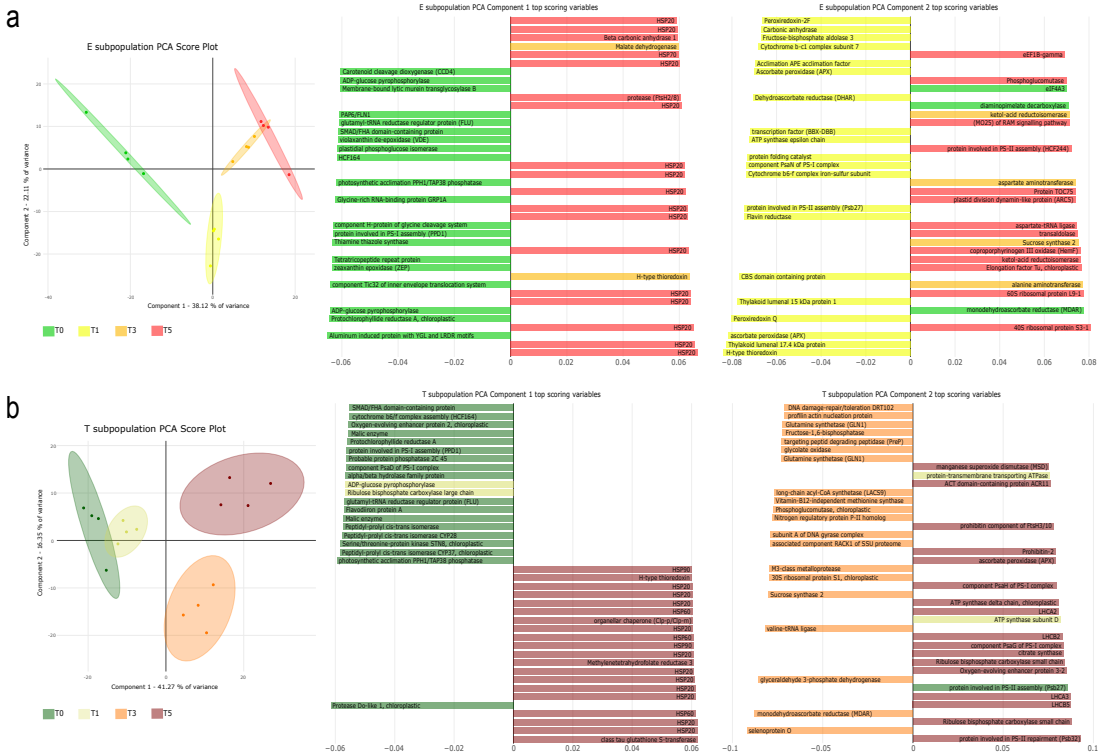


Fig. 3.5. PCA Score plot and PC1 and PC2 top-ranked proteins. (a) E progeny and, (b) T progeny chloroplast proteome. Top 40 scoring loadings (20 highest and 20 lowest) of PC1 and PC2 are shown by row for each PCA, bar colors indicate the experimental condition in which each top-scoring protein is more accumulated. Ellipses show a 90% confidence interval. Different colors indicate different experimental conditions (n = 4 biologically independent replicates)

3.4.4 Protein accumulation pattern allowed for the differentiation of subpopulations initially and during the heat stress

To further characterize the proteome patterns of both progenies a k-means clustering (Fig. 3.6, Table S3.6) was performed. This allowed to determine the shared and divergent changes in the proteome profiles of both plant sets along the stress. Shared profile clusters include 2, 6, and 9; and unrelated clusters 1, 3, 4, 5, 7 and 8.

Among common patterns, HSPs were found to be significantly enriched in ascendent trend cluster 2, along with chlororespiration related proteins; Cluster number 6 is made of plastid ribosomal proteins, which exhibited a drastic increase from T1 to T3, that was then maintained until T5, suggesting a deep proteomic rearrangement, probably also related to HSPs biosynthesis.

Switching the focus to contrasting patterns between the two progenies, Cluster 1, which gathers proteins related to PSII and ATP synthase, displayed opposite trends. While E offspring reached the maximum in T1 and then stabilized, T offspring proteins decreased until T3 and then re-raised their abundance; however, the greatest change in this cluster was found under basal conditions (T0), similarly to cluster 5 also related to PSII. These results are in accordance with the univariate analysis shown in the volcano plot (Fig. 3.4) where the highest variability between the two progenies was found in T0, before to start heat stress conditions.

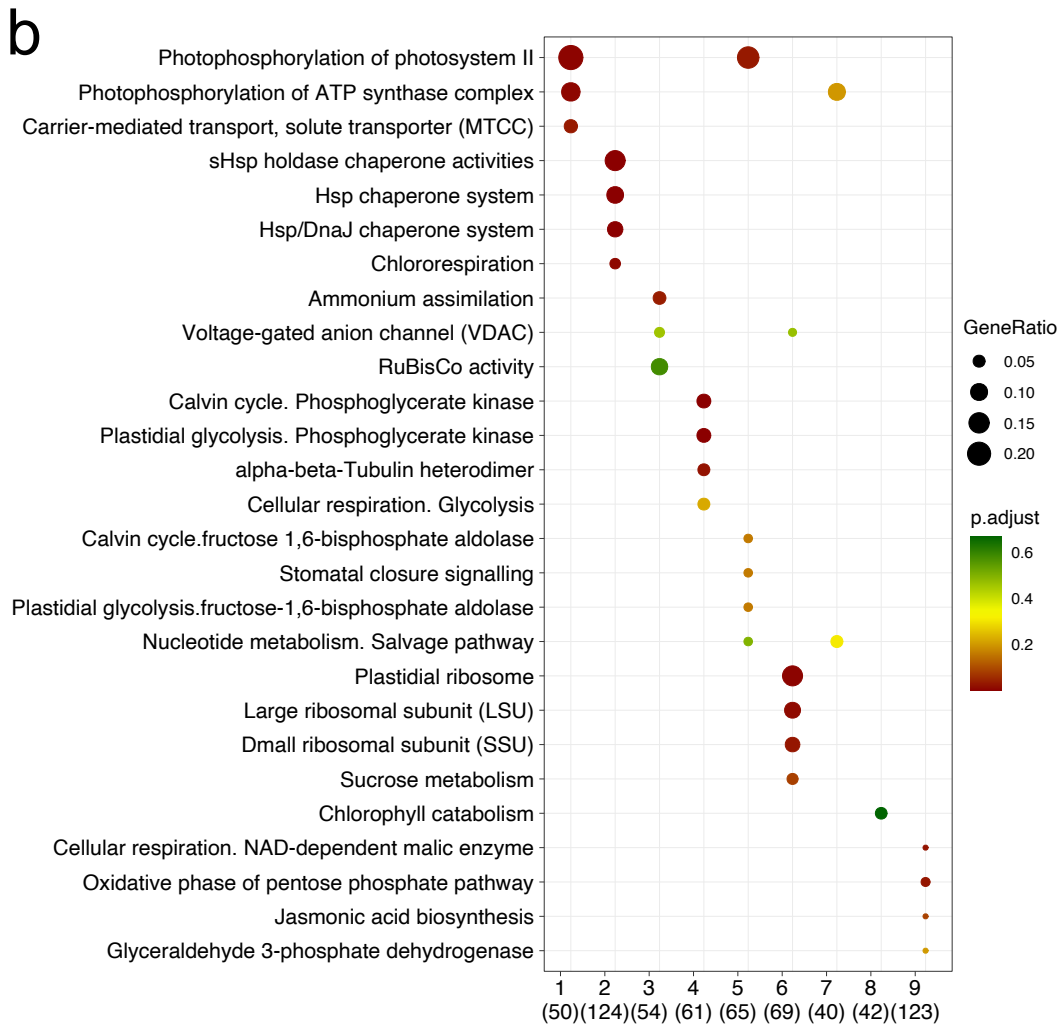


Fig. 3.6. Pattern-clustering and functional enrichment of chloroplast proteome for both progenies a) k-means analysis across the stress. Proteins for each progeny were, scaled using z-scores separately in each data set, and k-means clustered yielding 9 groups. Colors indicate the progeny. Ribbons indicate the 95% confidence interval of the mean values for each protein at each experimental condition (n = 4 biologically independent replicates) and bold lines indicate mean values for each cluster at each experimental condition and progeny. (b) Functional enrichment analysis for each k-means cluster. Different dot colors indicate different corrected pvalues according to Bonferroni method.

3.5 Discussion

Alterations in soil environment are known to induce several changes in the plants, which imply a reprogramming in the communications between shoots and roots, through hormone metabolism, damages in photosynthesis, ROS production, and secondary metabolism compound biosynthesis (Cheng *et al.*, 2011). Heat stress also entails photosynthesis impairment, ROS generation and heat stress responses rely as well in stomatal closure and secondary metabolism biosynthesis (Escandón *et al.*, 2015; Hu *et al.*, 2020). Since the response to a wide range of stresses share a set of common, master response mechanisms, the next step was to test whether ascendants mild repetitive stress events would represent an advantage to overcome a “new” or “not previously memorized” abiotic stress. This phenomenon relying on transgenerational stress cross-tolerance would help plants to adapt to challenging scenarios and may represent a useful tool for breeders to design climate change resilient forests.

Thus, in this work, it has been characterized firstly basal alterations in seedlings. And secondly, heat stress response differences between two progenies through biochemical profiling, and chloroplast proteomics. This experimental system, consisting on the progeny of two wild isogenic subpopulations, permitted to decipher stress cross-tolerance as a proof of principle to understudied mechanisms involving long-term memory in one generation and adaptation in its progeny; as well as deeper understand the natural mechanisms leading to population conformation.

3.5.1 Shared response mechanisms between heat and drought stresses may be the key for transgenerational abiotic stress cross tolerance

One of the most representative differences in the biochemical analysis in this work was the increased and retained concentration of soluble sugars in primed progeny needles coupled to a decrease in electrolyte leakage after five days of stress treatment. Soluble sugars are well-known osmolytes, signaling molecules, which also play an essential role in protecting membrane stability under heat stress (Wang et al. 2016). These are also reported to prevent impairments in other abiotic stresses as drought or salinity and, in some cases, stabilize biomolecules (Garg *et al.*, 2002). Additionally, the decreased electrolyte leakage, closely related to membrane integrity along with the chlorophyll content re-raise at long-term may indicate a better health status of T progeny plants at the end of the experiment, suggesting an enhanced tolerance in the drought primed progeny.

3.5.2 Different progenies developed different strategies based on short- and long-term heat stress tolerance acquisition

The applied stress had a sharp impact in E progeny, however, for T progeny, whose parents have passed through repetitive mild drought stress and nutrient deficiency, it was found a very similar proteome profiling when comparing basal conditions with the first six hours after heat stress (T1). This fact could be due to a rapid response, sufficient to contain the damage caused by the first shock, or a less sensitive heat stress phenotype. Both of them could be attributed to an inherited priming. These results support the inherited priming hypothesis and validate the aim of this work.

After five days of heat exposure, the analyses showed a greater abundance of HSP90 and HSP60 proteins in the T progeny, while a majority of sHSPs was found in E progeny. sHSPs are ATP-independent, and mechanistically, instead of refold unfolded or denaturalized proteins they bind early unfolded proteins to prevent them from irreversible aggregation, however, whether client proteins exceeded sHSP, the latter can be recruited to the insoluble fraction (Haslbeck and Vierling, 2015; Haslbeck *et al.*, 2019) the disassembling of insoluble substrate aggregates with incorporation sHSPs and substrate refolding can only be achieved by HSP60/HSP90 ClpB Chloroplast bichaperone system and require the consumption of ATP. Interestingly, even though both subpopulation exhibited large amounts of HSPs, it is noteworthy that “potentially primed subpopulation” T presented a shifted ratio HSPs/HSP in the longest stress exposure time. In addition, HSP90, and 70 are required together with calmodulin, a thermoprotective compound (Guihur *et al.*, 2022), to activate calcineurin signaling, which is a well-known regulator of the responses to a plethora of stressors (Someren *et al.*, 1999).

Small HSPs form a soluble complex with substrate proteins, creating a transient reservoir of substrates for subsequent refolding by ATP-dependent chaperone systems such as HSP60, 70, 90 ClpB and DnaJ DnaK among others (Nakamoto and Vígh, 2007). sHSP reconcentrate in the thylakoids and protect the PSII electron transport system. Also, they have a key role preventing the heat induced destabilization of lipid bilayers. On the other hand, several experiments have confirmed that HSP90 and 60 can induce the refolding of Calvin cycle enzymes, such as citrate synthase and malate dehydrogenase, chloroplast protein import machinery and GUN5, covering the main differences found when comparing the heat stress response of both subpopulations.

Thus, despite sHPS importance in heat stress response, these work as a first stage response to control protein aggregation and not a *de facto* strategy to overcome stress as other HSP family members. Altogether, it seems that the T progeny was able to further develop the UPR process even in the presence of the stressor and the energy (ATP) cost, showing an improved tolerance under the stress.

Despite of the potential relevance of epigenetic mechanisms in transgenerational memory as reviewed in (Bilichak and Kovalchuk, 2016), the data showed that, especially during the stress, this retained memory can be tracked through the chloroplasts, where it can be observed the traces of an altered retrograde signaling through ROS and gun genes, coupled to a photosystem II protein reinforcement, HSP90 and HSP60 families overaccumulation and increased chlorophyll content under the stress conditions. The results also indicated that primed plants were keener to recover the basal chloroplast-to-nucleus signaling and protein folding leading to an increased stress cross-tolerance.

3.6 References

Adrian-Kalchhauser I, Sultan SE, Shama LNS, Spence-Jones H, Tiso S, Keller Valsecchi CI, Weissing FJ. 2020. Understanding 'Non-genetic' Inheritance: Insights from Molecular-Evolutionary Crosstalk. Trends in Ecology & Evolution **35**, 1078–1089.

Almagro Armenteros JJ, Salvatore M, Emanuelsson O, Winther O, von Heijne G, Elofsson A, Nielsen H. 2019. Detecting sequence signals in targeting peptides using deep learning. Life Science Alliance **2**, e201900429.

Bäurle I. 2016. Plant Heat Adaptation: priming in response to heat stress. F1000Research **5**, 694.

Beltrán J, Wamboldt Y, Sanchez R, LaBrant EW, Kundariya H, Virdi KS, Elowsky C, Mackenzie SA. 2018. Specialized Plastids Trigger Tissue-Specific Signaling for Systemic Stress Response in Plants. Plant Physiology **178**, 672–683.

Bilichak A, Kovalchuk I. 2016. Transgenerational response to stress in plants and its application for breeding. Journal of Experimental Botany **67**, 2081–2092.

Briesemeister S, Rahnenführer J, Kohlbacher O. 2010. YLoc-an interpretable web server for predicting subcellular localization. Nucleic Acids Research **38**, 497–502.

Casimiro-Soriguer CS, Muñoz-Mérida A, Pérez-Pulido AJ. 2017. Sma3s: A universal tool for easy functional annotation of proteomes and transcriptomes. Proteomics **17**, 1700071.

Chen M, Galvão RM, Li M, Burger B, Bugea J, Bolado J, Chory J. 2010. Arabidopsis HEMERA/pTAC12 initiates photomorphogenesis by phytochromes. Cell **141**, 1230–1240.

Cheng J, He C-X, Zhang Z-W, Xu F, Zhang D-W, Wang X, Yuan S, Lin H-H. 2011. Plastid Signals Confer Arabidopsis Tolerance to Water Stress. **66**, 47–54.

Crisp PA, Ganguly D, Eichten SR, Borevitz JO, Pogson BJ. 2016. Reconsidering plant memory: Intersections between stress recovery, RNA turnover, and epigenetics. Science Advances **2**.

Dickinson PJ, Kumar M, Martinho C, *et al.* 2018. Chloroplast Signaling Gates Thermotolerance in Arabidopsis. Cell Reports **22**, 1657–1665.

EEA European Environmental Agency. 2015. Global and European temperatures.

Escandón M, Cañal MJ, Pascual J, Pinto G, Correia B, Amaral J, Meijón M. 2015. Integrated physiological and hormonal profile of heat-induced thermotolerance in *Pinus radiata*. Tree Physiology **36**, 63–77.

ç

Escandón M, Valledor L, Lamelas L, Álvarez JM, Cañal MJ, Meijón M. 2022. Multiomics analyses reveal the central role of nucleolus and nucleoid machinery during heat stress acclimation in *Pinus radiata*; bioRxiv, 2022.07.08.499117.

Fichman Y, Koncz Z, Reznik N, Miller G, Szabados L, Kramer K, Nakagami H, Fromm H, Koncz C, Zilberstein A. 2018. SELENOPROTEIN O is a chloroplast protein involved in ROS scavenging and its absence increases dehydration tolerance in Arabidopsis thaliana. Plant Science **270**, 278–291.

Foyer CH, Karpinska B, Krupinska K. 2014. The functions of WHIRLY1 and REDOX-RESPONSIVE TRANSCRIPTION FACTOR 1 in cross tolerance responses in plants: a hypothesis. Philosophical Transactions of the Royal Society B: Biological Sciences **369**, 20130226–20130226.

García-Campa L, Guerrero S, Lamelas L, Meijón M, Hasbún R, Cañal MJ, Valledor L. 2022. Chloroplast proteomics reveals transgenerational cross-stress priming in *Pinus radiata*. Environmental and Experimental Botany **202**, 105009.

Garg AK, Kim J-K, Owens TG, Ranwala AP, Choi Y Do, Kochian L V, Wu RJ. 2002. Trehalose accumulation in rice plants confers high tolerance levels to different abiotic stresses. Proceedings of the National Academy of Sciences **99**, 15898–15903.

Guihur A, Rebeaud ME, Goloubinoff P. 2022. How do plants feel the heat and survive? Trends in Biochemical Sciences **47**, 824–838.

Haslbeck M, Vierling E. 2015. A first line of stress defense: Small heat shock proteins and their function in protein homeostasis. Journal of Molecular Biology **427**, 1537–1548.

Haslbeck M, Weinkauf S, Buchner J. 2019. Small heat shock proteins: Simplicity meets complexity. The Journal of biological chemistry **294**, 2121–2132.

Herman J, Sultan S. 2011. Adaptive Transgenerational Plasticity in Plants: Case Studies, Mechanisms, and Implications for Natural Populations. Frontiers in Plant Science **2**, 102.

Hu S, Ding Y, Zhu C. 2020. Sensitivity and Responses of Chloroplasts to Heat Stress in Plants. Frontiers in Plant Science **11**, 375.

Inoue H, Li M, Schnell DJ. 2013. An essential role for chloroplast heat shock protein 90 (Hsp90C) in protein import into chloroplasts. Proceedings of the National Academy of Sciences of the United States of America **110**, 3173–3178.

Kobayashi K, Masuda T. 2016. Transcriptional Regulation of Tetrapyrrole Biosynthesis in Arabidopsis thaliana. Frontiers in Plant Science **7**.

Lamelas L, García L, Cañal MJ, Meijón M. 2020. Subcellular Proteomics in Conifers: Purification of Nuclei and Chloroplast Proteomes. Methods in molecular biology (Clifton, N.J.). Springer Nature, 69–78.

Lesk C, Rowhani P, Ramankutty N. 2016. Influence of extreme weather disasters on global crop production. Nature **529**, 84–87.

Locato V, Cimini S, De Gara L. 2018. ROS and redox balance as multifaceted players of cross-tolerance: epigenetic and retrograde control of gene expression. Journal of Experimental Botany **69**, 3373–3391.

Lohse M, Nagel A, Herter T, May P, Schroda M, Zrenner R, Tohge T, Fernie AR, Stitt M, Usadel B. 2014. Mercator: A fast and simple web server for genome scale functional annotation of plant sequence data. Plant, Cell and Environment **37**, 1250–1258.

López-Hidalgo C, Meijón M, Lamelas L, Valledor L. 2021. The rainbow protocol: A sequential method for quantifying pigments, sugars, free amino acids, phenolics, flavonoids and MDA from a small amount of sample. Plant, Cell & Environment **n/a**.

de Mendiburu F. 2017. Statistical Procedures for Agricultural Research.

Moczek AP, Sears KE, Stollewerk A, *et al.* 2015. The significance and scope of evolutionary developmental biology: a vision for the 21st century. Evolution & Development **17**, 198–219.

Nakamoto H, Vígh L. 2007. The small heat shock proteins and their clients. Cellular and molecular life sciences : CMLS **64**, 294–306.

O'Neill BC, Oppenheimer M, Warren R, *et al.* 2017. IPCC reasons for concern regarding climate change risks. Nature Climate Change **7**, 28–37.

Petrillo E, Godoy Herz MA, Fuchs A, *et al.* 2014. A chloroplast retrograde signal regulates nuclear alternative splicing. Science **344**, 427–430.

Razi K, Muneer S. 2021. Drought stress-induced physiological mechanisms, signaling pathways and molecular response of chloroplasts in common vegetable crops. Critical Reviews in Biotechnology **41**, 669–691.

Roces V, Lamelas L, Valledor L, Carbó M, Cañal MJ, Meijón M. 2022. Integrative analysis in Pinus revealed long-term heat stress splicing memory. The Plant Journal **112**, 998–1013.

Savojardo C, Martelli PL, Fariselli P, Profiti G, Casadio R. 2018. BUSCA: An integrative web server to predict subcellular localization of proteins. Nucleic Acids Research **46**, W459–W466.

Sofo A, Dichio B, Xiloyannis C, Masia A. 2004. Lipoxygenase activity and proline accumulation in leaves and roots of olive trees in response to drought stress. Physiologia Plantarum **121**, 58–65.

Someren JS, Faber LE, Klein JD, Tumlin JA. 1999. Heat Shock Proteins 70 and 90 Increase Calcineurin Activity in Vitro through Calmodulin-Dependent and Independent Mechanisms. Biochemical and Biophysical Research Communications **260**, 619–625.

Sperschneider J, Catanzariti AM, Deboer K, Petre B, Gardiner DM, Singh KB,

Dodds PN, Taylor JM. 2017. LOCALIZER: Subcellular localization prediction of both plant and effector proteins in the plant cell. Scientific Reports **7**, 1–14.

Stekhoven DJ, Bühlmann P. 2012. MissForest — non-parametric missing value imputation for mixed-type data. Bioinformatics **28**, 112–118.

Sung DY, Guy CL. 2003. Physiological and Molecular Assessment of Altered Expression of Hsc70-1 in Arabidopsis. Evidence for Pleiotropic Consequences. Plant Physiology **132**, 979–987.

Tamburino R, Vitale M, Ruggiero A, *et al.* 2017. Chloroplast proteome response to drought stress and recovery in tomato (*Solanum lycopersicum L.*). BMC Plant Biology **17**, 1–14.

Timperio AM, Egidi MG, Zolla L. 2008. Proteomics applied on plant abiotic stresses : Role of heat shock proteins (HSP) ☆. Journal of Proteome Research **71**, 391–411.

Valledor L, Weckwerth W. 2014. An improved Detergent-Compatible Gel-Fractionation LC-LTQ-Orbitrap-MS workflow for plant and microbial proteomics. Methods in Molecular Biology. United States, 347–358.

Walter J, Jentsch A, Beierkuhnlein C, Kreyling J. 2013. Ecological stress memory and cross stress tolerance in plants in the face of climate extremes. Environmental and Experimental Botany **94**, 3–8.

Wang X, Xin C, Cai J, Zhou Q, Dai T, Cao W, Jiang D. 2016. Heat Priming Induces Trans-generational Tolerance to High Temperature Stress in Wheat. Frontiers in Plant Science **7**.

Watson SJ, Sowden RG, Jarvis P. 2018. Abiotic stress-induced chloroplast proteome remodelling: A mechanistic overview. Journal of Experimental Botany **69**, 2773–2781.

Wibowo A, Becker C, Marconi G, *et al.* 2016. Hyperosmotic stress memory in Arabidopsis is mediated by distinct epigenetically labile sites in the genome and is restricted in the male germline by DNA glycosylase activity (S McCormick, Ed.). eLife **5**, e13546.

Yadav A, Singh J, Ranjan K, Kumar P, Khanna S, Gupta M, Kumar V, Wani SH, Sirohi A. 2020. Heat Shock Proteins. Heat Stress Tolerance in Plants.189–211.

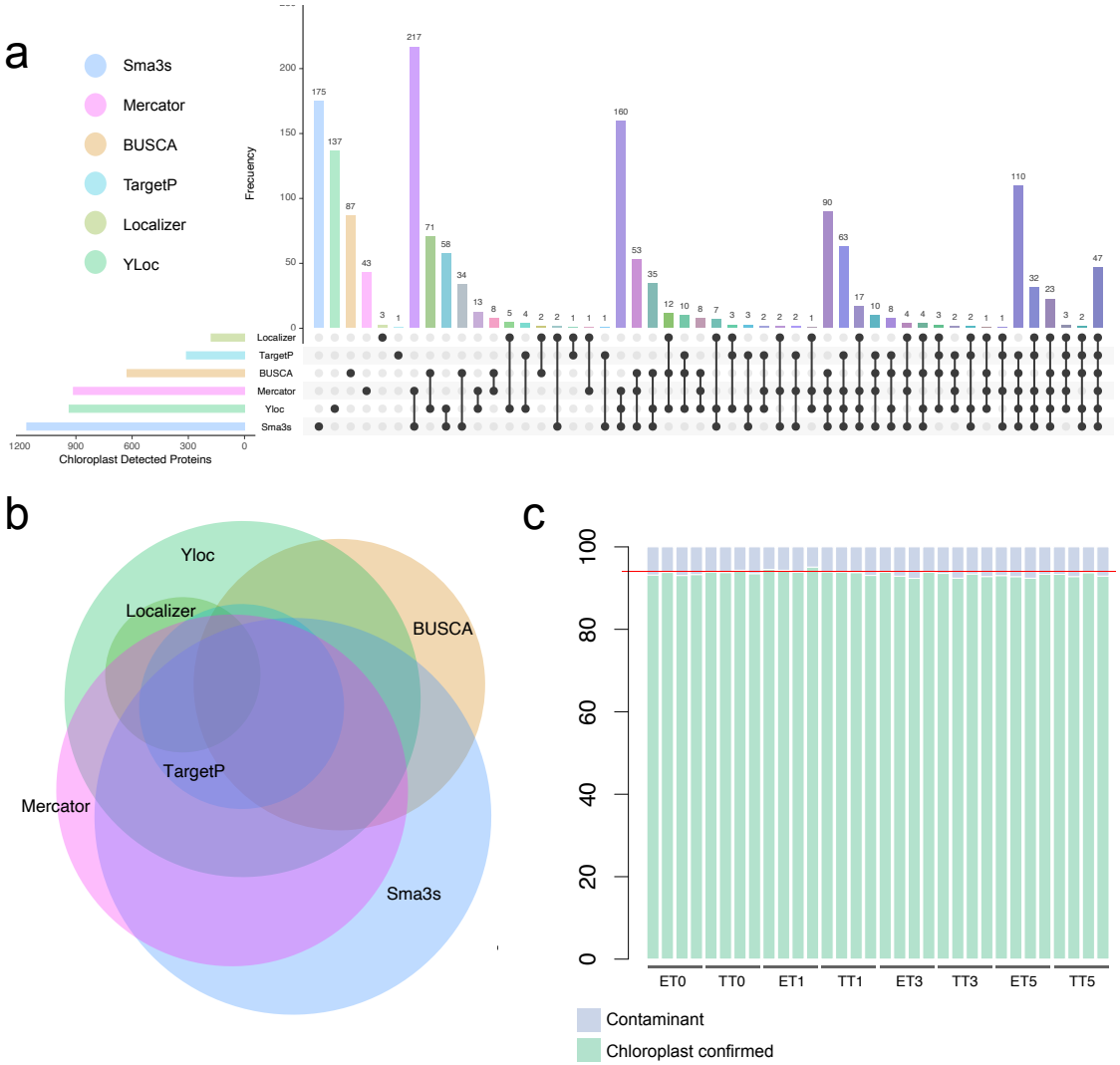
Zas R, Cendán C, Sampedro L. 2013. Mediation of seed provisioning in the transmission of environmental maternal effects in Maritime pine (*Pinus pinaster Aiton*). Heredity **111**, 248–255.

Zhao Q, Liu C. 2018. Chloroplast Chaperonin: An Intricate Protein Folding Machine for Photosynthesis . Frontiers in Molecular Biosciences **4**, 98.

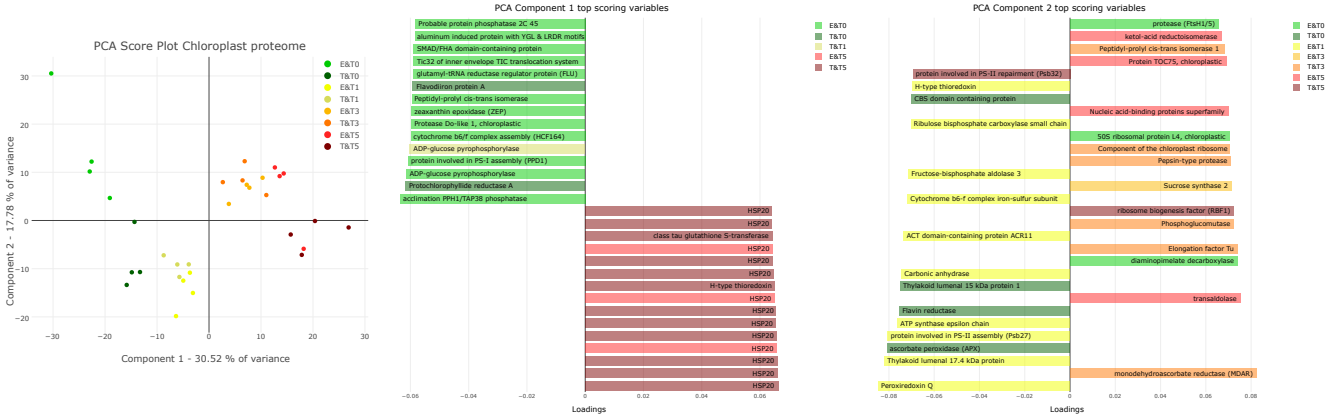
Zhou X, Hua D, Chen Z, Zhou Z, Gong Z. 2009. Elongator mediates ABA responses, oxidative stress resistance and anthocyanin biosynthesis in Arabidopsis. The Plant Journal **60**, 79–90.

3.7 Supplemental information

3.7.1 Supplemental Figures



Supplemental Figure S3.1. Protein contaminants assessment. A and b Venn diagram for chloroplast proteins detected according to each insilico subcellular location tool (Localizer, TargetP, BUSCA and YLoc) or annotation tool (Mercator and sma3s (Gene Ontology). Proteins not identified as chloroplastic by at least two location tools or one annotation tool were considered contaminants. c contaminant distribution (reads) along the samples, the horizontal red line indicates the average amount of contaminants for the data set and showed no bias within the dataset (P = 0.784).



