OXFORD UNIVERSITY PRESS | Molecular Biology and Evolution

Heat stress memory in *Pinus* requires divergent splicing in relation to other Spermatophyta

Journal:	*Molecular Biology and Evolution*
Manuscript ID	Draft
Manuscript Type:	Article
Date Submitted by the Author:	n/a
Complete List of Authors:	Fernández Roces, Víctor; University of Oviedo, Organisms and Systems Biology Lamelas, Laura; University of Oviedo, Department of Organisms and Systems Biology Valledor, Luis; University of Oviedo, Organisms and Systems Biology Carbo, María; University of Oviedo, Organisms and Systems Biology Cañal, María; University of Oviedo, Organisms and Systems Biology Meijón, Mónica; University of Oviedo, Department of Organisms and Systems Biology
Key Words:	long-term memory, splicing, acclimation, heat stress, Pinus, integrative approach

Article: Discoveries

Heat stress memory in *Pinus* requires divergent splicing in relation to other Spermatophyta

Víctor Roces[1], Laura Lamelas[1], María Carbó[1], Luis Valledor[1], María Jesús Cañal[1], Mónica Meijón[1*]

[1] Plant Physiology, Department of Organisms and Systems Biology, Faculty of Biology and Biotechnology Institute of Asturias, University of Oviedo, Asturias, Spain

[*] Correspondence: meijonmonica@uniovi.es

Short title: Long term heat splicing-memory in Pinus

Abstract

Current scenario of climate change has led to increase number of studies describing main drivers in abiotic stress. Recent findings suggest that temperature-responsive alternative splicing (AS) has a critical role in controlling plant response to high temperature at the molecular level. AS is a mechanism that allows organisms to create an assortment of RNA transcripts and proteins using single gene information. However, some of the most important insights suggested about stress AS could not be rigorously addressed because research has been focused on study model species which only covered a narrow phylogenetic and life-cycle spectrum. Thus, AS degree of diversification among more dissimilar taxa in heat response is still largely unknown. To fill this gap, the present study examines how AS landscape responds and 'remembers' from heat stress in the conifer *Pinus radiata,* a group which have received little attention even though their position can solve key evolutive and acclimation questions. Contrary to angiosperms, we found that potential intron retention may not be the most prevalent type of AS. Furthermore, our integrative analysis with metabolome and proteome data places splicing as the main source of variation during the response. Finally, we validated acquired long-term splicing-memory in a diverse subset of events and, although AS dynamics are divergent, splicing-memory seems to be conserved in seed plants. Our discoveries reveal the particular way of remembering past long-term temperature changes in conifers and open the door to include species with unique features to determine the extent of conservation in gene expression regulation.

Keywords: long-term memory, splicing, acclimation, heat stress, *Pinus,* integrative approach

Introduction

Plants high sensitivity to environment fluctuations together with the current scenario of climate change has led to an increasing number of studies describing main drivers implied in abiotic stress adaptation and acclimation. Recent findings suggest that temperature-responsive alternative splicing (AS) has a critical role in controlling plant response to high temperature at the molecular level (Capovilla et al. 2018). AS is a mechanism that allows organisms to create an assortment of RNA transcripts and proteins by using information from a single gene and functions in a wide range of physiological processes, such as growth and development (Smith et al. 2018; Huang et al. 2020). During the last decade, analysis of high-throughput RNA sequencing (RNA-seq) data facilitated the characterization of processing patterns even in non-sequenced species. Many of these works have contributed to the establishment of stress RNA biology general insights like intron retention (IR) as the most prevalent event in plants (Laloum et al. 2018). Despite IR lacks of functional impact, mainly because of intron high frequency of premature stop codons (PTCs) which disrupts protein translation through mRNA non-sense mediated decay (NMD); several systems have been proposed linking this AS-type with the acquisition and maintenance of stress tolerance like thermomemory, buffer to reduce metabolic cost of translation, NMD resistance by post-transcriptional splicing of introns and exitron-driven diversification of protein substrate specificity and enhanced regulatory capacity (Ling et al. 2018; Chaudhary et al. 2019; Jia et al. 2020) . These systems have an incredible potential for creating solutions to improve plant resources management and breeding programs. However, splicing interplay with other regulatory layers and its degree of diversification among taxa in a process as complex as heat response is still largely unknown.

Evolution and acclimation are continuous processes requiring coordinated changes in many traits according to environmental selective pressures. Considerable adaptation and acclimation may be mediated by modifications in AS, although, the patterns of gene-expression and splicing variation in response to abiotic stress are not conserved across species or even between genotypes within the same species (Smith et al. 2018; Hanemian et al. 2020; Meng et al. 2021). These reports highlight the importance of understanding degrees of splicing divergence among more dissimilar groups. Currently seed plants are represented by five lineages: one corresponding to the species-rich angiosperms and four to gymnosperms, among which conifers are the most widely distributed group (Wan et al. 2018). Conifers have received little attention even though their phylogenetic position can solve several key evolutive and environmental acclimation questions. Pinus, the largest genus of conifers and, arguably, the most

important genus of trees in the world; provides an ideal example to explore splicing differentiation due to its long evolutionary history and potentially unique genomic features (De La Torre et al. 2014; De La Torre et al. 2020; Jin et al. 2021).

Some of the most important general insights previously suggested could not be rigorously addressed because research has been mostly focused on the study of model species. These organisms have been crucial for in-depth study of cellular and molecular machinery but these findings might be far from being generalized from an ecological and evolutionary point of view due to the narrow phylogenetic and life-cycle spectrum covered by model species (Valledor et al. 2018). In contrast to this narrow but deep knowledge bottleneck, a broad but shallow approach could be adopted using orphan species to not only directly ask big questions but also give simple large-scale answers to derive ground truths that we would never have contemplated asking based on our molecular biology understanding of well-known organisms (Kliebenstein 2019). In the present study, and considering this knowledge trade-off, we comprehensively characterized high temperature response in the conifer *P. radiata* and described long-term splicing-memory and wide stress-responsive AS patterns in gymnosperms. Furthermore, the integration of splicing data with other layers revealed important sources of additional regulatory variation between distinct response stages.

Results

Heat-responsive differential splicing and/or differential expression characterization

Splicing and expression changes induced by high temperature were analyzed from RNA-seq data of *P. radiata* needles subjected to 40 ºC stress pulses (**Figure 1**; see Material and Methods). *De novo* calling of splicing events were performed using the KisSplice pipeline (Sacomoto et al. 2012), enabling the description of splicing landscape without using a reference genome which is a major bottle-neck in the study of non-model organisms like *P. radiata*. For each AS event (represented by a pair of isoforms) differential splicing (DS) was first analyzed using KissDE algorithm. We found that out of the 13980 isoforms, 5105 were differentially spliced (FDR-adjusted $P < 0.05$ in at least one contrast; **Figure 2A**). Secondly, differential expression (DE) using DESeq2 algorithm (Love et al. 2014) was conducted for each isoform. We detected 1568 DE isoforms (FDR-adjusted $P < 0.05$ in at least one contrast), of which 944 were also DS (regulated by both transcription and splicing; double differential (DD)) and 624 were not DS (only regulated by transcription). Consequently, 5729 isoforms exhibited significantly altered splicing or expression patterns (**Figure 2A**).

In order to assess the specificity of heat induced changes we next explored the differences between all the possible contrasts by comparing which isoforms were DE and DS. Around 22 % (1139) and 13 % (211) of DS and DE isoforms, respectively, were common to both treatment days. Major changes in AS and isoform-level expression occurred during the initial shock and throughout the heat period (**Figure 2B**). These dynamics reflected crucial aspects of the stress response such as high temperature perception, initial response (day 1 stress; T1) and acclimation (day 3 stress; T3). It is important to notice that isoforms, depending on their regulation level (DE, DS and DD), presented diverse patterns in functional categories previously described as important in stress response (**Figure 2C**) (Escandón et al. 2017; Lamelas et al. 2020). Thus, this functional discordance indicates that the role of isoforms and the variation produced by splicing may change according to their regulation or behavior.

AS can change transcripts structure in different ways and have multiple molecular consequences (Laloum et al. 2018). Thus, to verify that splicing variation effect was a driver on regulation and behavior, we investigated transcript models using CodAn (Nachtigall et al. 2020). This analysis was restricted to full models to ensure only one possible reading frame. Of all isoforms, 36.5 % (5067) AS events occurred in CDS, among which 65 % (3297) were predicted to introduce premature termination codons (PTCs) probably leading to non-sense-mediated transcript decay (NMD), while the remaining 35 % (1817) could make protein sequence changes. Alternatively, 39.8 % (5530) appeared in UTR regions, which might produce translation efficiency alterations and 23.4 % (3246) could not be classified because neither full model nor full-transcript sequences were predicted or mapped (**Figure 2D**). Furthermore, changes in this proportion were tested among regulation levels and non-supervised behavior clusters using hypergeometric tests ($P < 0.05$). Control vs stress DS and DD isoforms had a significant increase in 3'UTR AS events, having DS isoforms enrichments in CDS AS events as well (**Figure 2D**). At the behavior level, two clear non-mutually exclusive trends were found: 1) 3'UTR AS events are enriched in clusters where isoforms presented lower expression in T3 than C (2, 4, 5, 8, 11 and 12) and 2) 5'UTR AS events are enriched in clusters in which isoforms had higher expression in T3 than T1 (1, 3, 4, 6, 9 and 15) (**Supplementary Figure 2B**; **Supplementary Figure 3A**). In most cases, CDS AS events enrichments were due to an increase in the proportion of PTCs (**Figure 2D**; **Supplementary Figure 3B**). Hence, we propose that isoforms behavior and regulation are substantially influenced by the variation produced by splicing.

Intron morphology comparative analysis gives clues into intron retention splicing

Previous studies of stress-induced AS in seed plants are consistent with a higher intron retention (IR) (Laloum et al. 2018). Consequently, potential introns were sought in the variable sequences introduced by AS. As the lack of reference genome forced us to employ a mapping-free approach, variable sequences longer than 50 nucleotides (nt) and flanked by the conserved di-nt GT-AG were considered as potential introns. In contrast to aforementioned studies, we observed that only ~8 % of AS events could be IR.

We next tested whether this low potential IR percentage could be an artifact of our analysis (biases in flanking di-nts and presence of minor U12 introns which are surrounded by non-canonical sites) by comparing intron morphologies between model species (mainly angiosperms) and gymnosperms. *P. taeda* was selected as a proxy for *P. radiata* since it is the closest species with largest scaffolds and better high-quality gene models. In both U2- and U12-type introns, the most predominant flanking sites were those used in our analysis (**Figure 3**). Interestingly and in agreement with other studies (Wan et al. 2018), we noticed an increase in intron length effect size between angiosperms and gymnosperms. This trend remained constant even in *G. montanum* despite its small introns/genome. Unlike other model-species, our findings suggest a low IR percentage in *P. radiata* and places intron size as a possible key feature to understand splicing divergence between the two biggest groups of seed plants (**Supplementary Figure 6**).

Biological functions regulated at splicing and transcription levels

To investigate general heat-induced dynamics, we executed a heatmap/hierarchical clustering of total isoform expression levels according to Mercator4 functional annotation (**Figure 4A**). Four main expression profiles were revealed: 1) Isoforms whose expression decreased during heat stress reflected homeostasis loss of control related to cell wall-, chromatin- and cell cycle- organization, solute transport, nutrient uptake, amino acid metabolism and vesicle trafficking. 2) Expression increased during the treatments was associated to polyamine metabolism and external stimuli response that illustrated heat perception. 3) Expression peaked at T1 displayed stress damage mostly linked to protein homeostasis. 4) Isoforms that presented most of their expression focused at T3 were connected to DNA damage response, photosynthesis, secondary metabolism, phytohormone action and RNA-processing/biosynthesis functions and showed response or acclimation.

Further, to identify genes displaying large magnitude changes and discover biological functions regulated by splicing and transcription, volcano analysis and Mercator4 annotation enrichment were performed on isoforms of DD events, respectively (**Figure 4B and C; Supplementary Figure 4A and B**). In line with the previously described modulation, we observed enrichments in RNA-processing/phytohormone action and protein homeostasis at T3 and T1, respectively. According to the first isoform expression profile we also detected significative alterations in homeostasis-like categories for both stress vs control comparisons (**Figure 4C**). Stress-specific contrast (T3–T1) exhibited enrichments in secondary metabolism, phytohormone action and RNA biosynthesis congruent with the acclimation/response profile (**Supplementary Figure 4B**). These insights were supported by some of the most relevant isoforms for each comparison, such as protein homeostasis *HEAT SHOCK PROTEIN 21, CHLOROPLASTIC (HSP21)*, phytohormone action *ABCISIC STRESS-RIPENING PROTEIN 5 (ASR5),* cell wall organisation *GLYCINE-RICH CELL WALL PROTEIN (EMB31)* and secondary metabolism *FLAVONOID 3',5'-HYDROXYLASE (CYP75A2)* (**Figure 4B**; **Supplementary Figure 4A**).

