Pra-GE-ATLAS: empowering *Pinus radiata* stress and breeding research through a comprehensive multi-omics database

Víctor Roces[1], María Jesús Cañal[1], Juan Luis Mateo[2], Luis Valledor[1*]

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

[2] Department of Computer Science, University of Oviedo, Asturias, Spain

[*] Correspondence: valledorluis@uniovi.es

Summary

In recent decades, research on model organisms have significantly increased our understanding of core biological processes in plant science. However, this focus has created a substantial knowledge bottleneck due to the limited phylogenetic and ecological spectrum covered. Gymnosperms, especially conifers, represent a molecular and ecological diversity hotspot among seed plants. Despite their importance, research on these species is notably underrepresented, primarily due to a slower pace of investigation resulting from a lack of community-based resources and databases. To fill this gap, we developed P(inus)ra(diata)-G(ene)E(xpression)-ATLAS, which consists of several tools and two main modules: transcriptomics and proteomics, presented in this work for the forestry commercial and stress-sensitive species *Pinus radiata*. We summarised and centralised all the available information to provide a comprehensive view of the gene expression landscape. To illustrate how applications of the database lead to new biological insights, we integrated multiple regulatory layers across tissues and stressors. While stress favors the retention of small introns, harmonised alternative splicing analyses reveal that genes with conifers' iconic large introns tend to be under constitutive regulation. Furthermore, the degree of convergence between stressors differed between regulatory layers, with proteomic responses remaining highly distinctive even through intergenerational memory tolerance. Overall, Pra-GE-ATLAS aims to narrow the distance between angiosperms and gymnosperms resources, deepening our understanding of how characteristic pine features have evolved. Pra-GE-ATLAS is available at https://rocesv.github.io/Pra-GE-ATLAS.

Keywords: database, atlas, gymnosperms, systems biology, splicing.

Introduction

Model organisms have played a crucial role in deepening our understanding of core biological processes, shaping research topics in plant sciences. However, there is a strong bias in the taxa studied, with angiosperms, particularly Magnoliopsida, representing 93 % of the records (Shiu and Lehti-Shiu, 2023). This creates a huge knowledge bottleneck, mainly due to the narrow phylogenetic and ecological spectrum covered. Gymnosperms and angiosperms are the two major groups of extant seed plants, exhibiting extreme differences in life spans, species diversity and reproductive biology. Moreover, gymnosperms are an ancient clade that represents four of the five main lineages of seed plants and dominate boreal and temperate forests. Despite this, gymnosperms remain largely underrepresented in plant research, specially in molecular biology (Leebens-Mack et al., 2019; Niu et al., 2022). Therefore, the establishment of model organisms in gymnosperms becomes crucial, as minimal efforts could be translated into maximal plant community benefits, leveraging the evolutionary and ecological properties of this clade.

Among gymnosperms, conifers represent the most diverse group, comprising approximately 615 species that contribute to 39 % of the world's forests. *Pinus*, with 113 species, is the largest clade and one of the most important genus of trees (Jin et al., 2021), serving as a relevant model for exploring molecular divergence in seed plants. However, pines molecular evolutionary features pose a double-edge sword. While they provide valuable ecophysiological insights, their slow growth, long-lived nature, giant genomes and high repetitive elements content are far from those attributes proper of model species (De La Torre et al., 2020). Although recent incredible genomics efforts (Niu et al., 2022), the current post-genomic era has laid the groundwork for the emergence of other “-omics” and has challenged traditional views on how genes encode phenotypes, moving beyond a genic-centered perspective. Taking advantage of this data explosion, systems biology has gained relevance for its holistic approach to modeling complex biological processes (Argelaguet et al., 2020). Multi-omics profiling is becoming quite common, promising insights into the characterisation of unexplored species lacking reference genomes. In addition, recent RNA sequencing (RNA-seq) studies indicate that transcriptomes are often underestimated, even in model organisms (S. Zhang et al., 2020). Large-scale functional genomics data, such as transcriptomics and proteomics, can provide direct evidence for a high-resolution gene expression landscape. Nonetheless, the generation of curated databases and resources derived from cumulative research outputs becomes crucial to address the focus gap in this genus and facilitate future investigations.

To tackle these challenges, we constructed P(inus)ra(diata)-G(ene)E(xpression)-ATLAS, the most extensive pine multi-omics database to date, designating the forestry commercial and stress-sensitive species *Pinus radiata* as reference. We generated new datasets and centralised all the available transcriptomics and proteomics information in a single hub, encompassing various dimensions. To showcase how the results derived from the generated resources could be used to gain biological insights, we conducted in-depth characterisation and integrated multiple regulatory layers across tissues and stressors. Constitutive regulation of long introns was observed, while stress favoured the retention of smaller introns. Additionally, the agreement between stress responses varied between regulatory layers, with proteomics revealing highly unique responses maintained through intergenerational effects, potentially mediated by the translation of specialised members of gene families. We believe that Pra-GE-ATLAS will be a valuable database, not only supporting conifers research but also contributing to the assessment of the conservation of molecular plant discoveries across a broad range of dissimilar taxa.

Results

Construction and overview of Pra-GE-ATLAS

To gain a comprehensive understanding of *P. radiata* expression landscape, we obtained, uniformly processed and integrated multi-omics data, encompassing transcriptomics and proteomics, sourced from research articles and public repositories (**fig. 1, supplementary fig. S1**). The consolidated datasets, totaling 990 Gb and 1.89 billion high-quality reads from 141 RNA-seq transcriptomic samples, and 160 Gb and 202 RAW files from 155 MS-based proteomics samples, were analysed and summarised in the P(inus)ra(diata)-G(ene)E(xpression)-ATLAS database.

We generated a high quality reference transcriptome for *P. radiata*. The Benchmarking Universal Single Copy Ortholog (BUSCO) detected high completeness (>96 %) when compared against Embryophyta (**supplementary table S1**). This quality metric, comparable to other *de novo* high quality gymnosperms transcriptomes (Visser et al., 2023), alongside an average of 80 % reads mapping back, indicate a high-quality reference appropriate for downstream analyses. The final assembly served as database for the identification and quantification of proteins. A total of 7697 proteins met all the criteria for further characterisation (see **Methods**), significantly suparssing the number reported by previous proteomics studies in this organism (Pascual et al., 2016; Pascual et al., 2017; Escandón et al., 2017; Lamelas et al., 2020; Amaral et al., 2021; García-

Campa et al., 2022; Lamelas et al., 2022), and reinforcing the need for high-quality species-specific databases in proteomic approaches (Romero-Rodríguez et al., 2014).

In summary, Pra-GE-ATLAS database was constructed based on two modules, transcriptomics and proteomics, containing the largest amount of *P. radiata* – related data up to date. It provides access to various common online tools, enabling the extrapolation of findings from other species to our reference and establishing a foundation for in-depth research on this pine species.

Transcriptomics module: Core genes transcriptionally regulated and associated regulatory features

We characterised transcriptional module grouping changes in alternative splicing (AS) and gene expression (GE) into three core sets (see **Methods**): constitutively-alternative spliced/expressed (Pan), stress-specific (Stress), and tissue-specific (Tissue) events/genes.