Supplemental Figure S3.2. PCA Score plot and PC1 and PC2 top-ranked proteins. of chloroplast proteome for both progenies. Top 40 scoring loadings (20 highest and 20 lowest) of PC1 and PC2, bar colors indicate the experimental condition in which each top-scoring protein is more accumulated. Different colors indicate different experimental conditions (n = 4 biologically independent replicates)

3.7.2 Supplemental tables and files

Supplemental table legends

Supplemental Table S3.1. Environmental conditions of E and T subpopulations

Supplemental Table S3.2. Chloroplast proteins quantification, statistics and description according to Mapman annotation

Supplemental Table S3.3. List of qualitative and quantitative proteins for each sampling point

Supplemental Table S3.4. PCA explained variance and loading for both progenies

Supplemental Table S3.5. PCA explained variance and loading for (a) E progeny and, (b) T progeny

Supplemental Table S3.6. kmeans clustering of chloroplast proteins

Supplemental File legends

Supplemental File S3.1. Volcano interactive representation for T0 contrast between the chloroplast proteome of both progenies.

Supplemental File S3.2. Volcano interactive representation for T1 contrast between the chloroplast proteome of both progenies.

Supplemental File S3.3. Volcano interactive representation for T3 contrast between the chloroplast proteome of both progenies.

Supplemental File S3.4. Volcano interactive representation for T5 contrast between the chloroplast proteome of both progenies.

CHAPTER IV. Nuclear proteomics in response to heat stress and recovery in *Pinus radiata*[1]

4.1 Abstract

How do abiotic stresses affect nuclear proteome and mediate memory? Despite the relevance of this question in the present context of climate emergency, its answer remains unknown for most species. This chapter aims to define how *Pinus radiata* nuclei respond, acclimate, and remember heat stress. Seedlings were heat-stressed at 45 ºC in a 10-day-stress and recovery experiment. Nuclear proteins were isolated and quantified by nLC-MS/MS, and potential acquired memory was analyzed in recovered plants. Specific nuclear heat responsive proteins were identified, and its biological role evaluated employing a systems biology approach. In addition to HSP, several clusters involved in regulation processes, as epigenomic-driven gene regulation, some transcription factors and a variety of RNA-associated functions were discovered. Nuclei exhibited differential proteome profiles across experiment, being notably H2A histone and methyl cycle enzymes accumulated at the recovery step. These results suggest that epigenetic mechanisms play a key role in heat stress tolerance and priming mechanisms.

[1] **Lamelas L, Valledor L, Escandón M, Pinto G, Cañal MJ, Meijón M. 2020**. Integrative analysis of the nuclear proteome in Pinus radiata reveals thermopriming coupled to epigenetic regulation. *Journal of Experimental Botany* **71**: 2040–2057.

4.2 Introduction

Heat stress response implies changes in multiple mechanisms and metabolic pathways. And, although a few of the more important players in heat stress response and adaptation have been already depicted in *P. radiata*, as heat shock proteins, flavonoids or fatty acids (Escandón *et al.*, 2015, 2017, 2018), how gene regulation drives these biochemical and physiological responses remains elusive.

Nuclear proteome dynamics is crucial to increase our understanding of how both environmental and cytoplasmic signals are sensed and translated into molecular responses, mainly through the proteins that guide and control the gene expression. Nuclear proteomics is an useful approach not just for investigating the mechanisms underlying plant responses to abiotic stresses, including protein–protein interactions, enzyme activities, and post-translational modifications (Yin & Komatsu, 2016), but also with potential to create solutions to improve forest management and breeding programs.

Nuclear post-transcriptional regulatory mechanisms involving the processing of precursor mRNA as alternative splicing (Ling *et al.*, 2018), as well as structural variations in histone H2A and H2B dimers (Talbert & Henikoff, 2014) play an important role in relation to heat stress memory in Arabidopsis. Nevertheless, epigenetic factors are thought to play the main role in establishing this heat stress memory (Ling *et al.*, 2018).

In this respect, and in relation to nuclear proteins relevance when driving heat stress adaptation, it has been described in *Arabidopsis thaliana*, the key role of epigenetic regulation and histone modifications to remain memory of the stress (Bäurle, 2016; Lämke *et al.*, 2016) that lead to priming mechanism (Martinez-medina *et al.*, 2016). Priming involves a first training stress, a latent phase and a second stress event; in this later stress, the plant will be able to react in a more efficient way than previously, due to the information stored as chromatin structural changes and histone modifications (Gutzat and Scheid, 2012; Pastor *et al.*, 2013; Asensi-Fabado *et al.*, 2017). According to these findings, the epigenetic mechanisms, and particularly DNA methylation and nucleosome occupancy, seem to be the main players on priming establishment. The epigenetic mechanisms involve covalent modifications of DNA and histones, which affect transcriptional activity of chromatin (Valledor *et al.*, 2007). Since chromatin states can be propagated through cell divisions, epigenetic mechanisms are thought to provide heritable ‘cellular memory’ (Iwasaki & Paszkowski, 2014).

This work aims to provide a comprehensive knowledge of heat stress response and adaptation at nuclear level that will allow the depiction of the nuclear events involved in heat stress memory. Thus, the main goal consists of the characterization, quantification and biological interpretation of nuclear proteome of *P. radiata* needles in response to high temperature stress and the assessment of the recovery stage

The relevance of the discovered proteins related to chromatin reorganization will suppose a major advance in heat stress biology field, also providing a set of key nuclear elements in cytoplasmic proteome reorganization during the heat-stress response and recovery process.

4.3 Material and Methods

4.3.1 Plant material and experimental design

P. radiata seedlings were grown in a climate chamber under a photoperiod of 16 h (400 µmol m^{-2} s^{-1}) at 25 °C and 50 % relative humidity (RH), and 15 °C and 60 % RH during the night period (Fitoclima 1200, Aralab). The plants were watered with nutritive solution (N:P:K, 5:8:10), same conditions as stated in Chapter 3.

Eight-month-old seedlings (plant height 24 ± 0.4 cm) were sampled (T0) and then divided in two sets. The following day Plant Set I was heat-stressed as described in Chapter 3, while Set II was maintained under control conditions to test whether it was any long-lasting difference after the stress ended. Set I seedlings were sampled under control (T0) and heat stressed conditions according to the scheme detailed in Fig. 4.1a, and extending the experimental design of Chapter 3 to 10 days, plant material was collected at the end of the 6 h heat exposure on day 1 (T1), day 3 (T3), day 5 (T5) and day 10 (T10).

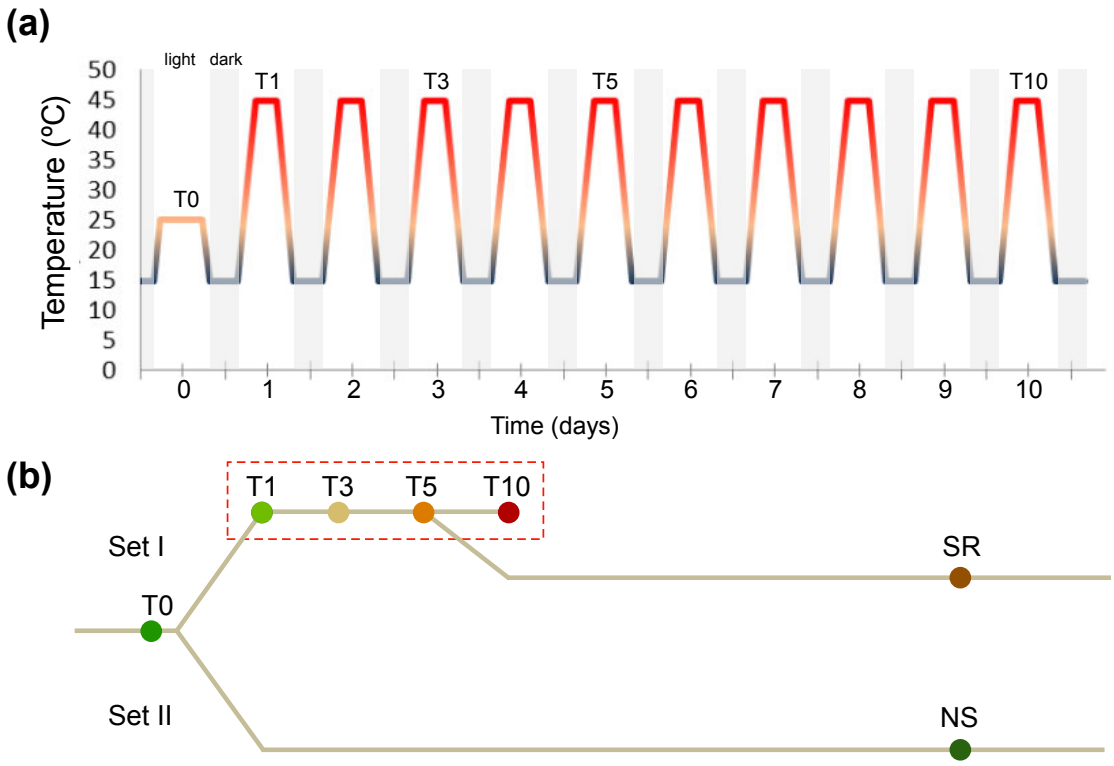


Fig.4.1. Experimental design. a) Detailed temperature profile of control (T0) and heat stress conditions (T1, T3, T5, and T10), all sampling procedures were performed after 6 hours at the maximum temperature of 45ºC. b) Outline of the experimental set up, heat treatments are squared in a red discontinuous line. SR (Stress recovered) and NS (not stressed) were maintained under optimum conditions over one month, and then sampled without any further stress treatment.

Since T10 plants were too damaged, T5 plants were selected to continue to test the possible memory acquisition process after the one-month recovery stage. Thus, T5 and Set II seedlings were maintained under control conditions and sampled after one month of the heat treatment, when the plants showed an apparent physiological recovery (Fig. 4.1b; SR, stress-recovered and NS, not-stressed respectively). Needles were frozen in liquid nitrogen immediately after sampling, and conserved at -80 ºC until nuclei and protein isolation were performed. Leaf gas-exchange parameters were quantified just prior to sampling. Electrolyte leakage (EL) was measured for heat stressed samples as described in Chapter 3.

In both sets, plants were divided into four pools, constituted of needles of three plants each. These pools constituted the four biological independent replicates that were analyzed.

4.3.2 Gas-exchange measurements

Net CO_2 assimilation rate (A, μmol CO_2 m^{-2} s^{-1}), stomatal conductance (g_s, mol H_2O $m^{-2}s^{-1}$), transpiration rate (E, μmol H_2O $m^{-2}s^{-1}$) and intercellular CO_2 concentration content (Ci, ppm) were measured in all plants and averaged for each biological replicate per each treatment in basal conditions and during the heat stress, using a portable infrared gas analyser (LCpro-SD, ADC BioScientific Ltd., UK) equipped with the broad leaf chamber. To find out the saturation light intensity A/PPFD (photosynthetic photon flux density; light response curves of CO_2 assimilation) curves were performed with the following PPFD: 2000, 1500, 1000, 750, 500, 250, 100, 50 and 0 μmol $m^{-2}s^{-1}$. After A/PPFD data analysis, punctual measurements at saturation light intensity were performed at 1000 μmol $m^{-2}s^{-1}$. The following conditions were maintained inside the chamber during all the measurements: air flux: 200 mol s^{-1}; block temperature: 25 ºC; and atmospheric CO_2 and H_2O concentration. Data were recorded when the measured parameters were stable. Instantaneous carboxylation efficiency (CE, μmol CO_2 m^{-2} s^{-1}) was calculated as the coefficient between A and Ci.

4.3.3 Nucleus isolation and protein extraction

Nuclear proteome was analysed the four biological replicates of phase I at all sampling times (T0, T1, T3, T5, T10, NS and SR). Nuclei were isolated following the protocol described by Lamelas *et al.*, (2020). Nuclei isolation and enrichment performance

was assessed by confocal microscopy (Leica TCS-SP2-AOBS) using propidium iodide (PI) dye (Fig. S4.1). The PI dye was excited by 488 nm diode laser and the fluorescence emission was recorded at 636 nm. Images were processed and analyzed using Fiji software (Schindelin *et al.*, 2012).

Once the nuclei were purified, samples were sonicated in 300 µL of 1 % SDS for 15 s at 60 % amplitude (HielcherUP200S) and then incubated in a vortex at maximum speed for 15 min at room temperature. Subsequently, 300 µL of extraction buffer (1.5 M sucrose, 10 mM DTT and 300 µL of phenol) were added to begin protein extraction. After mixing vigorously, tubes were centrifuged 5 min at 17,000 g and room temperature. After centrifugation, phenolic (upper) phase was saved and the lower phase was re-extracted by adding 300 µL of phenol. Both phenolic phases were collected and cleaned with extraction buffer in the same way to conserve the upper phase. Proteins were precipitated by adding 0.1 M ammonium acetate in methanol and incubated overnight at -20 °C. Tubes were centrifuged, and protein pellets washed twice with acetone. Dry pellets were dissolved in 1.5 % SDS, 8 M Urea. Protein content was quantified by BCA assay (Smith *et al.*, 1985). The enrichment in nuclear proteins was assessed by comparing running nuclear protein fraction and total protein in 1-DE SDS-PAGE.

4.3.4 Protein identification and quantitation by GeLC-Orbitrap/MS analysis

Proteins were in gel-digested with trypsin (Roche, Cat. No. 03 708 969 001), obtained peptides were extracted and desalted as described by Valledor and Weckwerth (2014).

Peptides were then analysed in Central Support Service for Research of the University of Cordoba (SCAI) employing one-dimensional (1D) nano-flow LC coupled to MS/MS Orbitrap Fusion (Thermo) spectrometer, using a 60 minutes gradient starting with 0.1 % formic acid and as mobile phase 80 % acetonitrile.

Three protein databases were used for protein identification: *Pinus taeda* genome v.1.01 (http://dendrome.ucdavis.edu), Uniprot/SwissProt, *Viridiplantae* and an in-house *Pinus radiata* transcriptome, following the recommendations described by Romero-Rodríguez *et al.*(2014).

Proteome Discoverer 2.2 was employed for the identification and quantification of proteins employing a 2 % false discovery rate (FDR), XCorr of 1.6, one unique or razor

peptide for identification, and one peptide (unique/razor) per protein for label free quantification. Lysine ubiquitination, methionine oxidation, acetylation of protein N-terminus and phosphorylation of serine, threonine and tyrosine were taken into account as dynamic modifications.

Identified protein sequences were analysed with PlantTFcat tool (Dai *et al.*, 2013), to identify transcription factors and nuclear regulators domains, and two independent and plant-specific subcellular localization tools, Localizer v1.0.4. (Sperschneider *et al.*, 2017) and YLoc (Briesemeister *et al*, 2010a, 2010b) using YLoc+ prediction model and plant version with a fixed probability greater than 0.75 and medium confidence score (at least 0.4). Proteins were also annotated using Mercator sequence annotation tool against TAIR, SwissProt/UniProt Plant Proteins, Clusters of orthologous eukaryotic genes (KOG) databases (Lohse *et al.*, 2014) and Uni-Prot KB/Swiss-Prot using sma3s v2 (Casimiro-Soriguer, 2017). Restrictive conditions were used to catalog nuclear proteins. Proteins not predicted to belong to the nucleus or endoplasmic reticulum, by at least of two from the five selected data sources, were dropped from the analysis.

4.3.5 Statistical and bioinformatics analysis

All statistical procedures were conducted with the R programming language running under the open-source computer software R v3.4.0 (R Core Team, 2017) and RStudio v1.1.456, available from http://www.rstudio.org/.

Four biological replicates were used for all statistical procedures. The agricolae package (de Mendiburu, 2017) was used for univariate one-way analysis of variance (ANOVA) followed by post-hoc multiple comparisons using Tukey's test (function HSD.test) to estimate the significance of the leaf gas-exchange and membrane damage data ($p \leq 0.05$).

Nuclear proteome was analyzed using pRocessomics R package (https://github.com/Valledor/pRocessomics) (described in Chapter 2), which was used to pre-filter and impute missing values according to Random Forest algorithm, MissForest package, (Stekhoven & Bühlmann, 2012), and consistency-based criteria with a threshold of 0.25. Then the three most stable protein abundances were identified with sLqPCR package, according to Vandesompele *et al*. (2002) to equalize inter-sample total intensity variation, whose mean values were used for normalizing each sample.

Most significant nuclear proteins were selected by performing an ANOVA analysis considering as cut off a q-value of 0.05 calculated using Benjamini-Hochberg model. Then, mixOmics R package (Rohart, 2017) was used to perform the statistical analysis including Principal Component Analysis (PCA) and sparse Partial Least Square (sPLS) providing two networks: first of them integrating physiologic and proteomics data, and a second network with identified transcription factors with significantly different nuclear proteins. Both networks were filtered by applying a net cut-off of 0.75.

4.4 Results

4.4.1 High temperature affects gas-exchange parameters and membrane integrity

Analysis of leaf gas-exchange parameters uncovered significant physiological changes in response to heat stress. Stressed plants showed significant decreases in net CO_2 assimilation rate (*A*), carboxylation efficiency (CE), stomatal conductance (g_s), and transpiration rate (*E*) Fig 4.2 a-d. The effects were more severe when the stress exposure time was increased, and showed a similar pattern in the four parameters. This, coupled with maintenance of the intercellular CO_2 content Fig. 4.2e, suggested a decrease of carbon fixation in the stressed plants in favor of photorespiration at sampling time T10, when the intercellular CO_2 concentration (C_i) was increased. In addition, the electrolyte leakage (EL) showed significant differences at T5 and it increased dramatically at T10, indicating severe disruption of cell membrane integrity (Fig. 4.2f).

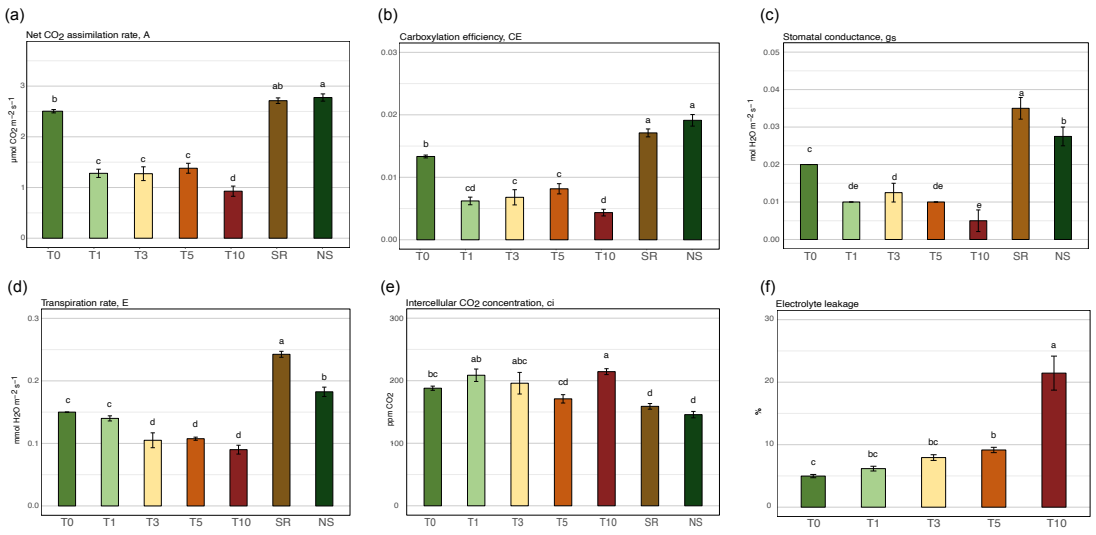


Fig. 4.2. Leaf gas-exchange parameters and electrolyte leakage. a) net CO_2 assimilation rate, A; b) Carboxylation efficiency, CE; c) stomatal conductance, g_s; d) Transpiration rate, E; e) Intercellular CO_2 concentration, Ci and f) Electrolyte leakage, EL; measurements in C, T1, T3, T5, T10, SR and NS. Different letters indicate statistically significate differences ($p < 0.05$).

After a one-month period of recovery in control conditions, stressed-recovered (SR) and non-stressed (NS) plants presented similar CO_2 values of *A* and CE Fig. 4.2 a and b, indicating that photosynthetic activity returned to normal. On the other hand, higher values of g_s and *E* Fig 4.2 c-d were observed in recovered stressed plants, which were probably related to some adaptation process or memory effect.

An age effect was also detected when comparing the initial control and NS groups, with increases in E, g_s, and CE and a decrease in C_i suggesting different physiological states that were possibly related to plant growth during the recovery period.

4.4.2 Identification and characterization of nuclear proteins across periods of exposure to heat stress

GeLC-Orbitrap/MS analysis of the nuclei-enriched fraction resulted in the identification of 3571 protein groups, of which 3328 could be reliably quantified. Given the limitations of non-model databases, very restrictive conditions were used for classifying nuclear proteins. Thus, after removal of the non-nuclear proteins based on their *in silico* annotations and YLoc and Localizer subcellular localization tools, 862 protein groups were identified as certainly being nuclear, of which 309 (q-value ≤0.05) were considered as differentially accumulated (Table S4.1; ANOVA, 5% FDR).

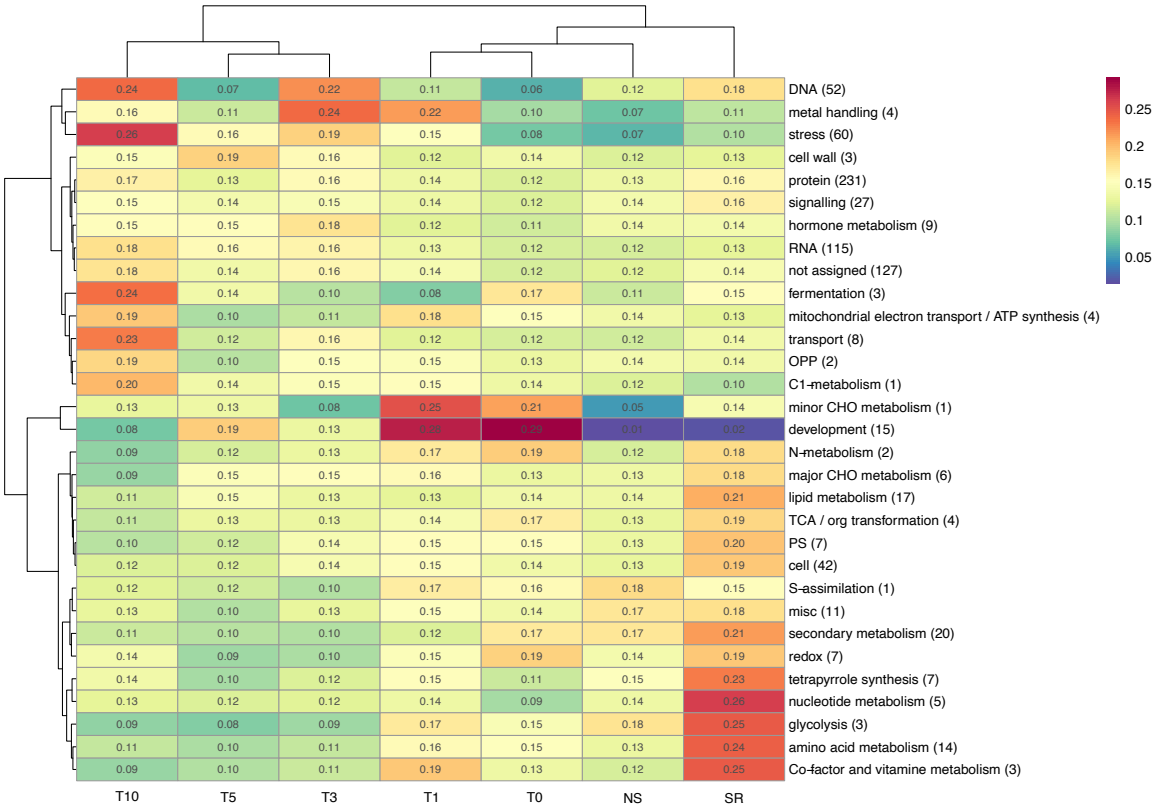


Fig. 4.3. Heatmap-Clustering analysis of Mapman categorization pathways in nuclear proteome. Numbers inside the cells indicate scaled abundance according to Mapman functional bin. Manhattan distance and Ward's aggregation method were used for hieratical clustering. Numbers in brackets indicate the proteins included in each category

MapMan functional classification of the identified nuclear proteins showed that the differentially accumulated pathways covered both primary and secondary metabolism (Fig. 4.3). The abundance of stress-related clusters (stress and metal handling) increased

under heat-stress conditions. In addition, T5 seedlings showed signs of recovery as their values for development and cell wall clusters showed increases. However, in T10 samples the development, cell wall, and protein metabolism clusters were found to be down-regulated; moreover, these samples showed considerable increases in fermentation and stress-related clusters, which highlighted cell impairment consistent with the physiological results shown in Fig 4.2a-f.

After the recovery period, SR plants exhibited differences compared to unstressed plants. Nucleotide, amino acid, cofactor and vitamin, and secondary metabolism were clearly up-regulated, as were the tetrapyrrole synthesis and glycolysis pathways. In addition, there were also found considerable changes in the development category between the T0–T1 and the SR–NS groups; however, these seemed to be linked to the age of the plants rather than being relevant for analyzing the heat-stress response. This highlighted the importance of maintaining control plants throughout the experiment.

In summary, the functional classification and heatmap clustering analysis of the nuclear proteome distinguished the different periods of heat exposure and allowed the determination of specific pathway clusters related to each exposure period.

4.4.3 A multivariate approach reveals the nuclear mechanisms involved in the heat-stress and memory responses

In order to reduce the complexity of the results, multivariate analyses including PCA and *K*-means were performed. PCA showed that the first two components accounted for 46% of the total variance (Fig. 4.4a). The treatments were separated into three mains groups: control plants (T0 and NS), heat-stressed plants (T1, T3, T5, and T10), and recovered plants (SR). The variance gathered for each component was explained by the proteins exhibiting the highest and lowest loadings for each component. Component 1 (PC1) seemed to be related to the heat-stress response (Table 4.1); thus, among the proteins in PC1 were heat-shock proteins (HSPs), proteins that have previously been related to cold stress such as LOS4 (Gong *et al.*, 2004) and PHOSPHOGLYCERATE KINASE (Bae *et al.*, 2003; Escandón *et al.*, 2017a) and elements involved in proteome and transcriptome reorganization like the ARGININE/SERINE-RICH SPLICING FACTOR (SRSF) or 40S RIBOSOMAL PROTEIN S30.

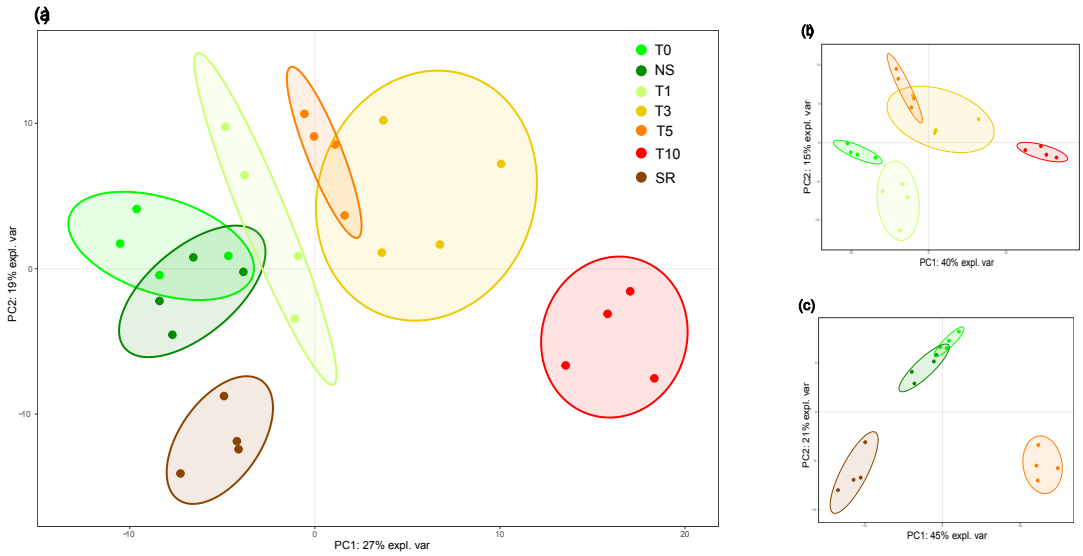


Fig. 4.4. Principal Component Analysis (PCA) of ANOVA filtered (q-value < 0.05) nuclear proteins of: a) all samples groups (C, T1, T3, T5, T10, SR and NS) together; b) Heat Treated samples groups (C, T1, T3, T5, T10); and c) Recovered samples groups (C, T5, SR and NS) analysing separately. Ellipses indicate a 0.90 confidence level.

Table 4.1 Top positive and negative loadings for PC1 and PC2 considering all treatments.

All treatments PC1 Top positive loadings	**All treatments PC2 Top positive loadings**
40S ribosomal protein S20-2	Importin-5
Temperature-induced lipocalin-1	Eukaryotic translation initiation factor 4B1
60S ribosomal protein L23a	Nucleic acid-binding, OB-fold
Heat shock protein SSA1	Putative mitochondrial ribosomal protein S1
Phosphoglycerate kinase	Eukaryotic initiation factor 4A
40S ribosomal protein S30	T-complex protein 1 subunit zeta 1
FIB4	Ras-related protein Rab-14
DnaJ protein ERDJ3A	Protein TRANSCRIPTIONALLY ACTIVE 14
RNA-binding protein 8A	RuvB-like 2
Small heat shock protein	Probable mediator of RNA polymerase II transcription subunit 37b
Sm-like protein LSM3A	Eukaryotic peptide chain release factor subunit 1
Ribosomal protein L34	Proline synthase co-transcribed bacterial homolog protein
17.5 kd heat shock family protein	Probable prolyl 4-hydroxylase 4
LOS4	DEAD-box ATP-dependent RNA helicase 52B
All treatments PC1 Top negative loadings	**All treatments PC2 Top negative loadings**
Flavin oxidoreductase	60S ribosomal protein L21
ATP-citrate (pro-S-)-lyase	NEDD8-activating enzyme E1 catalytic subunit
26S proteasome non-ATPase regulatory subunit 11	Phosphoenolpyruvate carboxylase
Coumarate 3-hydroxylase	Calcium-transporting ATPase
Protein plastid transcriptionally active 16, chloroplastic	glycosyltransferase STT3A
Delta(24)-sterol reductase	Sucrose-phosphatase 1
Regulator of transcription that contains myb domains	Sterol C7 reductase
PPIase	Cullin-1
26S proteasome non-ATPase regulatory subunit 12B	Coatomer subunit epsilon
Peroxisomal acyl-coenzyme A oxidase 1	Alpha/beta-Hydrolases superfamily protein
Arginine/serine-rich splicing factor RS34	ERBB-3 BINDING PROTEIN 1
Magnesium-protoporphyrin IX	LIM transcription factor homolog
DNA topoisomerase 6 subunit B	Signal recognition particle 54 kDa protein, chloroplastic
UDP-glucose 6-dehydrogenase 3	Ran-binding protein 1 homolog b

In the case of PC2, the biological interpretation remained unclear as it included proteins involved in a wide range of processes (signalization, endoplasmic reticulum transport, and DNA damage among others (Table 4.1). This was probably due to an excess of variability as a consequence of mixing stress-recovery (SR) and stress response (T1, T3, T5, T10) processes in the same multivariate analysis.