Although various terms converged between general dynamics and coupled splicing/transcription modulation, other enriched categories largely differed. Thus, while general RNA-processing isoform expression changes were not highlighted at T1 (**Figure 4A**), on DD level we saw a significative enrichment in this term ratified by *MULTIPLE ORGANELLAR RNA EDITING FACTOR 3, MITOCHONDRIAL (MORF3)* relevance (**Figure 4B and C**). General photosynthesis-associated expression increased at T3 (**Figure 4A**) supported by *INNER MEMBRANE PROTEIN ALBINO 3, CHLOROPLASTIC (ALB3)* importance in stress vs control volcano analysis (**Figure 4B**). Notwithstanding the above, DD functional analysis revealed greater expression of isoforms related to this category in control plants (**Figure 4C**). Similarly, solute transport also appeared as inconsistent between general dynamics and DD regulation in stress-specific contrast (**Supplementary Figure 4A and B**). The emergence of several isoforms such as *PROBABLE AQUAPORIN PIP1-4 (PIP1.4)* and *REF/SRPP-LIKE PROTEIN At3G05500 (At3G05500)* pointed to a possible water deprivation during the heat assay (**Figure 4B and Supplementary Figure 4A**). Collectively, these results prove a wide range of stress-responsive functional categories and elucidate the most important pathways involved in heat response because of global and DD modulation discrepancies.

Relative contribution of multiple regulatory layers to heat stress response

Heat is a critical and extensively studied abiotic stress and increasing literature supports lots of single –omics descriptions, yet the molecular determinants underlying high temperature responses are still not well understood for a lack of an holistic point of view. To fill this gap, we adopted a systems biology approach and expanded our splicing analysis with metabolome and proteome data previously generated by our group (**Figure 5A**) (Escandón et al. 2017). To further research the relationships between all three regulatory layers during heat response we next applied multiomics factor analysis (MOFA2) (Argelaguet et al. 2018), a statistical framework for unsupervised identification of principal sources of variation from multiple data types. We sought to identify distinct response stages based on one or more underlying molecular levels. By querying the latent factor scores and loadings calculated by MOFA2, the variation explained by each latent factor could be annotated and deciphered as reported below.

MOFA inferred four latent factors (LF1, LF2, LF3 and LF4) with common and unique contributions from each molecular level (**Figure 5B**). Overall, isoforms represented the majority of total variance (R^2=84 %), followed by metabolites (R^2=62 %) and proteins (R^2=43 %). We further inspected the top two factors (LF1 and LF2) sorted by variance explained, which defined control vs stress and stress intensities differences, respectively (**Figure 5C**).

We performed in-depth characterization of LF1, which allowed to distinguish between control and stress conditions and explained 47 % isoform-, 35 % protein- and 38 % metabolite- variance. Despite the greater splicing variance explained, top scoring features associated to LF1 were predominantly proteins (**Figure 5D**). To investigate if these LF1 top scoring features were enriched for functional terms, we performed Mercator4 enrichment analysis for isoform and protein layers using as background all input features (FDR adjusted P < 0.05). We tested for both positive and negative enrichments which means categories linked to samples with positive or negative factor values (**Figure 5E**; **Supplementary Figure 5A**). Interestingly, we observed some common terms with global dynamics and DD regulation (RNA-processing, solute transport, nutrient uptake, photosynthesis and cell-wall organization) but also new ones only revealed by the integration such as lipid metabolism at the isoform level (**Figure 5E**) and protein translocation at the protein level (**Supplementary Figure 5A**).

On the other hand, LF2 enabled all treatments discrimination and explained 21 % isoform-, 7 % protein- and 23 % metabolite- variance. In this case, top scoring features were principally metabolites, which is consistent with the variation explained (**Figure 5D**).

At the isoform level, the functional profile displayed was a mix between heat response (lipid metabolism and protein homeostasis) and acclimation (phytohormone action, solute transport, nutrient uptake, RNA biosynthesis, DNA damage response and lipid metabolism) terms. Additionally, we detected an enrichment in secondary metabolism functions which confirmed the prior role of metabolism layer in this latent factor (**Figure 5E**). Proteins related to response and acclimation were enriched in enzymes/protein translocation and lipid metabolism, respectively (**Supplementary Figure 5A**). Intriguingly, despite both enrichments of lipid metabolism, in isoforms it was related with heat response (positive) while in proteins it was linked to acclimation (negative).

It is worth mentioning that, although top absolute loading features above mentioned, all regulatory layers had strong biomarkers for each biologically relevant latent factor as shown in **Supplementary Figure 5B**. Our findings put splicing as a crucial regulatory layer. These analyses underline the need for an integrative approach since heat response more generally depends on the shared contribution of multiple molecular levels, each one may having greater relative contribution depending on the response stage. To try to answer the common belief that alternative splicing contributes to proteome diversity, we analyzed differential proteins and isoforms with volcanos (**Supplementary Figure 4C**). Only stress vs control comparisons showed medium, high or very high isoform-protein relations. These few direct or indirect isoform-protein relations could not be associated to variation produced by splicing and no relation was found between different proteins translated from same splicing event isoforms. Proteins projection over isoforms confirmed these results because no overlap was seen between both molecular levels (**Supplementary Figure 4D**).

Meta-network analysis reveals isoform co-function modules

In view of splicing importance, we focused on DD isoforms interactions and performed regulatory network analysis to identify co-function modules and their master regulators. We employed a “wisdom of crowds” approach using Seidr toolkit (Schiffthaler et al. 2019) to avoid method-biases towards certain types of biological interactions (Marbach et al. 2012). The resulting weighted undirected meta-network was mainly constituted by five highly-connected modules (**Figure 6A**) An interactive version of the meta-network can be accessed at https://rocesv.github.io/network/index.html.

Module nodes were ranked by Katz centrality to search the most regulatory isoforms. Interestingly, we found well-known regulators such as *SPLICING FACTOR U2AF SMALL SUBUNIT B (U2AF35B)* (module 2) (Laloum et al. 2018); isoforms earlier proposed in volcano and integrative analysis (**Figure 4B**; **Figure 5E**) like *ALB3* (module

1), *METALLOTHIONEIN-LIKE PROTEIN (EMB30)* (module 1 and 5), *SERINE/ARGININE-RICH SPLICING FACTOR RSZ22 (RSZ22)* (module 2), *CYP75A2* (module 4) and *ASR5* (module 3); and new ones for instance *50S RIBOSOMAL PROTEIN 5, CHLOROPLASTIC (PSRP5)* (module 1). Our analysis also enabled the inspection of the connections between same splicing event isoforms (intra-event), for example, local interaction network of *GENERAL TRANSCRIPTION AND DNA REPAIR FACTOR IIH SUBUNIT TFB5 (TFB5)* and *MAGNESIUM PROPORPHYRIN IX METHYLTRANSFERASE, CHLOROPLASTIC (CHLM*) (**Figure 6B**). Both functions of these isoforms and those involved in connecting them pointed to a nucleus-chloroplast coordination via regulation of transcription and response to heat by detoxification of reactive species (Bateman et al. 2021).

Topology enrichments of the two biggest modules (**Figure 6C**) revealed that both modules manifested different or even opposite functional profiles, specially looking at key response and acclimation terms like RNA-processing, polyamine metabolism and phytohormone action. These enrichments improve our understanding of the possible categories involved in the rewiring of inter-/intra-event interactions.

Stress-responsive AS events participate in the establishment of long-term thermomemory

Finally, in order to validate stress-responsive AS patterns identified and explore the role of splicing in long-term memory two assays were performed and several AS events were checked by RT-PCR (**Figure 1**). Eight events were selected based on the following criteria: differential trends, integrative analysis relevance, meta-network importance and event length of at least 80 nt. As expected, five examined events, including *RSZ22*, *GLYCINE-RICH RNA-BINDING PROTEIN RZ1A (RZ1A)*, *UNCHARACTERIZED PROTEIN DUF4050 (DUF4050)*, *HISTONE DEACETYLASE HDT2 (HDT2)* and *UBIQUITIN-CONJUGATING ENZYME E2 36 (UBC36)* (**Figure 7**), showed consistent splicing patterns with their profiles revealed by 40 ºC RNA-seq data, which therefore corroborated the accuracy of our bioinformatics workflow. Curiously, two events, *TFB5* and *CHLM*, presented the exact opposite trend. Lastly, although *ESCRT-RELATED PROTEIN CHMP1B (CHMP1B)* had a clear heat-specific pattern, it was only slightly congruent with the stress-induced long isoform dominance (**Figure 7**).

Recent research in Arabidopsis linked splicing memory to the ability of plants to survive subsequent and otherwise lethal heat stress (Ling et al. 2018). To further explore the potential acquisition of long-term stress memory, we reviewed these AS patterns in 45ºC heat-primed and non-primed plants (Phase II and Phase I; **Figure 1**). Of eight candidate

events, the existence of acquired memory could only be checked in seven as *UBC36* was not expressed at 45 ºC (**Figure 7**). It is important to note that the interpretation of T1 differs considerably according to stress intensity. In the RNA-seq based assay at 40 ºC we could already observe heat response in T1, however, in the memory assay at 45 ºC this sampling time reflected essentially damage. The response in the memory assay started at Phase I T3. Six AS events, comprising *CHLM, RZ1A, DUF4050, RSZ22, HDT2* and *CHMP1B*, displayed clear stress splicing-memory due to the response-specific isoform (Phase I T3) appearance at Phase II control. Large *TFB5* isoform disappearance at Phase II control seems to indicate thermomemory. However, the expression of the two isoforms at both T3 phases made this trend too complex to be able to undoubtedly confirm splicing memory. Detection of more than two isoforms reflected the effect of our stringent kmer (*CHMP1B*) and other 45 ºC-exclusive AS events (*RSZ22*). These findings provided evidence of stable acquired splicing-memory in heat-stressed *P. radiata* seedlings, being this mechanism a general feature of heat stress responses that could be particularly relevant in long-life cycle plants.

Discussion

RNA splicing is an essential process for plant development and intersects with many hallmarks of environmental responses. The link between AS and abiotic stress is well stablished (Laloum et al. 2018), but we only recently started to understand the molecular networks that drive RNA processing alterations in the acquisition and maintenance of heat stress tolerance, highlighting sophisticated mechanisms like splicing-memory, protein to RNA trans-regulation and NMD avoidance (Ling et al. 2018; Jia et al. 2020).

A major focus of this work was the analysis of splicing and expression differences between plants subjected to a moderate high temperature stress. We show that more than 4000 isoforms exhibiting a significant change in AS are not differentially expressed during heat stress. Moreover, functional categories covered by DS, DE and DD isoforms are not interlinked (**Figure 2**). Our results together with other works underline the importance of wide AS reprogramming as an additional layer of regulation to stress response (Calixto et al. 2018). In line with earlier research (Escandón et al. 2017), short-term stressed plants (T1) present the most relevant changes and display a functional profile defined by loss of control and damage categories like protein homeostasis (**Figure 2**; **Figure 4; Figure 5**), which is mainly involved in protein protection against heat-induced denaturation and aggregation through HSPs (Jacob et al. 2017). Interestingly, stress-specific contrast shows T3 expression recovery of photosynthesis- and water-related isoforms and enrichments in functions associated to control, as solute transport

(**Supplementary Figure 4**), thus, suggesting some degree of acclimation in mid-term stressed plants (T3). Nevertheless, these changes are not enough to detect a control-like recovery since homeostasis functions are still diminished in comparison with non-stressed plants (**Figure 4C; Figure 5E**). Collectively, these observations establish T3 as an essential point required to reveal key players and pathways in both response and acclimation such as potential alternative splicing feedback loop though RNA processing of several trans-acting RNA-binding proteins, for instance, *RSZ22* and *RZ1A* that have previously been described in stress tolerance and may be connected with splicing specificity (Reddy et al. 2013; Laloum et al. 2018); by the same token coordinated control of RNA biosynthesis (**Figure 5E; Figure 6B; Supplementary Figure 4B**) between nucleus and chloroplast is also emphasized including probably retrograde signaling, straight-forward regulation of transcription initiation and epigenetics illustrated by *CHLM*, *TFB5* and *HDT2* events, respectively (Bateman et al. 2021).