Global differences between GE and AS regulation were observed based on the number of shared genes/events between core sets (**fig. 2A**). PanGE, TissueGE, and their overlap constituted the biggest intersections, while most stress genes were shared with other sets. Conversely, each AS set specific events formed the largest intersections and the most substantial overlap occurred between StressAS and TissueAS. These findings suggested that GE could be the primary transcriptional mechanism, while AS seem to be more finely tuned in its regulatory role. Further inspection of AS sets trends was performed, checking the proportions between different AS types (**fig. 2B**). Consistent with previous studies (Martín et al., 2021), IR and AltAD were the most prevalent type of genome-wide AS. Nevertheless, the only prevalent type particularly enriched compared to Genome background was AltAD in AS-NR and PanAS. Thus, emphasising potential differences in functional impact and/or regulatory features associated with AS sets and types. Examining gene-level intersections (**fig. 2C**), the only set demonstrating a greater number of genes regulated by AS than GE, and with a lower overlap with the latter, was Stress.

To assess the functional relevance of AS, we researched their predicted impact on the canonical ORF (**fig. 2D**). Notably, for IR and AltAD events, we observed a significant enrichment in cases predicted to disrupt ORF for PanAS. Additionally, StressAS and TissueAS sets were predicted to significantly alter not-CDS regions, such as untranslated regions. Altogether, AS regulation appeared to be more linked to expression regulation and protein remodelling rather than functional variation in protein sequence.

To investigate genomic features related to AS regulation and type, exon and intron features were computed for each AS set (see **Methods**, **fig. 2E**). Genomic features showed a low degree of convergence across sets, with the unique exception of AltAD events significantly preferring smaller transcripts. IR PanAS events were enriched in transcripts with more and longer introns, higher upstream splice site GC content, and smaller flanking exons. Conversely, IR StressAS and TissueAS were preferentially presented in transcripts with less and smaller introns, and lower upstream GC content. In the case of ES, PanAS depicted the main hallmarks of exon definition, such as smaller target exons with longer upstream introns. Additionally, ES PanAS and TissueAS events presented lower transcript lengths, and lower and higher GC content in the target and downstream exon, respectively. Curiously, ES and AltAD exons across stresses and tissues were mainly located in the first exons. Overall, the different patterns emerged suggest that genomic features could be crucial for explaining specific regulation in splicing patterns in *Pinus*.

To provide a biological interpretation, functional enrichment analyses were conducted (**fig. 2F**). The functional terms covered by AS core sets exhibited a limited spectrum of pathways. PanAS and StressAS were the sets that shared most of the functional terms, emphasising RNA processing as an autoregulatory process. Remarkably, TissueAS showed enrichment only in redox homeostasis, representing the most divergent functional profile among sets. In contrast, GE sets included a broad range of functions, including terms essential for all types of regulation, such as phytohormones action. Briefly, the terms validated expected biological insights, such as photosynthesis enriched in PanGE and TissueGE, and secondary metabolism and redox homeostasis enriched in TissueGE and StressGE. In this case, PanGE stood out as the set with the most divergent profile.

Given the potential primary role of GE, we conducted a WGCNA (**fig. 2F**). In total, 20 co-expressed modules were identified and correlated with design factors. The largest modules tended to be related to tissues. This was illustrated by M03, which showed a positive correlation with bud and vascular tissues, and M04/M06 highlighting needle identity with some signals related to *P. pluvialis*/*F. circinatum*, respectively. However, stress-specific modules were also elucidated. Examples include M07, M09 and M10 representing *F. circinatum* stress. Using the previously introduced modules, both bud-vascular M03 and *F. circinatum* M10 revealed functional terms inherent to those design factors that were not represented in TissueGE and StressGE, such as DNA damage response, chromatin organisation and cell division.

Proteomics module: The landscape of protein information

The proteomic data generated in this study enabled us to inspect tissue signatures (**fig. 3A-C**). Enrichments of differential proteins revealed pathways involved in the functioning of each tissue, such as photosynthesis in needles, and RNA processing in buds (**fig. 3A**). Furthermore, attending to the size of the intersection between differential proteins, a decreasing trend was detected following the order needle>bud>root. These discoveries suggested that these tissues had different degrees of identity. Tissue hallmarks were complemented by volcano analyses (**fig. 3B**). The most relevant proteins remained consistent across comparisons and expanded previous mentioned pathways for roots and buds. Root tissue was mainly linked to energy metabolism (e.g ATP SYNTHASE SUBUNIT B). While functions in buds highlighted their role as a differentiating tissue with regulatory capabilities, the strongest markers were related to defense responses (e.g TERPENE SYNTHASE). Lastly, an evolutionary evaluation of protein abundance constraints was performed using Proteome Age Index (PAI) (**fig. 3C**). Although no differences were exhibited in root and bud, needles presented smaller PAI values, indicating a greater abundance of proteins with older evolutionary origins.

The stress diversity compiled in the proteomics module (**fig. 3D,E**) allowed the identification of shared functions across environmental clues, underscoring protein homeostasis and biosynthesis (**fig. 3D**). The largest intersections, which consisted of stress-stress rather than stress-control comparisons, revealed a low degree of convergence across stress proteins. Despite some common pathways being regulated for most stresses, the primary protein effectors appeared to diverge across conditions. Most relevant proteins in volcano analyses (**fig. 3E**) pointed to potential master features that were not significantly/consistently enriched at the pathway level, such as chromatin organisation (e.g HISTONE H2A) and RNA processing (e.g SM-LIKE PROTEIN LSM).

A total of 12 modules were detected by WGCNA, clustering proteins abundance across all conditions (**fig. 3F**). The largest module, M01, was related to tissues. M01 unveiled previously exposed functions and new ones such as vesicle trafficking and multiprocess/external-stimuli response (**fig. 3G**). The high resolution of protein modules revealed unknown stress dynamics. Heat stress presented modules related to each subcellular location and M09 correlating responses across nucleus and chloroplast. Nevertheless, UV stress was more specific, differing between chloroplast response and chloroplast response negatively correlated with nucleus, illustrated by M03 and M05. Despite both modules represented UV chloroplast response, their different relationship with the nucleus was also supported by the implication of distinct pathways such as protein translocation.

Application 1: Exploring unique and shared sources of transcriptional variation across multiple tissues and stressors

To exemplify applications of the generated resources, we investigated into the coordination of GE and AS in defining tissues using multi-omics factor analysis (MOFA) (**fig. 4**). Overall, MOFA inferred eight latent factors (LFs), with GE contributing to the majority of the total variance (**fig. 4A**). We examined the variance explained by the LFs and identified LF1 and LF3 as the most biologically relevant to discriminate between tissues (**fig. 4B**). Briefly, LF1 variance, mainly constituted by GE but also including remarkable AS variation, differed between needles and the rest of the tissues, while the GE-dominant LF3 variance mostly described the differences between buds/needles and vascular-related tissues.

These interpretations were supported by the top absolute loadings in each factor (**fig. 4C**). LF1 needle identity was reflected by photosynthetic required genes such as *RIBULOSE-PHOSPHATE 3-EPIMERASE* (*RPE)*. LF3 bud/needle identity was illustrated by cuticle related genes such as *CUTIN SYNTHASE2 (CUS2)*. Interestingly, LF1 specific enriched functions differed between regulatory layers and included cellular respiration and external stimuli response at the GE level, and redox homeostasis and secondary metabolism within AS layer (**fig. 4D**). LF3 specific enriched terms pointed to key divergent functions between xylem/phloem and needle/bud such as lipid metabolism and plant reproduction. From an evolutionary perspective, it seems that tissue identities described in LF1 had older origins than the distinctions covered by LF3 (**fig. 4D**). This was illustrated by enrichments in genes with younger origins (Phylostratum (PS), lower and higher PS denote older and younger origins) at the GE level in LF3 and very young gene family founder events (Phylostratum Family (PSF)) at the AS level. Additionally, these findings were further confirmed by Transcriptome Age Index (TAI) profiles which detected increasing TAI values across tissues, from older to younger origins, following the order needle>bud>xylem>phloem (**supplementary fig. S2A**).