Separating the treatments into two major categories, namely heat-treated (HT) consisting of T0, T1, T3, T5, and T10 (Fig 4.4b), and recovery (R) consisting of T0, T5, SR and NS (Fig. 4.4c), caused the explained variance to increase to 55% and 67%, respectively. Nevertheless, the PC2 negative loadings for all the treatments PCA Table 4.1 may have constituted a cluster that is mainly composed of recovery and heat-stress memory proteins. Neddylation, a post-transcriptional modification directly linked to histone H2A, seemed to be occurring, since the NEDD8-activating enzyme E1 catalytic subunit was over-accumulated, concomitantly with histone H2A replenishment above basal levels. NEDD8 is covalently conjugated to H2A, and that neddylation of H2A antagonizes its ubiquitylation (Li *et al.*, 2014).

PCA of HT plants and their control indicated that PC1 classified the samples into those consisting of no stress (T0), first and mid-term responses (T1, T3, and T5), and a long-term response (T10) (40% of explained variance; Fig 4.4b), while PC2 explained the variation among the mid-term stress treatments. The analysis of the proteins with the top-scoring loadings in PC1 (Table 4.2, Table S4.2)) identified small HSPs (sHSPs) and also proteins related to epigenetic and alternative-splicing regulation.

Adenosylhomocysteinase (SAHH) and its product–competitive inhibitor S-adenosylmethionine (SAM) synthase, proteins involved in DNA and histone methylation, and the SM-like protein LSM3A, which is directly linked to alternative-splicing regulation, all play critical roles in the regulation of development-related gene expression (Perea-Resa et al., 2012). Furthermore, two transcription factors belonging to NF-Y family involved in histones methyl marks (Donati *et al.*, 2008) were found to be up-regulated in response to heat stress.

Table 4.2 Heat stressed treatments PC1 and PC2 top positive and negative loadings

HT PC1 Top positive loadings	HT PC2 Top positive loadings
Hap3/NF-YB transcription factor	WD repeat-containing protein 5
Eukaryotic translation initiation factor 3 subunit G	Heme oxygenase 1, chloroplastic
40S ribosomal protein S20-2	Mitochondrial-processing peptidase subunit beta
60S ribosomal protein L23	Probable mediator of RNA pol II transcription subunit 37b
Small heat shock protein	arginine/serine-rich splicing factor
Low molecular weight heat shock protein	DNA/RNA-binding protein Alba-like protein
unnanotated	Trans-2,3-enoyl-CoA reductase
Sm-like protein LSM3A	Heterogeneous nuclear ribonucleoprotein U-like 1
Alba DNA/RNA-binding protein	Aldehyde dehydrogenase family 3 member F1
17.6 kDa class II heat shock protein	DEAD-box ATP-dependent RNA helicase 53
60S ribosomal protein L23a	Protein Fes1A
Hap3/NF-YB transcription factor	Protein HEAT-STRESS-ASSOCIATED 32
Heterogeneous nuclear ribonucleoprotein 27C	Novel plant SNARE 13
22.0 kDa class IV heat shock protein	Pre-mRNA cleavage factor Im 25 kDa subunit 2
HT PC1 Top negative loadings	**HT PC2 Top negative loadings**
DNA topoisomerase 6 subunit B	40S ribosomal protein S10-3
Salt tolerance protein 1	Dehydrogenase/reductase SDR family member 4
Thiamine thiazole synthase	EF1Bgamma class glutathione S-transferase
Delta(24)-sterol reductase	Ribosomal protein
S-adenosylmethionine synthase	Enoyl-CoA hydratase/3-hydroxyacyl-CoA dehydrogenase
Beta-glucosidase 42	DNA damage-inducible protein 1
fatty acid beta-oxidation protein	Chalcone synthase
UDP-glucose 6-dehydrogenase 3	Caffeic acid O-methyltransferase
S-adenosylmethionine synthase 5	Male gametophyte defective 1
Protein plastid transcriptionally active 16	Protein disulfide-isomerase
Calcium-dependent phosphotriesterase protein	ATP synthase subunit D
Peroxisomal acyl-coenzyme A oxidase 1	Aldehyde dehydrogenase
Coumarate 3-hydroxylase	Histone H2B
Adenosylhomocysteinase	Fructose-bisphosphate aldolase

These results suggested a fundamental change in cell organization leading to a new proteome profile. In addition, fatty acid metabolism and flavonoid biosynthesis also seemed to be altered, as indicated by changes found in PC1 to the proteins of the first steps of the phenylpropanoid biosynthesis pathway (coumarate 3-hydrolase, C3H) and PC2 (caffeic acid O-methyltransferase, COMT). Interestingly, in PC2, which separated the first response (T1) and the mid-term response (T3 and T5), a cluster of spliceosome-related proteins including heterogeneous nuclear ribonucleoprotein U-like protein 1 (HNRNP U-like1) and SRSF was found to be up-regulated proportionally to the exposure to heat stress (Table 4.2, PC2 positive loadings).

PCA of the recovery treatments (Fig. 4.4c) showed the nuclear proteins that did not return to the basal level after a 5-day heat treatment even after a month in control conditions. PC1 (45% of explained variance) distinguished stress-induced proteome

Table 4.3 Top positive and negative PC1 and PC2 of recovery groups

R PC1 Top positive loadings	R PC2 Top positive loadings
RNA-binding (RRM/RBD/RNP motifs) family protein	Salt tolerance protein 1
60S ribosomal protein L15-1	Late embryogenesis abundant protein LEA7-1
Glycine-rich RNA-binding protein RZ1B	Delta(24)-sterol reductase
DNA/RNA-binding protein Alba-like protein	Casein kinase II subunit alpha
DnaJ protein ERDJ3A	WD40-like transcription factor
Reticulon-like protein B2	Glutamate decarboxylase
Heterogeneous nuclear ribonucleoprotein 27C	unnanotated
SNF2 transcription factor	DNA topoisomerase 6 subunit B
Heat shock 90/70 organizing protein	Calcium-dependent phosphotriesterase superfamily protein
Heterogeneous nuclear ribonucleoprotein 27C	Dehydrogenase/reductase SDR family member 4
RuvB-like 2	CSC1-like protein ERD4
Peroxisomal targeting signal 2 receptor	Aldehyde dehydrogenase
Heterogeneous nuclear ribonucleoprotein U-like protein 1	Male gametophyte defective 1
Translocon-associated protein (TRAP), alpha subunit	Alpha/beta-Hydrolases superfamily protein
R PC1 Top negative loadings	**R PC2 Top negative loadings**
RS9, ribosomal protein 9	peptidase subunit beta
60S ribosomal protein L21	Serine-threonine kinase receptor-associated protein
40S ribosomal protein S15	Isoeugenol synthase 1
50S ribosomal protein L10, chloroplastic	Putative mitochondrial ribosomal protein S1
Transposon protein,	LOS4
Cucumisin	Riboflavin synthase
60S ribosomal protein L38	arginine/serine-rich splicing factor
NADH dehydrogenase 13-A	HPMS5 protein
Cinnamate 4-hydroxylase	AMPSase
60S ribosomal protein L26-1	T-complex protein 1 subunit zeta
Beta-glucosidase 42	Eukaryotic translation initiation factor 2 gamma subunit
17.5 kd heat shock family protein	AtSUFE
Histone H2A	PPIase
Histone H2B	UDP-sulfoquinovose synthase, chloroplastic

changes related to adaptation, with H2A and H2B (negative loadings; Table 4.3) being markers of the acquired memory (Kumar & Wigge, 2010; Liu *et al.*, 2015), and spliceosome-related proteins such as HNRNP U-like1 and HNRNP 27C (positive loadings) being markers of heat exposure. PC2 (21% of explained variance) showed stable induced markers in the modulation of the proteome profile, confirming the importance of molecular chaperones (sHSP and PPIase), ER–chloroplast crosstalk (e.g. UDP-sulfoquinovose synthase) (Higashi *et al.*, 2015) and DNA mismatch repair processes (HPMS5 HOMOLOG) (Horii *et al.*, 1994).

4.4.4 K-means clustering analysis identify the central role of the epigenome in heat-stress response

Co-accumulation patterns in heat treated and recovery groups (Fig. 4.5) were investigated employing a K-means clustering analysis. Proteins of the heat-treated

experimental conditions were clustered in 9 groups (Fig. 4.5a). The elements involved in the first response to heat were grouped in cluster 1 and included the peak-and-run class heat-shock factor HSFA6b, while proteins related to stress signaling and responses such as CALMODULIN-LIKE PROTEIN 3 (CML3), LSM3A, HSPs, and ribosomal proteins were found in cluster 6. Clusters 7 and 9 showed a pattern of continuous increases; these clusters included two of the most well-known proteins related to acquired thermotolerance, HSP101 and HEAT STRESS-ASSOCIATED 32 (HSA32) (Wu *et al.*, 2013). In contrast, cluster 8 showed a decreasing pattern; included in this cluster were proteins such as SAM synthase and SAHH, both key elements for epigenetic reorganization.

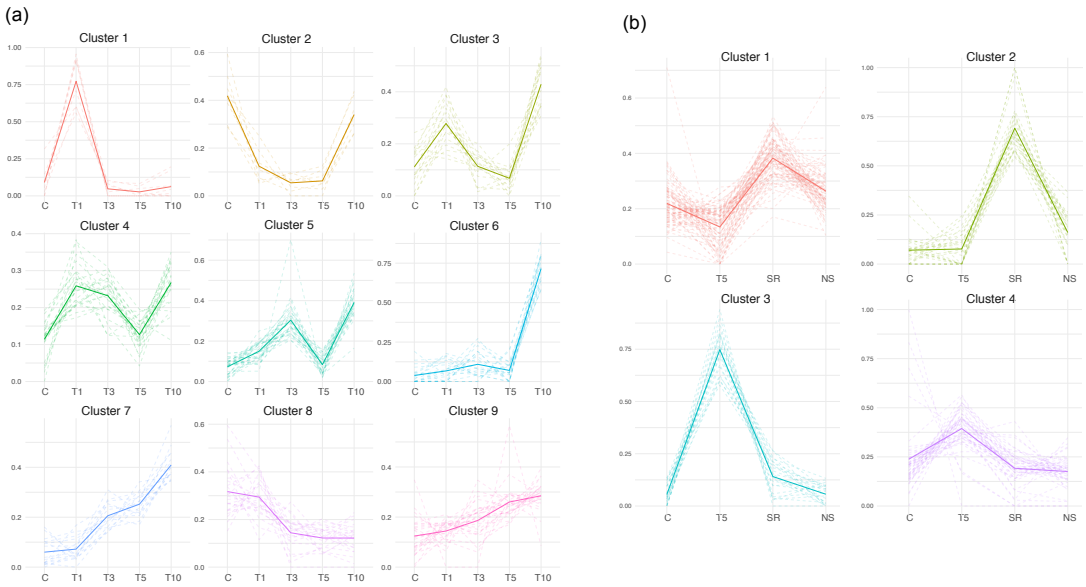


Fig. 4.5. k-means clustering of differentially accumulated (q<0.05) nuclear proteins in a) Heat Treated groups (C, T1, T3, T5, T10) and b) Recovered groups (C, T5, T5R and CR). Dashed lines show individual patterns and bold lines the mean for each cluster.

The nuclear proteomes of the recovery treatments were grouped into four different clusters (Fig. 4.5b). Clusters 1 and 2 showed a similar trend with a clear peak in SR, which indicated groups of proteins up-regulated 1month after the 5-d heat-stress treatment. While cluster 1 comprised histone H2A and some ribosomal proteins that were decreased at the stress point T5, cluster 2 showed over-accumulated proteins only in plants recovered from the stress. This cluster comprised key proteins including: SAM synthase that are strongly related to DNA, RNA, and histone methylation (Bender, 2004); calcium-binding protein CML13, a calcium sensor that is possibly involved in heat sensing and has previously been described as a cold-inducible nuclear protein in *Arabidopsis thaliana* (McCormack & Braam, 2003; Saidi *et al.*, 2011); Hap3/NF-YB transcription factors; NEDD8 E1 essential for DNA damage response modulation (Brown *et al.*, 2015), and histones H1

and H4. Cluster 4 included stress-related proteins (e.g. HSP101, FIB4, and LOS4, which were also identified in PCA HT loadings; Table 4.2) that returned to control levels when the heat exposure ended. Cluster 3 followed the same trend as the proteins related heat stress; however, in this group significant differences were detected between the levels in NS and SR plants, with a slight increase in SR protein abundance compared to the control NS (e.g. FES1A, HNRNP U-like 1, BAX inhibitor 1) (Table S4.3).

4.4.5 sPLS analysis reveals a complex network of nuclear protein interactions involved in the heat-stress response

Sparse partial least-square (sPLS) analysis produced networks that were constructed by considering transcription factors and regulators as the predictor matrix for the rest of the differentially accumulated nuclear proteins in heat-treated (HT) groups (Fig. 4.6a, Table S4.4a). This pinpointed the importance of sHSPs, ribosomal proteins, and spliceosome-related proteins during heat stress (e.g. LSM3A and the small nuclear ribonucleoprotein SmD3B). Histones (H2A, H2B, and H4) and eukaryotic translation initiation factor 3G (eIF3G) were shown to play major roles in managing the heat-stress response, being the nodes selected according to the 0.75 threshold to explain the relations among the nuclear proteins.

DnaJ ERDJ3A protein (a molecular chaperone) and glycine-rich protein were the unique nodes negatively related to net CO_2 assimilation rate and transpiration (Fig. S4.2, Table S4.4b) in an sPLS-based network where proteins predicted leaf gas-exchange parameters. Coumarate 3 hydroxylase, coumarate 4 hydroxylase, and DNA topoisomerases 6A and B were down-regulated in heat-treated plants and were positively related to the photosynthesis parameters shown in the network.

Additional biological functional analysis of protein–protein interactions was performed using STRING (Szklarczyk *et al.*, 2017) with *Arabidopsis thaliana* database. Proteins selected using the STRING database from the 20 highest positive and negative loadings of the first two components in the PC analysis of heat-treated plants Fig. 4.4b, (Table S4.2) are shown in the network in (Fig. 4.6b). Clear functional interactions were observed between the nuclear proteins identified, which frame the biological clusters in line with the mathematical correlations established using sPLS analysis.

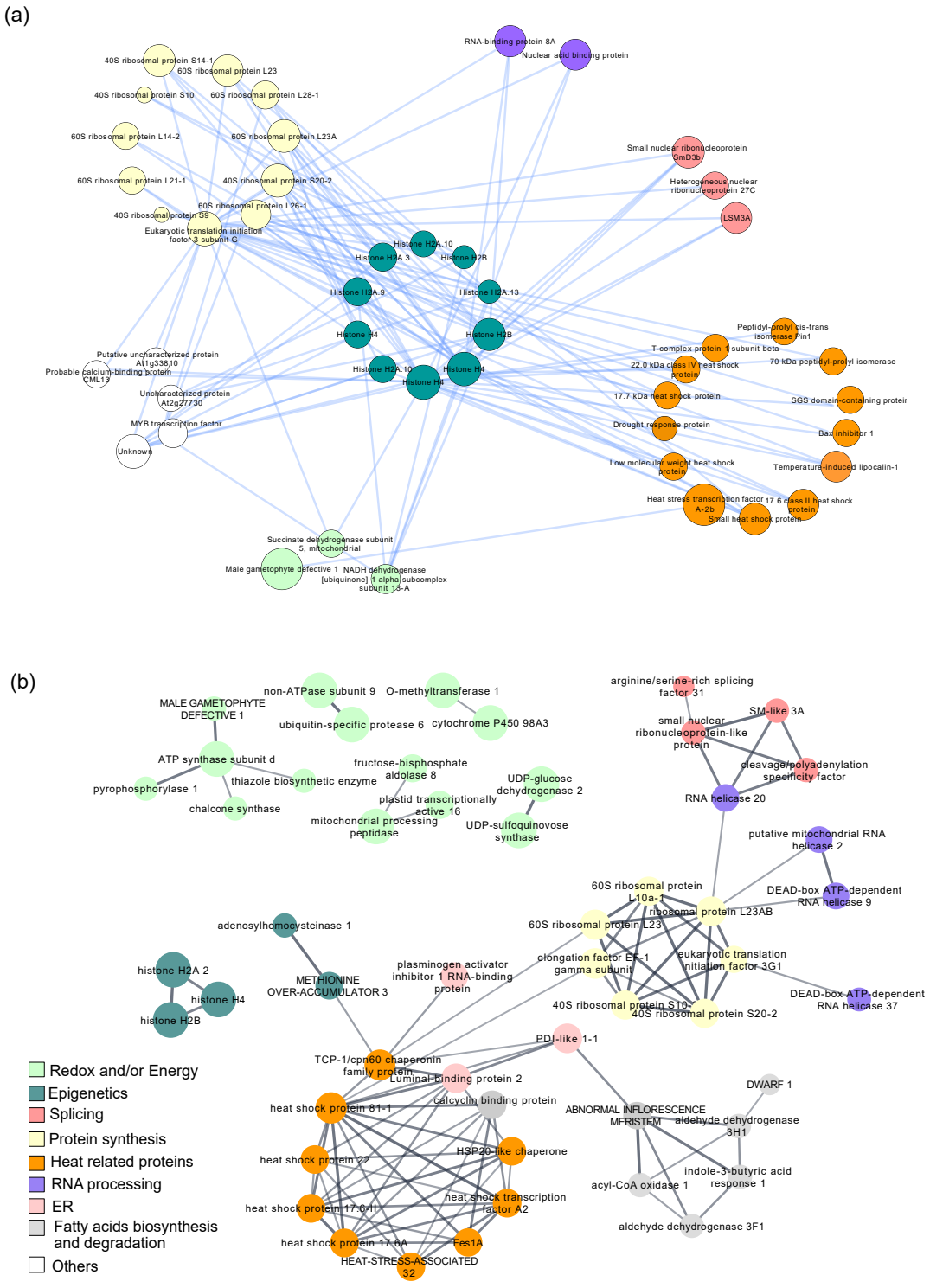


Fig. 4.6: Integrative analysis of nuclear proteins involved in heat stress and thermopriming process. a) sPLS-based network built using transcription factors and regulators identified with the TF predictor tool as the predictor matrix for changes in the rest of nuclear proteome. Correlation cut-off was 0.75, node size was calculated accordingly to radiality. b) STRING based network of 40 most relevant proteins (20 top scoring positive loadings and 20 top scoring negative loadings) in PCA components 1 and 2. Selected proteins were blasted against STRING database of the model species *Arabidopsis thaliana* and those with a minimum of homology of 60% were employed to build the network. Network edges indicate a biological correlation at least of 0.7 from experimental or curated databases resources.

The functional interaction analysis showed clusters that were devoted to HSPs: ribosomal activity, epigenetics, fatty acid metabolism, and RNA processing and splicing. This reflected the connection between spliceosome proteins and the ribosomal machinery, and between both of them and HSPs. In addition, HSPs were directly connected to epigenetic regulation (SAM synthase and SAHH), since epigenetic related changes are those intended to react in fastest way, the link among HSPs and methyl cycle enzymes may suggest the relevance of the quick regulation of HSP expression. Epigenetic proteins showed the highest values at T3 and T5, key points related to the acclimation process and adaptation, respectively (Fig. S4.3). An independent cluster related to redox, flavonoid biosynthesis, and energy processes was also found.

4.5 Discussion

4.5.1 Integrative analysis of the nuclear proteome confirms proteomic rearrangement and small HSPs as essential mechanisms to maintain plant function after initial heat stress

High temperature has a great impact on plant physiology (Escandón *et al.*, 2017), and many and complex processes are involved in heat-response signaling. Recent studies in plants have elucidated the complex transcriptional regulatory networks involved in high-temperature responses (Ohama *et al.*, 2017; Ling *et al.*, 2018). Our study was focused on studying the nuclear proteome under heat stress; the nucleus plays an essential role during genome organization, different phases of cellular development, and physiological responsiveness through regulated gene expression. Thus, identification of nuclear proteins represents an initial step towards gaining new insights into cell responses to heat stress in *P. radiata*. The nuclear proteome is highly dynamic, changing its composition in response to environmental and intracellular stimuli (Pascual *et al.*, 2016) in order to guide the subsequent remodulation of the global proteome (Narula *et al.*, 2013), and it provides useful information about the mechanisms underlying the heat-stress adaptation processes in *Pinus radiata* at transcriptional level.

In addition to HSPs, our study identified different regulation steps involved in epigenomic-driven gene regulation, several transcription factor families, and a variety of RNA-associated functions (spliceosome, proteasome, and mRNA surveillance). The results of all these changes were subsequently detectable in the over-accumulation of the translasome machinery (ribosomal proteins and eukaryotic initiation factors) needed to carry out the required cell reorganization. We also found differential responses to short-, mid-, and long-term heat exposure, as well as stable histone H2A-related heat-induced markers that were established during the recovery phase.

Photosynthetic activity was clearly impaired by heat stress in *P. radiata*, as observed from leaf gas-exchange parameters (Fig. 4.2) and heatmap clustering (Fig.4.3). These results, in accordance with Chapter 3 and those described in Buchner *et al.* (2015), showed that under heat stress photorespiration metabolism was favored, and in the extreme treatment for 10 d it led to fermentation. There were differences in the gas-exchange parameters such as transpiration rate and stomatal conductance, with values for recovered plants exceeded those of the controls. Together with the alterations that we

observed in the nuclear proteome, this provided a sign of epigenetic regulation; the biological processes were durable even in the absence of stress, indicating possible memory effects, which have already been described in other model species (Ling *et al.*, 2018).

Heatmapping revealed two groups associated with the stress treatments, namely those related to mid- (T3 and T5) and long-term (T10) exposure, and those related to short-term (T1), control, and recovery treatments, which emphasized the wide differences between stress and recovery events. Since the T10 treatment severely damaged the plants, the T5 plants were used for testing any possible primming effect (Fig. 4.1). A recent study (Ling *et al.*, 2018) refers to thermopriming as an event of non-lethal exposure to heat stress that allow plants to survive subsequent and otherwise lethal conditions.

Using multivariate analyses including PCA and *K*-means, together with integrative approaches and a comprehensive analysis of the generated sPLS networks with biological correlations and the STRING tool, we were able to uncover several links between key proteins in relation to both the stress and memory responses. As expected, we found high abundances of HSPs, ER molecular chaperones (DnaJ), and ribosomal machinery, as previously reported by Escandón *et al.* (2017), Several relevant players and pathways were identified that have also previously been described in heat and cold responses, such as LOS4 (Gong *et al.*, 2002, 2004), PHOSPHOGLYCERATE KINASE (Escandón *et al.*, 2017), and polyphenol biosynthesis through the alteration of COMT. These results not only validate the nuclear integrative approach used but also represent new findings, as discussed below.

4.5.2 Stress-response and memory effects in conjunction with differential and opposite epigenetic patterns involving hypo- and hypermethylation

DNA methylation is a well-known epigenetic marker of transcriptional gene silencing, but it also occurs in the establishment of heterochromatin, transposon control, and genomic imprinting (Galindo-Gonzalez *et al.*, 2018). Two of the key enzymes regulating the methylation cycle, SAM synthase and SAHH, were identified as central elements in our integrative analysis of the nuclear proteome, with both decreasing proportionally with the stress exposure time. This seemed to indicate that heat stress drives hypomethylation, since SAM is a methyl-group donor and an essential

methyltransferase co-factor, and SAHH is a methyl-cycle enzyme that is required for SAM regeneration and transcriptional gene silencing-mediated methylation. In addition, COMT, a precursor of fatty acid and flavonoid biosynthesis through phenylpropanoid biosynthesis, was found to be over-accumulated uniquely in the short-term response (T1). COMT is also a SAM-dependent methyltransferase. Hence, contributing to the depletion of SAM suggesting a possible DNA hypomethylation in the first heat stress shock. As showed in Escandón *et al.*, (2017), the overproduction of fatty acids is required for the short-term response to heat stress in *Pinus radiata*, and in the mid- and long term, enzymes implicated with flavonoid synthesis are essential for successful adaptation. This metabolites favor protein biosynthesis in the cell cytoplasm, which is of great importance at this point when the translasome machinery is clearly up-regulated and has a high demand for HSPs.

Interestingly, SAM synthase and SAHH were found to among the major differences between heat-stressed and stress-recovered plants (T5 versus SR); both were depleted by the stress, but higher levels were detected in the stress-recovered plants (Fig. 4.5b). Moreover, stress-induced demethylation has been found to relax chromatin structure, thereby allowing enhanced transcription and proteasomal rearrangement (Shilatifard, 2006; Santos *et al.*, 2011), which has been linked to heat-tolerant genotypes in other plant species (Gao *et al.*, 2014).

These results support the hypothesis that environmental factors (including temperature and other stresses) are more important in changes in DNA methylation than in those that occur spontaneously or that have a genetic basis (Dubin *et al.*, 2015). Over recent years it has been proposed that, as sessile organisms that must persist in the same location for a long time, plants may be particularly likely to exploit DNA methylation for rapid adaptation to changing environments (Valledor *et al.*, 2012; Kawakatsu *et al.*, 2017).

4.5.3 Nucleosome structure and spliceosome functioning are altered by heat stress in relation to the thermopriming process

Priming and the establishment of stress memory can help plants to survive a variety of abiotic stress conditions, including heat (Tanou *et al.*, 2012; Filippou *et al.*, 2012). The maintenance of acquired thermotolerance is crucial for successful priming and tolerance to subsequent exposure to otherwise lethal temperatures (Ling *et al.,* 2018). The rebound effect pattern showed through clusters 1 and 2 in protein k-means analysis of recovered

groups (Fig. 4.6b) seemed to reveal a priming process, since the highest levels of some key proteins were increased in recovered plants, such as histones H2A and H2B, and NEDD8 E1. The decrease of histone abundance during the first round of stress could be related to the formation of hexasomes, as described by Shaytan *et al*. (2015). Hexasomes are stable nucleosomes lacking H2A-H2B dimers, usually caused by RNA pol II transcriptional activity and DNA repair processes. Accordingly, RNA polymerase II kinetics are accelerated under heat exposure (Jonkers and Lis 2015, Chen et al. 2016) and some DNA mismatch repair proteins are over-accumulated such as HPMS5 HOMOLOG and NEDD8 E1. This hypothesis fits with the over-accumulation of some probable mediators of RNA polymerase II transcription proteins which were identified as the proteins in cluster 9 (Fig. 4.6a, Table S4.3) that were increased by the heat stress.

The loss of H2A–H2B dimers in the nucleosome was counteracted in the recovery step (1 month in control conditions), providing a sign of an epigenetic memory-related process. Interestingly, both histones and DNA methylation enzymes were found to be altered in order to enhance DNA accessibility by chromatin relaxation under heat stress. While chromatin relaxation marked the nucleosome status during the stress, a more compact nucleosome structure was found to occur after recovery, when the plant metabolism was again focused on development and growth pathways. This provide an insight into the importance of nucleosome occupancy and DNA accessibility in the non-primed and primed responses.

Spliceosome-related proteins were also significantly altered by heat stress, as shown in the PCA and, the sPLS and STRING based networks (Figs. 4 and 6). Spliceosomal activity was impaired under stress (Table 4.2) and then the abundance of spliceosomal proteins was increased beyond basal levels in recovered plants (Figure 5b, Table S4.4); as a consequence, mRNA surveillance and proteasome activities such as ubiquitination were found to have a pivotal role in the STRING and sPLS networks, as a link between the spliceosome and translasome machinery. This findings align with previous works, showing how thermopriming leads to alternative or impaired splicing events (Ling *et al.*, 2018).

This study shows for first time the dynamics of the nuclear proteome related to the heat-stress response and the recovery processes. The depth and complexity of this study in relation to the number of proteins identified and analyses performed allowed a detailed depiction of these routes, and revealed several crucial families of proteins that are involved in different key regulation steps such as proteasome reorganization, RNA-associated

functions, epigenomic-driven gene regulation, and specific transcription factors previously unconnected to heat stress. In addition, histone H2A, alternative splicing, and methyl-cycle enzymes seem to be directly linked to the induction of thermopriming, and the active remodeling of the transcriptome and proteome that triggers the crucial processes involved in high-temperature responses and adaptation. This newly discovered priming-induced epigenetic memory may represent a general feature of heat-stress responses in conifers, and it may facilitate the development of novel approaches to improving survival of pine trees under extreme heat stress in the current context of climate change.

4.6 References

Asensi-Fabado MA, Amtmann A, Perrella G. **2017**. Plant responses to abiotic stress: The chromatin context of transcriptional regulation. *Biochimica et Biophysica Acta - Gene Regulatory Mechanisms* **1860**: 106–122.

Bae MS, Cho EJ, Choi EY, Park OK. **2003**. Analysis of the Arabidopsis nuclear proteome and its response to cold stress. *Plant Journal* **36**: 652–663.

Bäurle I. **2016**. Plant Heat Adaptation: priming in response to heat stress. *F1000Research* **5**: 694.

Bender J. **2004**. DNA METHYLATION AND EPIGENETICS. *Annual Review of Plant Biology* **55**: 41–68.

Briesemeister S, Rahnenführer J, Kohlbacher O. **2010a**. Going from where to why-interpretable prediction of protein subcellular localization. *Bioinformatics* **26**: 1232–1238.

Briesemeister S, Rahnenführer J, Kohlbacher O. **2010b**. YLoc-an interpretable web server for predicting subcellular localization. *Nucleic Acids Research* **38**: 497–502.

Brown JS, Jackson SP, Jackson SP. **2015**. Ubiquitylation , neddylation and the DNA damage response. *Open biology* **150018**.

Buchner O, Stoll M, Karadar M, Kranner I, Neuner G. **2015**. Application of heat stress in situ demonstrates a protective role of irradiation on photosynthetic performance in alpine plants. *Plant, Cell and Environment* **38**: 812–826.

Carbó M, Iturra C, Correia B, Colina FJ, Meijón M, Álvarez JM, Cañal MJ, Hasbún R, Pinto G, Valledor L. **2019**. Epigenetics in forest trees: Keep calm and carry on. In: Álvarez-Venegas R, De la Peña C, Casas-Mollano JA, eds. Epigenetics in Plants of Agronomic Importance: Fundamentals and Applications: Transcriptional Regulation and Chromatin Remodelling in Plants: Second Edition. Springer, 381–3.

Casimiro-Soriguer CS, Muñoz-Mérida A, Pérez-Pulido AJ. **2017**. Sma3s: A universal tool for easy functional annotation of proteomes and transcriptomes. *Proteomics* **17**: 1700071.