An important goal of stress systems biology is to understand how different regulatory layers are integrated to contribute to shared or unique phases of the response. We observed that the underlying variation between control and stress involves coordinated programs of all molecular levels but it is mainly linked to splicing, while top loadings are principally proteomic features (**Figure 5B, C and D**). This result could reflect splicing and proteins greater participation in early stages of the response probably due to their implication in the initial signaling and perception (Lin et al. 2020). On the other hand, metabolites variation tends to discriminate between all response stages. We propose that this metabolome phase-specific contribution might be justified by distinct timescales for turnover between all molecular levels (Shamir et al 2016). Metabolites have the lowest turnover rate; hence they are more dynamic and accurately reflect current state of the plant. Despite greater splicing variance explained, we suggest that the recurrent presence of proteome and metabolome data in top loading features (**Figure 5B and D**) could be consequence of two main reasons: 1) Proteins and metabolites are directly functional components, therefore they are subjected to more layers of down/up-stream regulation that shift their abundances towards qualitative patterns and optimize energy investment in expensive processes such as metabolism or translation. 2) Splicing, because of its co-transcriptional nature, is one of the closest regulatory levels to the genome, lacking of direct functional effect and its abundance is not as controlled as the other molecular levels. Thus, splicing trends not only reveal effects and consequences, but also part of the artifacts and stochastic reprogramming that occur during the stress response (Chaudhary et al. 2019). Assuming trypsin digestion limitations, we detected little evidence of splicing contribution to proteome and poor coordination at functional,

abundance and variation-effect levels (**Figure 2D; Supplementary Figure 4C and D; Supplementary Figure 5A**). These analyses may constitute further evidence against AS as key mechanism to expand proteome complexity. More generally, our findings suggest that response stages are correlated to molecular levels contribution and place splicing as the main driver of heat stress variation.

The evolutionary history of gymnosperms involves a series of functional and morphological traits accompanied by well-defined genomic features (De La Torre et al. 2020) such as rarity of whole genome duplications, slow mutation rates, high repeat content, big genomes and long introns (**Figure 3**). Although these organisms are a clear plant molecular diversity hotspot, they remain underrepresented and unexplored in gene regulation research as a result of their long life cycle. Contrary to angiosperms, we found that IR may not be the most prevalent type of AS in conifers. Taking all the above into account, together with the limited capacity of KisSplice to assemble long introns which requires their fully coverage by reads to be correctly assembled (Ashraf et al. 2020), and earlier described accelerated RNA Polymerase II kinetics and nuclear DNA hypomethylation during high temperature treatment (Jonkers and Lis 2015; Lamelas et al. 2020), we propose that: transcription loss of control leads to simultaneous availability of weak and strong splice sites which could result in favoured retention of introns flanked by weak splice sites (Luco et al. 2011). However, short introns are significantly more retained because their flanking splice sites are recognized as a unit (Monteuuis et al. 2019). We hypothesize that the low potential IR exhibited could be caused by the fact that gymnosperms long introns might hinder intron recognition as single unit or could even prompt the appearance of strong cryptic splice sites, thus, probably forcing the retention of fragmented introns instead of complete sequences (**Supplementary Figure 6**). Considering the little previous evidence of splicing in gymnosperms, IR is reported as dominant in Gnetum but as a non-majority processing type in *G.biloba* (Deng et al. 2019; Sun et al. 2020). Although we cannot directly extrapolate these reports because they are not stress studies, these findings may fit with our hypothesis since the strange case of Gnetum could be justified due to its particularly smaller introns-genome compared to the rest of the group (Wan et al. 2018; De La Torre et al. 2020). Further full-length analyses become a priority to validate this fragmented intron retention hypothesis In summary, this insight pinpoints the importance of understanding the influence of clade-specific genomic features if we aim to generalize our knowledge about how plants model molecular mechanisms.

The next question to solve is: Is AS a key component in the maintenance of heat stress memory in *P.radiata*? .The maintenance of acquired thermotolerance is crucial for

successful tolerance to subsequent exposure to otherwise lethal temperatures (Ling et al. 2018). We validated the acquisition of heat stress-linked splicing-memory in a diverse subset of genes associated to aforementioned pathways and others previously described (Lamelas et al. 2020), such as proteins degradation illustrated by *UBC36* and *CHMP1B* events. The present report solves the initial question and address new ones such as: 1) Heat stress induced splicing-memory seems to be conserved between the two biggest groups of seed plants, therefore, underlining the importance of this mechanism to provide plants a versatile way to cope with changing environmental conditions. 2) All splicing events validated are intron complete retention-independent. Despite of its conservation, these results suggest that the processes that drive splicing-memory are substantially different between angiosperms and gymnosperms, probably because of genome architecture divergence. 3) This work constitutes the first evidence of both, long-term splicing-memory and its presence in long-life cycle species. The extension of memory maintenance from 1 month, which was the period originally described (Ling et al. 2018), to 6 months has an even greater impact for the development of tolerant plants in order to ensure food security and forest resources in the current context of climate change. Given the realization that chromatin structure can affect AS and the memory detected in *HDT2*, it is attractive to speculate about coupled epigenetic and splicing memory as a system to maintain acclimation-associated AS patterns over a longer period of time and it could be potentially implicated in transgenerational tolerance (Luco et al. 2011). So far, all validated events that impact CDS introduce PTCs, but they could be equally responsible of thermotolerance through earlier proposed models like buffer against the stress-responsive transcriptome to reduce metabolic cost of translating all AS transcripts and targeted decay-stability that may underpin memory and resetting (Crisp et al. 2016; Chaudhary et al. 2019). The concept of environment acclimation by transcript disfunction and splicing-memory has great potential to uncover the molecular diversity underlying stress responses, which will deepen our understanding of how plants phenotypic plasticity has fascinatingly evolved. Our study reveals how long-living conifers have developed a particular way of remembering past temperature changes over a longer period of time and opens the door to include plant species with unique features, like gymnosperms, to determine the extent of conservation in gene expression regulation.

Material and Methods

In silico analysis of AS events

Raw data

In this work we studied *P.radiata* response to heat (40ºC) employing already generated transcriptomics, proteomics and metabolomics (Escandón et al. 2017) datasets available at out laboratory.

RNA-seq data pre-processing and AS events identification

An overview of the bioinformatic workflow used in this work is shown in **Supplementary Figure 1.** The data pre-processing was performed following the recommendations of Sacomoto et al. (2012). *De-novo* AS events were identified using KisSplice (v.2.4.0-p1) with stringent parameters '-C 0.05' and testing different *k-mer* values between 51 and 61 nucleotides; 51 nucleotides *k-mer* was selected as it represented a good specificity-sensitivity balance.

Differential splicing and differential expression analysis

The KisSplice raw counts values were imported into R (v.3.6.0) using the Bioconductor (v.3.9) kissDE package (v.1.4.0) (Benoit-Pilven et al. 2019). For visualization and quality assessment, raw counts were normalized using a variance stabilizing transformation as implemented in DESeq2 (v.1.24.0) (Love et al. 2014). The biological importance of the data was evaluated by principal component analysis and other visualizations using pRocessomics R package (github.com/Valledor/pRocessomics), developed in our lab, and custom R scripts.

Statistical analysis of event differential splicing and isoform differential expression were carried out in R using Bioconductor KissDE and DESeq2 packages, respectively. For both differential analyses, all possible two-by-two contrasts were performed (T1-C; T3-C, T3-T1), allowing to reveal the contribution of (1) stress damage and also (2) acclimation/response. A common threshold of 5% FDR-adjusted p-value was used to assess significance. Only isoforms showing more than 1.8-fold-up/-down were considered as biologically relevant for volcano plots.

Annotation and functional enrichment analysis

Whole annotation pipeline is showed in **Supplementary Figure 2A.** KisSplice isoforms sequences were mapped to Trinity, Transabyss and Consensus assemblies (Escandón et al. unpublished data) using bowtie2 (v.2.4.2) (Langmead and Salzberg 2012) with the following parameters '--end-to-end -N 1 -t --no-unal -p 6'. Mapped transcripts sequences

were analyzed with InterPro (v.5.45-80.0) (Jones et al. 2014) and dammit (v.1.2.0) using all default databases, Uniref90 and a custom database containing all gymnosperms data from 1KP transcriptomes project (One Thousand Plant Transcriptomes Initiative, 2019).

Gene Ontology (GO) (Carbon et al. 2021) annotations were obtained using the RSQLite package (v.2.2.0) with custom R scripts. Gene Ontology tree maps were generated using REVIGO (Supek et al. 2011) changing the background database to '*Arabidopsis thaliana'*. Mapped transcript sequences were also annotated using Mercator4 (Schwacke et al. 2019) in order to obtain bin functional information. Gene set enrichment analysis was performed using the Bioconductor fgsea package (v.1.10.0). Not identified/annotated isoforms were completed with results of other assemblies (complementary assembler) and the information of the other isoform from the same AS event (complementary splicing).

Transcript models prediction, splicing variation effect and protein-isoform analysis

Transcript sequences coding regions (CDS) identification was performed using CodAn (v.1.0.0) (Nachtigall et al. 2020) with PLANTS_full model and protein sequences of 40 ºC assay proteome and transcriptome (Escandón et al. 2017) as blast database. Splicing variation effect was determined using custom R scripts with the following program: (1) bowtie2 coordinates and CodAn gtf transcript models were used to know if the variation was produced in CDS or untranslated regions (UTRs), (2) CDS affected isoforms were examined to detect if the variation had led to premature stop codons (PTCs) or CDS changes.

Statistical analysis of protein differential abundance was carried out in R using the Bioconductor limma (v.3.40.0) (Ritchie et al. 2015) package. All possible two-by-two contrasts were performed (T1-C; T3-C, T3-T1) and a threshold of 5% FDR-adjusted p-value was used to assess significance. Isoforms corresponding proteins were identified using BLASTp with an E-value cutoff of 1e-5 and 80% identity threshold. Protein/isoforms showing more than 0.5, 1 and 1.8 -fold-up/-down were considered as weak, strong and very strong associations, respectively. To reveal isoforms-proteins relationships, UMAP training/projection were performed with isoforms/proteins data using uwot (v.0.1.8) R package.

Multi-omics factor analysis (MOFA)

Inference of shared sources of variation from combined data types was performed in R (v.4.0.3) using the Bioconductor (v.3.12) MOFA2 (v.1.0.1) package. To prepare for model training, three sets of 40 ºC heat stress assay regulatory layers were used: isoforms,

proteins and metabolites (Escandón et al. 2017). One replicate for each treatment was imputed for isoforms data using the pRocessomics R package. Training of the model was carried out using the following options: maxiter = 60000, convergence_mode = "slow". Each biologically relevant latent factor underwent over-representation analysis to determine the magnitude to which factors represented different Mercator4 functional categories.

Meta-network analysis

To avoid biases towards specific interaction patterns, a meta-network was created using Seidr network toolkit (v.0.14.2) (Schiffthaler et al. 2019). Double differential VST-normalized isoform expression data were used as input for network construction, keeping the same number of samples as in MOFA. Inference methods applied were ARACNE, CLR, Elastic Net and SVM ensemble, Partial Correlation, NARROMI, GENIE3, PLSNET, TIGRESS, TOM similarity and Spearman correlation. The outputs were aggregated into a meta-network using the irp method and the resulting network was filtered executing network backbonning with -F 1. Infomap (v.1.3.0) (Rosvall and Bergstrom. 2008) was used to detect highly connected modules. Neat (v.1.1.3) R package was used to discover network modules enriched for Mercator4 functional categories (alpha = 0.05).

Intron comparisons

All genome data sources are showed in **Supplementary Table 1**. Among all sequenced conifer genomes, *Pinus taeda* was selected as the most appropriate reference species for genomic comparisons with *Arabidopsis thaliana, Zea mays, Oryza sativa, Amborella trichopoda, Ginkgo biloba, Gnetum montanum, Pseudotsuga menziesii* and *Pinus lambertiana*. IntronIC (v.1.1.2) (Moyer et al. 2020) was applied to classify U12/U2 introns and obtain intron length and terminal dinucleotides information taking into account all isoforms. Boxplot construction and effect size calculation were performed according to Wan et al.(2018).

Experimental validation of AS events

Plant material and growth conditions

Plant material was generated for (1) validating RNA-Seq based results and also for (2) testing the potential implication of AS in stress memory and checking AS patterns coherency at a more lethal temperature (**Figure 1**).

One-year-old *P. radiata* seedlings (height 33 ± 4 cm) were kept in 1 L pots under a photoperiod of 16 h (400 µmol m−2 s−1) at 25 ºC and 50% relative humidity (RH) and 8 h at 15 ºC and 60% RH during the night period in a climate chamber under controlled

conditions (Fitoclima 1200, Aralab) where stress assays were also conducted. Prior assays, plants were acclimatized for over a 1-month period time inside the controlled chamber, watered at field capacity every day and fertilized weekly with nutritive solution (N:P:K 5:8:10).