Finally, to demonstrate that applications derived from Pra-GE-ATLAS could be translated into new biological insights, we evaluated splicing of potential isoform markers across tested tissues with different ages in an exploratory fashion (**fig. 4E**). Three genes were chosen based on differential contributions: *SIGNAL RECOGNITION PARTICLE 43 KDA, CHLOROPLASTIC (CAO), SUGAR TRANSPORTER ERD6 (ERD6),* and *COMPONENT OF CIRCADIAN EVENING COMPLEX CLOCK ELF4 (ELF4-like)*. All genes preferentially expressed the smallest isoforms in adult needles. While the largest *CAO* and medium-sized *ERD6* isoforms were common for buds and juvenile needles, most of the medium-sized *CAO* isoforms were juvenile-specific and large *ERD6* isoform was

juvenile needle-specific. Lastly, while the *ELF4-like* bud isoform diversity did not reflect a clear pattern, large *ELF4-like* isoforms appeared to be juvenile-specific.

Next, we employed MOFA multigrouped framework to evaluate the degree of convergence in *P. radiata* transcriptional responses to multiple stressors (**fig. 4F-H**). Due to a stress-related higher prevalence of particular types, such as IR (Laloum et al., 2018), we decided to split AS by type. A total of eight LFs were detected, with most of them being uniquely related to GE, the layer contributing to the majority of variance (**fig. 4F**). We identified the top three LFs as biologically relevant (**fig. 4G**). LF1 exhibited significant GE activity across all biotic stimuli, primarily associated with high stress damage. This was illustrated by high positive scores in most susceptible genotypes and stress samples under severe *D. septosporum*/*F. circinatum* and *P. pluvialis*, respectively. LF2 showed remarkable GE activity across *D. septosporum*, *F. circinatum* and heat, linked to control-stress differences. LF3 captured *F. circinatum*–specific susceptible-resistant genotype and stress-control variation for samples without genotype information. Furthermore, LF3 detected changes across all molecular layers, with higher variance explained by AS than GE.

The provided definitions were affirmed by the top loadings and functions for each LF (**fig. 4H,I**). For enrichment analyses, IR was selected as the AS representative (**fig. 4I**). LF1 shared biotic stress damage, represented by fungal-specific factors such as *ENDOCHITINASE 2 (CHTB2)* (**fig. 4H**), and specifically enriched in redox homeostasis at both transcriptional levels (**fig. 4I**). Due to the control samples tissue composition, some of the genes illustrated by LF2 were shared with tissue LF3 bud/needle (**fig. 4C**). However, new genes exclusively linked to LF2 cross-stress control-associated variation were also suggested, such as *TRANSCRIPTION FACOR BHLH62 (BHLH62)*. LF3 stress/susceptible-genotype vs control-damaged/resistant-genotype pointed to different members of *CYSTEINE-RICH RECEPTOR-LIKE PROTEIN KINASE (CRK)* stress-responsive family (Y. Zhang et al., 2023). Evolutionary-related analyses revealed that LF2 was significantly enriched in younger phylostrata than LF1 at the GE level (**fig. 4I**). This notion was further inspected by TAI profiles (**supplementary fig. S2B**), which detected significantly younger transcriptomes in earlier stress phases for heat and *D. septosporum*.

Given that GE dominant role in transcriptional variation could mask AS differential contributions between tissues and stressors, we compared the relative PSI variation of stress versus tissues in our reference, *P. radiata*, and in *A. thaliana*, *D. melanogaster* and *H. sapiens* data produced by Martín et al. (2021) (see **Methods**, **supplementary**

fig. S2C). Strikingly, we observed a contribution skewed towards stress and tissues in plant and metazoan species, respectively.

Application 2: Uncovering proteomic cross-talk among stresses, subcellular locations, and intergenerational memory

Using MOFA, we identified shared and unique sources of proteomic variation across stressors (**fig. 5A-D**), subcellular locations (**fig. 5E-H**), and intergenerational memory (**fig. 5I-L**).

In the cross-stress total proteomes framework, four LFs were identified, all considered biologically meaningful (**fig. 5A**). These LFs disentangled stress-specific variance, with heat-specific LF1 discriminating between the earliest stress timepoint, and heat-specific LF2, *F. circinatum*-specific LF3, and UV-specific LF4 showing control/recovery-stress differences (**fig. 5B**). These results were supported by top absolute loadings (**fig. 5C**). Examples include several proteome remodelling features for LF1 and LF4, photosynthetic proteins and chaperones reflecting high temperatures main targets for LF2, and defense mechanisms illustrated by OXALATE OXIDASE 1 (OXO1) for LF3. The model captured common functions such as protein biosynthesis-homeostasis, chromatin organisation and photosynthesis (**fig. 5D**). Despite non meaningful constrained abundance patterns detected by PAI profiles (**supplementary fig. S2D**), enrichments revealed shared evolutionary origin signatures among abiotic stressors, with LF4 UV-related features being relatively younger at the gene level (**fig. 5D**).

Next, we integrated abiotic stressors total, nucleus and chloroplast proteomes and 11 LFs were identified. We selected LF2, LF6, LF7 and LF9 for further biological description (**fig. 5E**). LF2 explained the most variance and was associated with stress-independent subcellular location, highlighting functions such as protein modification and chloroplast-localised features through RHO-N DOMAIN-CONTAINING PROTEIN 1, CHLOROPLASTIC (RHON1) (**fig. 5F-H**). UV-specific LF6 characterised chloroplast response, while heat-specific LF7 and LF9 involved nucleus- and chloroplast-specific stressess, respectively. Stress- and localisation-specific LFs expanded previous pathways considering subcellular information. UV-specific LF6 unveiled chloroplast protein synthesis through ATP-DEPENDENT CLP PROTEASE ADAPTER PROTEIN CLPS1, CHLOROPLASTIC (CLPS1). Heat-specific LF7 and LF9 reflected nucleus coordination and chloroplast response with HEAT SHOCK FACTOR-DNA BINDING (HSF-DNA BIND) domains, and a wide set of small heat shock proteins, respectively. Stress-specific LFs responses still shared above mentioned terms and new ones, for instance RNA processing. PAI profiles did not detect constrained abundance patterns,

but trends were appreciated when considering PAI profiles and enrichments together (**fig. 5H**, **supplementary fig. S2E**). The biggest differences between PAI values were observed between subcellular locations instead of stress timepoints. Additionally, the subcellular location with the younger profile also diverged between stressors, being the nucleus for heat and the chloroplast for UV. The latter was further emphasized with younger gene origins and family founder events enriched in UV-specific LF6 compared to heat-specific LF9.