Chen Y, Müller F, Rieu I, Winter P. **2016**. Epigenetic events in plant male germ cell heat stress responses. *Plant Reproduction* **29**: 21–29.

Donati G, Gatta R, Dolfini D, Fossati A, Ceribelli M, Mantovani R. **2008**. An NF-Y-dependent switch of positive and negative histone methyl marks on CCAAT promoters. *PLoS ONE* **3**.

Dubin MJ, Zhang P, Meng D, Remigereau M-S, Osborne EJ, Paolo Casale F, Drewe P, Kahles A, Jean G, Vilhjálmsson B, *et al.* 2015. DNA methylation in Arabidopsis has a genetic basis and shows evidence of local adaptation. *eLife* **4**: e05255–e05255.

Escandón M, Cañal MJ, Pascual J, Pinto G, Correia B, Amaral J, Meijón M. **2015**. Integrated physiological and hormonal profile of heat-induced thermotolerance in Pinus radiata. *Tree Physiology* **36**: 63–77.

Escandón M, Meijón M, Valledor L, Pascual J, Pinto G, Cañal MJ. **2018**. Metabolome Integrated Analysis of High-Temperature Response in Pinus radiata. *Frontiers in Plant Science* **9**: 1–15.

Escandón M, Valledor L, Pascual J, Pinto G, Cañal MJ, Meijón M. **2017a**. System-wide analysis of short-term response to high temperature in Pinus radiata. *Journal of Experimental Botany* **68**: 3629–3641.

Escandón M, Valledor L, Pascual J, Pinto G, Cañal MJ, Meijón M. **2017b**. System-wide analysis of short-term response to high temperature in Pinus radiata. *Journal of experimental Botany* **68**: 3629–3641.

Filippou P, Tanou G, Molassiotis A, Fotopoulos V. **2012**. Plant Acclimation to Environmental Stress Using Priming Agents. In: Plant Acclimation to Environmental Stress. 1–28.

Galindo-Gonzalez L, Sarmiento F, Quimbaya MA. **2018**. Shaping Plant Adaptability , Genome Structure and Gene Expression through Transposable Element Epigenetic Control : Focus on Methylation. *Agronomy* **8**.

Gao G, Li J, Li H, Li F, Xu K, Yan G, Chen B, Qiao J, Wu X. **2014**. Comparison of the heat stress induced variations in DNA methylation between heat-tolerant and heat-sensitive rapeseed seedlings. *Breeding science* **64**: 125–133.

Gong Z, Dong C, Lee H, Zhu J, Xiong L, Gong D. **2004**. A DEAD Box RNA Helicase Is Essential for mRNA Export and Important for Development and Stress Responses in Arabidopsis. *Plant Cell* **17**: 1–12.

Gong Z, Lee H, Xiong L, Jagendorf A, Stevenson B, Zhu J-K. **2002**. RNA helicase-like protein as an early regulator of transcription factors for plant chilling and freezing tolerance. *Proceedings of the National Academy of Sciences of the United States of America* **99**: 11507–11512.

Gutzat R, Mittelsten Scheid O. **2012**. Epigenetic responses to stress: Triple defense? *Current Opinion in Plant Biology* **15**: 568–573.

Higashi Y, Okazaki Y, Myouga F, Shinozaki K, Saito K. **2015**. Landscape of the lipidome and transcriptome under heat stress in Arabidopsis thaliana. *Scientific Reports* **5**: 1–7.

Horii A, Han HJ, Sasaki S, Shimada M, Nakamura Y. **1994**. Cloning, characterization and chromosomal assignment of the human genes homologous to yeast PMS1, a member of mismatch repair genes. *Biochemical and Biophysical Research Communications* **204**: 1257–1264.

Iwasaki M, Paszkowski J. **2014**. Epigenetic memory in plants. *The EMBO Journal* **33**: 1987–1998.

Jonkers I, Lis JT. **2015**. Getting up to speed with transcription elongation by RNA polymerase II. *Nature Reviews Molecular Cell Biology* **16**: 167.

Kawakatsu T, Nery JR, Castanon R, Ecker JR. **2017**. Dynamic DNA methylation reconfiguration during seed development and germination. *Genome Biology* **18**: 171.

Kumar SV, Wigge PA. **2010**. H2A.Z-Containing Nucleosomes Mediate the Thermosensory Response in Arabidopsis. *Cell* **140**: 136–147.

Lämke J, Brzezinka K, Altmann S, Bäurle I. **2016**. A hit-and-run heat shock factor governs sustained histone methylation and transcriptional stress memory. *The EMBO Journal* **35**: 162–175.

Li T, Guan J, Huang Z, Hu X, Zheng X. **2014**. RNF168-mediated H2A neddylation antagonizes ubiquitylation of H2A and regulates DNA damage repair. *Journal of Cell Science* **127**: 2238–2248.

Ling Y, Serrano N, Gao G, Atia M, Mokhtar M, Woo YH, Bazin J, Veluchamy A, Benhamed M, Crespi M, *et al.* 2018. Thermopriming triggers splicing memory in Arabidopsis. *Journal of Experimental Botany* **69**: 2659–2675.

Liu J, Feng L, Li J, He Z. **2015**. Genetic and epigenetic control of plant heat responses. *Frontiers in Plant Science* **06**: 1–21.

Lohse M, Nagel A, Herter T, May P, Schroda M, Zrenner R, Tohge T, Fernie AR, Stitt M, Usadel B. **2014**. Mercator: A fast and simple web server for genome scale functional annotation of plant sequence data. *Plant, Cell and Environment* **37**: 1250–1258.

Martinez-medina A, Flors V, Heil M, Mauch-mani B, Pieterse CMJ, Pozo MJ, Ton J, Dam NM Van. **2016**. Recognizing Plant Defense Priming. *Trends in Plant Science* **21**: 2–5.

McCormack E, Braam J. **2003**. Calmodulins and related potential calcium sensors of Arabidopsis. *New Phytologist* **159**: 585–598.

de Mendiburu F. **2017**. Statistical Procedures for Agricultural Research.

Narula K, Datta A, Chakraborty N, Chakraborty S. **2013**. Comparative analyses of nuclear proteome : extending its function. *Frontiers in Plant Science* **4**: 1–14.

Ohama N, Sato H, Shinozaki K, Yamaguchi-Shinozaki K. **2017**. Transcriptional Regulatory Network of Plant Heat Stress Response. *Trends in Plant Science* **22**: 53–65.

Pascual J, Alegre S, Nagler M, Escandón M, Annacondia ML, Weckwerth W, Valledor L, Cañal MJ. **2016**. The variations in the nuclear proteome reveal new transcription factors and mechanisms involved in UV stress response in Pinus radiata.

Journal of Proteomics **143**: 390–400.

Pastor V, Luna E, Mauch-Mani B, Ton J, Flors V. **2013**. Primed plants do not forget. *Environmental and Experimental Botany* **94**: 46–56.

Perea-Resa C, Hernandez-Verdeja T, Lopez-Cobollo R, Castellano M d. M, Salinas J. **2012**. LSM Proteins Provide Accurate Splicing and Decay of Selected Transcripts to Ensure Normal Arabidopsis Development. *The Plant Cell* **24**: 4930–4947.

R Core Team. R Foundation for Statistical Computing. **2017**. R: A language and environment for Statistical computing.

Rohart F, Gautier B, Singh A, Lê Cao K-A. **2017**. mixOmics: An R package for 'omics feature selection and multiple data integration. *PLOS Computational Biology* **13**: e1005752.

Romero-Rodríguez MC, Pascual J, Valledor L, Jorrín-Novo J. **2014**. Improving the quality of protein identification in non-model species. Characterization of Quercus ilex seed and Pinus radiata needle proteomes by using SEQUEST and custom databases. *Journal of Proteomics* **105**: 85–91.

Saidi Y, Finka A, Goloubinoff P. **2011**. Heat perception and signalling in plants: A tortuous path to thermotolerance. *New Phytologist* **190**: 556–565.

Santos AP, Ferreira L, Maroco J, Oliveira MM. **2011**. Abiotic Stress and Induced DNA Hypomethylation Cause Interphase Chromatin Structural Changes in Rice rDNA Loci. *Cytogenetic and Genome Research* **132**: 297–303.

Schindelin J, Arganda-Carreras I, Frise E, Kaynig V, Longair M, Pietzsch T, Preibisch S, Rueden C, Saalfeld S, Schmid B, *et al.* 2012. Fiji: An open-source platform for biological-image analysis. *Nature Methods* **9**: 676–682.

Shaytan AK, Landsman D, Panchenko AR. **2015**. Nucleosome adaptability conferred by sequence and structural variations in histone H2A-H2B dimers. *Current Opinion in Structural Biology* **32**: 48–57.

Shilatifard A. **2006**. Chromatin Modifications by Methylation and Ubiquitination: Implications in the Regulation of Gene Expression. *Annual Review of Biochemistry* **75**: 243–269.

Smith PK, Krohn RI, Hermanson GT, Mallia AK, Gartner FH, Provenzano MD, Fujimoto EK, Goeke NM, Olson BJ, Klenk DC. **1985**. Measurement of protein using bicinchoninic acid. *Analytical Biochemistry* **150**: 76–85.

Sperschneider J, Catanzariti AM, Deboer K, Petre B, Gardiner DM, Singh KB, Dodds PN, Taylor JM. **2017**. LOCALIZER: Subcellular localization prediction of both plant and effector proteins in the plant cell. *Scientific Reports* **7**: 1–14.

Stekhoven DJ, Bühlmann P. **2012**. MissForest — non-parametric missing value imputation for mixed-type data. *Bioinformatics* **28**: 112–118.

Szklarczyk D, Morris JH, Cook H, Kuhn M, Wyder S, Simonovic M, Santos A, Doncheva NT, Roth A, Bork P, *et al.* 2017. The STRING database in 2017: quality-controlled protein-protein association networks, made broadly accessible. *Nucleic acids research* **45**: D362–D368.

Talbert PB, Henikoff S. **2014**. Environmental responses mediated by histone variants. *Trends in Cell Biology* **24**: 642–650.

Tanou G, Fotopoulos V, Molassiotis A. **2012**. Priming against environmental challenges and proteomics in plants: Update and agricultural perspectives . *Frontiers in Plant Science* **3**: 216.

Valledor L, Cañal MJ, Pascual J, Rodríguez R, Meijón M. **2012**. Early induced protein 1 (PrELIP1) and other photosynthetic, stress and epigenetic regulation genes are involved in Pinus radiata D. don UV-B radiation response. *Physiologia Plantarum* **146**: 308–320.

Valledor L, Hasbún R, Meijón M, Rodríguez JL, Santamaría E, Viejo M, Berdasco M, Feito I, Fraga MF, Cañal MJ, *et al.* 2007. Involvement of DNA methylation in tree development and micropropagation. *Plant Cell, Tissue and Organ Culture* **91**: 75–86.

Valledor L, Weckwerth W. **2014**. An improved Detergent-Compatible Gel-Fractionation LC-LTQ-Orbitrap-MS workflow for plant and microbial proteomics. In: Methods in Molecular Biology. United States, 347–358.

Vandesompele J, De Preter K, Pattyn F, Poppe B, Van Roy N, De Paepe A, Speleman F. **2002**. Accurate normalization of real-time quantitative RT-PCR data by geometric averaging of multiple internal control genes. *Genome biology* **3**.

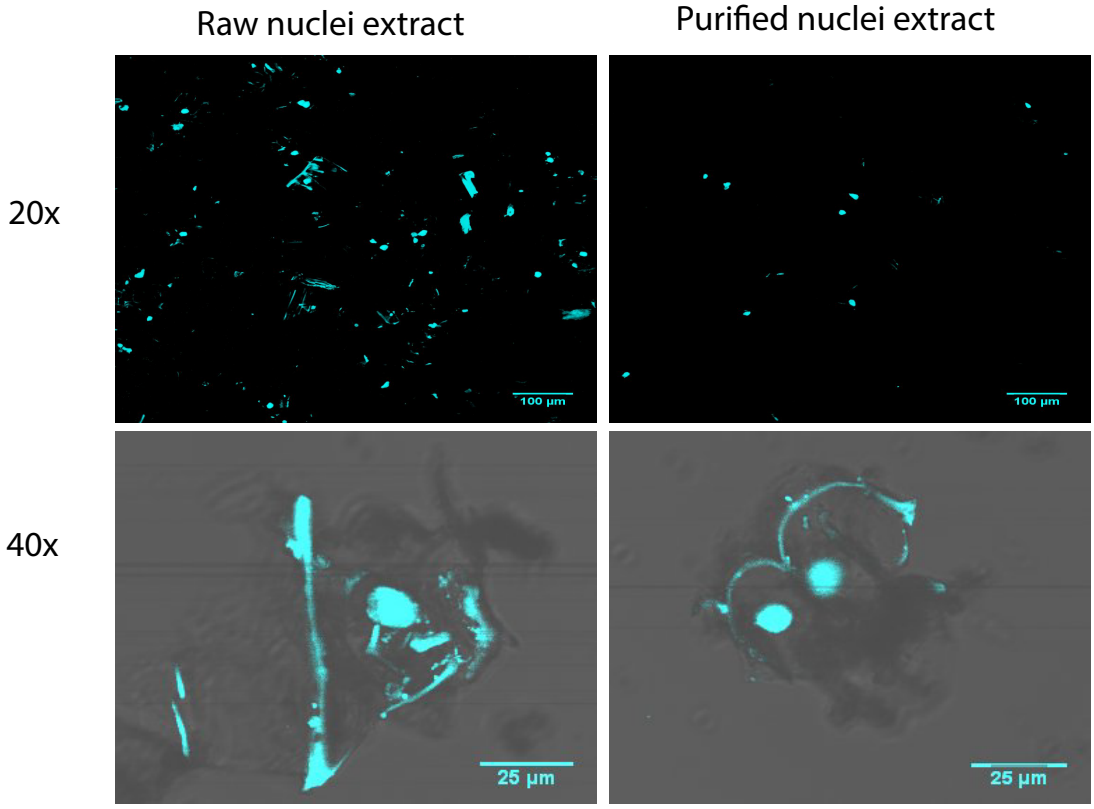
Wu T -y., Juan Y -t., Hsu Y -h., Wu S -h., Liao H -t., Fung RWM, Charng Y -y. 2013. Interplay between Heat Shock Proteins HSP101 and HSA32 Prolongs Heat Acclimation Memory Posttranscriptionally in Arabidopsis. *Plant Physiology* **161**: 2075–2084.

Yin X, Komatsu S. **2016**. Plant nuclear proteomics for unraveling physiological function. *New Biotechnology* **33**: 644–654.

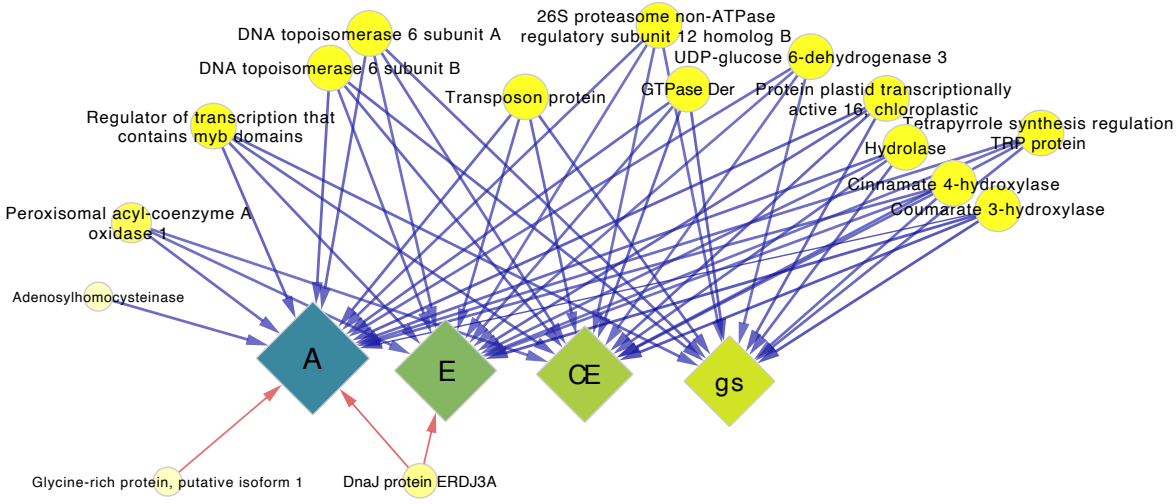
Zhao PX, Dai X, Sinharoy S, Udvardi M, Zhao PX. **2013**. PlantTFcat : An online plant transcription factor and transcriptional regulator categorization and analysis tool PlantTFcat : an online plant transcription factor and transcriptional regulator categorization and analysis tool. *BMC Bioinformatics* **14**.

4.7 Supplemental information

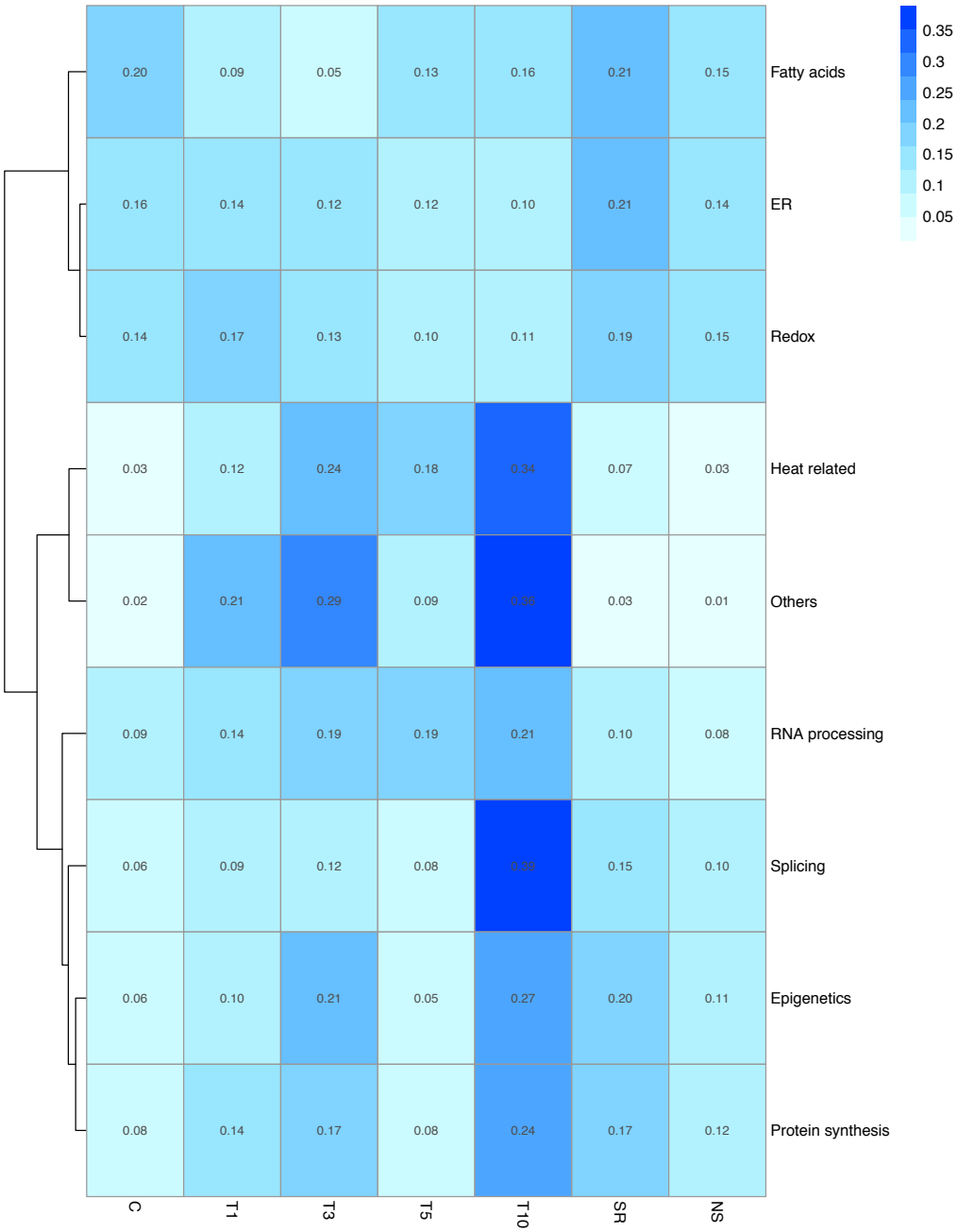
4.7.1 Supplemental Figures



Supplemental Figure S4.1. Nuclear raw and purified factions images stained with DAPI obtained by confocal microscopy.



Supplemental Figure S4.2: sPLS based network of IRGA vs. nuclear proteins



Supplemental Figure S4.3. Heatmap-Clustering analysis of most relevant categories of nuclear proteins species depicted in protein-protein interaction networks

4.7.2 Supplemental tables

Supplemental table legends

Supplemental Table S4.1. Nuclear proteins identification, quantification, annotation and univariate analysis.

Supplemental Table S4.2. PCA loadings considering the nuclear proteins in a) all the treatments (C, T1, T3, T5, T10, SR and NS); b) Heat-treated groups (C, T1, T3, T5 and T10) and c) Recovered groups (C, T5, SR and NS).

Supplemental Table S4.3. Kmeans clustering of anova filtered nuclear proteins (qvalue ≤ 0.05) of a) Heat-treated groups (C, T1, T3, T5 and T10) and b) recovered groups(C, T5, SR, NS).

Supplemental Table S4.4. Spare Partial Least Square regression of a) transcription regulators network and b) IRGA and nuclear proteins network.

CHAPTER V. Integration of nuclear and chloroplast proteomes to decipher signaling mechanisms under heat stress in *Pinus radiata*[1].

5.1 Abstract

The recovery and maintenance of plant homeostasis under stressful environments are complex processes involving organelle crosstalk for a coordinated cellular response. Here, we revealed through nuclear and chloroplast subcellular proteomics, biochemical cell profiles and targeted transcriptomics how chloroplasts and nuclei developed their responses under increased temperatures in a long-lived species, such as *Pinus radiata*. Parallel to photosynthetic impairment and reactive oxygen species production in the chloroplast, a DNA damage response was triggered in the nucleus followed by an altered chromatin conformation. In addition, in the nuclei, we found several proteins, such as HEMERA or WHIRLY, which change their locations from the chloroplasts to the nuclei carrying the stress message. Additionally, our data showed a deep rearrangement of RNA metabolism in both organelles, revealing this process as a potential regulator of the acclimation mechanisms. Altogether, our study highlights the synchronization among the different stages required for thermotolerance acquisition in *P. radiata*, pointing out the role of chromatin conformation and posttranscriptional gene regulation in overcoming heat stress and assuring plant survival for the following years.

[1] **Lamelas L, Valledor L, López-Hidalgo C, Cañal MJ, Meijón M. 2022**. Nucleus and chloroplast: A necessary understanding to overcome heat stress in Pinus radiata. *Plant Cell and Environment* **45**: 446–458.

5.2 Introduction

Human-induced climate change represents a fundamental challenge for vegetation dynamics, being a concern for crop productivity and the sustainability of some ecosystems. Moreover, climate models predict average temperatures to rise and heat waves to be increasingly frequent (Lesk *et al.*, 2016; O'Neill *et al.*, 2017). Fortunately, as sessile organisms, plants have evolved sophisticated molecular mechanisms to perceive and cope with environmental stresses like high temperatures and even "learn" from them (Bäurle, 2016; Ling *et al.*, 2018). This molecular-based memory represents an essential strategy for plants, especially for those species with long life cycles, such as forest trees; "with age comes wisdom". Despite the importance of high temperatures as one of the key stressors of forest ecosystems, the cellular mechanisms leading to its perception, signaling and molecular memory acquisition are still poorly understood.

Stress acclimation mechanisms rely on the precise coordination between nuclei and endosymbiotic organelles (Pfannschmidt *et al.*, 2020). While the nucleus is the main regulatory hub of the cell; the chloroplast is the plant-cell defining organelle, which houses photosynthesis and the biosynthesis of secondary metabolite precursors (Dobrogojski *et al.*, 2020). Its genome codes around 100 proteins (Daniell *et al.*, 2016). Nevertheless, thousands of genes now encoded in the nuclear genome were transferred from the chloroplast genome during evolution (Dobrogojski *et al.*, 2020). The proteins coded by these genes are still required to be imported into the chloroplast for its proper development in a process called anterograde regulation (Unal *et al.*, 2020).

On the other hand, the chloroplast can regulate nuclear gene expression via organelle-to-nucleus retrograde signaling (Jung & Chory, 2010). Although the coordinated expression of chloroplast and nuclear genes regulated by retrograde signaling is indispensable for plant growth and development, how the organelle-to-nucleus communication takes place is largely unknown (Zhao *et al.*, 2020). Recent studies point out that chloroplasts act as thermosensors or thermal alarms (Dickinson *et al.*, 2018; Zhao *et al.*, 2020). While photosynthesis is damaged and photorespiration increases (Hu *et al.*, 2020), ROS and other signals including carotenoid derivatives such as tocopherols, isoprenoid precursors, phospho-nucleotides and heme are released into the cytoplasm (Zhao *et al.*, 2019, 2020). These signals travel to the nucleus where the main cell response can be triggered by, among others, alterations in nuclear gene expression. This

coordination is essential to attempt to reach the stress-adapted cell homeostasis required for overcoming the stress.

To get a deeper understanding of these signals and how these processes are synchronized in this chapter they are integrated two untargeted subcellular proteomics studies (nuclear and chloroplast, independently) in a time-course experiment with high temperature stressed *P. radiata* plants.

5.3 Material and Methods

5.3.1 Data acquisition

The spectra obtained in the previous Chapters (3 and 4) were used for the integrative analysis presented in this Chapter. In brief, nuclear and chloroplast proteomes in basal conditions and under a 5-day heat wave with three sampling points at the end of the six hour 45ºC stress treatment (Fig S5.1) have been re-analyzed and integrated as described below, due to the new availability of the assembly of the *Pinus radiata de novo* transcriptome under heat stress, which was used for an enhanced protein identification.

5.3.2 Proteomics re-analysis

Protein identification was performed with Proteome Discoverer v2.2 (ThermoFisher). A combined database was compiled with three protein databases and used for protein identification, including the *Pinus taeda* genome v.1.01 (https://bioinformatics.psb.ugent.be/plaza/versions/gymno-plaza/), UniProt/SwissProt Viridiplantae and a novel *P. radiata* transcriptome under heat stress (Escandón *et al.*, 2022). The mass spectrometry proteomics data including RAW, msf and pepXML files have been deposited to the ProteomeXchange Consortium via the PRIDE (Perez-Riverol *et al.*, 2019) partner repository with the dataset identifier PXD029114.

Identified protein sequences were blasted using the following *in silico* localization tools: BUSCA (Savojardo *et al.*, 2018), Localizer (Sperschneider *et al.*, 2017), YLoc (Briesemeister *et al.*, 2010), TargetP (v2.0) (Almagro Armenteros *et al.*, 2019), NucPred (Brameier *et al.*, 2007), NLStradamus (Nguyen Ba *et al.*, 2009), NLSdb (Nair *et al.*, 2003) and Plant TFDB (v4.0) (Jin *et al.*, 2017). Then, proteins were annotated with sma3s (Casimiro-Soriguer *et al.*, 2017) and Mercator MapMan (Lohse *et al.*, 2014) tools.

Finally, we addressed contamination issues by dropping for downstream analyses those proteins with less than two matches for their cellular organelle considering *in silico* localization tools or with no positive subcellular location in the annotation according to Mercator Mapman or sma3s.

5.3.3 Statistical analyses

All statistical analyses were performed in R (v 4.0.2) (R Core Team, 2020); proteomics datasets were analyzed using pRocessomics R package, (available at htttps://github.com/Valledor/pRocessomics) and described in Chapter 2, to perform data pre-processing, univariate (Venn) and multivariate analyses (PCA, kmeans, sPLS). Self-organizing maps (SOM) were built using kohonen R package (Wehrens & Buydens, 2007); and T-distributed stochastic neighbor embedding (t-SNE) was calculated with Rtsne package (van der Maaten & Hinton, 2008).

In brief, each proteomics dataset was pre-processed independently, keeping those proteins that were present in at least 15% of the samples or in all the replicas that constituted a treatment; missing values were imputed using Random Forest algorithm. After data pre-processing, univariate analyses were performed, and then both datasets were z-scaled for multivariate analysis.

Correlation-based networks were inferred using sPLS multivariate analysis, independently for each cellular compartment, by splitting each dataset in nuclear- or chloroplast-encoded proteins and gathering the correlations between both groups of proteins, using a cut-off value of 0.7, networks were depicted using Cytoscape (v3.7.2) tool (Shannon et al., 2003), following the recommendations of Escandón et al., (2020).

5.4 Results

5.4.1 Proteomics rearrangement as stress indicator

A 5-day heatwave was mimicked, as described in Chapters 3 and 4, in two time-series experiment to elucidate the coordinated response in nuclei and chloroplasts, mainly through their proteomes. The shared experimental design between the previous chapters, enables to track the crosstalk between these two organelles from both “points of view

To characterize the molecular mechanisms driving this response, a systems biology analysis combining two subcellular bottom-up proteomics assays was performed. Spectra were identified employing a new *P. radiata* database built after heat-stress RNA-Seq reads (Escandón *et al.*, 2022) allowing a deep characterization of chloroplast proteome (belonging to the not-primmed progeny, E, described in Chapter 3,) with 1182 quantifiable proteins. This database was also employed for reanalyzing available nuclei spectra from Chapter 4 enhancing previous protein identification in nucleus proteome, obtaining 1451 quantifiable proteins (Table S5.1). Qualitative proteome distributions across the time points are shown in the Venn diagram (Fig. 5.1a), showing subtle but consistent differences across both organelles.

The biological meaning of proteome quantitative distribution was further studied with SOM. The SOMs of both organelles clustered the biological replicates according to sampling points in the same cells (Fig. 5.1b). The nuclear proteome was remodeled after the first stress exposure, with an increased DNA damage response and multi-process regulation MapMan categories. Following the bibliography and previous Chapters, as the stress time increased, the nuclear proteome dynamics turned to increased chromatin organization (Chapter 4), protein translocation (Krause *et al.*, 2012; Li *et al.*, 2017) and vesicle trafficking (Wang *et al.*, 2020) categories. Finally, after 5 days under stress, more biological processes became relevant, especially those related to RNA and proteins. Meanwhile, in the chloroplast, SOM clusters corresponding to Control (T0) and T1 were in non-adjacent cells, indicating a drastic initial response, mainly described by a photosynthesis-related proteins depletion in T1. The multi-process regulation category was firstly triggered by the stress along with coenzyme and lipid metabolism, which was maintained in T3 when secondary metabolism and RNA clusters became more abundant until T5, where protein transport and trafficking classifications were the most relevant.