For the first experiment, *P. radiata* seedlings were subjected to 40 ºC heat stress assay as described by Escandón et al. (2017). Briefly, heat stress was applied during the central hours of the day, employing a temperature ramp from 15 ºC to 40 ºC over 5 h. High temperature was hold for 6 h and then ramped down again to night values. The stress was applied for 3 consecutive days, and plants were sampled at the end of the 40 ºC period on first (T1) and last day (T3) (**Figure 1**).

On the other hand, in the second experiment *P. radiata* seedlings were subjected to 45 ºC heat stress memory assay as described by Lamelas et al. (2020). Summing up, to test long-lasting memory acquisition related to an acclimation process, assay was divided in two similar phases 6 months apart from each other. In each phase, treatment was applied during the central hours of the day, employing a temperature ramp from 15 ºC to 45 ºC over 5 h. High temperature was hold for 6 h and then ramped down again to night values. This procedure was repeated for 5 consecutive days, and plants were sampled at the end of the 45 ºC period on day 1 (T1), day 3 (T3) and day 5 (T5). Thus, Phase I was stablished as the first 45 ºC treatment and Phase II as the second treatment after 6 months (**Figure 1**).

Control samples (C) were collected the first day of each experiment before starting the heat exposure treatment. The three biological replicates of each treatment were pools of three plant each. The individuals conforming each pool were kept during the experiments. All samples were immediately frozen in liquid nitrogen and stored at -80 ºC until RNA was extracted.

RT-PCR analysis

RNA of the samples was extracted as described by Valledor et al.(2014) and then quantified in a Nabi UV/Vis Nano Spectrophotomer. RNA integrity was checked by agarose gel electrophoresis, and potential DNA contamination by PCR. cDNA was obtained from 500 ng of RNA using the RevertAid kit (ThermoFisher Scientific) and random hexamers as primers following the manufacturer's instructions. RT-PCR was performed following indications of Ling et al. (2018). Primers for each AS event (**Supplementary Table 2**) were designed to amplify both two splice variants in a single reaction. Sequence of different splice variants were validated by Sanger sequencing (Stab Vida, Lda; Portugal).

Data availability

The data and code underlying this article are available in github, at https://github.com/RocesV/AS_heat_Pra.

Acknowledgments

This work was supported by the Spanish Ministry of Economy and Competitiveness (PID2019-1071076B-100). VR was supported by a fellowship from Spanish Ministry of Science, Innovation and Universities (FPU18/02953). LL was supported by a fellowship from Spanish Ministry of Science, Innovation and Universities (BES-2017-082092). MC was supported by Severo Ochoa Predoctoral Program (BP19-137). LV was supported by the Spanish Ministry of Economy and Competitiveness (RYC-2015-17871).

References

Argelaguet R, Velten B, Arnol D, Dietrich S, Zenz T, Marioni JC, Buettner F, Huber W, ,Stegle O. 2018. Multi-Omics Factor Analysis—a framework for unsupervised integration of multi-omics data sets. *Mol Syst Biol*. 14(6): e8124.

Ashraf U, Benoit-Pilven C, Navratil V, Ligneau C, Fournier G, Munier S, Sismeiro O, Coppée JY, Lacroix V, Naffakh N. 2020. Influenza virus infection induces widespread alterations of host cell splicing. *NAR Genomics and Bioinformatics 2*(4): 1–13.

Bateman A, Martin MJ, Orchard S, Magrane M, Agivetova R, Ahmad S, Alpi E, Bowler-Barnett EH, Britto R, Zhang J, et al. 2021. UniProt: The universal protein knowledgebase in 2021. *Nucleic Acids Research* 49(D1): D480–D489.

Calixto CPG, Guo W, James AB, Tzioutziou NA, Entizne JC, Panter PE, Knight H, Nimno HG, Zhang R, Brown JW. S. 2018. Rapid and dynamic alternative splicing impacts the arabidopsis cold response transcriptome[CC-BY]. *Plant Cell* 30(7): 1424–1444.

Capovilla G, Delhomme N, Collani S, Shutava I, Bezrukov I, Symeonidi E, Amorim M, Laubinger S, Schmid M. 2018. PORCUPINE regulates development in response to temperature through alternative splicing. *Nature Plants* 4(8): 534–539.

Carbon S, Douglass E, Good BM, Unni DR, Harris NL, Mungall CJ, Basu S, Chisholm RL, Dodson RJ, Hartline E, et al. 2021. The Gene Ontology resource: Enriching a GOld mine. *Nucleic Acids Research* 49(D1): D325–D334.

Chaudhary S, Jabre I, Reddy ASN, Staiger D, Syed NH. 2019. Perspective on Alternative Splicing and Proteome Complexity in Plants. *Trends in Plant Science* 24(6): 496–

506.

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(2): 1–14.

De La Torre AR, Birol I, Bousquet J, Ingvarsson PK, Jansson S, Jones SJM, Keeling CI, Mackay J, Nilsson O, Ritland K, et al. 2014. Insights into Conifer Giga-Genomes. *Plant Physiology* 166(4): 1724–1732.

De La Torre AR, Piot A, Liu B, Wilhite B, Weiss M, Porth I. 2020. Functional and morphological evolution in gymnosperms: A portrait of implicated gene families. *Evolutionary Applications* 13(1): 210–227.

Deng N, Hou C, Ma F, Liu C, Tian Y. 2019. Single-molecule long-read sequencing reveals the diversity of full-length transcripts in leaves of gnetum (Gnetales). *International Journal of Molecular Sciences* 20(24): 1–16.

Escandón M, Valledor L, Pascual J, Pinto G, Cañal MJ, Meijón M. 2017. System-wide analysis of short-term response to high temperature in Pinus radiata. *Journal of Experimental Botany* 68(13): 3629–3641.

Hanemian M, Vasseur F, Marchadier E, Gy I, Violle C, Loudet O. 2020. Natural variation at FLM splicing has pleiotropic effects modulating ecological strategies in Arabidopsis thaliana. *Nature Communications* 11(1): 1–12.

Huang J, Lu X, Wu H, Xie Y, Peng Q, Gu L, Wu J, Wang Y, Reddy ASN, Dong S. 2020. Phytophthora Effectors Modulate Genome-wide Alternative Splicing of Host mRNAs to Reprogram Plant Immunity. *Molecular Plant* 13(10): 1470–1484.

Jacob P, Hirt H, Bendahmane A. 2017. The heat-shock protein/chaperone network and multiple stress resistance. *Plant Biotechnology Journal* 15(4): 405–414.

Jia J, Long Y, Zhang H, Li Z, Liu Z, Zhao Y, Lu D, Jin X, Deng X, Xia R, et al. 2020. Post-transcriptional splicing of nascent RNA contributes to widespread intron retention in plants. *Nature Plants* 6(7): 780–788.

Jin W, Gernandt DS, Wehenkel C, Xia X, Wei X, Wang XQ. 2021. Phylogenomic and ecological analyses reveal the spatiotemporal evolution of global pines. *PNAS* 118(20): 1–11.

Jones P, Binns D, Chang HY, Fraser M, Li W, McAnulla C, McWilliam H, Maslen J, Mitchell A, Nuka G, et al. 2014. InterProScan 5: Genome-scale protein function

classification. *Bioinformatics* 30(9): 1236–1240.

Jonkers I, Lis JT. 2015. Getting up to speed with transcription elongation by RNA polymerase II. *Nature Reviews Molecular Cell Biology* 16(3): 167–177.

Kliebenstein DJ. 2019. Questionomics: Using big data to ask and answer big questions. *Plant Cell* 31(7): 1404–1405.

Laloum T, Martín G, Duque P. 2018. Alternative Splicing Control of Abiotic Stress Responses. *Trends in Plant Science* 23: 140–150.

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(6): 2040–2057.

Langmead B, Salzberg SL. 2012. Fast gapped-read alignment with Bowtie 2. *Nature Methods* 9(4): 357–360.

Lin J, Xu Y, Zhu Z. 2020. Emerging Plant Thermosensors: From RNA to Protein. *Trends in Plant Science* 25(12): 1187–1189.

Ling Y, Serrano N, Gao G, Atia M, Mokhtar M, Woo YH, Mahfouz MM. 2018. Thermoprimming Triggers Splicing Memory in Arabidopsis. *Journal of Experimental Botany* 69(10): 2659-2675.

Love MI, Huber W, Anders S. 2014. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biology* 15(12): 1–21.

Luco RF, Allo M, Schor IE, Kornblihtt AR, Misteli T. 2011. Epigenetics in alternative pre-mRNA splicing. *Cell* 144(1): 16–26.

Marbach D, Costello JC, Küffner R, Vega NM, Prill RJ, Camacho DM, Allison KR, the DREAM5 Consortium, Kellis M, Collins JJ, et al. 2012. Wisdom of crowds for robust gene network inference. *Nature Methods* 9(8): 796–804.

Meng X, Liang Z, Dai X, Zhang Y, Mahboub S, Ngu DW, Roston RL, Schnable JC. 2021. Predicting transcriptional responses to cold stress across plant species. *PNAS* 118(10): 1–9.

Monteuuis G, Wong JJL, Bailey CG, Schmitz U, Rasko JEJ. 2019. The changing paradigm of intron retention: Regulation, ramifications and recipes. *Nucleic Acids Research* 47(22): 11497-11513.

Moyer DC, Larue GE, Padgett RA, Hershberger CE, Roy SW. 2020. Comprehensive

database and evolutionary dynamics of U12-type introns. *Nucleic Acids Research* 48(13): 7066–7078.

Nachtigall PG, Kashiwabara AY, Durham AM. 2020. CodAn : predictive models for precise identification of coding regions in eukaryotic transcripts. *Briefings in Bioinformatics* 22(3): 1–11.

Reddy ASN, Marquez Y, Kalyna M, Barta A. 2013. Complexity of the Alternative Splicing Landscape in Plants. *The Plant Cell* 25(10): 3657–3683.

Ritchie ME, Phipson B, Wu D, Hu Y, Law CW, Shi W, Smyth GK. 2015. limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Research* 43(7): e47.

Rosvall M, Bergstrom CT. 2008. Maps of random walks on complex networks reveal community structure. *PNAS* 105(4): 1118-1123.

Sacomoto GAT, Kielbassa J, Chikhi R, Uricaru R, Antoniou P, Sagot MF, Peterlongo P, Lacroix V. 2012. KISSPLICE : de-novo calling alternative splicing events from RNA-seq data. *BMC Bioinformatics* 13(Suppl 6): S5.

Schiffthaler B, van Zalen A,Serrano A, Street N, Delhomme N. 2021. Seidr : a gene meta-network calculation toolkit. *BioRxiv* 2–20.

Schwacke R, Ponce-soto GY, Krause K, Bolger AM, Arsova B, Hallab A, Gruden K, Stitt M, Bolger ME, Usadel B. 2019. MapMan4 : A Refined Protein Classification and Annotation Framework Applicable to Multi-Omics Data Analysis. *Molecular Plant* 12: 879–892.

Shamir M, Bar-On Y, Phillips R, Milo R. 2016. SnapShot: Timescales in Cell Biology. *Cell* 164(6): 1302-1302.e1.

Smith CCR, Tittes S, Mendieta JP, Collier-zans E, Rowe HC, Rieseberg LH, Kane NC. 2018. Genetics of alternative splicing evolution during sunflower domestication. *PNAS* 115(26): 6768–6779.

Sun S, Li Y, Chu L, Kuang X, Song J, Sun C. 2020. Full-length sequencing of ginkgo transcriptomes for an in-depth understanding of flavonoid and terpenoid trilactone biosynthesis. *Gene* 758: 144961.

Supek F, Bosnjak M, Skunca N, Smuc T. 2011. REVIGO Summarizes and Visualizes Long Lists of Gene Ontology Terms. *PLoS ONE* 6(7): e21800.

One Thousand Plant Transcriptomes Initiative. 2019. One thousand plant transcriptomes

and the phylogenomics of green plants. *Nature* 574: 679-685.

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. *Progress in Botany* 81: 353-375.

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(1): 173–180.

Wan T, Liu ZM, Li LF, Leitch AR, Leitch IL, Lohaus R, Liu ZJ, Xin HP, Gong YB, Liu Y, et al. 2018. A genome for gnetophytes and early evolution of seed plants. *Nature Plants* 4: 82–89.