Lastly, we interrogated whether shared cross-stress intergenerational memory variation could be detected at the protein level. To answer this, we integrated abiotic stressors chloroplast-enriched proteomes from two populations with similar genetic-backgrounds but different local-environment histories (**supplementary fig. S2F**) (García-Campa et al., 2022; Lamelas et al., 2022). Among six LFs, LF2, LF3, LF5 and LF6 were retained for subsequent analyses (**fig. 5I**). LF2 explained shared variance across populations but was only associated with UV (**fig. 5I,J**). Furthermore, LF2 displayed population-specific differences, as population E (PopE) could discriminate between all intensities, while samples under more severe conditions were merged for population T (PopT). While heat-specific LF3 also explained shared variance for both populations, LF3 did not detect population differential contributions because both populations mainly discriminated between control and severe heat intensities. Following the same pattern, LF5 and LF6 were UV- and heat-specific, respectively, discriminating the earlier timepoints in both cases. However, UV-specific LF5 reflected population differences mainly linked to earlier stress samples in PopE. Top absolute loadings (**fig. 5K**) and functional enrichments (**fig. 5L**) validated the results, illustrating protein homeostasis term shared among all LFs, and LF3 features being equivalent to previous abiotic stressors LF9 (**fig. 5G**). Since differential contributions to populations could reveal clues into intergenerational memory, we further inspected UV-specific LFs loadings. Interestingly, loadings highlighted a protein that could interact with RNA POLYMERASE SIGMA FACTOR (SIGA), essential for photosystem stoichiometry, and lignin biosynthesis reflected by CYNNAMYL ALCOHOL DEHYDROGENASE 3 (CAD3) (Bateman et al., 2021). Meaningful constrained patterns were not detected by PAI profiles (**supplementary fig. S2G**); however, UV-specific LFs enrichments uncovered slightly younger originated gene families in LF2 compared to LF5.

Discussion

In this study, we constructed Pra-GE-ATLAS, the most extensive pine multi-omics database to date (**fig. 1**). Despite pine species constituting a clear hotspot of plant

348 molecular diversity, their research remains largely underrepresented (De La Torre et al.,
349 2020; Shiu and Lehti-Shiu, 2023). To fill this gap, Pra-GE-ATLAS offers data resources
350 and tools aimed not only to assist *P. radiata* research but also to determine the extent of
351 plant biology discoveries considering more dissimilar taxa.

352 Our research presents the most exhaustive AS analysis conducted in a pine species so
353 far. Despite the divergent genomic architecture of conifers, characterised by long introns
354 (Niu et al., 2022), IR is the most prevalent AS type, consistent with prior studies (Laloum
355 et al., 2018) (**fig. 2B**). Notably, non-IR events, such as AltAD, underrated in plant
356 science, are identified as widespread and overrepresented in more sets than IR,
357 indicating potentially greater functional relevance (**fig 2B, D**). Overall, AS sequence
358 variation appears to play a more significant role in regulating gene expression and
359 protein abundance than introducing functional sequence changes. In line with earlier
360 research indicating that pine genes with longer introns are constitutively expressed (Niu
361 et al., 2022), the specific genic structure of conifers seems related to their AS regulation.
362 Long introns surrounded by small exons and small exons surrounded by long introns are
363 preferentially retained and skipped, respectively, in constitutively alternatively spliced
364 transcripts (**fig. 2E**). Conversely, stress-induced IR appears to affect small introns at the
365 beginning of the transcript. These observations highlight the innovative molecular
366 strategies adopted by conifers to keep transcription efficacy. Finally, leveraging the
367 phylogenetic position of pines, we extend the previously reported favoured regulation of
368 AS under stress in *A. thaliana* (Martín et al., 2021) as a potential general feature in seed
369 plants, contrasting with animal AS controlled in a tissue-specific manner
370 (**supplementary fig. S2C**).

371 To illustrate applications of the generated resources, we integrated multiple regulatory
372 layers across tissues and stressors. Tissue emerged as the primary driver of variation in
373 the data. Our analyses revealed distinctive patterns according to tissues' evolutionary
374 origin, following the order from more to less conserved: needle>bud/root>xylem/phloem
375 (**fig. 3C and fig. 4D; supplementary fig. S2A**). This trend aligns with the notion of
376 needle identity being more constrained, supporting expected tissue-function acquisition
377 during plant evolution (first photosynthesis), land colonisation (roots, tissue-transitions),
378 and radiation of vascular plants (xylem/phloem) (Clark et al., 2023). Next, we examined
379 tissue AS patterns, given the limited exploration of this aspect in plants. While GE
380 predominantly dictated tissue variation and could differentiate between heterogeneous
381 tissues on its own, splicing is required to distinguish between more dynamic definitions
382 (**fig. 4A, B**). Additionally, we evaluated the splicing patterns of selected potential isoform
383 markers across tissues with different ages. Interestingly, our observations extended

beyond tissue-specific patterns to include age-specific trends, such as adult tissues preferentially expressing fewer and lower isoforms (**fig. 4E**). This highlights differences in the regulation of tissue-related functions, such as photosynthesis, *CAO*, solute transport, *ERD6*, and environmental perception, *ELF4-like*, through AS. Our findings underscore how the resources provided by Pra-GE-ATLAS can be utilised to generate novel biological insights.

Stress biology is a crucial aspect of plant science; however, the convergence among stress mechanisms remains poorly characterised. The transcriptional integration revealed shared variation across stressors (**fig. 4F,G**), while the proteomic integration depicted highly unique responses (**fig. 5**). A thorough examination of total proteomes revealed evolutionary signatures shared among abiotic stressors, including similar gene family founder events. The only shared proteomic variation across abiotic stressors described stress-independent subcellular locations, with stress-linked variation remaining highly distinctive. Despite the absence of recent whole-genome duplications in pines, large-scale dispersed duplications are prevalent, and expanded gene families are associated with stress responses (Niu et al., 2022). Considering the distinct stress compositions between both modules, our discoveries suggest that the higher transcriptional convergence may be explained because transcription, as one of the closest regulatory levels to the genome, lacks direct functional effects, and its variation is associated with response, duplication-derived redundancy and stochastic stress reprogramming. In contrast, proteins, being functional components, are modulated only in specialised members of gene families due to the expensive energy investment in translation. Shared variance across stressors was exclusively linked to GE, as AS only explained variance associated with resistant/susceptible genotypes under *F. circinatum* (**fig. 4F-H**). This underscores the relevance of AS in detecting stress-related changes at smaller scales, such as genotypes, suggesting the *CRK* family, known for anti-fungal activity, as novel targets for *F. circinatum* tolerance (Amaral et al., 2022; Y. Zhang et al., 2023). Considering a broader evolutionary context, our data supported the hypothesis that earlier/mild timepoints/intensities could be related with the regulation of younger genes (**fig. 4F and fig. 5J; supplementary fig. S2B**). However, these effects are partially masked by stronger constraints detected in tissues and subcellular locations.

P. radiata, due to its long-lived nature, provides an ideal example to explore intergenerational memory. To disentangle memory, we integrated two independently published matched assays describing chloroplast-enriched proteomes under heat and UV in two populations with similar genetic-backgrounds but different local-environmental histories (García-Campa et al., 2022; Lamelas et al., 2022) (**fig. 5I-L**). Thus, variation

with differential contribution among populations could be defined as intergenerational memory consequences. We found memory evidence only under UV, associated with a higher sensitivity of PopE. Two potential non exclusive hypothesis could be highlighted. On one hand, chloroplasts could be more responsive to UV than heat stress, illustrated by younger PAI profiles and a greater variation explained by LF6 than LF9 (**fig. 5E-F; supplementary fig. S2E, G**). Therefore, depending on the organelle, certain stress modifications may be more proned to be remembered. On the other hand, given the specificity of proteomic responses, it is probable that UV range across locations was more divergent and/or plants were more sensitive to those changes (**supplementary fig. S2F**). The elevation range, which is related with UV exposure, have been described as a selective pressure on pine evolution, shifting their distribution and species diversity (Jin et al., 2021). Hence, our results suggest that the intergenerational features detected among populations may be originated from a greater susceptibility to elevation range rather than a cross-stress memory.