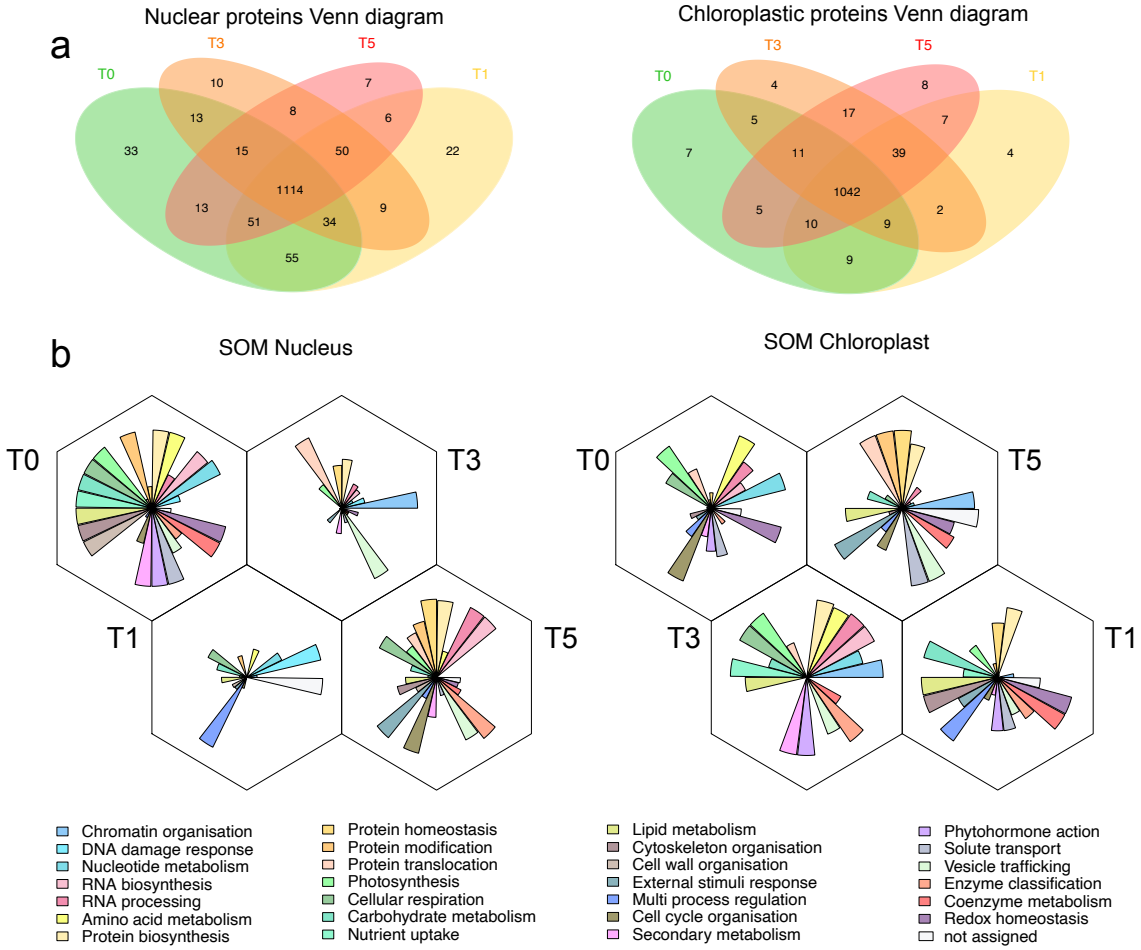


Figure 5.1. Proteomics exploratory analysis (a) Venn diagram representing the overlap of nuclear and chloroplast proteomes across the experiment. (c) Self-organizing map (SOM) of nuclear and chloroplast proteomes. Pie charts inside the cells indicate the relevance of protein functional categories according to Mercator MapMan. All procedures were done with four biologically independent.

5.4.2 Chloroplast proteome response stopped evolving at midterm heat exposure, while nuclear proteome kept changing

After exploring the main mechanisms leading the heat stress response in both organelles, Multivariate analyses were employed with a focus on identifying those concrete protein candidates that were the most representative of each stress stage. To this end, sample clustering unsupervised analysis we performed over the proteome datasets, including PCA and t-SNE.

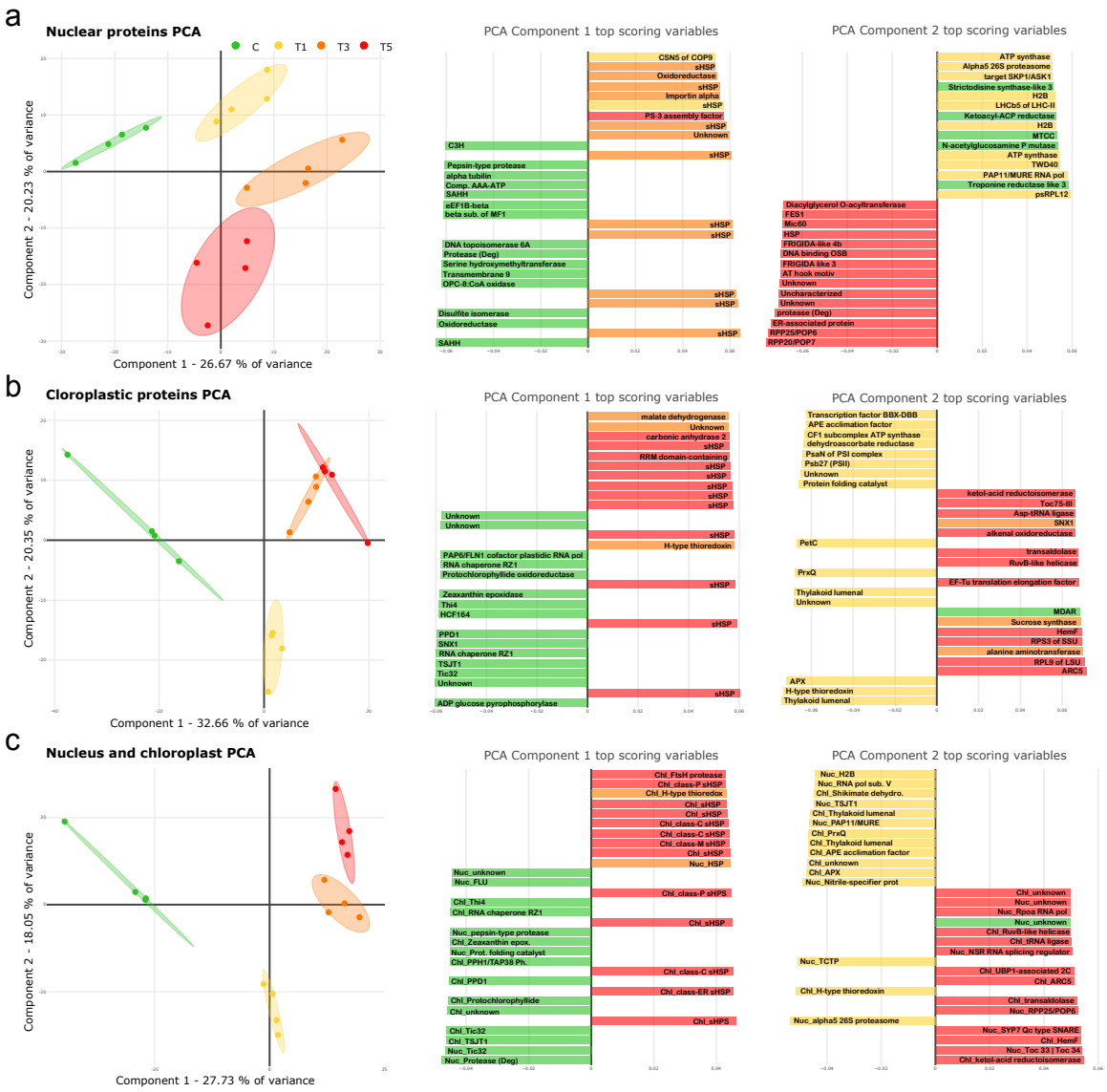


Fig. 5.2. PCA Score plot and PC1 and PC2 top-ranked proteins. (a) Nuclear proteome, (b) chloroplast proteome and (c) their combination. (a–c) Top 30 scoring loadings (15 highest and 15 lowest) of PC1 and PC2 are shown by row for each PCA, bar colours indicate the experimental condition in which each top-scoring protein is more accumulated. Ellipses show a 75% confidence interval. Different colours indicate different experimental conditions ($n = 4$ biologically independent replicates).

PCA allowed the determination of the main sources of variation among our sampling points in both proteomic datasets and their combination. In each case (Fig. 5.2, Table S5.2), the sum of Principal Components 1 and 2 (PC1, PC2) explained more than 45% of the variance. In addition, in all analyses, PC1 separated control from stressed plants, while PC2 highlighted different processes depending on the organelle. Nuclear proteome PC2 gathered the variance related to stress exposure (Fig. 5.2a), whereas chloroplast and combined proteomes separated the first stress shock (Fig. 5.2b and c) from the other two sampling points. In addition, chloroplast proteome PCA (Fig. 5.2b) was

unable to differentiate between T3 and T5 samples. This indicated that, in terms of proteome variation, chloroplasts reached in T3 a stationary state, which was maintained in T5. To further check sample distribution, t-SNE analysis (Fig. 5.3) was performed, known to be insensitive to data collinearity and overfitting. t-SNE corroborated chloroplast PCA clustering showing no differences between T3 and T5 and revealed similarities (by cluster overlap) between T0 and T5 in the nuclear data set.

Once PCA sample distribution was validated by t-SNE, top-scoring loadings of PC1 and 2. The proteins showing top positive loadings in PC1 (overaccumulated along with the stress) showed, as expected, the relevance of heat shock proteins (HSP) as a common heat stress response in both organelles. However, a closer look into the proteins, which were more abundant under control conditions, revealed relevant organelle-specific processes. These included S-adenosylhomocysteine hydrolase (SAHH), related to the methyl cycle in the nucleus (Figure 2a). As well as redox, RNA binding and photosynthesis-related proteins in the chloroplast (Figure 2b) and in the combined datasets (Figure 2c), such as PPD1, PROTOCHLOROPHYLLIDE OXIDASE, ZEAXANTHIN EPOXIDASE, FTSH protease 2 and HCF164 proteins, all of them linked to D1 chloroplast protein synthesis and assembly (Schult *et al.*, 2007) and to the PSII repair cycle (Kato & Sakamoto, 2018), known to be sensitive to stress.

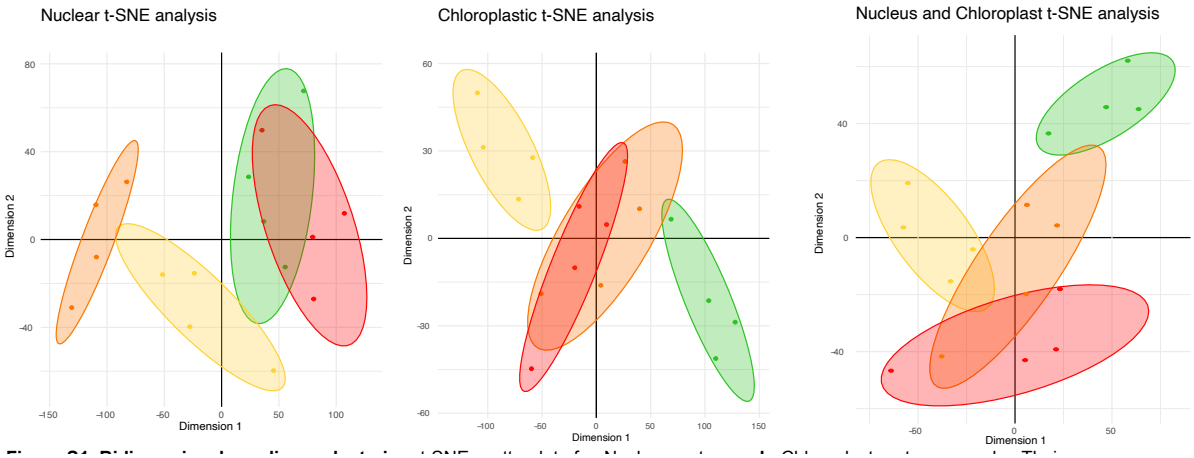


Fig. 5.3. Bidimensional non-linear clustering. t-SNE scatterplot of a, Nuclear proteome, b, Chloroplast proteome, and c, Theircombination. (a-c) Ellipses show a 75% confidence interval. Different colours indicate different experimental conditions (n = 4 biologicallyindependent replicates).

Regarding PC2, in the chloroplast (Fig.5.2b) and combined datasets (Fig. 5.2c), samples were clustered distinguishing first versus subsequent stress exposure days. Acclimation of photosynthesis to the environment (APE) acclimation factor and other proteins related to photosynthesis, such as PSB27, previously linked to PSII-independent adaptation to light stress (Hou *et al.*, 2015), showed high negative loadings and an overaccumulation in T1 samples. Contrastingly, carbon metabolism proteins (such as sucrose synthase or transaldolase), EF-TU translation elongation factor, related to heat tolerance (Ristic *et al.*, 2007), ARC5 essential for chloroplast division and biogenesis (Gao *et al.*, 2003) and RUVB-LIKE HELICASE, also known as heat-responsive and related to thermotolerance in rice (Saifi *et al.*, 2018), presented their highest positive values and peaked in longer exposure times (T3 and T5).

The combined analysis of both datasets besides revealed a cluster of proteins with high scoring values related to RNA metabolism, such as NSR RNA splicing regulator, RPOA RNA polymerase or RNA polymerase V. The latter is a multisubunit plant-specific nuclear RNA polymerase required for the normal function and biogenesis of small interfering RNA (siRNA) and is involved in the regulation of gene expression by siRNA-directed DNA methylation (Xie & Yu, 2015).

5.4.3 Dual-located proteins as key players in the coordination of stress response mechanisms.

An increasing number of proteins are found dually localized in the plastids and the nucleus (Krupinska *et al.*, 2020). Many nuclear transcription factors were shown to be controlled by signals generated in the organelles. In addition to the metabolites involved in retrograde signaling (Zhao *et al.*, 2020), there is accumulating evidence suggesting a role for proteins in plastid-to-nucleus communication. Indeed, several proteins exhibiting a dual localization in the plastids and the nucleus are promising candidates for direct signal transduction involving regulatory protein storage in the plastids (Krause *et al.*, 2012). In this work, 340 proteins were detected in both organelles (Table S5.3).

An overview of the changes in these double agents across the stress in both organelles was performed using the k-means algorithm, obtaining 25 different clusters according to their abundance profile (Fig. 5.4a). Several clusters showed a similar trend in both organelles. An increasing accumulation pattern during stress conditions was shown in Clusters 2 and 22. Low levels in T1 and high levels in T3–T5 for Clusters 1 and 19 and

a decreasing accumulation in Clusters 18 and 23, where we found HEMERA protein (also known as pTAC12 and PAP5). This protein exhibits different functions depending on its cellular location; in the nucleus, it is related to phytochrome signaling (Chen *et al.*, 2010), and in the chloroplasts, it is required for gene expression (Steiner *et al.*, 2011).

The most recurring trend was an opposite profile (Clusters 5, 7, 8, 11, 12 and 17). These organelle-specific accumulated clusters included proteins related to carbohydrate metabolism, cell cycle organization, chromatin organization, photosynthesis and cellular respiration (Fig. 5.4b), all of these essential processes in plant metabolism. Within these groups, DNA/RNA binding protein WHIRLY1 was noteworthy (Cluster 8, Table S4), since this protein has been proposed to move from the chloroplast to the nucleus in response to environmental cues such as high light intensity, in which a WHIRLY1-dependent increase of nuclear microRNAs was reported (Świda-Barteczka *et al.*, 2018). Now, through this work, WHIRLY1 was found to follow the same pattern in high-temperature response as well.

Strictly increasing and decreasing clusters were mainly composed of redox homeostasis and photosynthesis proteins, respectively (Fig. 5.4b); while the profile of Clusters 1 and 19, with low levels in T1 and high levels in T3–T5, which could be connected to acclimation processes and possibly to signaling or memory acquisition. These clusters essentially collected proteins from RNA metabolism and protein biosynthesis (Fig. 5.4b), as RNA splicing regulator, RH3 plastid RNA basal splicing factor and ribosomal proteins (Table S5.3).

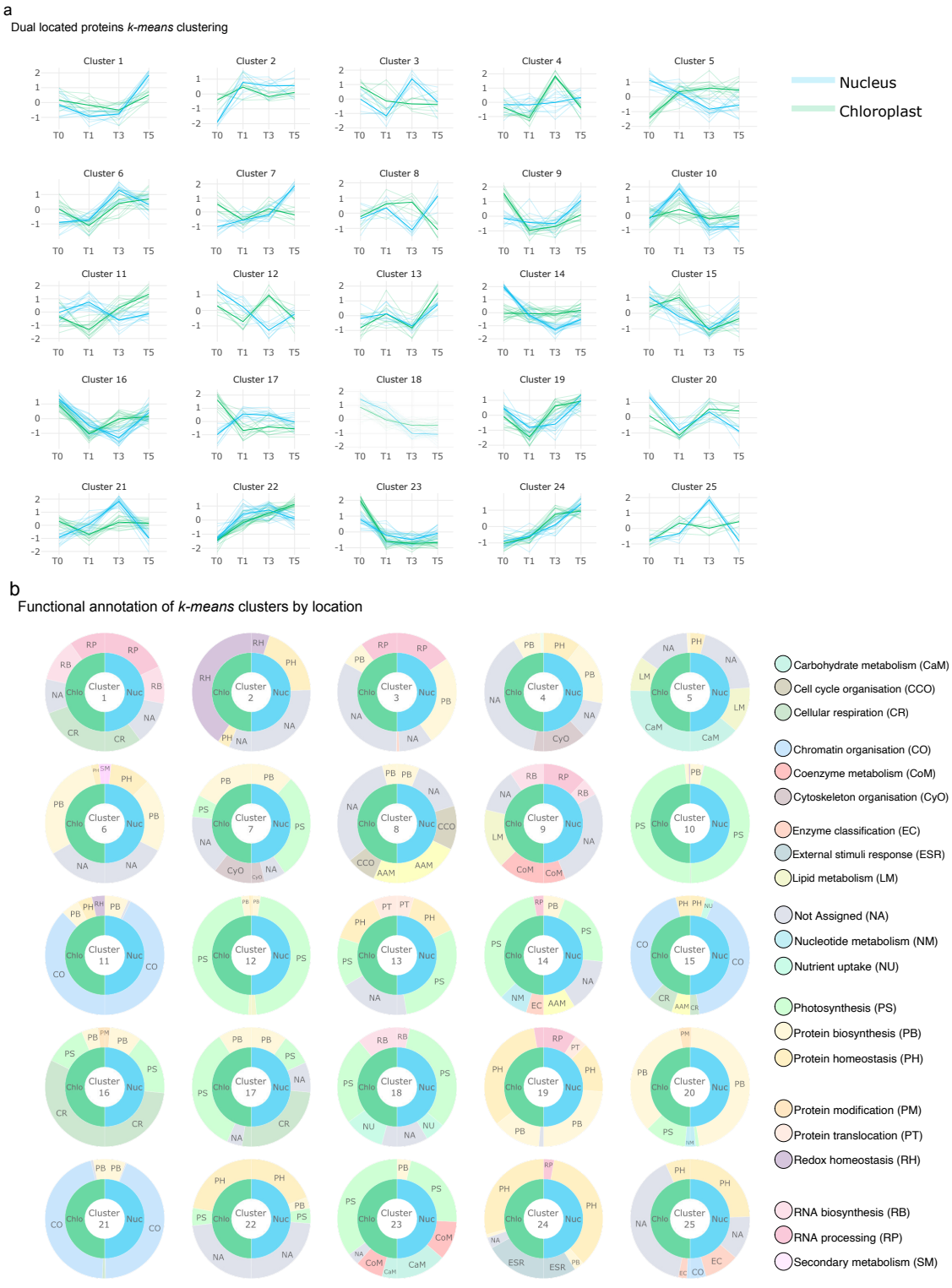


Fig. 5.4. Pattern-clustering of dual localized proteins. (a) *k*-means analysis of the abundance of dual located proteins across the stress. Three hundred forty proteins showing dual localization were identified, scaled in each data set, and *k*-means clustered yielding 25 groups. Colours indicate the cellular location; nuclear in light blue and chloroplast in green. Continuous lines indicate mean values for each protein at each experimental condition ($n = 4$ biologically independent replicates) and bold lines indicate mean values for each cluster at each experimental condition. (b) Pie charts of most abundant function annotation classifications for each cluster and cellular location. Different colours indicate different protein functional annotation classification according to Mercator Mapman.

5.4.4 Nuclear-chloroplast crosstalk, a two-way road

Both nuclei and chloroplasts contain proteins whose DNA is encoded in the genomes of other organelles. These ‘non-native’ proteins are known to play a role in organellar communication along with the dually located proteins already identified. To further explore this crosstalk, we performed an integrative analysis by evaluating the correlations among the proteins quantified. To do so, each data set was independently divided into ‘native’ (encoded in this organelle genome) and ‘nonnative’ proteins, and the correlations among them were evaluated using sPLS algorithm and depicted as two protein–protein networks (Fig. 5.5, Table S5.4). The relations between nuclear-encoded proteins found in the chloroplast proteome might give insights related to anterograde communication, and the relations between chloroplast-encoded proteins found in the cell nucleus might be related to retrograde communication.

Anterograde communication traces (Fig. 5.5a) were found in the chloroplasts, where the ubiquitous HSPs and RNA chaperones were transcribed in the cell nuclei and sent to the chloroplasts, probably as the desirable consequence of the gene expression alterations at cell nuclei as a defense shield against stress. These proteins are known to be essential to maintain chloroplast functionality, along with redox enzymes (APX, thioredoxins, SOD) also highlighted in chloroplast PCA loadings (Fig. 5.2b, PC2), which have a role in counteracting the electron transfer flux disruption and redox imbalance triggered by the hyperthermal stress (Fang *et al.*, 2019). Furthermore, this network provided potential candidates and indicated that RNA metabolism rearrangement is also required in the chloroplasts and was at least partially driven from the cell nucleus, with the opposite abundance changes of organelle RRM domain-containing protein 6 (Fu *et al.*, 2007) and RNA binding protein CP29B linked to photosynthesis and RNA metabolism (Hackett *et al.*, 2017).

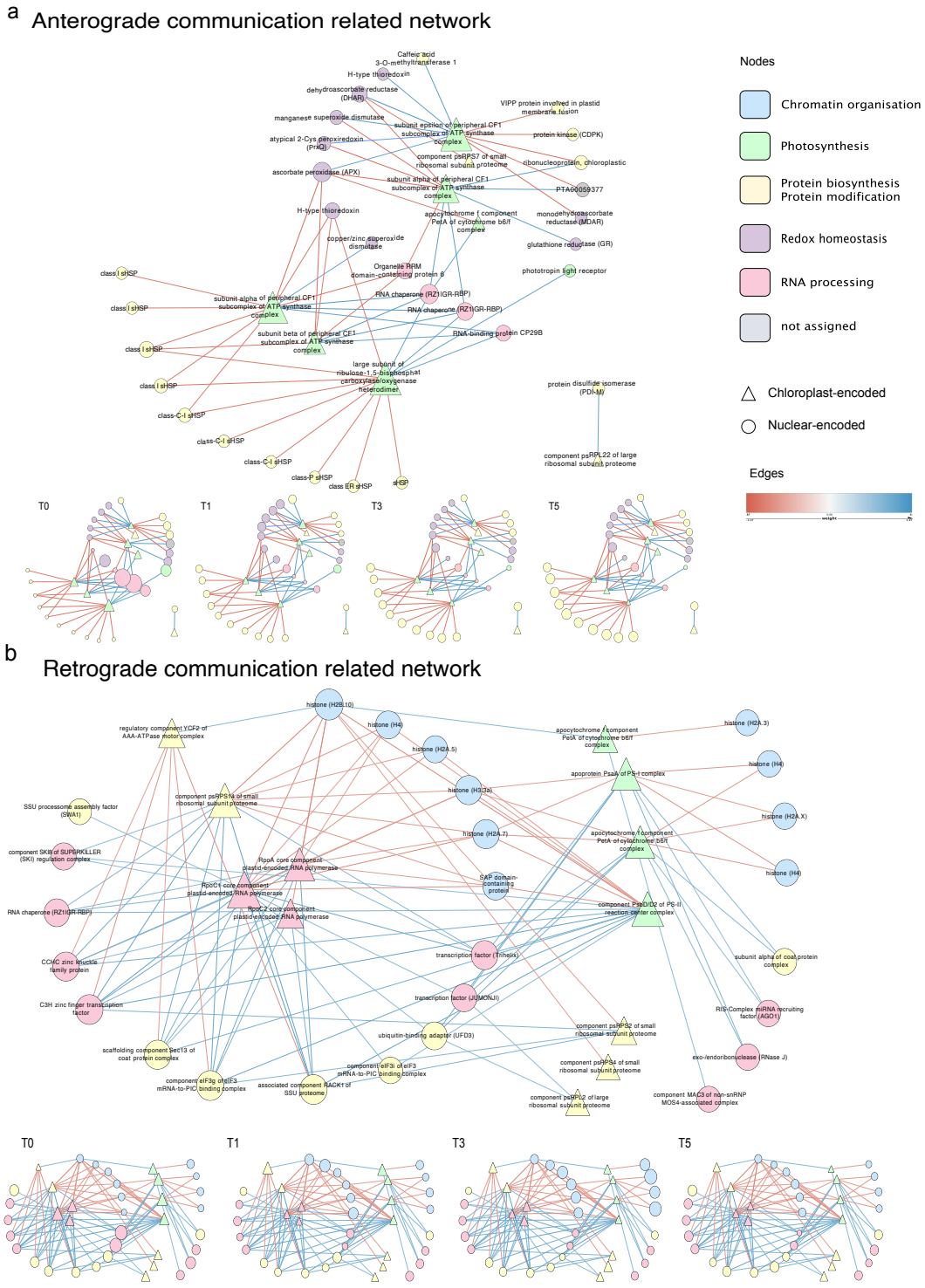


Figure XX. Reconstruction of nuclei-chloroplasts communication by sPLS networks. (a) Anterogr ade communication-related network, was built using chloroplast targeted proteins, which were divided according to their gene localization (nuclear or chloroplast genome), network links depict the correlations found among nuclear-encoded proteins and chloroplast-encoded proteins. (b) Retrograde communication-related network, was built following the same schema, being this time nucleus targeted proteins divided according to their gene localization. Triangles indicate chloroplast-encoded proteins and circles nuclear-encoded proteins. Node colours indicate Mercator MapMan classification according to the legend. Edge colour indicated inversely proportional (negative) relations in red and proportional (positive) relations in blue. Edges below a 0.75 cutoff were removed. T0, T1, T3 and T5 subnetworks in the bottom represent the proteins' abundance distribution across the control and heat-stress treatments. All models were built using four biologically independent replicates

On the other hand, retrograde communication (Fig. 5.5b) seems to be linked to chloroplast-encoded RNA polymerases, whose expression were altered concomitantly with histones and RNA splicing factor MAC3, transcription factors (C3H Zinc Finger, Triple Helix and jumonji) and protein metabolism (processome components and ribosomal subunits such as SWA1(Shi *et al.*, 2005) and RACK1 (Guo *et al.*, 2011)).

Taken together, both networks covered potential players of the nucleus-chloroplast communication, involving chromatin organization, photosynthesis, redox, RNA and protein metabolisms.

5.5 Discussion

An important unanswered question in stress plant biology is how signaling coordination between organelles takes place and what experimental approach can be used to address these molecular mechanisms. Here, we show that subcellular untargeted proteomics is a powerful strategy to provide insights into how plants orchestrate physiological responses.

5.5.1 The acclimation stage seems to be reached firstly in the chloroplast and then in the nucleus

The employed experimental design allowed to monitor organelle responses after one, three and five consecutive days of heat stress. Chloroplasts exhibited a more drastic proteomic variation after the first day of stress (Fig. 5.1b), while their proteomic profiles varied slightly from T3 to T5 (Figs. 5.2b and 5.3), suggesting that a new homeostatic state was reached at mid-term stress exposure. Taken globally, these results showed various pieces of evidence indicating an acclimation process at mid-term stress exposure. The APE acclimation factor, triggered in T1, may be key to survive to the following days, given that *Arabidopsis thaliana ape* mutants showed a defective acclimation response (Walters *et al.*, 2003). Additionally, the chloroplast protein synthesis elongation factor EF-TU raised in T5 was found to be implicated in heat tolerance in maize (Ristic *et al.*, 2007) revealing a possible long-term thermotolerance acquisition process placed in the chloroplast. Also, the over-accumulation of ARC5 in T5, which is essential for chloroplast division (Gao *et al.*, 2003) gave insights into chloroplast heat-stress-overcoming mechanisms.

Meanwhile, the nucleus triggered after the first heat shock a DNA damage response (Fig. 5.1b). The third stress exposure caused an accumulation of chromatin organizing proteins as histone isoforms (Fig. 5.1b and, 5.5b), and after five days transcription and translation metabolism-related proteins were over-accumulated (Figs. 5.2a and, 5.2c), coupled to a whole-cell FAA increase, suggesting a deep proteomic rearrangement towards thermotolerance acquisition. These processes were coupled to an upregulation of PTGS complexes through microRNA and AGO1 over-accumulation in the fifth consecutive exposure day (Figs. 5.5b), when plants finally seemed to be acclimated to the applied stress. This co-occurrence might imply an up-regulation of microRNAs metabolism as a result of an acclimated state; Post Transcriptional Gene Silencing (PTGS)

could function as a protection mechanism against an exaggerated response, which is no longer needed in the new homeostatic state.

5.5.2 Heat and light stresses seem to reprogram nucleus-chloroplast crosstalk through pigments in a similar way

In previous studies, it was observed that heat stress provoked a moderate and maintained decrease in PSII maximum efficiency. Interestingly, the decreased abundance of photosynthetic pigments, such as chlorophyll and carotenoids, matched the decrease Fv/Fm only in short and midterm heat exposures (Chapter 3). The mismatch found after five days between these measurements lead us to the chloroplast proteome, where it was found that D1 protein was depleted along with the enzymes related to the PSII repair cycle. Hence, PSII impairment likely explains the Fv/Fm drop. Additionally, since photosynthetic machinery is in part controlled by nuclear genome gene expression (Unal *et al.*, 2020), these may indicate a reprogramming of basal communications between the two organelles during the stress.

The chlorophyll depletion in short- and midterm stress can be due to different phenomena mainly including chlorophyll being used as a source for other metabolites, chlorophyll biosynthesis pathway downregulation or a combination of both processes. This latter option was likely to occur in different phases of each day stress exposure. As chlorophyll has been proposed to be a source for the biosynthesis of tocopherol through the phytol recycling pathway (Muñoz & Munné-Bosch, 2019), which is a signaling metabolite (Jung & Chory, 2010; Fang *et al.*, 2019; Serrano *et al.*, 2019), the pigment decrease could be linked to a heat-warning signal. In addition, tocopherol has an essential role in avoiding the propagation of lipid peroxidation (Muñoz & Munné-Bosch, 2019), so the MDA content maintenance along with the stress further supports a phytol recycling pathway-based-signaling towards stress. In addition, carotenoids serve as precursors of a wide variety of signaling molecules (Moreno *et al.*, 2021); their depletion after six hours of stress on the first and third days seems to indicate that they acted as a source for other metabolites.