Figures

Figure 1. Experimental design. Two different heat assays were conducted. In 40 ºC assay (upper part), plants were stressed and sampled at intervals up to 3 days to characterize short and medium responses: T1, T3. Then, RNA extracted from treated plants was sequenced in order to perform splicing analysis. 45 ºC assay (lower part) was divided in two phases. In Phase I, plants were stressed and sampled at intervals up to 5 days to characterize long responses: T1, T3, T5. At 6 months after the end of Phase I, the plants were subjected to another round of treatment (Phase II) to evaluate the potential acquisition of long-term splicing memory. C = Control; T1, T3, T5: stressed plants samples after 1,3 or 5 days of stress.

Figure 2. Summary of alternative splicing events and DE, DS and DD analysis of *P. radiata* response to high temperature. (a) Flow chart showing the distribution of the 5105 DS and 1568 DE isoforms detected. DS = differential splicing; DE = differential expression; DD = double differential. (b) Matrix layout for all intersections of DS and DE comparisons, sorted by decreasing isoforms number. Dark circles in the matrix indicate sets that are part of the intersection. C/T1, C/T3 and T1/T3 refer to T1-C, T3-C and T3-T1 comparisons, respectively. (c) Differential expression trends using Mercator4 categorization terms. The numbers indicate scaled expression according to the Mercator4 functional bin. (d) The distribution of distinct types of AS events by differential level and contrast. From bottom to top, 100 % refers to all CDS AS events and all AS events, respectively. Change = make protein sequence changes; NoRes = neither full model nor full-transcript sequences were predicted/mapped so events could not be classified. The sampling times correspond to the 40 ºC assay shown in **Figure 1**.

Figure 3. Intron morphologies comparative analysis. Nine representative seed plants species were included: *A.thaliana, Arabidopsis thaliana; O.sativa, Oryza sativa; Z.mays, Zea mays; A.trichopoda, Amborella trichopoda; G.biloba, Ginkgo biloba; G.montanum, Gnetum montanum; P.menziesii, Pseudotsuga menziesii; P.lambertiana, Pinus lambertiana; P.taeda, Pinus taeda*. For all isoforms of each species we computed (from left to right): genome sizes (bp, scaled to *P. taeda* genome), all-introns length distributions, U12-introns length distributions, all-introns flanking di-nts proportions, U12-introns flanking di-nts proportions and all-/U12-introns length effect size setting *P. taeda* as reference. Different colors in flanking di-nt proportions barplots indicate distinct types of di-nts. Gold and purple in effect size lineplots indicate all- /U12-introns, respectively.

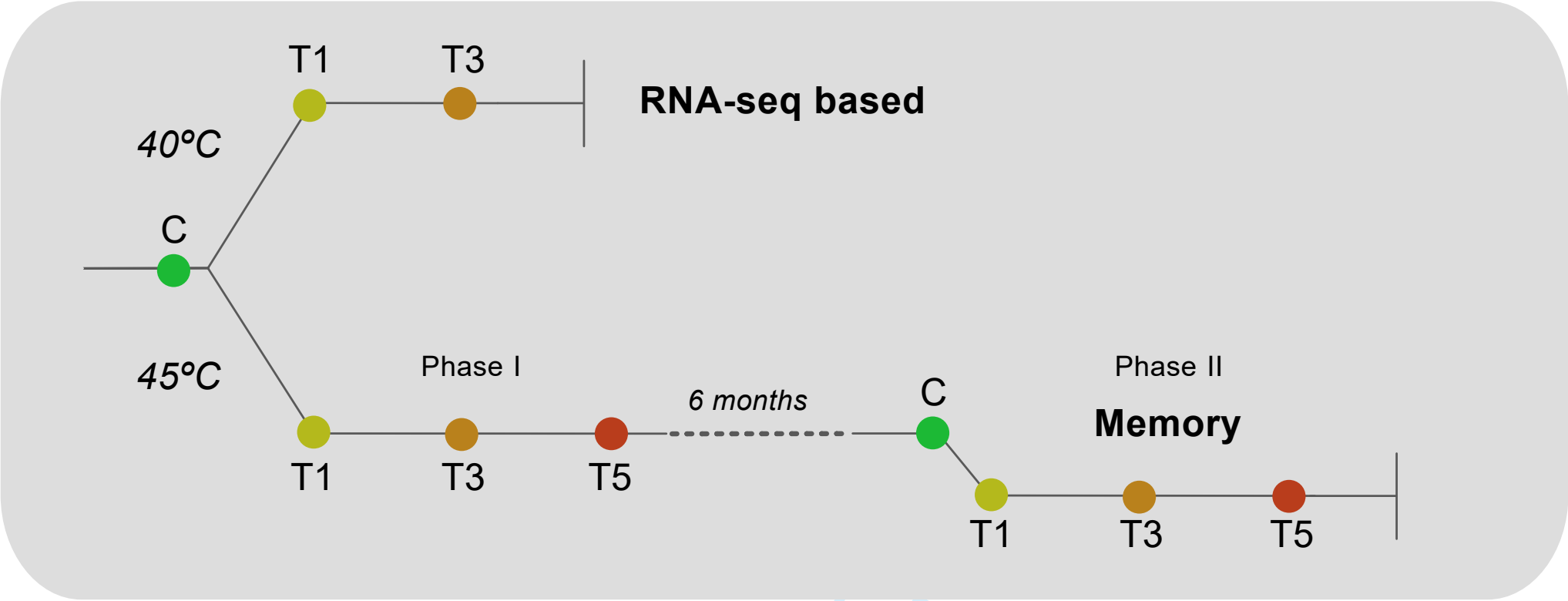
Figure 4. Global dynamics and DD modulation functional description. (a) Heatmap/hierarchical clustering using Mercator4 categorization terms. The numbers indicate scaled expression, by rows, according to the Mercator4 functional bin. (b) Volcano analysis of DD event isoforms. Not significant = FDR-adjusted P > 0.05 and absolute log fold change < 1.8; Significant = FDR-adjusted P < 0.05 and absolute log fold change < 1.8; Fold Change = FDR-adjusted P > 0.05 and absolute log fold change > 1.8; Significant & Fold Change = FDR-adjusted P <0.05 and absolute log fold change > 1.8. (c) Gene set enrichment analysis of DD event isoforms using Mercator4 functional categories. Blue and gold indicate significative and non-significative enriched terms, respectively. The sampling times correspond to the 40 ºC assay shown in **Figure 1**.

Figure 5. Relative contributions of distinct regulatory layers to the definition of heat response. (a) Several layers of coordinated molecular levels collectively and differentially participate in heat response definition. (b) Percentage of variance explained (R^2) by each MOFA2 factor (columns) across regulatory layers (rows) (left pannel). Total contributions of each layer are summarized in the bar-plot (right panel). (c) Scatter-plot of latent factor 1 (x axis) and latent factor 2 (y axis) illustrating the variation described. Samples are colored according to treatment (C, T1, T3). (d) Lollipop-plots showing top loading latent factor 1 (top) and latent factor 2 (bottom) features in descending order. Red and green indicate metabolome and proteome features, respectively. The sampling times correspond to the 40 ºC assay shown in **Figure 1**. (e) Enriched Mercator4 functional categories and top weight isoforms associated to latent factor 1 (top) and latent factor 2 (bottom). Positive = functional terms linked to samples with > 0 scores; Negative = functional terms linked to samples with < 0 scores.

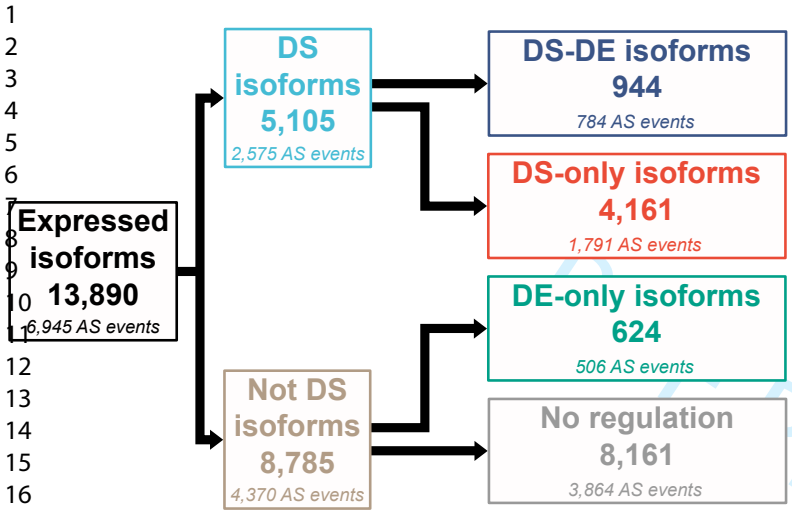
Figure 6. Meta-network regulatory analysis. (a) Meta-network constituted by five highly-connected modules (different colours). Nodes represent isoforms and node size

reflect Katz centrality. Network edges indicate biological interaction. (b) Local interaction network connecting both isoforms from two AS events. Nodes and edges highlighted in red illustrate the shortest path between intra-event isoforms. (c) Radar-plots showing topology over- (red) and under-enrichment (blue) of mercator4 functional terms for the two biggest modules of the meta-network. Peak size = -log10 (adjusted-P).

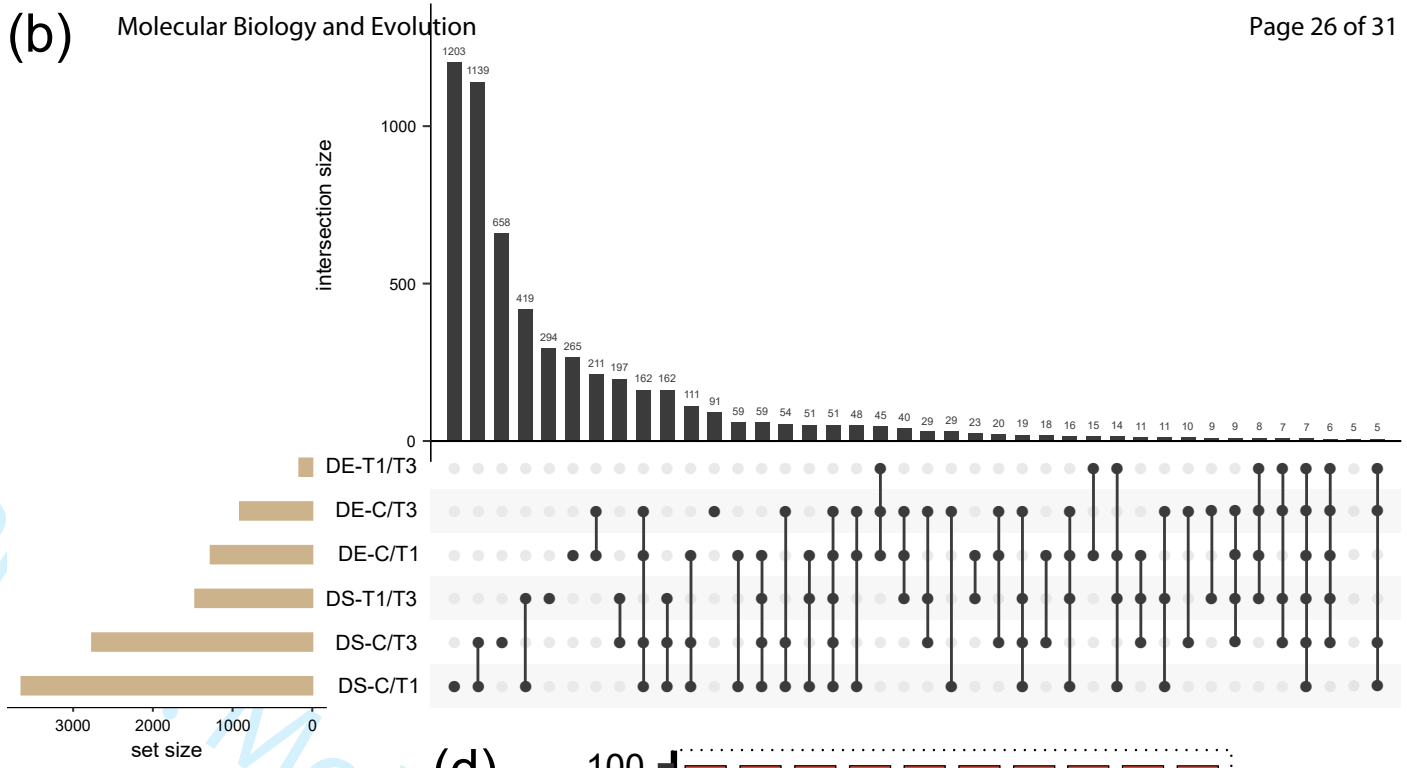
Figure 7. Experimental validation of heat-induced AS events and splicing-memory by RT-PCR. Each AS event is represented by a block constituted by 3 panels. Top panel: transcript models of the small isoform displaying event length, PTCs introduction and where the variation is produced. Mid-left and mid-right panels: bar-plots showing relative expression of isoform 1 (blue) and isoform 2 (yellow) of AS events revealed by RNA-seq and RT-PCR AS patterns in the sampling times corresponding to the 40 ºC assay shown in **Figure 1**, respectively. Bottom panel: RT-PCR AS patterns in the sampling times corresponding to the 45 ºC memory assay shown in **Figure 1**. Phase I and II correspond to two different 45 ºC heat exposures 6 months apart from each other. The primers used allows the amplification of both splice variants.