While we expect Pra-GE-ATLAS to be useful, we acknowledge certain limitations. As pines are considered non-model species, datasets covered a wide temporal range. Therefore, newly reported datasets, taking advantage of recent technological improvements, increased analytical resolution of MS and enhanced performance of sequencers, will significantly improve the resources presented, owing to higher throughputs. The results promoted the potential application of Pra-GE-ATLAS to test new hypothesis in both intra-species, breeding targets, and inter-species, evolutionary stress studies, contexts. Here, we focused on transcriptomics and proteomics, which are closely linked to gene expression. Given the increasing availability of -omics data, the utility of Pra-GE-ATLAS will continue to grow, providing long-term support with annual updates. Our next steps involve the establishment of variation and metabolomic modules, and, once the genome of *P. radiata* is released, compute high-quality gene models. In summary, Pra-GE-ATLAS aims to narrow the distance between angiosperms and gymnosperms resources and designates the commercial and stress-sensitive species *P. radiata* as a reference for understanding the intriguing evolutionary features of pines.

Experimental procedures

An overview of the bioinformatic workflow used in this study is shown in **supplementary fig. S1**.

Plant materials

To generate the tissues proteomic dataset, we sampled seedlings (one year-old) and adult trees which are maintained under routine management at Plant Physiology Laboratory of the University of Oviedo. Roots (growing tips), young (growth period one cm length) and adult (> 12 cm, mature) needles, and stem (less lignified and mature), apical floral buds were collected. Three biological replicates for each tissue were constituted pooling two different plants.

Protein extraction, digestion, fractionation and MS acquisition

Protein extraction was performed following phenol-sodium dodecyl sulfate (SDS) protocol according to Valledor et al. (2014). Initial amount varied from 75 to 250 mg of fresh weight depending on the processed tissue. As protein samples were dissolved with the detergent SDS, sixty µg of proteins were in gel fractionated and digested as described by Valledor and Weckwerth (2014). Peptides were cleaned, extracted and desalted as previously described (Valledor and Weckwerth, 2014). Peptides were analysed in a HPLC-MS/MS Orbitrap Fusion spectrometer (ThermoFisher Scientific), employing a 60-min gradient starting with 0.1 % formic acid and with 80 % acetonitrile as the mobile phase.

RT-PCR analysis

Total RNA was extracted following Valledor et al. (2014). RNA concentration was determined by a Navi UV/Vis Nano Spectrophotometer and its integrity was checked by agarose gel electrophoresis. Next, cDNA was obtained by RevertAid kit (ThermoFisher Scientific) using random hexamers as primers following manufacturer's instructions. RT-PCR was performed with BesTaq polymerase (**supplementary table S2**). Primers for each AS event were designed to amplify multiple splice variants in a single reaction.

Data collection

We collected all transcriptomic data from *P. radiata* (term: “Pinus radiata”) available from the NCBI Short Read Archive with associated published reference to ensure high quality data (**supplementary table S1**, last: February 2022). The transcriptomic data collection covered five tissues (bud, xylem, phloem, needle and megagametophyte), one abiotic stress (heat), and three biotic stresses (*Fusarium circinatum*, *Dothistroma septosporum* and *Phytophthora pluvialis*).

We collected all proteomic data from *P. radiata* (term: “Pinus radiata”) available based on PRIDE and PubMed search. Publication was required to ensure high quality data (**supplementary table S3**, last: October 2023). The proteomic data collection covered

three tissues generated in this study (root, needle and bud), one biotic stress (*F. circinatum*), and two abiotic stresses (heat and ultraviolet (UV)) over three different subcellular locations (total proteins, nucleus and chloroplast).

Transcriptomics data processing

Trimmomatic v0.39 (Bolger et al., 2014), SortMeRNA (Kopylova et al., 2012) and Rcorrector (Song and Florea, 2015) were applied to remove adapters and low-quality reads, filter rRNA and correct sequencing errors, respectively. *Fusarium circinatum* reads were discarded mapping to FSP34 genome using bowtie2 (Langmead and Salzberg, 2012). Each condition was assembled independently and reads were normalised for those conditions exceeding 200 million reads using Trinity v2.15.1 (Grabherr et al., 2011). Cleaned reads were assembled using Trinity v2.15.1 and rnaSPADES v3.14 (Bushmanova et al., 2019). Lastly, assemblies were concatenated through EvidentialGene tr2aacds v2017.12.21 pipeline to reduce redundancy and select for the optimal assembled transcripts. The consensus assembly, based on EvidentialGene primary transcripts, was evaluated using BUSCO v5.2.2 (Simão et al., 2015), Trinity v2.15.1 Ex90N50, and backmapping (**supplementary table S1**).

For subsequent procedures, a final assembly was created concatenating EvidentialGene primary transcripts with alternate transcripts. This was achieved after applying cd-hit -c 0.905 (Fu et al., 2012) within the alternate set. The final assembly was functionally annotated by EggNOG-mapper v2 (Cantalapiedra et al., 2021), Mercator4 v6 (Schwacke et al., 2019), Interproscan v5.44.79 (Jones et al., 2014) and dammit v1.

Salmon v1.5.2 (Patro et al., 2017) was employed to quantify expression levels against *Pinus taeda* v2.0.1 (the closest species with an available genome), obtained from TreeGenes (Falk et al., 2018; Jin et al., 2021). *De novo* splicing events were identified, classified, and quantified using KisSplice v2.6.2 (-k 51 -C 0.05) (Sacomoto et al., 2012), Kiss2refgenome v2.0.8, and kissDE v1.4.0, respectively. sva v3.48.0 (Leek et al., 2012) was employed to remove raw counts unwanted variation derived from study/sequencing-type. GenEra v1.4 (Barrera-Redondo et al., 2023) was then applied to identify gene families, their founder events, and determine the ages of *P. taeda* genes. NR database was completed adding gymnosperms data (*Abies alba, Ginkgo biloba, Gnetum montanum, Picea abies, Pinus lambertiana, Pseudotsuga menziesii* and *Sequoiadendron giganteum;* TreeGenes).

Definitions of core AS and GE sets

Gene expression (GE) and alternative splicing (AS) trends were grouped into three core sets. To define the three core sets, we followed a similar approach as described by Martín et al., (2021):

Pan core set referred to genes/events that are expressed/alternatively spliced across most sample types. For PanAS set, we required sufficient read coverage in at least 20 % of the total samples. AS read coverage was defined based on kissDE default. We then defined the PanAS events as those with a Percent-Splice-In (PSI) between 0.1 and 0.9 (alternatively spliced) in > 70 % of samples with sufficient read coverage. For PanGE set only genes with an expression level of at least 20 normalised counts in at least 70 % of samples were considered.

Tissue core set referred to genes/events that are up/down regulated across tissues. Megagametophyte was excluded and phloem-xylem samples were grouped as vascular tissue due to the low number of samples. TissueAS required events with sufficient read coverage in at least two replicates for all tissue types, and the absolute difference in PSI between the target tissue and the average of the other tissues must be of at least 0.25. Then, genes with a median expression level of at least 5 normalised counts in at least one tissue type and a fold change of at least 3 in the same direction with related to all other tissues types were kept as TissueGE. DESeq2 v1.40.1 (Love et al., 2014) was applied to compute fold change.