On the other hand, in the proteome, we found some clues about the pigment changes, such as the under accumulation of protochlorophyllide oxidoreductase (Fig. 5.2b, PC1), an enzyme implicated in chlorophyll biosynthesis. In any case, free chlorophyll usually generates phototoxic catabolites; thus, the decrease in this pigment content may

prevent a further toxic effect. Simultaneously, zeaxanthin epoxidase was also depleted under heat stress (Fig. 5.2b, PC1), which may imply a xanthophyll shift towards an increased zeaxanthin amount, one well-known photoprotective pigment under high-light conditions (Jahns *et al.*, 2009).

Interestingly, several candidates involved in chloroplast heat reprogramming, such as the APE acclimation factor, PSB27, HEMERA and WHIRLY were previously linked to light adaptation (Walters *et al.*, 2003; Chen *et al.*, 2010; Krause *et al.*, 2012; Foyer *et al.*, 2014; Hou *et al.*, 2015; Krupinska *et al.*, 2020). However, these candidates have not been related so far to heat stress or heat tolerance. The relevance of these proteins in both stresses could be explained by the assumption that in natural conditions light and heat stress are often simultaneous, and that these candidates are related to the signaling of photosynthetic damage through ROS, which is a shared signaling pathway for several abiotic stressors. Among them, WHIRLY and HEMERA showed dual localization, probably acting as messengers between the two organelles.

5.5.3 RNA metabolisms seem to be involved in organelle communication

As depicted in the anterograde network (Fig. 5.5a), plastid-encoded RNA polymerases seem to be actively involved in stress response. Unexpectedly, we found plastid-encoded RNA polymerases along with other ten chloroplast-encoded proteins in the cell nucleus (Fig. 5.5b, Table S5.4). This finding should be taken cautiously, as to date there is no precedent record of chloroplast-encoded proteins targeted to the cell nucleus. Despite this fact, the presence of nuclear localization signals in these protein sequences may suggest them as potential dual-targeted proteins, which may enter the cell nucleus after chloroplast membrane disruption or fluidization, which is known to happen in heat stress conditions (Hu *et al.*, 2020). Another plausible option is the direct protein export from the chloroplasts to the nucleus, which has been proven to be possible through stromules, which are tubular channels that allow metabolite and protein exchange (Köhler *et al.*, 1997; Hanson & Sattarzadeh, 2013). Moreover, ROS production in the chloroplasts triggers plastid movement towards the nucleus and stimulates stromules formation (Kwok & Hanson, 2004; Brunkard *et al.*, 2015; Hu *et al.*, 2020; Mullineaux *et al.*, 2020), which may be intended to communicate the stress signal avoiding cytoplasmic diffusion. However, further studies are required to test the extent and functionality of this physical communication.

In the nucleus, RNA metabolism also exhibited a characteristic profile in the last stress exposure (Figs 5.1b and, Fig.5.2). The main role of PTGS and microRNA previously linked to organelle crosstalk under stress conditions (Fang *et al.*, 2019; Zhao *et al.*, 2019, 2020), as well as our results related to dual located proteins (Figs. 5.4a and, b), support the idea of a central role of RNA metabolism in heat stress response. Besides, apocytochrome proteins were found to be correlated to RNA metabolism through microRNA biogenesis, exoribonuclease and AGO1 proteins, which are shown to be the main players in stress signaling by directing gene repression (Fang *et al.*, 2019) from the central regulatory hub of the cell.

To sum up, in this Chapter, untargeted -omics approaches were combined to decipher the biochemical signals relative to stress acclimation, revealing how finely-tuned these sequential mechanisms are, covering ROS detoxifying, chromatin remodeling and their downstream consequences as RNA and protein metabolism reprogramming, which lead to stable changes that allow plants to survive to heat stress. These results increase our understanding of how plants adapt to challenging environments in long-lived species, such as *Pinus radiata*, which need to acclimate and survive for years to endure as species.

5.6 References

Almagro Armenteros JJ, Salvatore M, Emanuelsson O, Winther O, von Heijne G, Elofsson A, Nielsen H. **2019**. Detecting sequence signals in targeting peptides using deep learning. *Life Science Alliance* **2**: e201900429.

Bäurle I. **2016**. Plant Heat Adaptation: priming in response to heat stress. *F1000Research* **5**: 694.

Brameier M, Krings A, MacCallum RM. **2007**. NucPred—Predicting nuclear localization of proteins. *Bioinformatics* **23**: 1159–1160.

Briesemeister S, Rahnenführer J, Kohlbacher O. **2010**. YLoc-an interpretable web server for predicting subcellular localization. *Nucleic Acids Research* **38**: 497–502.

Brunkard JO, Runkel AM, Zambryski PC. **2015**. Chloroplasts extend stromules independently and in response to internal redox signals. *Proceedings of the National Academy of Sciences* **112**: 10044 LP – 10049.

Casimiro-Soriguer CS, Muñoz-Mérida A, Pérez-Pulido AJ. **2017**. Sma3s: A universal tool for easy functional annotation of proteomes and transcriptomes. *Proteomics* **17**: 1700071.

Chan KX, Crisp PA, Estavillo GM, Pogson BJ. **2010**. Chloroplast-to-nucleus communication: Current knowledge, experimental strategies and relationship to drought stress signaling. *Plant Signaling and Behavior* **5**: 1575–1582.

Chen M, Galvão RM, Li M, Burger B, Bugea J, Bolado J, Chory J. **2010**. Arabidopsis HEMERA/pTAC12 initiates photomorphogenesis by phytochromes. *Cell* **141**: 1230–1240.

Daniell H, Lin CS, Yu M, Chang WJ. **2016**. Chloroplast genomes: Diversity, evolution, and applications in genetic engineering. *Genome Biology* **17**: 1–29.

Dickinson PJ, Kumar M, Martinho C, Yoo SJ, Lan H, Artavanis G, Charoensawan V, Schöttler MA, Bock R, Jaeger KE, *et al.* 2018. Chloroplast Signaling Gates Thermotolerance in Arabidopsis. *Cell Reports* **22**: 1657–1665.

Dobrogojski J, Adamiec M, Luciński R. **2020**. The chloroplast genome: a review. *Acta Physiologiae Plantarum* **42**: 1–13.

Escandón M, Lamelas L, Roces V, Guerrero-Sanchez VM, Meijón M, Valledor L. **2020**. Protein Interaction Networks: Functional and Statistical Approaches. In: Methods in Molecular Biology. 21–56.

Escandón M, Valledor L, Lamelas L, Álvarez JM, Cañal MJ, Meijón M. **2022**. Multiomics analyses reveal the central role of nucleolus and nucleoid machinery during heat stress acclimation in *Pinus radiata*; *bioRxiv*: 2022.07.08.499117.

Fang X, Zhao G, Zhang S, Li Y, Gu H, Li Y, Zhao Q, Qi Y. **2019**. Chloroplast-to-Nucleus Signaling Regulates MicroRNA Biogenesis in Arabidopsis. *Developmental Cell* **48**: 371-382.e4.

Foyer CH, Karpinska B, Krupinska K. **2014**. The functions of WHIRLY1 and REDOX-RESPONSIVE TRANSCRIPTION FACTOR 1 in cross tolerance responses in plants: a hypothesis. *Philosophical Transactions of the Royal Society B: Biological Sciences* **369**: 20130226–20130226.

Fu ZQ, Guo M, Jeong B, Tian F, Elthon TE, Cerny RL, Staiger D, Alfano JR. **2007**. A type III effector ADP-ribosylates RNA-binding proteins and quells plant immunity. *Nature* **447**: 284–288.

Gao H, Kadirjan-Kalbach D, Froehlich JE, Osteryoung KW. **2003**. ARC5, a cytosolic dynamin-like protein from plants, is part of the chloroplast division machinery. *Proceedings of the National Academy of Sciences of the United States of America* **100**: 4328–4333.

Guo J, Wang S, Valerius O, Hall H, Zeng Q, Li J-F, Weston DJ, Ellis BE, Chen J-G. **2011**. Involvement of Arabidopsis RACK1 in protein translation and its regulation by abscisic acid. *Plant physiology* **155**: 370–383.

Hackett JB, Shi X, Kobylarz AT, Lucas MK, Wessendorf RL, Hines KM, Bentolila S, Hanson MR, Lu Y. **2017**. An Organelle RNA Recognition Motif Protein Is Required for Photosystem II Subunit psbF Transcript Editing. *Plant physiology* **173**: 2278–2293.

Hanson MR, Sattarzadeh A. **2013**. Trafficking of Proteins through Plastid Stromules. *The Plant Cell* **25**: 2774–2782.

Hou X, Fu A, Garcia VJ, Buchanan BB, Luan S. **2015**. PSB27: A thylakoid protein enabling Arabidopsis to adapt to changing light intensity. *Proceedings of the National Academy of Sciences* **112**: 1613 LP – 1618.

Hu S, Ding Y, Zhu C. **2020**. Sensitivity and Responses of Chloroplasts to Heat Stress in Plants . *Frontiers in Plant Science* **11**: 375.

Jahns P, Latowski D, Strzalka K. **2009**. Mechanism and regulation of the violaxanthin cycle: the role of antenna proteins and membrane lipids. *Biochimica et biophysica acta* **1787**: 3–14.

Jin J, Tian F, Yang D-C, Meng Y-Q, Kong L, Luo J, Gao G. **2017**. PlantTFDB 4.0: toward a central hub for transcription factors and regulatory interactions in plants. *Nucleic Acids Research* **45**: D1040–D1045.

Jung HS, Chory J. **2010**. Signaling between chloroplasts and the nucleus: Can a systems biology approach bring clarity to a complex and highly regulated pathway? *Plant Physiology* **152**: 453–459.

Kato Y, Sakamoto W. **2018**. FtsH protease in the thylakoid membrane: Physiological functions and the regulation of protease activity. *Frontiers in Plant Science* **9**: 1–8.

Köhler RH, Jun C, R. ZW, W. WW, R. HM. **1997**. Exchange of Protein Molecules Through Connections Between Higher Plant Plastids. *Science* **276**: 2039–2042.

Krause K, Oetke S, Krupinska K. **2012**. Dual targeting and retrograde translocation: Regulators of plant nuclear gene expression can be sequestered by plastids. *International Journal of Molecular Sciences* **13**: 11085–11101.

Krupinska K, Blanco NE, Oetke S, Zottini M. **2020**. Genome communication in plants mediated by organelle–nucleus-located proteins. *Philosophical Transactions of the Royal Society B: Biological Sciences* **375**.

Kwok EY, Hanson MR. **2004**. Plastids and stromules interact with the nucleus and cell membrane in vascular plants. *Plant Cell Reports* **23**: 188–195.

Lesk C, Rowhani P, Ramankutty N. **2016**. Influence of extreme weather disasters on global crop production. *Nature* **529**: 84–87.

Li Y, Williams B, Dickman M. **2017**. Arabidopsis B-cell lymphoma2 (Bcl-2)-associated athanogene 7 (BAG7)-mediated heat tolerance requires translocation, sumoylation and binding to WRKY29. *New Phytologist* **214**: 695–705.

Ling Y, Serrano N, Gao G, Atia M, Mokhtar M, Woo YH, Bazin J, Veluchamy A, Benhamed M, Crespi M, *et al.* 2018. Thermopriming triggers splicing memory in Arabidopsis. *Journal of Experimental Botany* **69**: 2659–2675.

Lohse M, Nagel A, Herter T, May P, Schroda M, Zrenner R, Tohge T, Fernie AR, Stitt M, Usadel B. **2014**. Mercator: A fast and simple web server for genome scale functional annotation of plant sequence data. *Plant, Cell and Environment* **37**: 1250–1258.

van der Maaten L, Hinton G. **2008**. Visualizing Data using t-SNE. *Journal of Machine Learning Research* **9**: 2579–2605.

Moreno JC, Mi J, Alagoz Y, Al-Babili S. **2021**. Plant apocarotenoids: from retrograde signaling to interspecific communication. *The Plant Journal* **105**: 351–375.

Mullineaux PM, Exposito-Rodriguez M, Laissue PP, Smirnoff N, Park E. **2020**. Spatial chloroplast-to-nucleus signalling involving plastid–nuclear complexes and stromules. *Philosophical Transactions of the Royal Society B: Biological Sciences* **375**: 20190405.

Muñoz P, Munné-Bosch S. **2019**. Vitamin E in Plants: Biosynthesis, Transport, and Function. *Trends in Plant Science* **24**: 1040–1051.

Nair R, Carter P, Rost B. **2003**. NLSdb: database of nuclear localization signals. *Nucleic Acids Research* **31**: 397–399.

Nguyen Ba AN, Pogoutse A, Provart N, Moses AM. **2009**. NLStradamus: a simple Hidden Markov Model for nuclear localization signal prediction. *BMC bioinformatics* **10**: 202.

O'Neill BC, Oppenheimer M, Warren R, Hallegatte S, Kopp RE, Pörtner HO, Scholes R, Birkmann J, Foden W, Licker R, *et al.* 2017. IPCC reasons for concern regarding climate change risks. *Nature Climate Change* **7**: 28–37.

Perez-Riverol Y, Csordas A, Bai J, Bernal-Llinares M, Hewapathirana S, Kundu DJ, Inuganti A, Griss J, Mayer G, Eisenacher M, *et al.* 2019. The PRIDE database and related tools and resources in 2019: Improving support for quantification data. *Nucleic Acids Research* **47**: D442–D450.

Pfannschmidt T, Terry MJ, Van Aken O, Quiros PM. **2020**. Retrograde signals from endosymbiotic organelles: A common control principle in eukaryotic cells. *Philosophical Transactions of the Royal Society B: Biological Sciences* **375**.

R Core Team. **2020**. R: A Language and Environment for Statistical Computing.

Ristic Z, Momčilović I, Fu J, Callegari E, DeRidder BP. **2007**. Chloroplast protein synthesis elongation factor, EF-Tu, reduces thermal aggregation of rubisco activase. *Journal of Plant Physiology* **164**: 1564–1571.

Saifi SK, Passricha N, Tuteja R, Tuteja N. **2018**. Stress-induced Oryza sativa RuvBL1a is DNA-independent ATPase and unwinds DNA duplex in 3′ to 5′ direction. *Protoplasma* **255**: 669–684.

Savojardo C, Martelli PL, Fariselli P, Profiti G, Casadio R. **2018**. BUSCA: An integrative web server to predict subcellular localization of proteins. *Nucleic Acids Research* **46**: W459–W466.

Schult K, Meierhoff K, Paradies S, Töller T, Wolff P, Westhoff P. **2007**. The nuclear-encoded factor HCF173 Is involved in the initiation of translation of the psbA mRNA in Arabidopsis thaliana. *Plant Cell* **19**: 1329–1346.

Serrano N, Ling Y, Bahieldin A, Mahfouz MM. **2019**. Thermopriming reprograms metabolic homeostasis to confer heat tolerance. *Scientific Reports* **9**.

Shannon P, Markiel A, Owen Ozier 2, Baliga NS, Wang JT, Ramage D, Amin N, Schwikowski B, Ideker T. **2003**. Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Research*: 2498–2504.

Shi D-Q, Liu J, Xiang Y-H, Ye D, Sundaresan V, Yang W-C. **2005**. SLOW WALKER1, essential for gametogenesis in Arabidopsis, encodes a WD40 protein involved in 18S ribosomal RNA biogenesis. *The Plant cell* **17**: 2340–2354.

Sperschneider J, Catanzariti AM, Deboer K, Petre B, Gardiner DM, Singh KB, Dodds PN, Taylor JM. **2017**. LOCALIZER: Subcellular localization prediction of both plant and effector proteins in the plant cell. *Scientific Reports* **7**: 1–14.

Steiner S, Schröter Y, Pfalz J, Pfannschmidt T. **2011**. Identification of Essential Subunits in the Plastid-Encoded RNA Polymerase Complex Reveals Building Blocks for Proper Plastid Development. *Plant Physiology* **157**: 1043–1055.

Świda-Barteczka A, Krieger-Liszkay A, Bilger W, Voigt U, Hensel G, Szweykowska-Kulinska Z, Krupinska K. **2018**. The plastid-nucleus located DNA/RNA binding protein WHIRLY1 regulates microRNA-levels during stress in barley (Hordeum vulgare L.). *RNA Biology* **15**: 886–891.

Unal D, García-Caparrós P, Kumar V, Dietz KJ. **2020**. Chloroplast-associated molecular patterns as concept for fine-tuned operational retrograde signalling. *Philosophical Transactions of the Royal Society B: Biological Sciences* **375**.

Wahid A, Gelani S, Ashraf M, Foolad MR. **2007**. Heat tolerance in plants: An overview. *Environmental and Experimental Botany* **61**: 199–223.

Walters RG, Shephard F, Rogers JJM, Rolfe SA, Horton P. **2003**. Identification of mutants of arabidopsis defective in acclimation of photosynthesis to the light environment. *Plant Physiology* **131**: 472–481.

Wang X, Xu M, Gao C, Zeng Y, Cui Y, Shen W, Jiang L. **2020**. The roles of endomembrane trafficking in plant abiotic stress responses. *Journal of Integrative Plant Biology* **62**: 55–69.

Wehrens R, Buydens LMC. **2007**. Self- and Super-organizing Maps in R: The kohonen Package. *JSS Journal of Statistical Software* **21**.

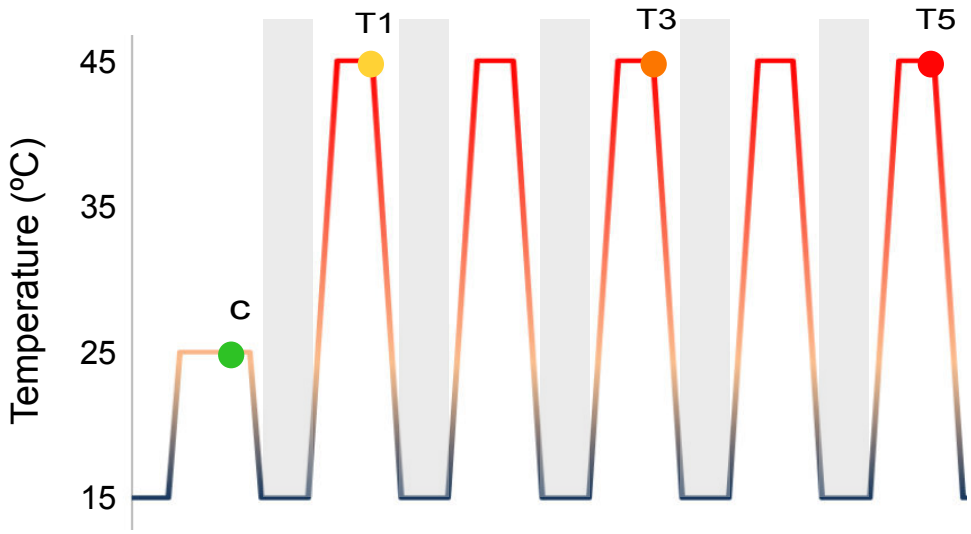
Xie M, Yu B. **2015**. siRNA-directed DNA Methylation in Plants. *Current Genomics* **16**: 23–31.

Zhao X, Huang J, Chory J. **2019**. GUN1 interacts with MORF2 to regulate plastid RNA editing during retrograde signaling. *Proceedings of the National Academy of Sciences of the United States of America* **116**: 10162–10167.

Zhao X, Huang J, Chory J. **2020**. Unraveling the Linkage between Retrograde Signaling and RNA Metabolism in Plants. *Trends in Plant Science* **25**: 141–147.

5.7 Supplemental information

5.7.1 Supplemental figures



Supplemental Figure S5.1. Experimental design mimicking a heatwave. Temperature tanked from 15 ºC to 25 ºC (night/day) in control conditions and from 15 ºC to 45 ºC under stress, night periods are shadowed. Selected sampling points from previous chapters to conform the integrative analysis between nuclei and chloroplast are indicated with circles.

5.7.2 Supplemental tables

Supplemental table legends

Supplemental Table S5.1. Protein re-identification, subcellular location, preprocessed proteome abundances, univariate analysis and annotations of nucleus and chloroplast proteome datasets

Supplemental Table S5.2. PCA explained variance, loadings and annotations considering nuclear proteome, chloroplast proteome and the concatenation of both them.

Supplemental Table S5.3. Mean abundances, description, functional annotation and kmeans clustering of proteins showing dual location in the nucleus and the chloroplast

Supplemental Table S5.4. Explained variance and multiblock loadings of anterograde and retrograde communication sPLS networks.

CHAPTER VI. Epigenetic (thermo)memory in *Pinus radiata*

6.1 Abstract

Plants are sessile organisms that usually need to face recurrent weather extremes in order to persist. Fortunately, molecular memory represents a promising mechanism for plant survival. However, despite its relevance, how plants remember stress remains poorly understood, particularly in long-lived species as *Pinus radiata*. To fill this gap, in this chapter, molecular memory has been explored from three epigenetic perspectives covering DNA methylation, studied through SAM synthase and SAHH gene expression and 5-methylcytosineC (5mC) detection; histone variants H2A and H2B, and microRNAs as miR160 and miRNA396. The evaluated candidates were selected accordingly to the previous chapters of this thesis and tracked along different experimental designs covering trans- and intra- generational memory development in seeds and seedlings. This newly discovered priming-induced epigenetic memory may represent a general feature of heat-stress responses in forests species Thus, providing a panel of heat memory biomarkers that enhance our current knowledge on (thermo)primming field, and could represent a major advance for tolerant genotypes selection to be further studied by breeders.

.

6.2 Introduction

It has been established that environmental challenges to a maternal plant can affect the quantity and composition of starch reserves, epigenetic status, microRNAs, mRNAs, proteins, hormones, and other primary and secondary metabolites packaged into seeds (Herman & Sultan, 2011) leading to altered seed provisioning (Zas *et al.*, 2013). These changes are expected to produce a more resistant phenotype showing an adaptational mechanism. This effect is called transgenerational memory (Liu *et al.*, 2022a), and it has been shown that this kind of transmittable memory persists over generations and may explain the differences between isogenic subpopulations based, at least partially, on epigenetic mechanisms in model species (Adrian-Kalchhauser *et al.*, 2020) as well as in forest trees (Carbó *et al.*, 2019; García-Campa *et al.*, 2022). Regarding epigenetic mechanism, microRNAs have been shown to play a main role in the maintenance of heritable molecular information from the parents to their offspring (Locato *et al.*, 2018; Houri-Zeevi *et al.*, 2021).

In recent years, it has been reported that chloroplast retrograde signaling could involve the regulation of microRNA biogenesis in *Arabidopsis thaliana* (Lin *et al.*, 2018; Świda-Barteczka *et al.*, 2018; Fang *et al.*, 2019; Ravichandran *et al.*, 2019). However, the link between both processes is quite little-known, and even less so in long-lived species such as forest trees. MicroRNAs have been also established as key players in vital cellular aspects through microRNA-guided post-transcriptional gene silencing (PTGS). A few studies have claimed the relevance of PTGS in the heat stress response in other plant species, mainly through alterations in antioxidant activity and redox homeostasis (Guerra *et al.*, 2015; Fang *et al.*, 2019; Ravichandran *et al.*, 2019), although it remains unknown whether this process is related to damage, signaling, or acclimation. However, it is still unclear, especially for woody plants, how this molecular process reprograms gene expression and whether they have a role in inherited memory.

Besides microRNAs, other epigenetic components in the development of the memory are Histones and DNA methylation (Iwasaki & Paszkowski, 2014; Vriet *et al.*, 2015; Wibowo *et al.*, 2016; Liu *et al.*, 2022b). Hence, and in relation to nuclear proteins relevance when driving heat stress adaptation, it has been described the key role of epigenetic regulation and histone modifications to keep the memory of the stress (Bäurle, 2016; Lämke *et al.*, 2016) that lead to priming mechanism (Martinez-medina *et al.*, 2016).

Priming involves a first training stress, a latent phase and a second stress event; in this later stress, the plant will be able to react in a more efficient way than previously, due to the information stored as chromatin structural changes and histone modifications (Gutzat and Scheid, 2012; Pastor *et al.*, 2013; Asensi-Fabado *et al.*, 2017). According to these findings, the epigenetic mechanisms, and particularly DNA methylation and nucleosome occupancy, seem to be main players on priming establishment. The epigenetic mechanisms involve covalent modifications of DNA and histones, which affect transcriptional activity of chromatin (Valledor *et al.*, 2007). Since chromatin states can be propagated through cell divisions, epigenetic mechanisms are thought to provide heritable 'cellular memory' (Iwasaki & Paszkowski, 2014).

In this chapter, several types of molecular memory have been explored, including transgenerational memory, short term memory and stable memory acquisition processes by the combination of different experimental schemes. To characterize these changes leading to more resistant plants, throughout this chapter, the relative accumulation of microRNAs was analyzed in parallel to targeted transcriptomics analysis to elucidate the role of microRNAs in retrograde signaling and memory acquisition and maintenance processes. In addition, DNA methylation dynamics in basal conditions, during the stress and after plants were recovered was studied in the seedlings. And finally, in a second round of stress DNA methylation and histone isoforms related candidates to play a key role in the thermopriming processes were tested by RT-qPCR (reverse transcription quantitative real-time PCR). This approach included seeds, seedlings and saplings so a deep coverage of the epigenetic mechanism was achieved to enhance the understanding of these vital processes given the critical environmental situation we are currently facing.

6.3 Material and Methods

6.3.1 Plant material and heat stress treatments

Field-collected seeds belonging to the isogenic subpopulations described in Chapter 3 (E and T), were weighted, their area was measured using imageJ (Schindelin *et al.*, 2012) and, dissected; embryos were used for (micro)RNA extraction; and the endosperms were lyophilized and used for physiological profiling (see below). Seeds from each progeny were divided into four pools that constituted the independent biological replicates in this study.

Additionally, seeds belonging to both subpopulations were grown in 1 dm^3 pots in a climate chamber (Fitoclima 1200, Aralab) under the following control day/night conditions: 16/8 h photoperiod (400 µmol m^{-2} s^{-1}), 25/15 °C and 50/60% relative humidity (RH). Eight-month-old seedlings of the two progenies (height 25 ± 0.3 cm, showing no growth difference between them, t test, P>0.05) were used for mi(RNA) extraction and physiological profiling.

The non-primmed progeny (E) was selected for the heat stress treatment, which was applied in an identical manner to previous chapters. In brief, plants were exposed to a 45 ºC maximum temperature, during five days, mimicking a heat wave in a climate chamber (Fitoclima 1200, Aralab). Needles from these plant (four independent pools, made of three plants each) were used for micro(RNA) extraction.

After the heat stress, E progeny seedlings were allowed to recover in control conditions (described above) during one month, and sampled again, sections of needles (from T0, T1, T3, T5 and SR) were fixed in 4% paraformaldehyde for further 5-mdC immunolocalization analysis. Five months later, (six month after the first heat stress) E progeny saplings were heat-stressed again, together with plants of same age (14 months) without any prior stress exposure. The application of the heat stress was done in an identical manner as the first time (five consecutive days, 45 ºC, 6 hours/day). Needles were sampled for RNA extraction and physiological profiling.

6.3.2 Physiological profiling

Quantification of Chlorophyll a, Chlorophyll b, Carotenoids, Malondyaldehyde (MDA), Free Amino Acids (FAA), Total Soluble Sugars (TSS), Starch (STA), Total Flavonoids (TFL), and Total Phenolics Compounds (TPC) were performed according to López-Hidalgo *et al.* (2021). This analysis was performed for T0 sampling points and both progenies starting from 10 mg of lyophilized needles. FAA, TSS, STA, TFL and TPC were also quantified in seeds of both progenies, starting from 20 mg of lyophilized endosperms.

6.3.3 Immunolocalization of 5-mdC

Methylated DNA was monitored in samples T0, T1, T3, T5, and SR (Stress Recovered) by 5-mdC immunolocalization according to the procedure described by (Meijón *et al.*, 2010). Briefly, fixed needles were sectioned at 50 µm thickness using a CH 1510-1 cryomicrotome (Leica Microsystems). The samples were permeabilized, blocked with bovine serum albumin and incubated with anti-5-mdC mouse antibody (Eurogentec, Belgium) diluted 1:50 in 1% blocking solution. Alexa Fluor 488-labelled anti-mouse polyclonal antibody (Invitrogen) diluted 1:25 was used as secondary antibody for detection of 5-mdC. The slides were counterstained with DAPI. Fluorescence was visualized using a TCS-SP2-AOBS confocal microscope (Leica). Maximal projection from a stack of six slides per sample was acquired using the Fiji software (Schindelin *et al.*, 2012).

6.3.4 (micro)RNA extraction, quantification, cDNA synthesis and quantitative reverse transcription PCR

A set of common microRNA sequences previously identified in *Pinus pinaster* (Rodrigues *et al.*, 2019), were blasted against miRbase (Release 22.1) (Kozomara *et al.*, 2019). Those sequences conserved in other tree species were kept and blasted against the transcripts sequences corresponding to the nuclear and chloroplast proteins, which were found to be significantly accumulated when comparing control to stress conditions as described in Chapter 5, using psRNATarget tool (Dai *et al.*, 2018).

The (micro)RNA extraction was performed according to Valledor *et al.*, (2014), with minor modifications to enrich microRNA concentration: absolute ethanol was added to RNA-containing supernatant and then passed through the silica columns. RNA and microRNA were quantified using Qubit Assay Kits (Thermo Scientific, RNA cat. no.

Q33223; miRNA cat. no. Q32880), and a total amount of 500 ng were retro-transcribed using mi-X-RNA kit (Takara, cat. no. 638315).

The qPCR reactions were performed in a CFX Connect Real-Time PCR machine (Bio-Rad) with SsoAdvanced Universal SYBR Green Supermix (Bio-Rad), using three biological and three analytical replicates. Normalized Relative Quantities (NRQ) and Standard Errors of RQ were determined according to Hellemans et al., (2007). Expression levels of *ACTIN* (*ACT*) and *GLYCERALDEHYDE 3-PHOSPHATE DEHYDROGENASE* (*GAPDH*) were used as endogenous controls for mRNA quantification and U6 snRNA expression levels for microRNA quantification as recommended in mi-X-RNA kit (Takara, cat. no. 638315). Detailed information about the primers used for RT-qPCR experiments is available in Table S6.1.