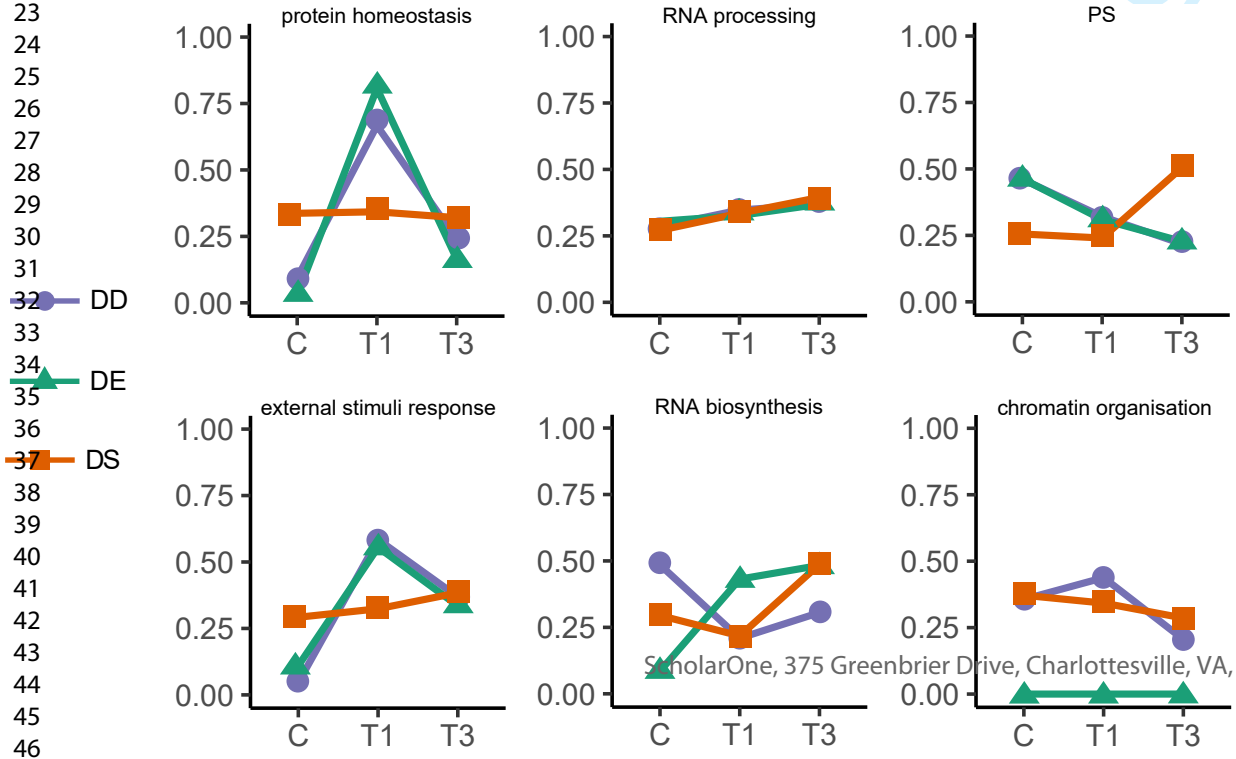
(a)



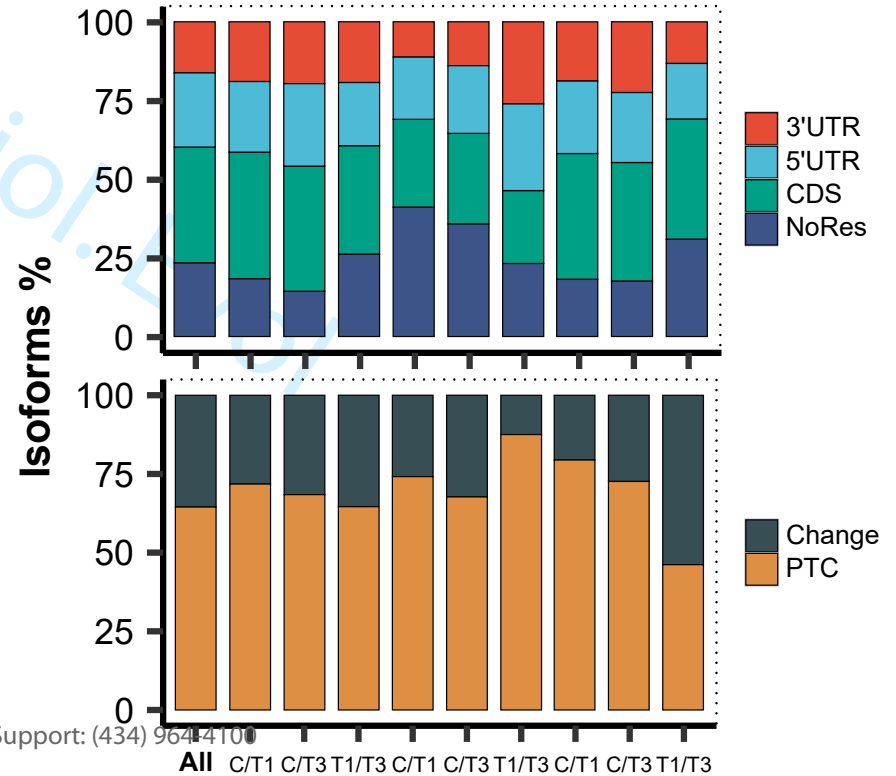
(b)

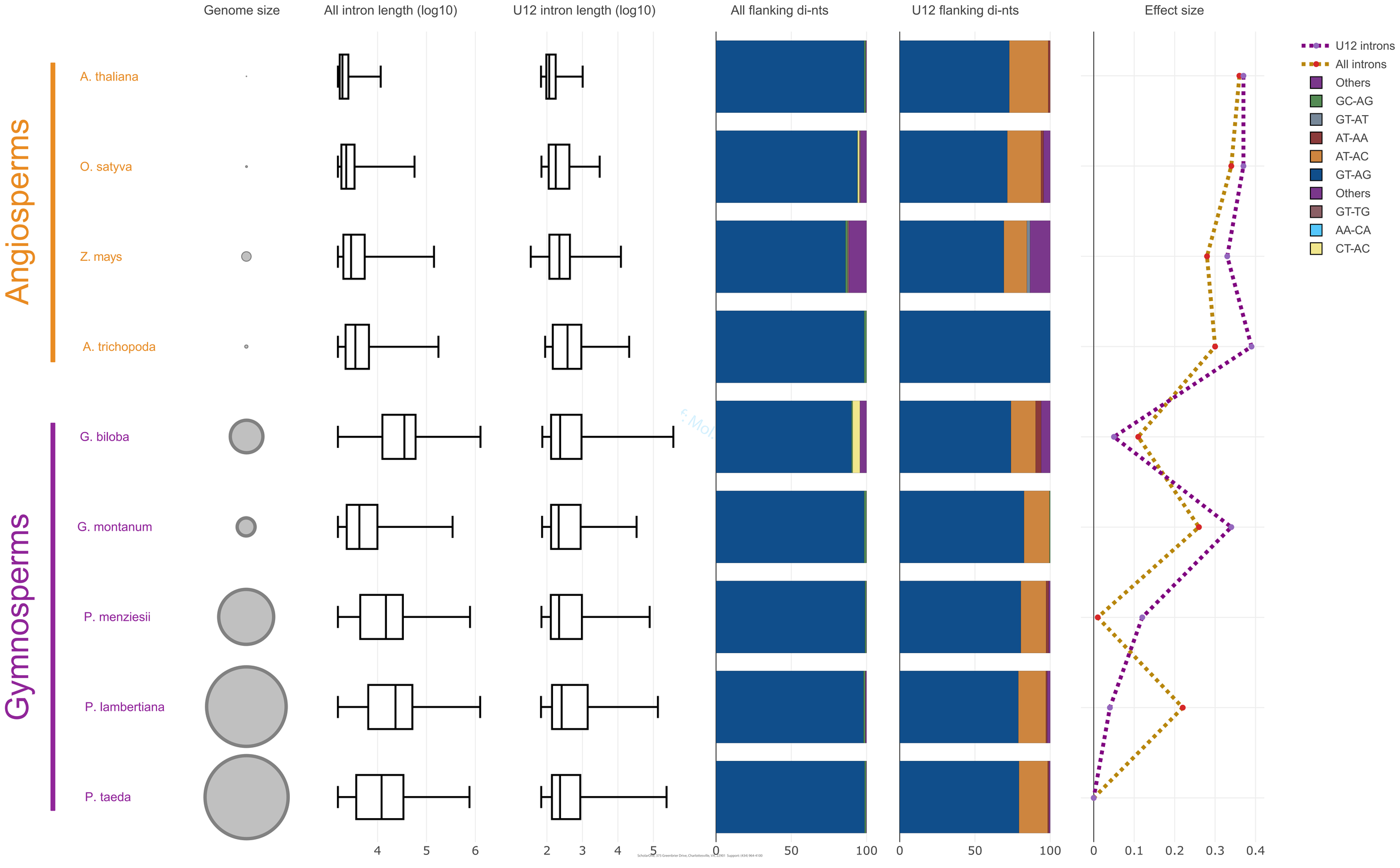


(c)

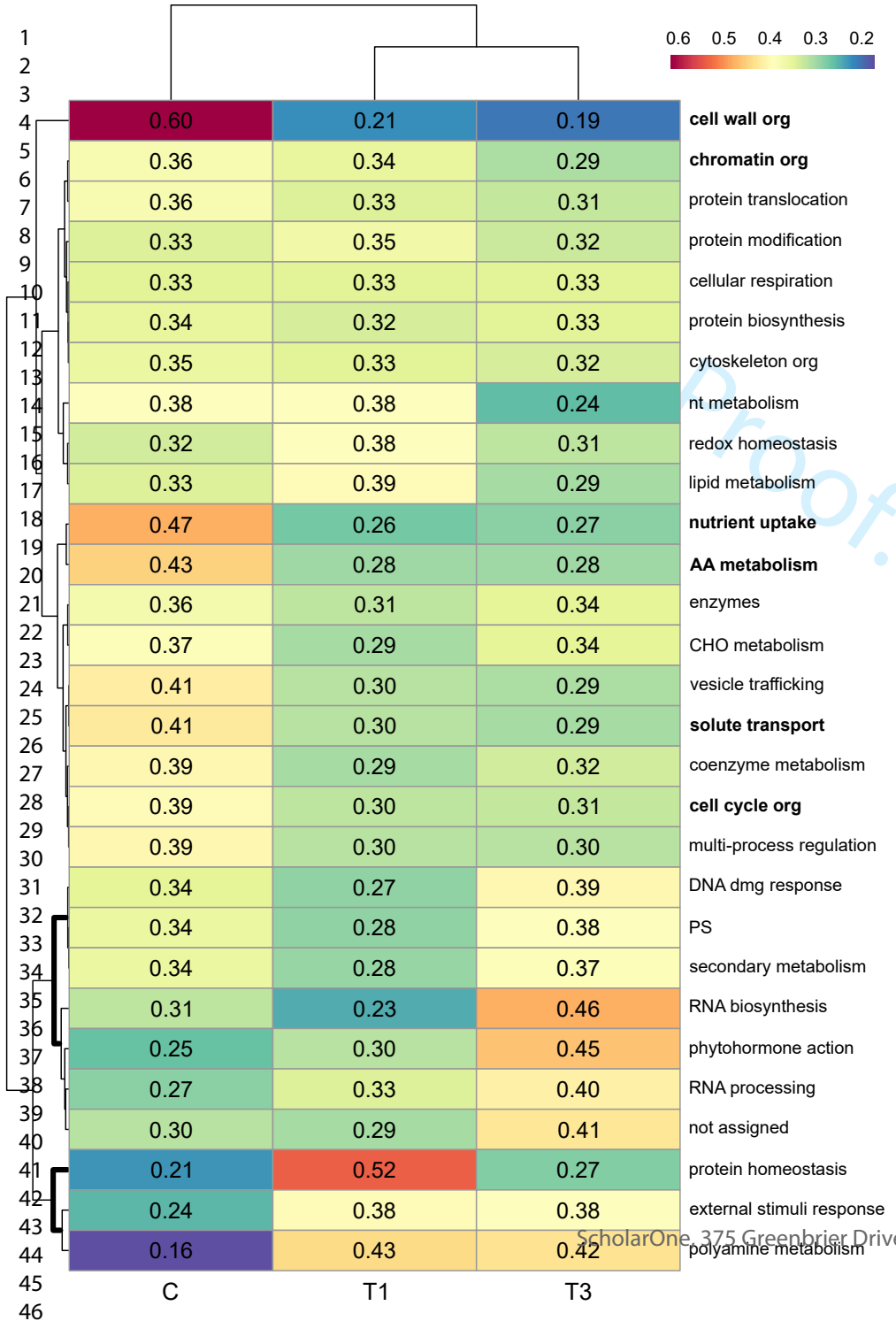


(d)