To identify stress-regulated AS and GE, each stress experiment was compared against its respective matched control. Since the majority of the stress transcriptomic and proteomic experiments involved sampling similar phases, we uniformly renamed the different time points based on stress duration/intensity. AS events needed to have sufficient read coverage in at least two stress and control replicates for each of the five stress experiments studied. Then, only events with an absolute PSI difference of at least 0.15 in the same direction between stress and control conditions for at least two out of five stress experiments were retained as StressAS. Regarding StressGE, the same criteria was required considering at least 5 normalised counts and a fold-change of at least 2 as coverage and magnitude thresholds, respectively. Thus, ensuring that features are expressed/spliced and avoiding ambiguous regulation across stresses in opposite directions.

We established control groups for set comparison: background (Genome) and non-regulated (NR). Genome comprised events and genes that met the same coverage criteria and filters as those used to define each core set, but without any PSI-/fold

554 change-related requirements. AS-NR group was determined on basis of each AS core
555 set. For TissueAS, AS-NR events were those alternatively spliced and with an absolute
556 PSI difference <0.05 for each tissue versus the rest. For StressAS, AS-NR events were
557 those alternatively spliced in at least one sample and with an absolute PSI difference
558 <0.05 in at least one stress experiment. Finally, to obtain a common AS-NR, we retained
559 events that were part of both AS-NR sets. The intersections between genes and events
560 were assessed using nVennR v0.2.3 (Pérez-Silva et al., 2018).

561 ## Predicted protein impact and genomic regulatory feature analysis

562 Splicing variation effect were determined using custom scripts employing the following
563 approach: (i) Kiss2refgenome v2.0.8 coordinates and GTF annotations were used to
564 determine if the variation occurred inside/outside of coding-sequence (CDS), (ii) CDS-
565 affecting isoforms were examined to detect if variation led to the introduction of
566 premature termination codons (PTCs), (iii) CDS-affecting isoforms without PTCs
567 underwent further evaluation to check if the variation disrupted the open reading frame
568 (ORF) frameshift.

569 To compare exon and intron features associated with different AS core sets, Matt v1.3.1
570 (Gohr and Irimia, 2019) was employed. Briefly, Matt *cmpr_introns*, for intron retention
571 (IR) events, and *cmpr_exons*, for exon skipping (ES) and alternative acceptor/donor site
572 (altAD, both 5' and 3'), commands were employed to extract and compare multiple intron
573 and exon genomic features associated with AS regulation. Statistical significance was
574 addressed by comparing each set to Genome.

575 ## Proteomics data processing

576 Proteome Discoverer 2.2 (Thermo Fisher Scientific, USA) along with the Sequest-HT
577 and MS-Amanda algorithms, were employed for peptide processing, and protein
578 identification-quantification, establishing at least one high-confidence unique peptide
579 umbral for protein identification and one peptide (unique/razor) per protein for label-free
580 quantification. The final assembly underwent six-frame translation, and peptides
581 exceeding 50 amino acids were retained and used as database.

582 Each proteome underwent preprocesing using pRocessomics v.0.1.13
583 (github.com/Valledor/pRocessomics). In summary, missing values and additional
584 replicates for the *Fusarium circinatum*, heat stress total, and UV nucleus proteomes,
585 were imputed using random forest method, with a threshold of 34 %. Variables present
586 in less than 50 % of samples were dropped out. Abundance values were normalised by
587 sample-centric approach and multiplied by the average intensity of all samples. Protein
588 abundances were transformed with a log10(+1.1) for subsequent analyses. sva v3.48.0

was employed to remove abundance unwanted variation. GenEra v1.4 was employed, as mentioned above, using *P. radiata* proteins as query.

Proteins differential analyses

Statistical analyses of protein-level differential abundance were carried out using the sva v3.48.0 coupled to limma v3.56.2 (Ritchie et al., 2015) employing FDR < 0.05 as threshold. For volcano, proteins were required to exhibit a log2(fold change)>1.5 to be considered biologically relevant. The intersections between differential proteins were assessed using UpSetR v1.4.0 (Conway et al., 2017).

Co-expression network analyses

Weighted Gene Co-expression Analysis (WGCNA) was conducted using WGCNA v1.72-1 (Langfelder and Horvath, 2008) to identify highly co-expressed genes (DESeq2 VST) and proteins (log10(+1.1)). A signed-hybrid type of adjacency matrix was constructed, with $\beta = 7/9$ for proteins/genes, using biweight midcorrelation. Hierarchical clustering was performed, and co-expression modules were identified using dynamic tree cut height of 0.3 and a minimum module size of 30. Modules were named based on their size. Module eigengenes were employed to compute correlations between modules and design factors (traits). Only correlations with an adjusted-P<0.05 were considered. Module membership was computed based on the correlation between genes and module eigengenes for each module.

Enrichment analyses

Enrichment analyses using Mercator4 terms were conducted using fgsea v1.26.0. Briefly, for the transcriptomics module, we applied an overrepresentation analysis (adjusted-P < 0.1). Meanwhile, for the proteins, gene set enrichment analyses (adjusted-P < 0.1) were performed using limma-derived statistics and modules membership.

Evolutionary transcriptomics and proteomics

To investigate the potential existence of evolutionary constraints, we employed myTAI v0.9.3 (Drost et al., 2018). For evolutionary transcriptomics analyses, *P. taeda* gene ages and VST expression data were employed. For evolutionary proteomics analyses, *P. radiata* protein ages and log10(+1.1) abundance data were used. In both cases, the Transcriptome/Proteome Age Index (TAI/PAI) approach was followed for gene/protein age evaluation. The significance of evolutionary constraint was assessed using the FlatLineTest.

Relative contribution of tissues and stress conditions to global PSI variation

For the comparisons of the relative contribution to the total PSI variation of tissue versus stress, we adopted a similar approach as described by Martín et al. (2021). We incorporated data from Martín et al. (2021) for *Arabidopsis thaliana, Drosophila melanogaster* and *Homo sapiens.* Due to the limited number of abiotic stress transcriptomic experiments in *P. radiata* and to find general stress trends, we chose to merge abiotic and biotic experiments. We required that AS events must have read coverage in all tissue types and three stress experiments, with a global PSI variation exceeding 10.

Inference of hidden factors from multiple stresses and tissues sources

The Inference of sources of variation was carried out using MOFA2 (docker latest image: 2e858d684c5f) (Argelaguet et al., 2020). To characterise transcriptional variation in tissues, an ungrouped framework was executed, considering expression (VST) and splicing (PSI) as two distinct regulatory layers. Only the top 10,000 features with the highest variance (HVF) were considered. For the assessment of transcriptional variation between stresses, a grouped framework was employed, splitting AS by type and considering the top 10,000 and 5,000 HVFs for expression and splicing-related layers, respectively. To evaluate proteomic (log10(+1.1)) variation between stresses, three different grouped frameworks were computed, removing low variance features in each model. In all cases, model training was performed with maxiter = 100,000 and convergence_mode = “slow”. Each biologically relevant latent factor underwent enrichment analysis (adjusted-P < 0.1).

Database resource

We developed P(inus)ra(diata)-G(ene)E(xpression)-ATLAS database, a comprehensive multi-omics hub aimed to provide public access to the information generated in this work. Pra-GE-ATLAS features three main tools: 1) Search section with interactive tables and heatmaps for quick retrieval of protein-, transcript-, splicing event-information. 2) Diamond BLASTP sequence alignment (Buchfink et al., 2021). 3) Shiny-application to compute *P. radiata* orthologs based on our consensus assembly using orthologr (Drost et al., 2015). Pra-GE-ATLAS is available at https://rocesv.github.io/Pra-GE-ATLAS.