6.4 Results

6.4.1 Seeds belonging to two wild isogenic subpopulations already exhibited differences in weight, shape, composition and gene expression according to their maternal provenance

In Chapter 3, two subpopulations were compared under heat stress, but interestingly, Fv/Fm, EL, TSS, and STA exhibited differences prior to the stress exposure (T0; $P < 0.05$, t test, Fig S6.1). To check whether these differences found in basal conditions could be detected already in the seeds belonging to each subpopulation, their weight, area, and several physiological biomarkers of the seed endosperms (FAA, TSS, STA, TPC, and TFL) were measured. In addition, to further explore the underlying causes of these differences a targeted transcriptome profile including mRNAs (SAM synthase, SAHH and H2A) and miRNAs (miRNA 947, miRNA 396, miRNA 160, and miRNA 1131) was performed in seed embryos.

As shown in Fig. 6.1a, the results confirmed the differences in seed provisioning between the progenies of both subpopulations. E progeny presented higher weight and lower area. For this reason, the rest of the measured parameters were scaled to dry weight. E progeny showed a higher amount of FAA and TFL, and potentially primed progeny (belonging to T subpopulation) showed an increase in starch and total phenolics content.

The expression levels of three genes related to epigenetics: S-Adenosylmethionine (SAM) Synthase and S-Homocysteine Hydrolase (SAHH), required for DNA methylation, and the histone variant H2A.X linked to heat stress memory and recovery, as described in Chapter 4, as well as, biotic stress tolerance in pine species (Amaral *et al.*, 2021) were measured. The expression of the heat-responsive microRNAs miR947, miR396, miR160 and miR1131 was also evaluated (Fig. 6.1b). These microRNAs were found to be involved in the inhibition of H2A, APX, eIF3G and GUN4, covering epigenetics, ROS signaling, translation, and chloroplast-to-nucleus signaling (Table S6.2). Additionally, miR160 is related to plant development and heat stress response triggering (Lin *et al.*, 2018).

All mRNAs exhibit significant differences between seeds belonging to the two subpopulations. Those differences were however buffered in the seedlings, eight months after germination (Fig. S6.1b) when there were no growth differences between the two

progenies. On the other hand, miRNA expression also shifted from a lower abundance in the seeds to increased levels in T subpopulation seedlings for miR396, miR160 and miR1311 (Fig. S6.1b).

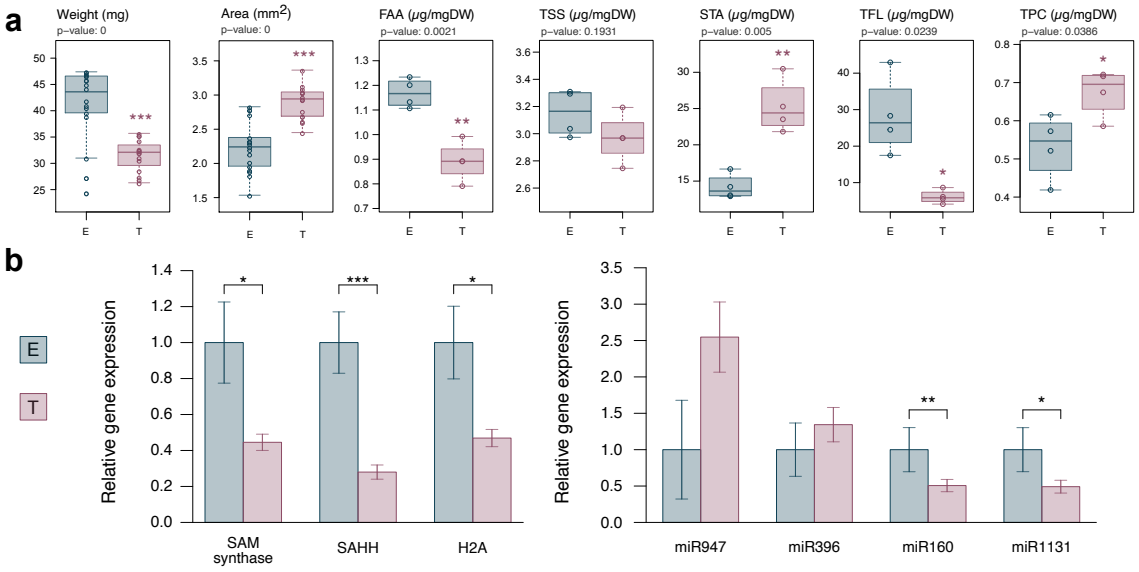


Fig. 6.1. Seeds characterization. a) Seeds weight, area, and physiology profile including Free Amino Acids (FAA), Total Soluble Sugars (TSS), starch (STA), Total Flavonoids (TFL) and Total Phenolic compounds (TPC) and b) mRNA and microRNA relative expression of selected candidates. For both progenies. Asterisks indicate significance according to Wilcox non-parametric test; * pvalue < 0.05; ** pvalue < 0.01; *** pvalue < 0.001.

6.4.2 microRNA contents raised at long-term heat stress exposure

To further investigate the role of RNA modulation in the stress response, microRNA abundances were quantified in basal and under stress conditions (Fig 6.2a). The obtained results showed an increasing trend along with stress and a significant change in T5.

In addition, we aimed to find concrete microRNAs and mRNA pairs relevant to heat stress acclimation mechanisms. To do so, transcripts from nuclear-encoded proteins, which exhibited significant variations in nuclear or chloroplast proteomes, were blasted as targets against conserved microRNA (Table S6.2). The in silico identified target-microRNA pairs were filtered to keep the most probable pairs.

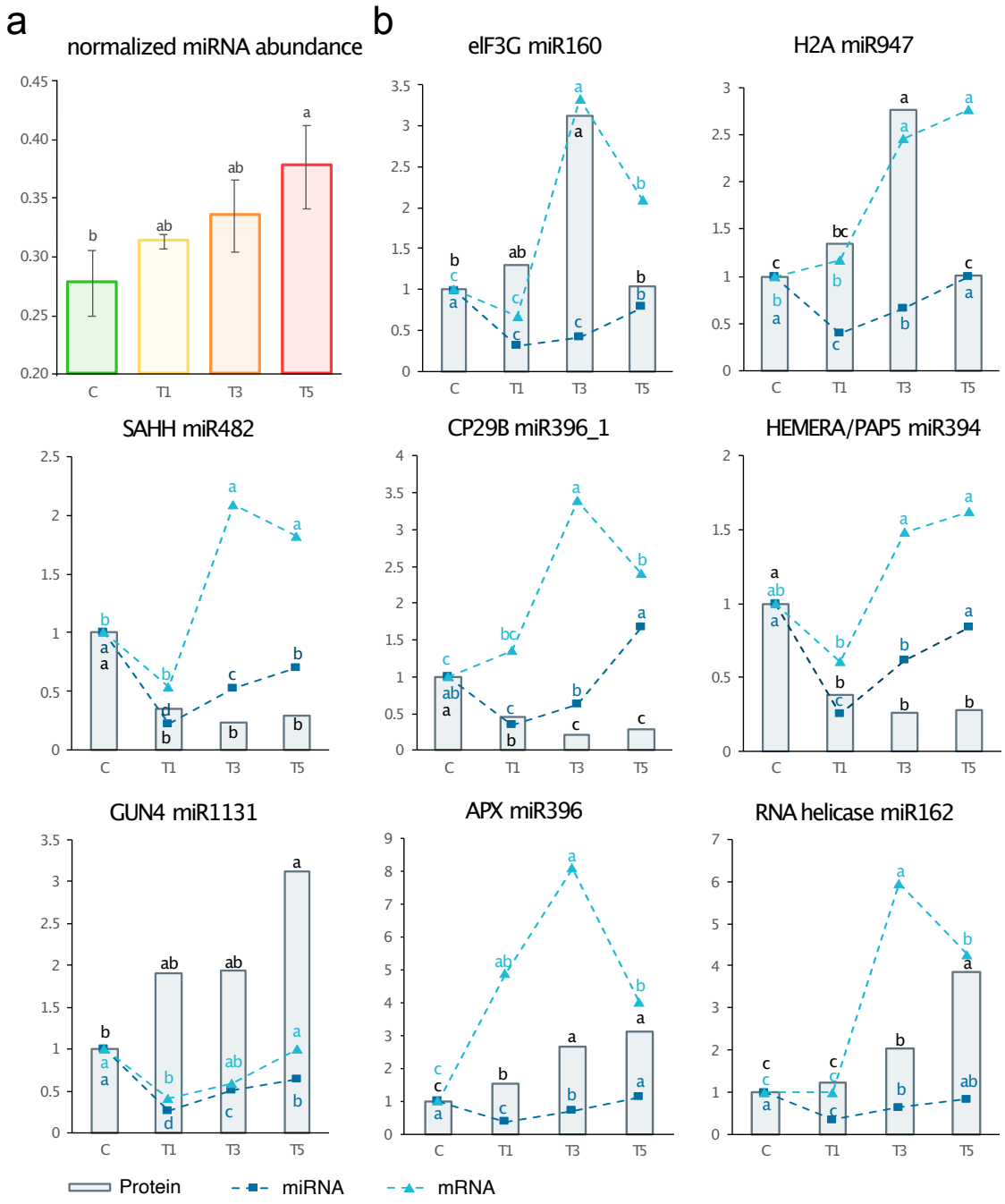


Fig. 6.2. microRNAs biogenesis across heat stress and correlation to candidate target mRNA and relative protein levels. (a) Normalized microRNA abundance, (b) candidate microRNAs levels, fold change of their targeted mRNA expression and correlative protein abundance. Proteins are represented in bars (grey) and mRNA (light blue) and microRNA (dark blue) in lines. (a,b) Different letters indicate significant changes according to HSD test ($p < 0.05$) four biologically independent replicates were used for proteomics analysis and, three independent biological replicates were used for gene expression measurements.

Among them, we selected biological relevant targets covering translation (eIF3G; miR160), chromatin conformation (H2A; miR947), genome methylation (SAHH; miR482), photosynthesis (CP29B; miR396_1), retrograde communication (HEMERA/PAP5; miR394 and GUN4; miR1131), redox homeostasis (APX; miR396) and RNA metabolism (CP29B;

miR396_1, RNA helicase; miR162). miR160 has been previously linked to acquired thermotolerance (Lin et al., 2018) and miR162, miR394, miR396 and miR482 have been linked to heat stress, with no common pattern among different species or experimental designs, and none of them has been monitored for more than 24 h under stress (reviewed by Liu et al., 2015).

Proteins, mRNAs and their potential microRNAs change folds are shown in Fig 6.2b. As expected, in some cases the discrepancy between protein and mRNA abundances can be explained with microRNA expression changes, especially in long-term heat-treated samples. eIF3G, H2A, SAHH, CP29B, HEMERA transcripts were upregulated along with their proposed microRNA pairs, and the protein abundance decreased, while for the other pairs, their relation is not clear. These results indicate possible additional regulatory layers that complicate the interpretation of the results and reveal a complex network involved in their regulation.

6.4.3 5-mdC immunolocalization in needles under heat stress and recovery showed hypomethylation during the heat stress response and hypermethylation in the latent phase

Epigenetic reorganization related to adaptation was further studied by immunolocalization analysis of 5-mdC in the needles collected during the heat stress experiments (Fig. 6.3, negative control showed in Fig. S6.2). From T0 to T3, the signal of 5-mdC massively decreased in cell nucleus, focusing the scarce signal in the vascular tissue. In T5 signal of 5-mdC start to rebound, reaching in recovered plants (SR) the highest levels of DNA methylation. These results align with the accumulation patters of SAM synthase and SAHH obtained in Chapter 4, corroborating the hypothesized relaxed chromatin status under the applied heat stress.

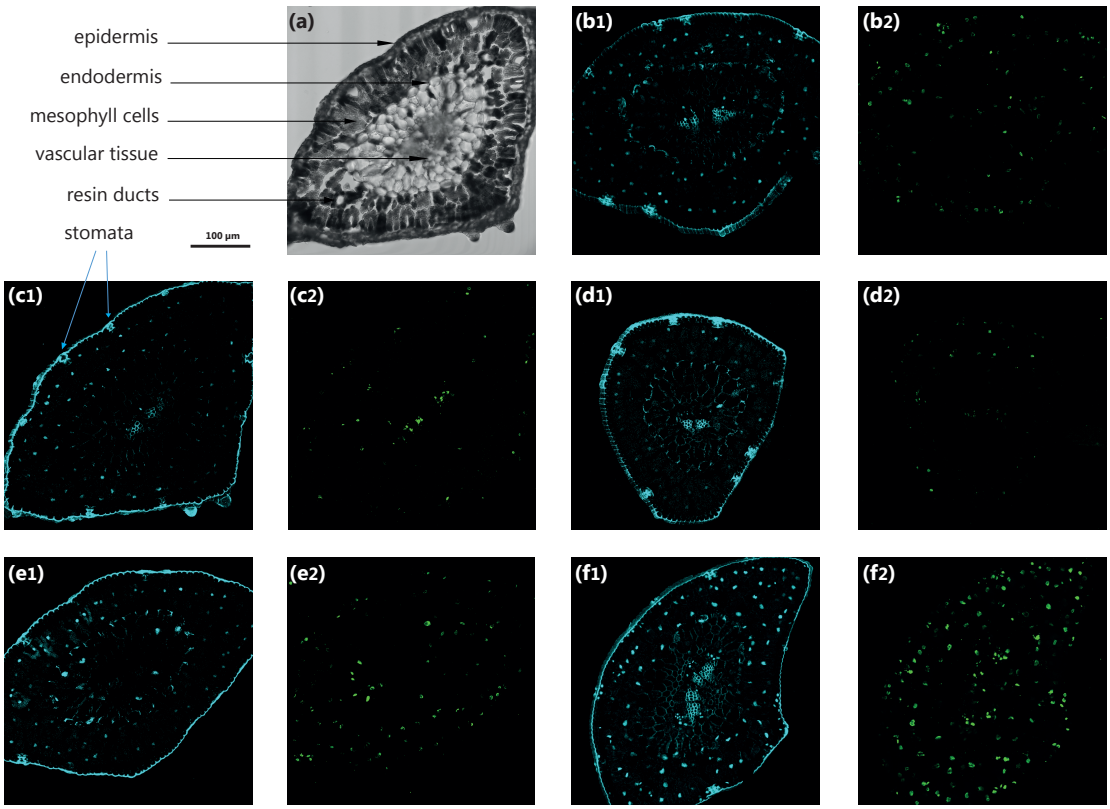


Fig. 6.3. 5-methylcytosine (5-mC) immunolocalization in transversal sections of needles. Differential interference contrast (DIC) image of a transversal needle section is showed in the first image; fluorescence signal from confocal microscope across the treatments is showed in the following images: **a)** T0; **b)** T1; **c)** T3; **d)** T5; **e)** SR. Sub-numbers Labeling: **(1)** nuclei mark from DAPI (blue signal) and **(2)** 5-mC mark (green signal). Black bar = 100 µm.

6.4.4 Targeted transcriptome analysis and photosynthesis measurements supported the better performance of primed plants upon a second round of stress.

To further validate the potential acquisition of memory after stress exposure, the photochemical performance, some physiological biomarkers and the gene expression of nuclear candidates highlighted in Chapter 4 and 5 in heat primed and non-primed plants were measured following the experimental design showed in Fig. 6.4.

Heat primed and non-primed plants showed significant differences during and after the heat stress (Fig. 6.5, Fig. S6.3). The plants subjected to first stress exposure suffered higher photosynthetic damage in comparison with thermoprimed plants. On the other hand, TSS and total phenolics showed significant higher levels in primed plants which indicates a physiological preconditioning to stress (Jesus *et al.*, 2015). This provided insights of a stable acquired thermotolerance, that lasted at least six months.

Molecular responses of these plants were further studied analysing expression levels of genes coding for some of the key proteins previously described according to multivariate analyses, covering DNA methylation (*SAM SYNTHASE* and *SAHH*), splicing (*LSM3A*), DNA repair (*NEDD 8 E1*), and nucleosome assembly (*H2A* and *H2B*). First stressed and primed plants (second stress) followed the same gene expression profile across the stress time points, with the exception of *SAM SYNTHASE*. However, higher values of gene expression were found in first stressed plants for most of the genes and times. *SAM SYNTHASE* showed a particular behaviour, required for DNA methylation, exhibited control-like expression in ST5 treatment in thermoprimed plants, while in first stress its levels continued dropping.

The better performance of thermoprimed plants when looking to the photosynthetic activity coupled to the lower expression of candidate genes and the recovery of methyl cycle enzymes expression during the stress supports the hypothesis of stable acquired thermotolerance guided by epigenetic events.

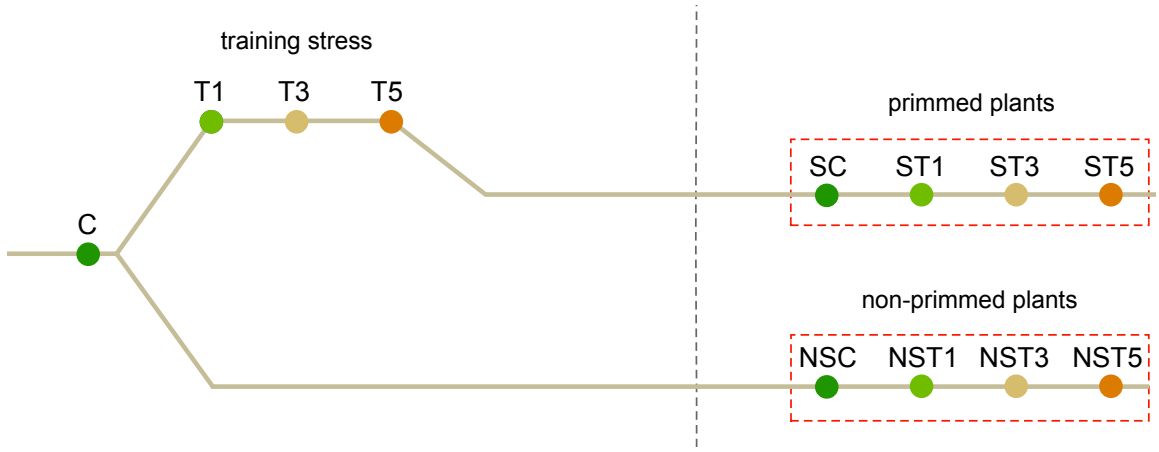


Fig. 6.4 Outline of experimental set up for long term memory testing. Control plants were divided in two plant sets. Primmed plants were heat stressed at 45 ºC during 6 hours/day, during five days. Meanwhile, the rest of the seedlings were allowed to grow in control conditions, as a control line. Both sets were kept under control conditions during six months. Finally, both plant sets were heat treated (equally to the previous stress for primmed plants), in order to check whether long-term memory had been developed by trained plants by comparing previously Stressed plants, (primmed plants) SC, ST1, ST3 and ST5, and (non-primed plants) Not previously Stressed plants denoted as NSC, NST1, NST3 and NST5.

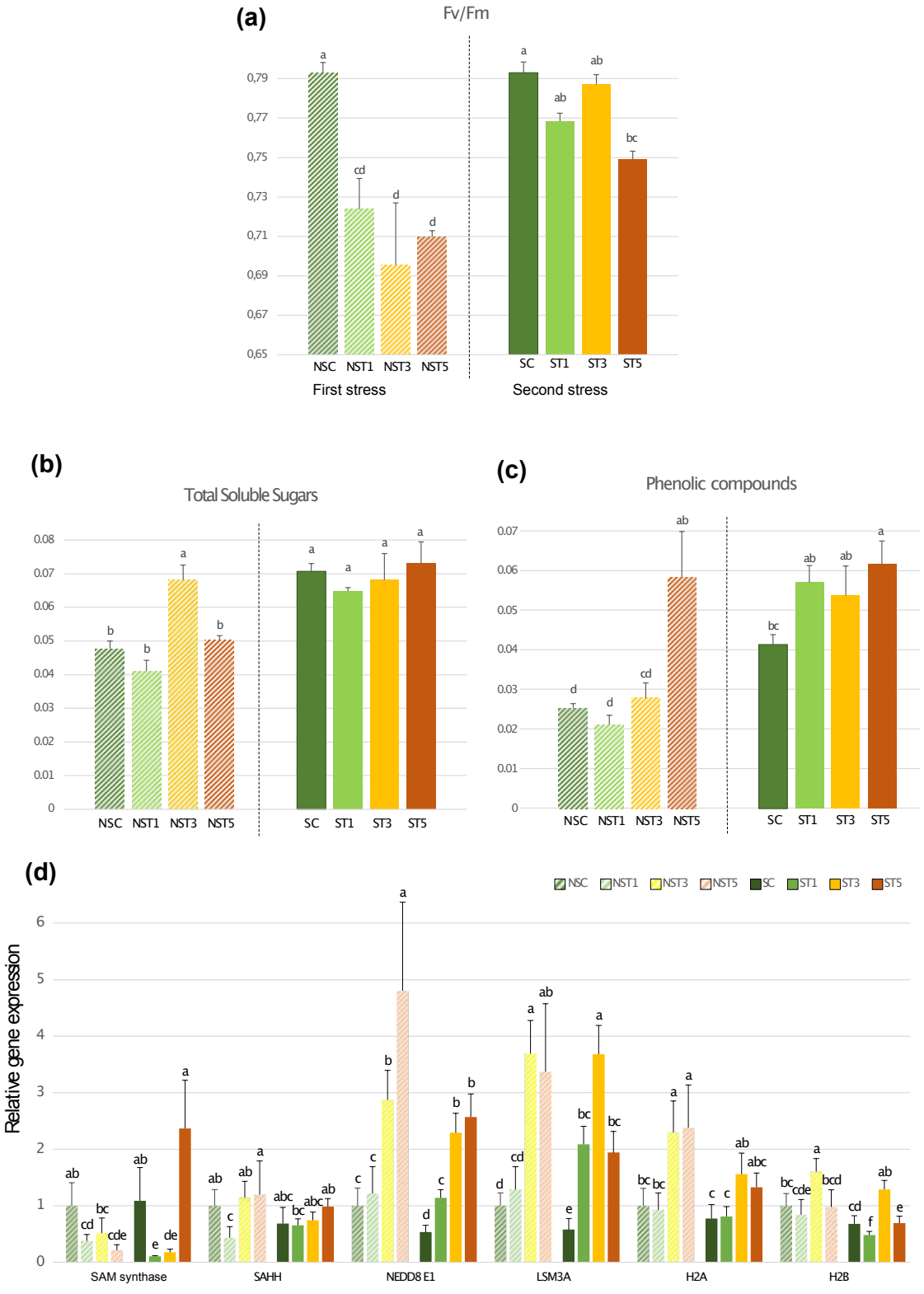


Fig. 6.5 a)Maximum yield of Photosystem II (PSII) Fv/Fm b) TSS content c) Phenolic compounds content d) Relative quantification of gene expression of candidate genes selected according to multivariate and integrative analysis. All the procedures were measured on heat primed and non-primed plants of experimental phase II. Different letters indicate statistically significate differences ($p < 0.05$).

6.5 Discussion

Basal conditions in seeds and seedlings exhibited transgenerational memory

The physiological analysis of seed endosperms showed a shift between STA and TPC content. Interestingly, there is a direct relation between these two parameters as it is known that phenolic compounds act as inhibitors of starch digestion (Kandil *et al.*, 2012; Zhu, 2015). Concomitantly with this, the FAA raise in non-primed progeny seeds suggested that in these seeds (E progeny) was higher metabolic activity. This matches the weight difference found, since T seeds were lighter, hence probably drier since water content is indispensable for metabolic activity. These effects may then be transmitted to the progeny, making them less keen to germinate and more tolerant to long term perdurance, maybe in order to ensure a prolonged time in a suitable soil environment (as water content) to start the germination process.

Additionally, we found differences in miR160 expression profile, which was higher in E progeny in the seeds while in T progeny in the seedlings. Overexpression of miR160 provokes improved germination in the seeds and better heat stress tolerance in the seedlings in Arabidopsis (Lin *et al.*, 2018). This fact aligns with our hypothesis of difficulties in seed germination and improved seedling heat stress performance in T progeny. Also, miR160 is known to target AUXIN RESPONSE TRANSCRIPTION FACTORS (ARF) (Mallory *et al.*, 2005; Dai *et al.*, 2021; Hao *et al.*, 2022), suggesting a complex interplay in the heritable memory including hormones and highlighting the role of miR160 as a valuable biomarker for tolerance testing.

Despite the differences found in seed weight and composition, seedlings grew to the same height. Therefore, under optimal conditions (before the stress treatment) we could not detect any energy cost associated to the memory retained according to Herman & Sultan, (2011), maybe due to the enhanced FvFm and increased accumulation of photosystems related proteins verified in T progeny.

Hence, it was found that the changes between the more sensitive and more tolerant isogenic subpopulations may be previously detected and predicted by analyzing the seeds. This strategy may represent an advance for early classifying seed primming

status without the requirement of grow the seedlings, representing a cost and time efficient screening method.

6.5.1 miRNAs as the main heritable epigenetic mechanisms in the memory acquisition process

Plant transgenerational memory seems to provide general coping mechanisms to overcome previously known and/or unknown stresses in a more effective way, allowing primed plants to adapt to their environment and constitute populations. The relevance of epigenetic mechanisms such as DNA methylation, histone variants and microRNAs has been determined during the establishment of life-time memory under stress in different plant species (Dubin *et al.*, 2015; Li *et al.*, 2019).

However, there are still some open questions in regard to its heritability. It has been suggested and discussed that this acquired tolerance is rooted epigenetically through methylome for some abiotic stresses (Kou *et al.*, 2011; Wibowo *et al.*, 2016), but DNA methylation studies show no clear arguments in response to heat or drought stress (Ganguly *et al.*, 2017). We focused on the understudied miRNA-based mechanisms, providing evidence of a molecular reprogramming in basal conditions in the dry seeds that is then shifted in the seedlings. These altered paths are known to control gene expression and according to be plausible players and subjects of the memory acquisition related processes (Locato *et al.*, 2018).

6.5.2 Life-time thermomemory seems to be rooted epigenetically, by DNA methylation and thermolabile histones

DNA hypermethylation levels in long-term treatments and recovered plants, whose adaptation to heat stress had started, also correlated with the high expression level of *SAM synthase* quantified in primed plants. Moreover, according to bibliography, stress-induced demethylation has been found to relax chromatin structure, thereby allowing enhanced transcription and proteasomal rearrangement (Shilatifard, 2006; Santos *et al.*, 2011), which has been linked to heat-tolerant genotypes in other plant species (Gao *et al.*, 2014). This behaviour remarks the pivotal role of epigenetics.

Changes in gene expression and alternative splicing in primed and non-primed plants revealed that alternative splicing functions as a novel component of heat shock memory in *Arabidopsis thaliana* (Ling *et al.*, 2018). Similar results were found through this

work where primed seedlings also showed higher levels of *LSM3A* expression and faster evolution in the heat stress response.

Additionally, in the targeted transcriptome analysis those thermolabile histones were found to be upregulated, providing an insight of the importance of nucleosome occupancy and DNA accessibility in non-primed and primed response. However, thermoprimed plants showed a more moderated increase of H2A y H2B expression during the second round of stress confirming this the essential role of nucleosome regulation in priming process

6.5.3 The epigenetics cornerstones seem to be deeply and tightly linked among them

The nucleosome stoichiometric changes that were tracked between the primmed versus non-primmed plants seemed lead to thermomemory acquisition, as stated in other works (Bäurle, 2016; Lämke *et al.*, 2016), driving the cell to a new primed status. This epigenetic memory based in methylation changes (SAHH) and H2A histone variant could be regulated by microRNA (Figure 5b) since their profile of mRNA expression is not sufficient to explain their protein abundance. Those mismatches provide a proof of concept to the wide variety of functions that microRNA can regulate. In this work, we provided several mRNA-microRNA potential pairs and a strong evidence of the relevance of this post-transcriptional gene modulation mechanism, which seems to be relevant to come back to “the new normal” after the first days of stress and to acquire heat stress tolerance and long-term memory. We delved into microRNA-mRNA balance and provided for the first time in this species a set of validated microRNA sequences, a new family of heat-sensitive microRNAs, including miR947 (H2A), and the time-series pattern that these heat-sensitive miRNAs followed during the stress stages.

The newly discovered priming-induced epigenetic memory may represent a general feature of heat stress responses in conifers. Furthermore, this finding could facilitate the development of novel approaches to improve pine survival under extreme heat stress in the current context of climate change.

6.6 References

Adrian-Kalchhauser I, Sultan SE, Shama LNS, Spence-Jones H, Tiso S, Keller Valsecchi CI, Weissing FJ. **2020**. Understanding 'Non-genetic' Inheritance: Insights from Molecular-Evolutionary Crosstalk. *Trends in Ecology & Evolution* **35**: 1078–1089.

Amaral J, Lamelas L, Valledor L, Castillejo MÁ, Alves A, Pinto G. **2021**. Comparative proteomics of Pinus–Fusarium circinatum interactions reveal metabolic clues to biotic stress resistance. *Physiologia Plantarum* **173**: 2142–2154.

Asensi-Fabado MA, Amtmann A, Perrella G. **2017**. Plant responses to abiotic stress: The chromatin context of transcriptional regulation. *Biochimica et Biophysica Acta - Gene Regulatory Mechanisms* **1860**: 106–122.

Bäurle I. **2016**. Plant Heat Adaptation: priming in response to heat stress. *F1000Research* **5**: 694.

Carbó M, Iturra C, Correia B, Colina FJ, Meijón M, Álvarez JM, Cañal MJ, Hasbún R, Pinto G, Valledor L. **2019**. Epigenetics in forest trees: Keep calm and carry on. In: Álvarez-Venegas R, De la Peña C, Casas-Mollano JA, eds. Epigenetics in Plants of Agronomic Importance: Fundamentals and Applications: Transcriptional Regulation and Chromatin Remodelling in Plants: Second Edition. Springer, 381–3.

Dai X, Lu Q, Wang J, Wang L, Xiang F, Liu Z. **2021**. MiR160 and its target genes ARF10, ARF16 and ARF17 modulate hypocotyl elongation in a light, BRZ, or PAC-dependent manner in Arabidopsis: miR160 promotes hypocotyl elongation. *Plant science : an international journal of experimental plant biology* **303**: 110686.

Dai X, Zhuang Z, Zhao PX. **2018**. psRNATarget: a plant small RNA target analysis server (2017 release). *Nucleic Acids Research* **46**: W49–W54.

Dubin MJ, Zhang P, Meng D, Remigereau M-S, Osborne EJ, Paolo Casale F, Drewe P, Kahles A, Jean G, Vilhjálmsson B, *et al.* 2015. DNA methylation in Arabidopsis has a genetic basis and shows evidence of local adaptation. *eLife* **4**: e05255–e05255.

Fang X, Zhao G, Zhang S, Li Y, Gu H, Li Y, Zhao Q, Qi Y. **2019**. Chloroplast-to-Nucleus Signaling Regulates MicroRNA Biogenesis in Arabidopsis. *Developmental Cell* **48**: 371-382.e4.

Ganguly DR, Crisp PA, Eichten SR, Pogson BJ. **2017**. The Arabidopsis DNA Methylome Is Stable under Transgenerational Drought Stress . *Plant Physiology* **175**: 1893–1912.