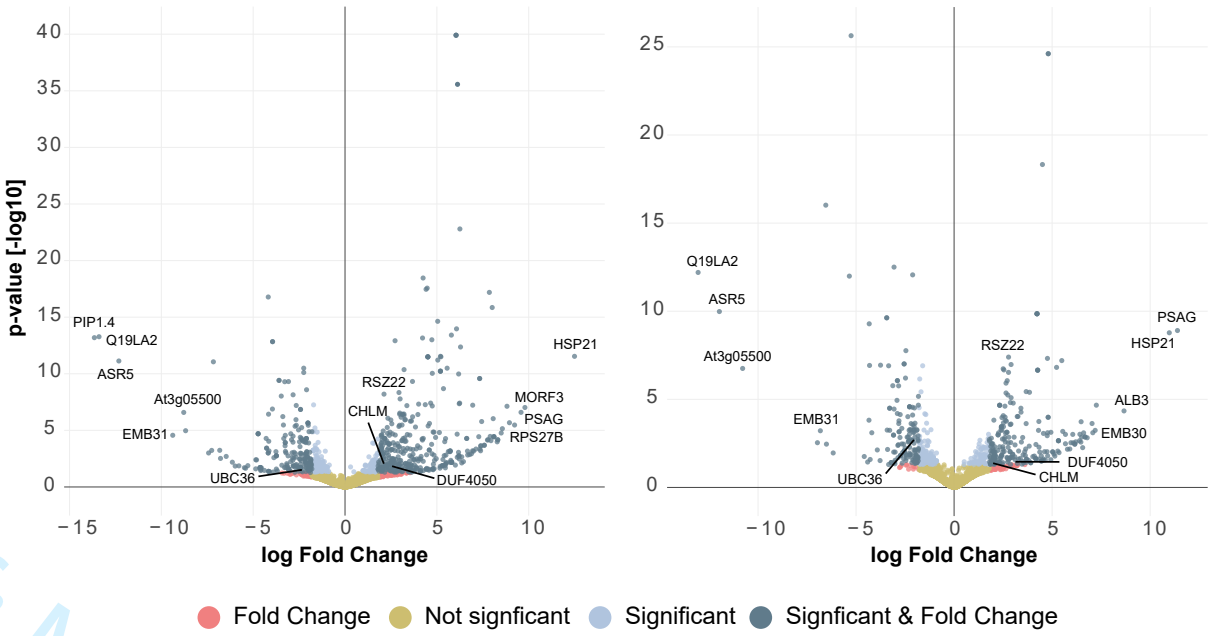




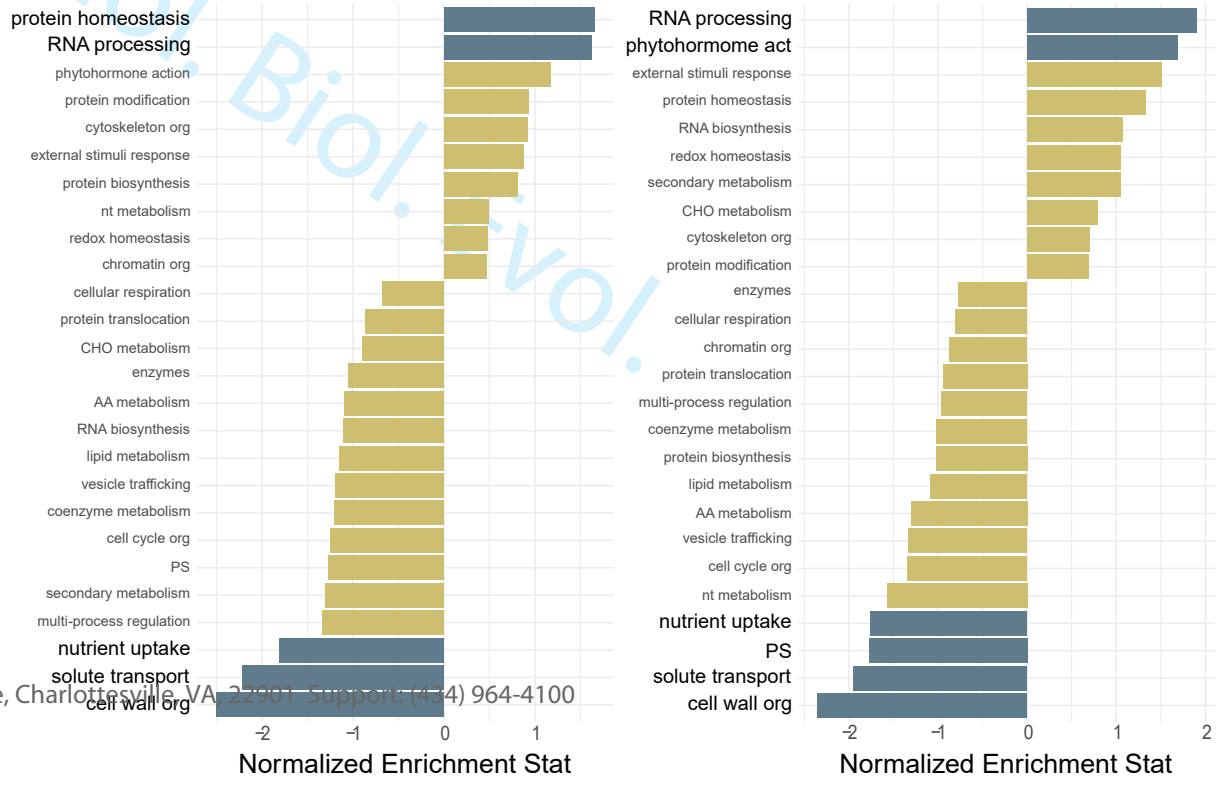
(a)

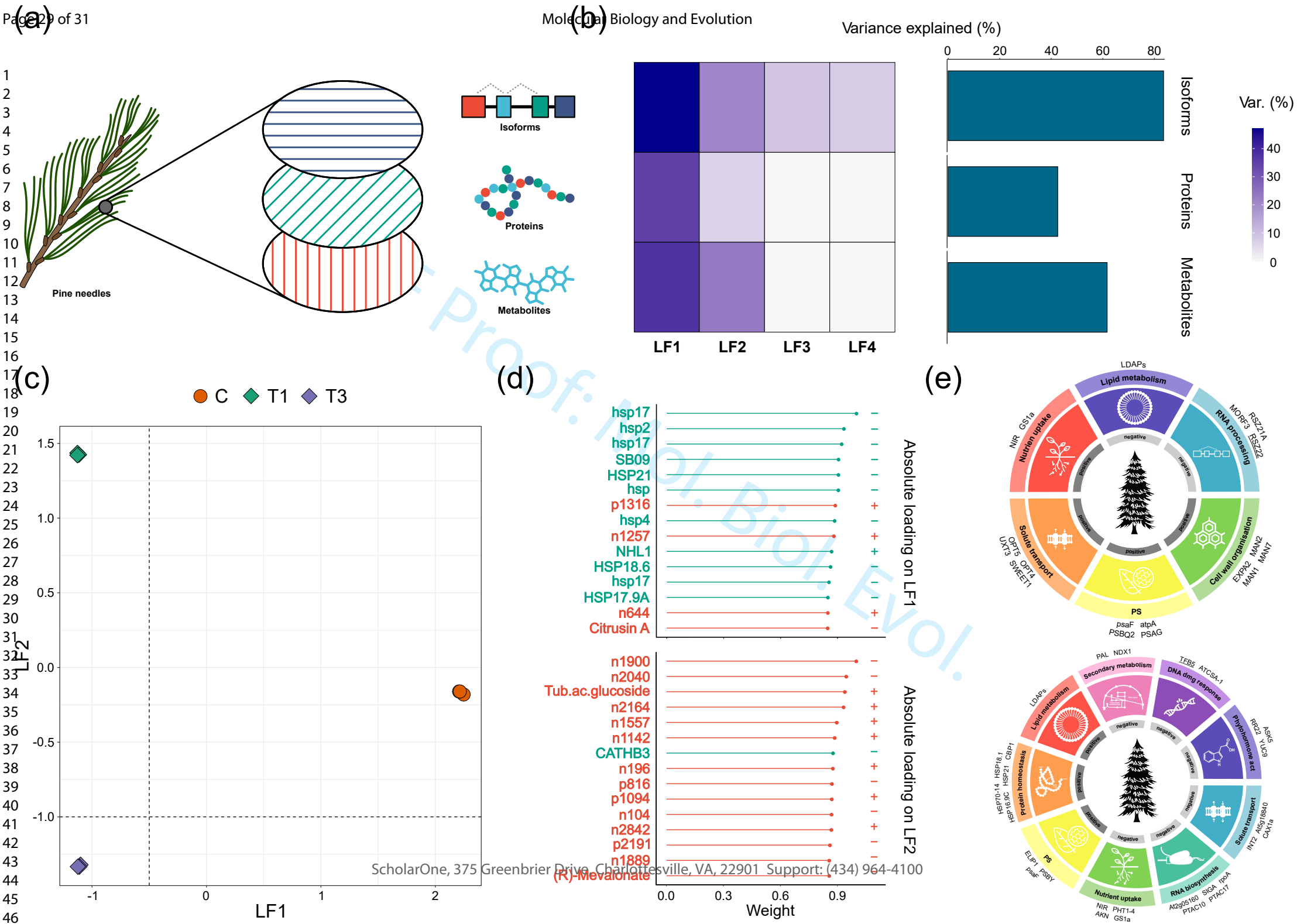


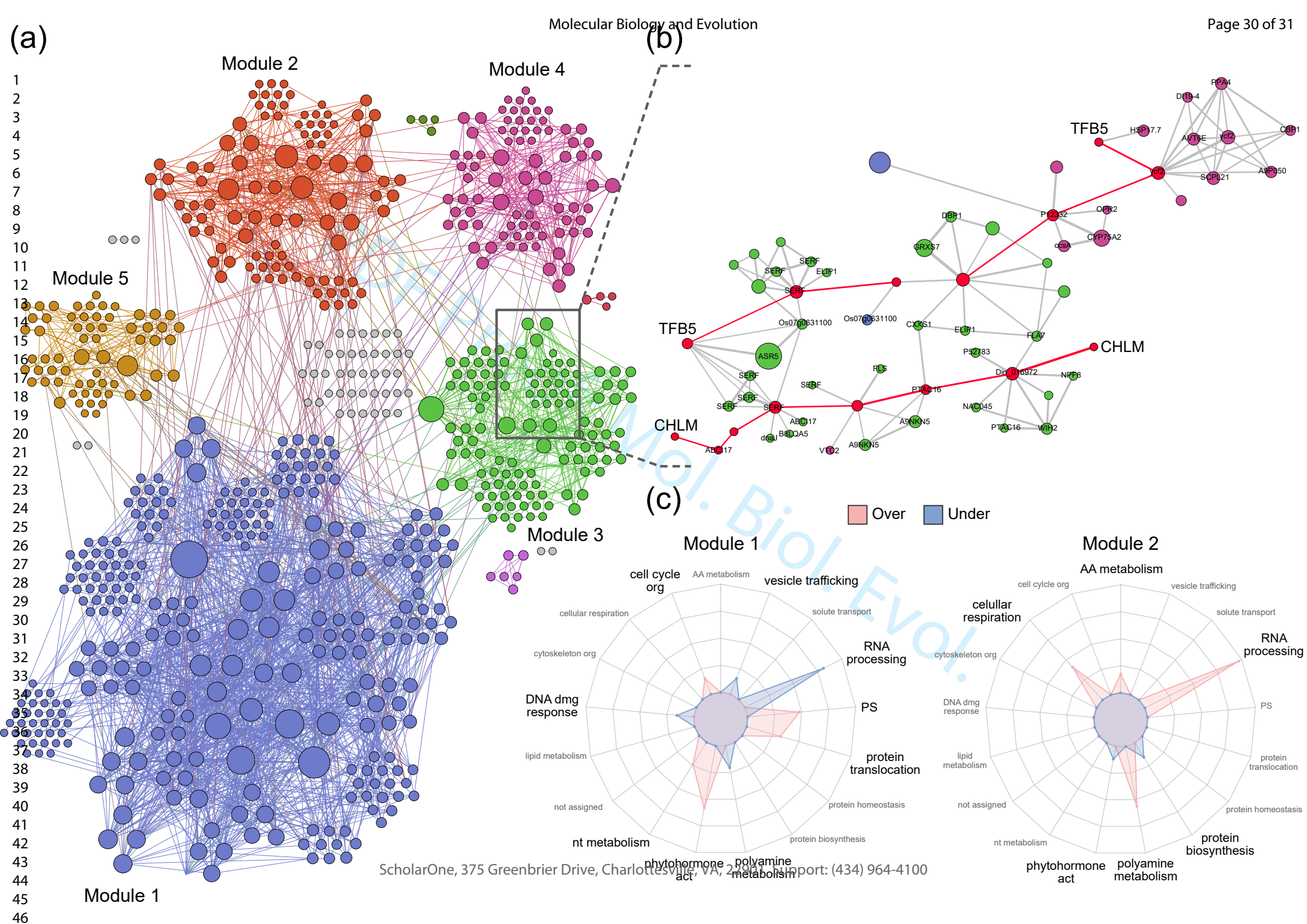
(b)