Accesion numbers

All the data generated in this study are available at Pra-GE-ATLAS database https://rocesv.github.io/Pra-GE-ATLAS and https://doi.org/10.5281/zenodo.10494507. The code used in this work is available at https://github.com/RocesV/Pra-GE-

ATLAS manuscript. The mass spectrometry data have been deposited to the ProteomeXchange Consortium via the PRIDE (Perez-Riverol et al., 2022) partner repository with the dataset identifier PXD047869 (Reviewer account details: Username: reviewer_pxd047869@ebi.ac.uk; Password: wL7XdldN).

Aknowledgements

VR was supported by a fellowship from Spanish Minitry of Universities (FPU18/02953). We are grateful to Picasso, the supercomputing cluster of the University of Málaga, for kindly providing computational resources.

Short legends for Supporting Information

Figure S1. Overview of the methods workflow.

Figure S2. Evolutionary transcriptomics and proteomics patterns, tissues vs stress contribution to global PSI variation in different species and populations experimental design. A) Transcriptomic Age Index (TAI) of tissues corresponding high values to younger genes. Flat line test p-value < 0.05 highlights a significant evolutionary pattern. **B)** Transcriptomic Age Index (TAI) corresponding high values to younger genes. HS = heat; DO = *Dothistroma*; PH = *Phytophthora*; FU = *Fusarium*. Individual stress experiments with significant evolutionary patterns are highlighted with "*". **C)** Comparison of the relative contribution to the total PSI variation of the tissue samples vs stress experiments in each species. The total PSI variation for each AS event is calculated as the sum of two relative contributions: (i) the PSI range across tissues, (ii) the maximum difference between PSI among stress experiments (see **Methods**). Colours represent the number of AS events found on each intersection between the relative contributions (in percentage) for each set of samples. **D)** Proteomic age index (PAI) of all stresses total proteomes corresponding high values to younger protein genes. **E)** Proteomic age index (PAI) of abiotic stresses all proteomes corresponding high values to younger protein genes. **F)** Intergenerational stress populations experimental design. The divergent local-environment conditions involved, setting Population T as reference, +50 meters elevation, +44 mm mean rainfall, and +1.72 mean ºC. Nevertheless, PopE plants were fertirrigated during the dry months. **G)** Proteomic age index (PAI) of abiotic stresses chloroplast proteomes corresponding high values to younger protein genes.

Table S1. Transcriptomic data collection and consensus assembly evaluation.

Table S2. Primers used for the validation of tissues/age-induced alternative splicing.

Table S3. Proteomic data collection.

Conflict of interest

The authors declare there is no conflict of interest.

Author contributions

VR and LV conceived the study. VR and JLM designed the research. LV and performed proteomic experiments. VR and LV collected the data. VR performed computational analyses, built the database and figures, analysed-interpreted the data and wrote the manuscript draft under supervision of LV, JLM and MJC. All authors revised, read, and approved the final manuscript.

References

Amaral, J., Lamelas, L., Valledor, L., Castillejo, MA., Alves, A., and Pinto, G. (2021). Comparative proteomics of Pinus-Fusarium circinatum interactions reveal metabolic clues to biotic stress resistance. *Physiologia Plantarum*, *173(4), 2142-2154*.

Amaral, J., Valledor, L., Alves, A., Martín-García, J., and Pinto, G. (2022). Studying tree response to biotic stress using a multi-disciplinary approach: The pine pitch canker case study. In *Frontiers in Plant Science* (Vol. 13). Frontiers Media S.A.

Argelaguet, R., Arnol, D., Bredikhin, D., Deloro, Y., Velten, B., Marioni, J. C., and Stegle, O. (2020). MOFA+: A statistical framework for comprehensive integration of multi-modal single-cell data. *Genome Biology*, *21*(1).

Barrera-Redondo, J., Lotharukpong, J. S., Drost, H. G., and Coelho, S. M. (2023). Uncovering gene-family founder events during major evolutionary transitions in animals, plants and fungi using GenEra. *Genome Biology*, *24*(1).

Bateman, A., Martin, M. J., Orchard, S., Magrane, M., Agivetova, R., Ahmad, S., Alpi, E., Bowler-Barnett, E. H., Britto, R., Bursteinas, B., Bye-A-Jee, H., Coetzee, R., Cukura, A., Silva, A. da, Denny, P., Dogan, T., Ebenezer, T. G., Fan, J., Castro, L. G., … Zhang, J. (2021). UniProt: The universal protein knowledgebase in 2021. *Nucleic Acids Research*, *49*(D1), D480–D489.

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

Buchfink, B., Reuter, K., and Drost, H. G. (2021). Sensitive protein alignments at tree-of-life scale using DIAMOND. *Nature Methods*, *18*(4), 366–368.

719 Bushmanova, E., Antipov, D., Lapidus, A., and Prjibelski, A. D. (2019). RnaSPAdes: A
720 de novo transcriptome assembler and its application to RNA-Seq data. *GigaScience*,
721 *8*(9).

722 Cantalapiedra, C. P., Herɲandez-Plaza, A., Letunic, I., Bork, P., and Huerta-Cepas, J.
723 (2021). eggNOG-mapper v2: Functional Annotation, Orthology Assignments, and
724 Domain Prediction at the Metagenomic Scale. *Molecular Biology and Evolution*, *38*(12),
725 5825–5829.

726 Clark, J. W., Hetherington, A. J., Morris, J. L., Pressel, S., Duckett, J. G., Puttick, M. N.,
727 Schneider, H., Kenrick, P., Wellman, C. H., and Donoghue, P. C. J. (2023). Evolution of
728 phenotypic disparity in the plant kingdom. *Nature Plants*, *9*(10), 1618–1626.

729 Conway, J. R., Lex, A., and Gehlenborg, N. (2017). UpSetR: An R package for the
730 visualization of intersecting sets and their properties. *Bioinformatics*, *33*(18), 2938–2940.

731 de La Torre, A. R., Piot, A., Liu, B., Wilhite, B., Weiss, M., and Porth, I. (2020). Functional
732 and morphological evolution in gymnosperms: A portrait of implicated gene families. In
733 *Evolutionary Applications* (Vol. 13, Issue 1, pp. 210–227). Wiley-Blackwell.

734 Drost, H. G., Gabel, A., Grosse, I., and Quint, M. (2015). Evidence for active
735 maintenance of phylotranscriptomic hourglass patterns in animal and plant
736 embryogenesis. *Molecular Biology and Evolution*, *32*(5), 1221–1231.

737 Drost, H. G., Gabel, A., Liu, J., Quint, M., and Grosse, I. (2018). MyTAI: Evolutionary
738 transcriptomics with R. *Bioinformatics*, *34*(9), 1589–1590.

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

742 Falk, T., Herndon, N., Grau, E., Buehler, S., Richter, P., Zaman, S., Baker, E. M.,
743 Ramnath, R., Ficklin, S., Staton, M., Feltus, F. A., Jung, S., Main, D., and Wegrzyn, J. L.
744 (2018). Growing and cultivating the forest genomics database, TreeGenes. *Database*,
745 *2018*(2018).

746 Fu, L., Niu, B., Zhu, Z., Wu, S., and Li, W. (2012). CD-HIT: Accelerated for clustering the
747 next-generation sequencing data. *Bioinformatics*, *28*(23), 3150–3152.

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

751 Gohr, A., and Irimia, M. (2019). Matt: Unix tools for alternative splicing analysis.
752 *Bioinformatics*, *35*(1), 130–132.