Gao G, Li J, Li H, Li F, Xu K, Yan G, Chen B, Qiao J, Wu X. **2014**. Comparison of the heat stress induced variations in DNA methylation between heat-tolerant and heat-sensitive rapeseed seedlings. *Breeding science* **64**: 125–133.

García-Campa L, Guerrero S, Lamelas L, Meijón M, Hasbún R, Cañal MJ, Valledor L. **2022**. Chloroplast proteomics reveals transgenerational cross-stress priming in Pinus radiata. *Environmental and Experimental Botany* **202**: 105009.

Guerra D, Crosatti C, Khoshro HH, Mastrangelo AM, Mica E, Mazzucotelli E. **2015**. Post-transcriptional and post-translational regulations of drought and heat response in plants: a spider's web of mechanisms. *Frontiers in Plant Science* **6**: 1–14.

Gutzat R, Mittelsten Scheid O. **2012**. Epigenetic responses to stress: Triple defense? *Current Opinion in Plant Biology* **15**: 568–573.

Hao K, Wang Y, Zhu Z, Wu Y, Chen R, Zhang L. **2022**. miR160: An Indispensable Regulator in Plant . *Frontiers in Plant Science* **13**.

Hellemans J, Mortier G, De Paepe A, Speleman F, Vandesompele J. **2007**. qBase relative quantification framework and software for management and automated analysis of real-time quantitative PCR data. *Genome biology* **8**: R19.

Herman J, Sultan S. **2011**. Adaptive Transgenerational Plasticity in Plants: Case Studies, Mechanisms, and Implications for Natural Populations . *Frontiers in Plant Science* **2**: 102.

Houri-Zeevi L, Teichman G, Gingold H, Rechavi O. **2021**. Stress resets ancestral heritable small RNA responses. *eLife* **10**: 1–31.

Iwasaki M, Paszkowski J. **2014**. Epigenetic memory in plants. *The EMBO Journal* **33**: 1987–1998.

Jesus C, Meijón M, Monteiro P, Correia B, Amaral J, Escandón M, Cañal MJ, Pinto G. **2015**. Salicylic acid application modulates physiological and hormonal changes in Eucalyptus globulus under water deficit. *Environmental and Experimental Botany* **118**: 56–66.

Kandil A, Li J, Vasanthan T, Bressler DC. **2012**. Phenolic Acids in Some Cereal Grains and Their Inhibitory Effect on Starch Liquefaction and Saccharification. *Journal of Agricultural and Food Chemistry* **60**: 8444–8449.

Kou HP, Li Y, Song XX, Ou XF, Xing SC, Ma J, Von Wettstein D, Liu B. **2011**. Heritable alteration in DNA methylation induced by nitrogen-deficiency stress accompanies enhanced tolerance by progenies to the stress in rice (Oryza sativa L.). *Journal of Plant Physiology* **168**: 1685–1693.

Kozomara A, Birgaoanu M, Griffiths-Jones S. **2019**. miRBase: from microRNA sequences to function. *Nucleic Acids Research* **47**: D155–D162.

Lämke J, Brzezinka K, Altmann S, Bäurle I. **2016**. A hit-and-run heat shock factor governs sustained histone methylation and transcriptional stress memory. *The EMBO Journal* **35**: 162–175.

Li P, Yang H, Wang L, Liu H, Huo H, Zhang C, Liu A, Zhu A, Hu J, Lin Y, *et al.* 2019. Physiological and Transcriptome Analyses Reveal Short-Term Responses and Formation of Memory Under Drought Stress in Rice . *Frontiers in Genetics* **10**: 55.

Lin JS, Kuo CC, Yang IC, Tsai WA, Shen YH, Lin CC, Liang YC, Li YC, Kuo YW, King YC, *et al.* 2018. MicroRNA160 modulates plant development and heat shock protein gene expression to mediate heat tolerance in Arabidopsis. *Frontiers in Plant Science* **9**: 1–16.

Ling Y, Serrano N, Gao G, Atia M, Mokhtar M, Woo YH, Bazin J, Veluchamy A, Benhamed M, Crespi M, *et al.* 2018. Thermopriming triggers splicing memory in Arabidopsis. *Journal of Experimental Botany* **69**: 2659–2675.

Liu H, Able AJ, Able JA. **2022a**. Priming crops for the future: rewiring stress memory. *Trends in Plant Science* **27**: 699–716.

Liu X, Quan W, Bartels D. **2022b**. Stress memory responses and seed priming correlate with drought tolerance in plants: an overview. *Planta* **255**: 1–14.

Locato V, Cimini S, De Gara L. **2018**. ROS and redox balance as multifaceted players of cross-tolerance: epigenetic and retrograde control of gene expression. *Journal of Experimental Botany* **69**: 3373–3391.

López-Hidalgo C, Meijón M, Lamelas L, Valledor L. **2021**. The rainbow protocol: A sequential method for quantifying pigments, sugars, free amino acids, phenolics, flavonoids and MDA from a small amount of sample. *Plant, Cell & Environment* **n/a**.

Mallory AC, Bartel DP, Bartel B. **2005**. MicroRNA-directed regulation of Arabidopsis AUXIN RESPONSE FACTOR17 is essential for proper development and modulates expression of early auxin response genes. *The Plant cell* **17**: 1360–1375.

Martinez-medina A, Flors V, Heil M, Mauch-mani B, Pieterse CMJ, Pozo MJ, Ton J, Dam NM Van. **2016**. Recognizing Plant Defense Priming. *Trends in Plant Science* **21**: 2–5.

Meijón M, Feito I, Valledor L, Rodríguez R, Cañal MJ. **2010**. Dynamics of DNA methylation and Histone H4 acetylation during floral bud differentiation in azalea. *BMC Plant Biology* **10**.

Pastor V, Luna E, Mauch-Mani B, Ton J, Flors V. **2013**. Primed plants do not forget. *Environmental and Experimental Botany* **94**: 46–56.

Ravichandran S, Ragupathy R, Edwards T, Domaratzki M, Cloutier S. **2019**. MicroRNA-guided regulation of heat stress response in wheat. *BMC Genomics* **20**: 1–16.

Rodrigues AS, Chaves I, Costa BV, Lin YC, Lopes S, Milhinhos A, Van de Peer Y, Miguel CM. **2019**. Small RNA profiling in Pinus pinaster reveals the transcriptome of developing seeds and highlights differences between zygotic and somatic embryos. *Scientific Reports* **9**: 1–14.

Santos AP, Ferreira L, Maroco J, Oliveira MM. **2011**. Abiotic Stress and Induced DNA Hypomethylation Cause Interphase Chromatin Structural Changes in Rice rDNA Loci. *Cytogenetic and Genome Research* **132**: 297–303.

Schindelin J, Arganda-Carreras I, Frise E, Kaynig V, Longair M, Pietzsch T, Preibisch S, Rueden C, Saalfeld S, Schmid B, *et al.* 2012. Fiji: An open-source platform for biological-image analysis. *Nature Methods* **9**: 676–682.

Shilatifard A. **2006**. Chromatin Modifications by Methylation and Ubiquitination: Implications in the Regulation of Gene Expression. *Annual Review of Biochemistry* **75**: 243–269.

Świda-Barteczka A, Krieger-Liszkay A, Bilger W, Voigt U, Hensel G, Szweykowska-Kulinska Z, Krupinska K. **2018**. The plastid-nucleus located DNA/RNA binding protein WHIRLY1 regulates microRNA-levels during stress in barley (Hordeum vulgare L.). *RNA Biology* **15**: 886–891.

Valledor L, Escandón M, Meijón M, Nukarinen E, Cañal MJ, Weckwerth W. **2014**. A universal protocol for the combined isolation of metabolites, DNA, long RNAs, small RNAs, and proteins from plants and microorganisms. *Plant Journal* **79**: 173–180.

Valledor L, Hasbún R, Meijón M, Rodríguez JL, Santamaría E, Viejo M, Berdasco M, Feito I, Fraga MF, Cañal MJ, *et al.* 2007. Involvement of DNA methylation in tree development and micropropagation. *Plant Cell, Tissue and Organ Culture* **91**: 75–86.

Vriet C, Hennig L, Laloi C. **2015**. Stress-induced chromatin changes in plants: Of memories, metabolites and crop improvement. *Cellular and Molecular Life Sciences* **72**: 1261–1273.

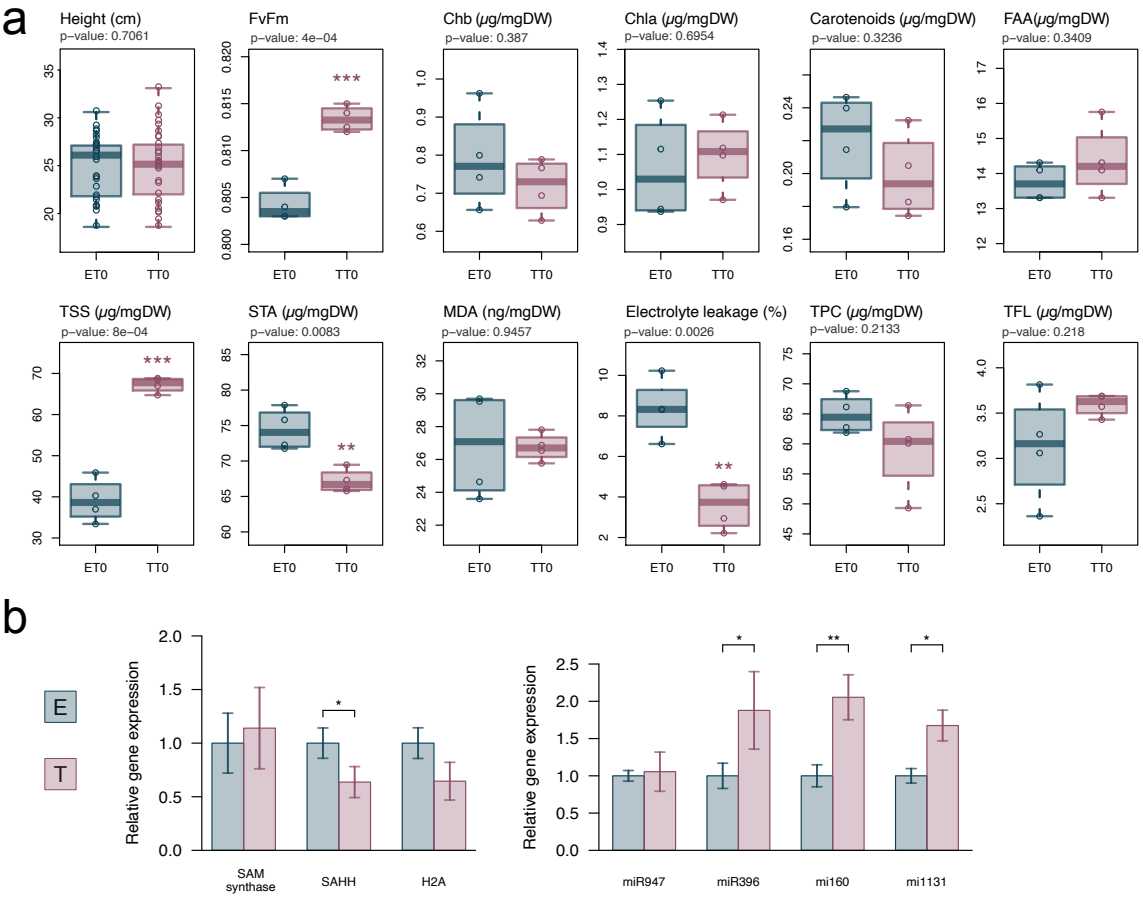
Wibowo A, Becker C, Marconi G, Durr J, Price J, Hagmann J, Papareddy R, Putra H, Kageyama J, Becker J, *et al.* 2016. Hyperosmotic stress memory in Arabidopsis is mediated by distinct epigenetically labile sites in the genome and is restricted in the male germline by DNA glycosylase activity (S McCormick, Ed.). *eLife* **5**: e13546.

Zas R, Cendán C, Sampedro L. **2013**. Mediation of seed provisioning in the transmission of environmental maternal effects in Maritime pine (Pinus pinaster Aiton). *Heredity* **111**: 248–255.

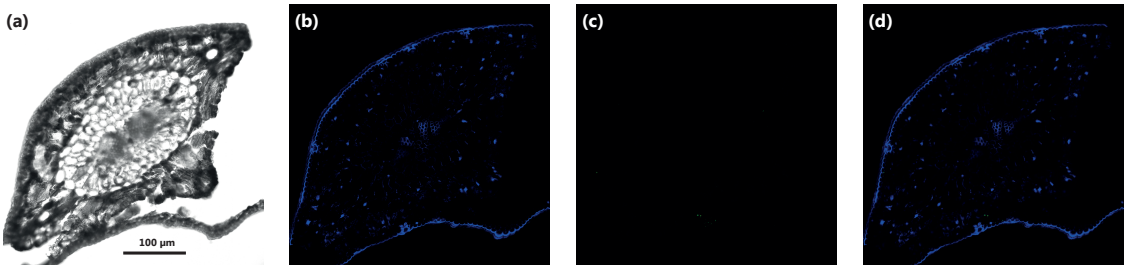
Zhu F. **2015**. Interactions between starch and phenolic compound. *Trends in Food Science & Technology* **43**: 129–143.

6.7 Supplemental information

6.7.1 Supplemental figures



Supplemental Figure S6.1. Seedlings characterization in basal conditions. a)Height, FvFm and, physiology profile including Chlorophyll b, Chlorophyll a, Carotenoids, Free Amino Acids (FAA), Total Soluble Sugars (TSS), starch (STA), malonaldehide (MDA), Electrolyte leakage, Total Flavonoids (TFL) and Total Phenolic compounds (TPC) and b) mRNA and microRNA relative expression of selected candidates. For E and T progenies. Asterisks indicate significance according to Wilcox non-parametric test; * pvalue < 0.05; ** pvalue < 0.01; *** pvalue < 0.001.



Supplemental Figure S6.2. Negative control of 5-mC immunolocalization analysis. a) Differential interference contrast (DIC) in control plants; b) Blue signal of DAPI; c) Green signal of 5-mC monoclonal antibody and d) DAPI and 5-mC merged in transversal needle section.

Supplemental Fig S6.3. Representative pine seedlings from previously primmed and non-primmed plants before and after the heat stress corresponding to SC, ST5, NSC and NST5 sampling points. Apical bud of not primed plants showed a decay after the stress exposure (NST5) along with damaged needle tips and dry needles, while previously stressed plants (ST5) manifested no signs of severe impairment, showing damage only in few needle tips.

6.7.2 Supplemental tables

Supplemental table legends

Supplemental Table S6.1. List of the primers used in this chapter.

Supplemental Table S6.2. miRNA - mRNA pairs in silico identification. miRNA and mRNA target accesion, target description, alignment and inhibition type.

CHAPTER VII. Conclusions

7.1 Conclusions

- pRocessomics, the developed R package for single or multiple omics data analysis coupled to the proposed pipeline enabled the exploration of the experimental data obtained in this thesis, in order to deeply characterize *P. radiata* subcellular heat stress responses.

- Chloroplast proteome in basal and under heat stress conditions allowed to distinguish between isogenic *P. radiata* subpopulations whose parents were exposed to different environments, providing a molecular proof of heritable memory and adaptation mechanisms that are able to avoid the epigenetic resetting.

- Nuclear response to heat stress is in short-term driven by a major DNA methylation loss tracked by methyl cycle enzymes depletion and validated by 5mC immunolocalization; while in the long term heat seemed to be directed by post transcriptional gene silencing since the occurrence of an overaccumulation of AGO1 and an upregulation in microRNAs abundance.

- Acclimation to high temperature required a broad and synchronized remodeling of the subcellular proteomes that was triggered by specific photosynthetic impairment in the chloroplasts provoking disturbances in the redox signals which were transmitted to the nucleus in order to reprogram the regulation of the transcription via splicing and ribosomal rearrangement leading to a new homeostatic state.

- Intragenerational memory *'learned'* during a first stress exposure endured long-term and enhanced the performance of *P. radiata* seedlings during a second exposure via changes in the methylation enzymes expression and histone thermolabile isoform upregulation.

- Transgenerational memory shaped the transcriptional regulation through several interrelated epigenetic mechanisms as microRNAs and DNA methylation, that could be already assessed in the seeds through microRNA160 and, SAM SYNTHASE and SAHH expression levels.

- The combination of the subcellular proteomics approach together with developed analysis tool and the targeted transcriptomics employed along this thesis, permitted the definition of a reliable biomarker panel for the selection of heat primmed plants. The accuracy of this approach was endorsed by the identification of SAM SYNTHASE, one of the most promising biomarker proved to be a main player in trans- and intra- generational memory acquisition as well as a driver of the stress response.

CHAPTER VIII. Resumen

8.1 Introducción

El cambio climático ha provocado un aumento en la frecuencia y severidad de los llamados eventos climatológicos extremos, entre ellos, destacan las olas de calor, que se definen como días especialmente cálidos en los que se supera para lugar determinado la temperatura habitual. El estrés por altas temperaturas es uno de los más perjudiciales para las plantas, limitando tanto su crecimiento como su producción, por lo que la respuesta de las plantas al estrés térmico por alta temperatura es considerado un tema de gran relevancia.

Afortunadamente, y debido a su naturaleza sésil, las plantas han desarrollado mecanismos moleculares de un alto grado de sofisticación y complejidad, a menudo redundantes entre sí, para hacer frente a los diversos estreses o eventos climáticos a los que se enfrentan a lo largo de su ciclo de vida.

A pesar de lo ampliamente estudiado del tema, la respuesta al estrés por alta temperatura en plantas es muy compleja y variada, incluyendo varios niveles de organización diferentes y pobremente comprendida en especies forestales, cuyos ciclos de vida son largos y puede que por sus diferencias con otras plantas no compartan los mismos componentes en la respuesta y especialmente en el proceso de adquisición de memoria.

Los bosques son esenciales en los ecosistemas, proporcionando alimentos y materias primas para las industrias. Para satisfacer las demandas actuales y futuras de estos productos, las especies de rápido crecimiento, como *Pinus radiata* D. Don, son una clara opción para aumentar la productividad de las plantaciones y la sostenibilidad a largo plazo. Sin embargo, varios informes recientes han descrito que los efectos potenciales del cambio climático antropogénico podrían poner en entredicho la viabilidad comercial de las plantaciones en varios países. Se han logrado progresos significativos en la elucidación de los mecanismos fisiológicos y moleculares subyacentes a esta respuesta térmica, aunque en especies leñosas, su respuesta y adaptación está poco estudiada.

Los árboles han desarrollado sofisticados mecanismos de percepción del estrés y transducción de señales. En respuesta a estos estreses, se activan multitud de procesos que permiten a las plantas hacer frente a la nueva situación. Estos incluyen la activación de procesos para la protección frente al daño oxidativo, cambios en la fluidez de la membrana, estimulación del metabolismo secundario, activación de la señalización (hormonal, ROS, lipídica…), la alteración de la expresión de los genes relacionados con estrés y la producción de proteínas de estrés, entre muchos otros.

Otro factor crucial y de los más ampliamente estudiados en la respuesta al calor es la inducción de proteínas de choque térmico (HSPs). Las HSP son una familia de proteínas muy conservada evolutivamente, que se inducen en todos los organismos en respuesta a estrés ambiental y también en procesos de desarrollo. En plantas, las HSP pueden clasificarse en cinco grupos en base a su masa molecular: HSP20 que incluye las más pequeñas entre 16-42 kDa (también conocida como sHSPs), la familia HSP60, HSP70, HSP90 y HSP100. Estas familias de HSP desempeñan una función esencial como chaperonas que controlan el plegamiento y desplegamiento de la estructura terciara de las proteínas. Las HSP ayudan a la célula a conservar las proteínas desnaturalizadas, uniéndose a ellas para evitar que se agreguen, o manteniéndolas desplegadas en un estado semi-funcional, para que una vez que el estrés haya cedido, puedan volver al plegamiento inicial y recuperar por lo tanto su función.

8.2 Planteamiento y objetivos

El actual marco del rápido cambio climático subraya la urgencia de generar conocimiento aplicable que tenga un alcance global, y que contraste y describa la capacidad de adaptación de los organismos y ecosistemas frente a las olas de calor. Para poder comprender la magnitud, el alcance y la vulnerabilidad de las plantas, se deben realizar los ensayos imitando las olas de calor de una forma estandarizada y controlada. Es por ello de suma importancia establecer un sistema experimental que tenga en cuenta tanto las condiciones climáticas futuras (como las olas de calor) como los actores más relevantes en el desencadenamiento de la respuesta y en el establecimiento de la posible memoria consecuencia de una o varias exposiciones al estrés térmico en una especie de ciclo de vida largo como *Pinus radiata*.

Por ello, en esta tesis se ha establecido un tratamiento que imita una ola de calor aplicándose este estrés de la misma manera en todos los experimentos realizados. Dicha exposición consistió en al menos 5 días con una temperatura máxima de 45 ºC mantenida durante 6 horas al día, y representa un tratamiento de estrés térmico lo suficientemente severo como para permitir la caracterización del respuesta al estrés por calor.

Para estudiar más a fondo estos mecanismos, se ha empleado un enfoque de proteómica no dirigida a nivel nuclear y cloroplástico, ambos subproteomas se demostraron útiles debido a la importancia central de los cloroplastos como transducción de señales de calor y la relevancia de los núcleos en la orquestación de respuestas y los procesos de adquisición de memoria. Además, debido a los múltiples componentes y tipos de memoria molecular, es de gran importancia enfocar el análisis en una población de *Pinus radiata* sin un componente preexistente de primado al calor u otros estreses, que pueda introducir ruido y producir resultados irreproducibles en poblaciones no primadas.

En conclusión, el objetivo principal de esta tesis es identificar los mecanismos de adquisición de memoria y señalización de respuesta a alta temperatura más relevantes en *Pinus radiata* empleando un enfoque de proteómica subcelular y usando como diseño experimental un escenario realista de aumento de temperatura alta para proporcionar un conjunto de indicadores confiables. y biomarcadores útiles para la selección temprana de árboles y semillas termotolerantes o primados. Para lograr este objetivo principal, se definió el siguiente conjunto de objetivos parciales:

6 Desarrollo de un protocolo universal para el análisis e integración de datos (prote)ómicos (Capítulo 2).

7 Caracterización del proteoma del cloroplasto en dos poblaciones isogénicas silvestres cultivadas en diferentes ambientes en condiciones óptimas y bajo ola de calor controlada. Exploración de la memoria transgeneracional a través del cloroplasto. (Capítulo 3).

8 Caracterización del proteoma nuclear en respuesta al estrés por calor, antes, durante y después de la exposición al estrés por calor en una población de *P. radiata* no cebada. (Capítulo 4)

9 Integración y evaluación de la sincronización de proteomas nucleares y cloroplásticos en respuesta al estrés calórico previo y durante una ola de calor controlada. (Capítulo 5)

10 Identificación de un panel de biomarcadores de termotolerancia y termomemoria en semillas y plántulas en respuesta a altas temperaturas en Pinus radiata (Capítulo 6).

8.3 Resultados y discusión

El sistema experimental de esta tesis fue diseñado para caracterizar la respuesta subcelular de *Pinus radiata* al estrés térmico por altas temperaturas, incluyendo el uso de técnicas proteómicas, que a pesar de su amplia aplicabilidad y crecimiento en los últimos años, aún presentan algunas barreras de entrada. Probablemente, la más frecuente sea la dificultad de analizar los conjuntos de datos generados, cada vez mayores, lo que representa un gran obstáculo. Por ello y con el objetivo de poder analizar los datos experimentales de esta Tesis, se ha desarrollado un paquete de R, pRocessomics, que puede utilizarse para preprocesar, analizar (incluyendo análisis estadísticos uni y multivariante) así como integrar conjuntos de datos ómicos.

Además, teniendo en cuenta el enfoque subcelular, también se ha desarrollado un protocolo de purificación de núcleos y cloroplastos (Apéndice I) compatible con espectrometría de masas, que ha sido necesario para la elaboración de esta Tesis.

El análisis del perfil fisiológico y proteoma del cloroplasto de las progenie de dos subpoblaciones clonales de *Pinus radiata* que provenían de diferentes entornos, permitió distinguir las diferentes estrategias adoptadas por las dos subpoblaciones. Una de las diferencias más representativas del análisis bioquímico fue el aumento y el mantenimiento de la concentración de azúcares solubles en las acículas de la progenie primada tras cinco días de tratamiento de estrés por alta temperatura. Los azúcares solubles son osmolitos, moléculas de señalización y también desempeñan un papel esencial en la protección de la estabilidad de la membranas frente al estrés térmico. Además, previenen los daños causados en estreses abióticos como la sequía o la salinidad y, en algunos casos, estabilizan las biomoléculas.

Además, se encontró que la integridad de la membranas se mantenía más estable en la progenie de la subpoblación primada, este hecho junto con el aumento del contenido de clorofila a largo plazo, puede indicar un mejor estado de salud de las plantas de la progenie primada al final del experimento, lo que sugiere una mayor tolerancia.

A pesar de observar en ambas progenies un aumento de las proteínas choque térmico en ambos conjuntos de plantas, se encontraron diferencias en las subfamilias más acumuladas tras cinco días de estrés, en las que en la progenie de la subpoblación primada predominaban las HSP90 y en la progenie no primada las sHSP. Las sHPS son

las encargadas de prevenir la agregación proteica mientras que las HSP90 intervienen en el proceso de volver a plegar las proteínas desnaturalizadas y consumen ATP. Por ello, este resultado también apunta a una mejor respuesta de la progenie primada avalando la hipótesis de la memoria transgeneracional heredada.

Dicha memoria transgeneracional parece proporcionar estrategias y mecanismos para superar estreses previamente conocidos y/o desconocidos de una manera más eficaz, lo que permite a las plantas preparadas adaptarse a su entorno y constituir poblaciones. Se ha determinado la relevancia de mecanismos epigenéticos como la metilación del ADN, variantes de histonas y microARNs durante el establecimiento de la memoria transgeneracional bajo estrés en diferentes especies vegetales. Sin embargo, aún quedan algunas cuestiones abiertas en cuanto a su heredabilidad.

A pesar de la relevancia potencial de los mecanismos epigenéticos en la memoria transgeneracional, los datos obtenidos mostraron que, especialmente durante el estrés, esta memoria retenida se puede rastrear a través de los cloroplastos, donde observamos los rastros de una señalización retrógrada alterada a través de ROS y genes *GUN*, junto con un refuerzo de la proteína fotosistema II, la sobreacumulación de las familias HSP90 y HSP60 y el aumento del contenido de clorofila en las condiciones de estrés. Los resultados también indicaron que las plantas de la progenie primada eran más proclives a recuperar la señalización basal cloroplasto-núcleo y el plegamiento de proteínas, lo que conducía a una mayor tolerancia cruzada al estrés. Más adelante, se comprobó que los cambios entre las subpoblaciones isogénicas más sensibles y más tolerantes pueden detectarse y predecirse previamente mediante el análisis de las semillas. Esta estrategia puede representar un avance en la clasificación del estado de imprimación de las semillas sin necesidad de cultivar las plántulas de una manera eficiente en cuanto a costes y tiempo.

Para poder estudiar la memoria intrageneracional, se seleccionó la subpoblación no primada y se realizó el análisis del proteoma nuclear, durante el estrés y después de la recuperación. Mostrando por primera vez la dinámica del proteoma nuclear relacionada con la respuesta al estrés térmico y el proceso de primado térmico a lo largo del ciclo de vida de la planta. La profundidad que se consiguió con el enfoque subcelular, permitió la descripción de este proceso, revelando varias familias cruciales de proteínas implicadas en diferentes pasos clave de regulación, como la reorganización del proteasoma, las funciones asociadas al ARN, la regulación génica impulsada por la epigenómica y factores de transcripción específicos previamente desconectados del estrés térmico y asociados

a los cambios lumínicos. Además, la histona H2A, el splicing alternativo y las enzimas del ciclo de la metilación parecen estar directamente relacionados con la inducción del primado térmico. En último término, estas remodelaciones activas del transcriptoma y el proteoma detectadas desencadenan los procesos cruciales implicados en la respuesta y adaptación a altas temperaturas. La memoria epigenética inducida por el primado puede representar una característica general de las respuestas al estrés térmico en coníferas. Además, este hallazgo podría facilitar el desarrollo de nuevos enfoques para mejorar la supervivencia de los pinos en condiciones de estrés térmico severo en el contexto actual de cambio climático.

8.4 Conclusiones

- pRocessomics, el paquete de R desarrollado para el análisis de datos de una o multiples capas ómicas unido a la esquema propuesto permitió la exploración de los datos experimentales obtenidos en esta tesis, para caracterizar en profundidad las respuestas subcelulares de *Pinus radiata* frente al estrés térmico.

- El proteoma del cloroplasto en condiciones basales y de estrés térmico permitió distinguir entre dos poblaciones isogénicas de *Pinus radiata* cuyos padres fueron expuestos a diferentes entornos, proveyendo una prueba molecular de memoria heredable y mecanismos de adaptación que son capaces de evitar el reseteado epigenético.

- La respuesta nuclear al estrés térmico está dirigida a corto plazo por una importante pérdida de metilación del ADN monitorizada por la disminución de las enzimas de ciclo de la metilación y validada por la inmunolocalización de 5mC; mientras que a calor a largo plazo parecía estar dirigida por le silenciamiento génico postranscripcional debido a la sobreacumulación de AGO1 y la sobrerregulación en la abundancia de microARNs.

- La aclimatación a las altas temperaturas requirió una remodelación amplia y sincronizada de los proteomas subcelulares que fue desencadenada por un deterioro fotosintético específico en los cloroplastos provocando alteraciones en las señales redox que se transmitían al núcleo para reprogramar la regulación de la transcripción vía splicing y reordenamiento ribosomal que condujo a un nuevo estado homeostático .

- La memoria intrageneracional 'aprendida' durante una primera exposición al estrés perduró a largo plazo y mejoró el rendimiento de las plántulas de *P. radiata* durante una segunda exposición a través de cambios en la expresión de enzimas de metilación y la regulación positiva de isoformas termolábiles de histonas

- La memoria transgeneracional dio forma a la regulación transcripcional a través de varios mecanismos epigenéticos interrelacionados como los microARNs y la metilación del ADN, que pudo evaluarse en las semillas a través de los niveles de expresión de microARN160 y SAM SINTASA y SAHH..

- La combinación del enfoque de la proteómica subcelular junto con la herramienta de análisis desarrollada y la transcriptómica dirigida empleada a lo largo de esta tesis, permitió la definición de un panel de biomarcadores certero para la selección de plantas primadas con calor. La precisión de este enfoque fue respaldada por la identificación de SAM SINTASA, uno de los biomarcadores más prometedores que demostró ser un actor principal en la adquisición de memoria trans e intrageneracional, así como un impulsor de la respuesta al estrés.