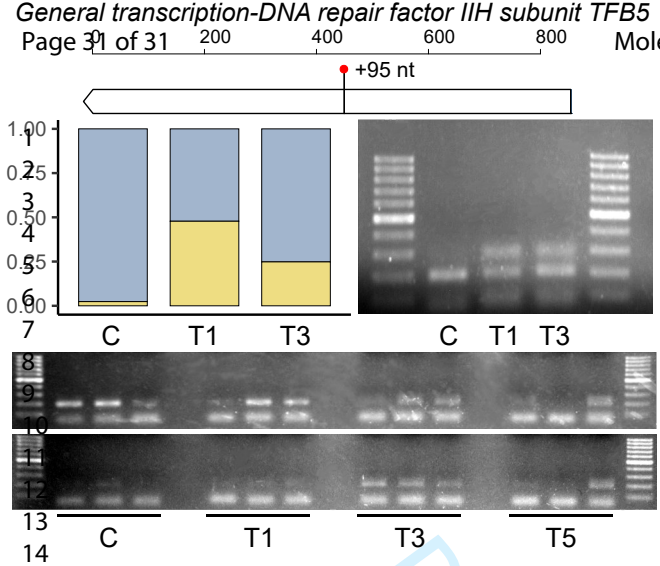
(c)



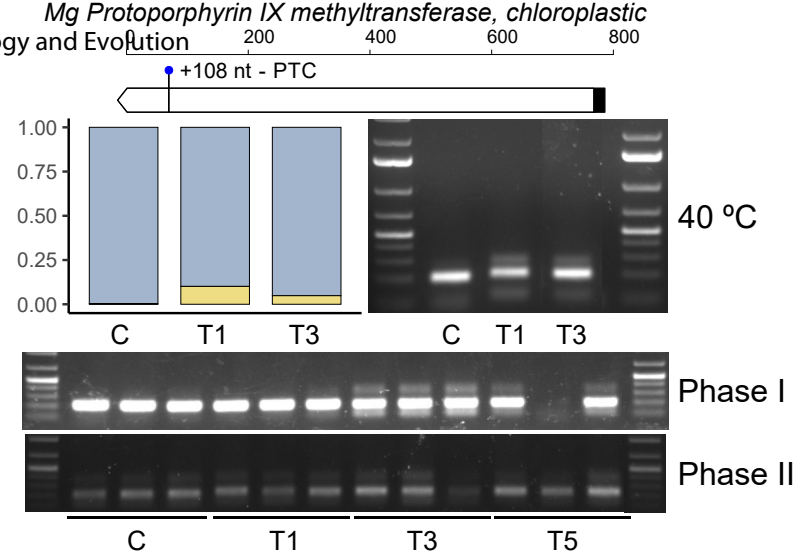




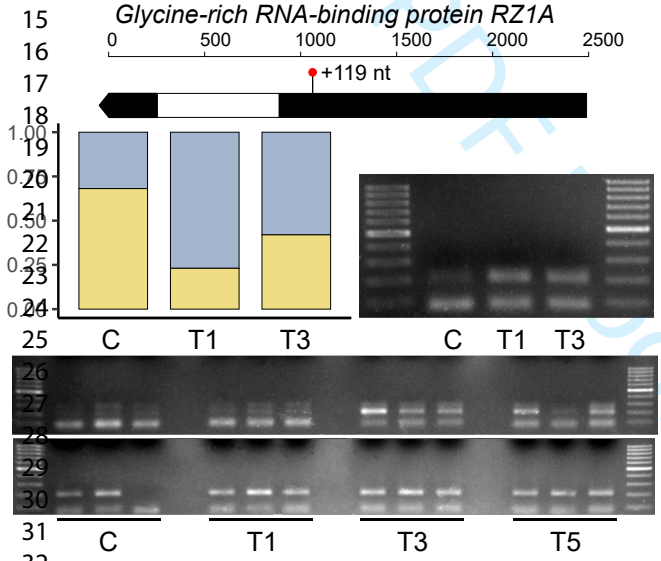
General transcription-DNA repair factor IIH subunit TFB5



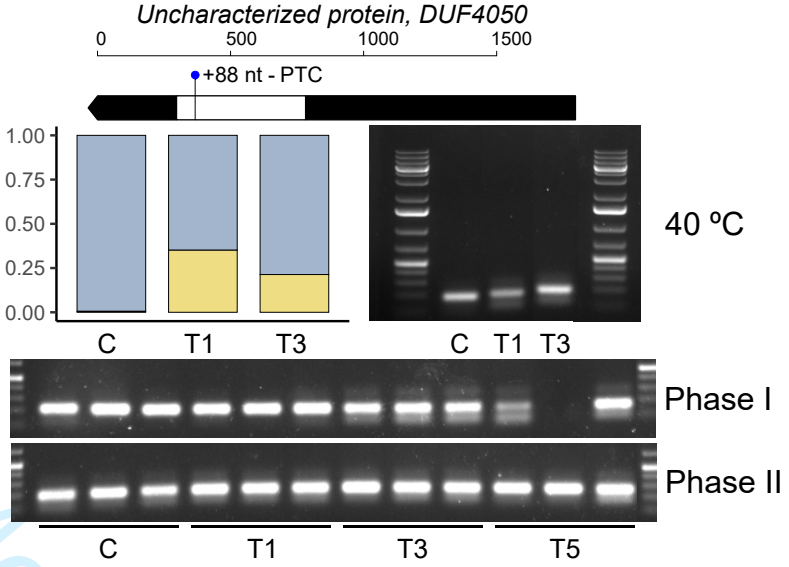
Mg Protoporphyrin IX methyltransferase, chloroplastic



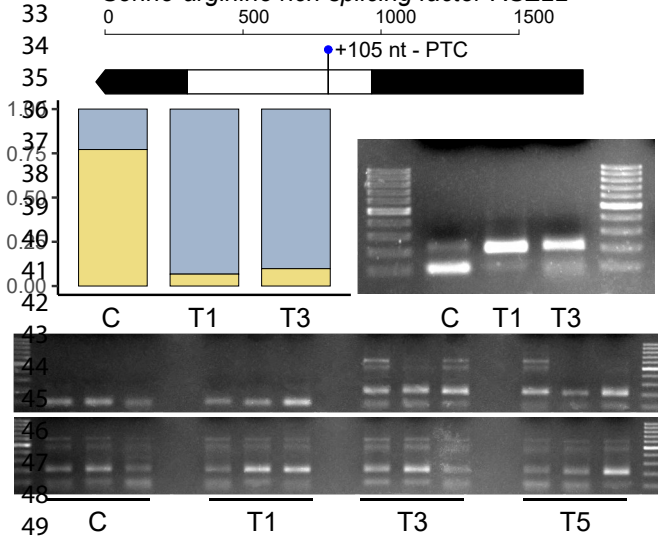
Glycine-rich RNA-binding protein RZ1A



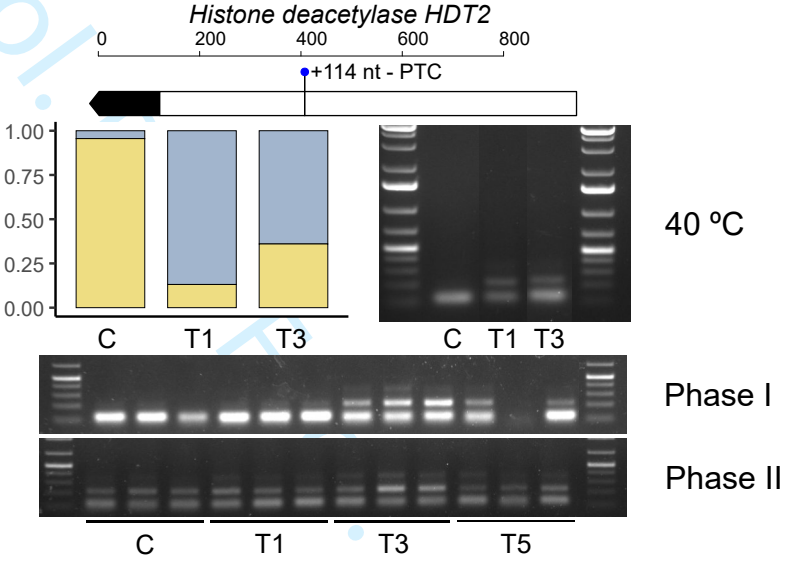
Uncharacterized protein, DUF4050



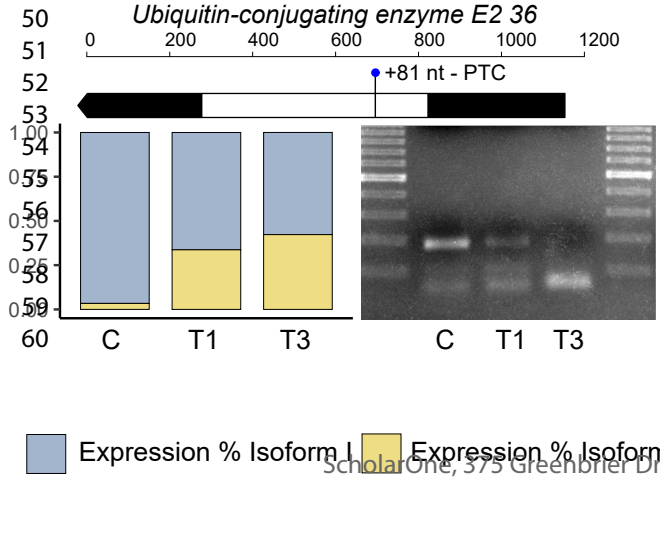
Serine-arginine rich splicing factor RSZ22



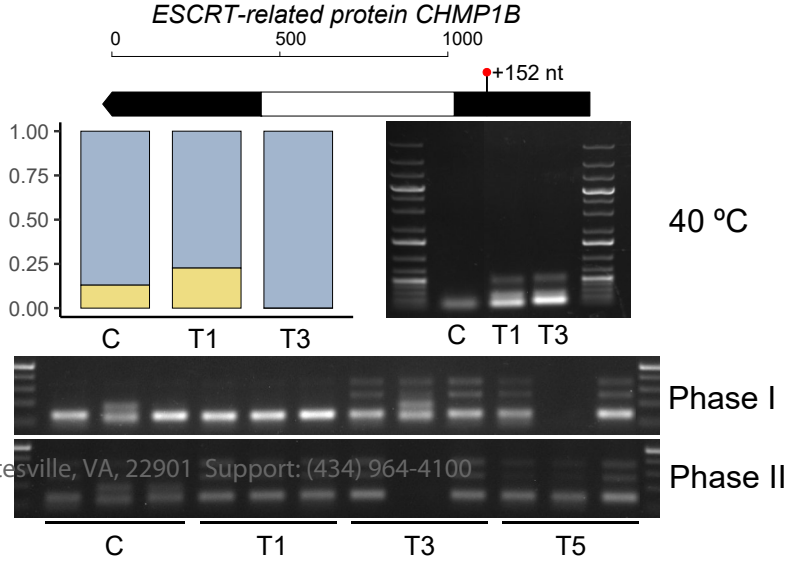
Histone deacetylase HDT2



Ubiquitin-conjugating enzyme E2 36



ESCRT-related protein CHMP1B



Expression % Isoform I Expression % Isoform II