753 Grabherr, M. G., Haas, B. J., Yassour, M., Levin, J. Z., Thompson, D. A., Amit, I.,
754 Adiconis, X., Fan, L., Raychowdhury, R., Zeng, Q., Chen, Z., Mauceli, E., Hacohen, N.,
755 Gnirke, A., Rhind, N., di Palma, F., Birren, B. W., Nusbaum, C., Lindblad-Toh, K., …
756 Regev, A. (2011). Full-length transcriptome assembly from RNA-Seq data without a
757 reference genome. *Nature Biotechnology*, *29*(7), 644–652.

758 Jin, W.-T., Gernandt, D. S., Wehenkel, C., Xia, X.-M., Wei, X.-X., and Wang, X.-Q.
759 (2021). Phylogenomic and ecological analyses reveal the spatiotemporal evolution of
760 global pines. *PNAS*, *118*, 2022302118.

761 Jones, P., Binns, D., Chang, H. Y., Fraser, M., Li, W., McAnulla, C., McWilliam, H.,
762 Maslen, J., Mitchell, A., Nuka, G., Pesseat, S., Quinn, A. F., Sangrador-Vegas, A.,
763 Scheremetjew, M., Yong, S. Y., Lopez, R., and Hunter, S. (2014). InterProScan 5:
764 Genome-scale protein function classification. *Bioinformatics*, *30*(9), 1236–1240.

765 Kopylova, E., Noé, L., and Touzet, H. (2012). SortMeRNA: Fast and accurate filtering of
766 ribosomal RNAs in metatranscriptomic data. *Bioinformatics*, *28*(24), 3211–3217.

767 Laloum, T., Martín, G., and Duque, P. (2018). Alternative Splicing Control of Abiotic
768 Stress Responses. In *Trends in Plant Science* (Vol. 23, Issue 2, pp. 140–150).

769 Lamelas, L., Valledor, L., Escandón, M., Pinto, G., Cañal, M. J., and Meijón, M. (2020).
770 Integrative analysis of the nuclear proteome in Pinus radiata reveals thermopriming
771 coupled to epigenetic regulation. *Journal of Experimental Botany*, *71*(6), 2040–2057.

772 Lamelas, L., Valledor, L., López-Hidalgo, C., Cañal, M. J., and Meijón, M. (2022).
773 Nucleus and chloroplast: A necessary understanding to overcome heat stress in Pinus
774 radiata. *Plant Cell and Environment*, *45*(2), 446–458.

775 Langfelder, P., and Horvath, S. (2008). WGCNA: An R package for weighted correlation
776 network analysis. *BMC Bioinformatics*, *9*.

777 Langmead, B., and Salzberg, S. L. (2012). Fast gapped-read alignment with Bowtie 2.
778 *Nature Methods*, *9*(4), 357–359.

779 Leebens-Mack, J. H., Barker, M. S., Carpenter, E. J., Deyholos, M. K., Gitzendanner, M.
780 A., Graham, S. W., Grosse, I., Li, Z., Melkonian, M., Mirarab, S., Porsch, M., Quint, M.,
781 Rensing, S. A., Soltis, D. E., Soltis, P. S., Stevenson, D. W., Ullrich, K. K., Wickett, N. J.,
782 DeGironimo, L., … Wong, G. K. S. (2019). One thousand plant transcriptomes and the
783 phylogenomics of green plants. *Nature*, *574*(7780), 679–685.

784 Leek, J. T., Johnson, W. E., Parker, H. S., Jaffe, A. E., and Storey, J. D. (2012). The
785 SVA package for removing batch effects and other unwanted variation in high-throughput
786 experiments. *Bioinformatics*, *28*(6), 882–883.

787 Love, M. I., Huber, W., and Anders, S. (2014). Moderated estimation of fold change and
788 dispersion for RNA-seq data with DESeq2. *Genome Biology*, *15*(12).

789 Martín, G., Márquez, Y., Mantica, F., Duque, P., and Irimia, M. (2021). Alternative
790 splicing landscapes in Arabidopsis thaliana across tissues and stress conditions highlight
791 major functional differences with animals. *Genome Biology*, *22*(1).

792 Niu, S., Li, J., Bo, W., Yang, W., Zuccolo, A., Giacomello, S., Chen, X., Han, F., Yang,
793 J., Song, Y., Nie, Y., Zhou, B., Wang, P., Zuo, Q., Zhang, H., Ma, J., Wang, J., Wang,
794 L., Zhu, Q., … Wu, H. X. (2022). The Chinese pine genome and methylome unveil key
795 features of conifer evolution. *Cell*, *185*(1), 204-217.e14.

796 Pascual, J., Alegre, S., Nagler, M., Escandón, M., Annacondia, M. L., Weckwerth, W.,
797 Valledor, L., and Cañal, M. J. (2016). The variations in the nuclear proteome reveal new
798 transcription factors and mechanisms involved in UV stress response in Pinus radiata.
799 *Journal of Proteomics*, *143*, 390–400.

800 Pascual, J., Canal, M. J., Escandon, M., Meijon, M., Weckwerth, W., and Valledor, L.
801 (2017). Integrated physiological, proteomic, and metabolomic analysis of ultra violet (UV)
802 stress responses and adaptation mechanisms in pinus radiata. *Molecular and Cellular*
803 *Proteomics*, *16*(3), 485–501.

804 Patro, R., Duggal, G., Love, M. I., Irizarry, R. A., and Kingsford, C. (2017). Salmon
805 provides fast and bias-aware quantification of transcript expression. *Nature Methods*,
806 *14*(4), 417–419.

807 Perez-Riverol, Y., Bai, J., Bandla, C., García-Seisdedos, D., Hewapathirana, S.,
808 Kamatchinathan, S., Kundu, D. J., Prakash, A., Frericks-Zipper, A., Eisenacher, M.,
809 Walzer, M., Wang, S., Brazma, A., and Vizcaíno, J. A. (2022). The PRIDE database
810 resources in 2022: A hub for mass spectrometry-based proteomics evidences. *Nucleic*
811 *Acids Research*, *50*(D1), D543–D552.

812 Pérez-Silva, J. G., Araujo-Voces, M., and Quesada, V. (2018). NVenn: Generalized,
813 quasi-proportional Venn and Euler diagrams. *Bioinformatics*, *34*(13), 2322–2324.

814 Ritchie, M. E., Phipson, B., Wu, D., Hu, Y., Law, C. W., Shi, W., and Smyth, G. K. (2015).
815 Limma powers differential expression analyses for RNA-sequencing and microarray
816 studies. *Nucleic Acids Research*, *43*(7), e47.

Romero-Rodríguez, M. C., Pascual, J., Valledor, L., and 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.

Sacomoto, G. A. T., Kielbassa, J., Chikhi, R., Uricaru, R., Antoniou, P., Sagot, M. F., Peterlongo, P., and Lacroix, V. (2012). Kissplice: De-novo calling alternative splicing events from RNA-seq data. *BMC Bioinformatics*, *13*.

Schwacke, R., Ponce-Soto, G. Y., Krause, K., Bolger, A. M., Arsova, B., Hallab, A., Gruden, K., Stitt, M., Bolger, M. E., and Usadel, B. (2019). MapMan4: A Refined Protein Classification and Annotation Framework Applicable to Multi-Omics Data Analysis. *Molecular Plant*, *12*(6), 879–892.

Shiu, S.-H., and Lehti-Shiu, M. D. (2023). Evolution of research topics and paradigms in plant sciences. *BioRxiv*.

Simão, F. A., Waterhouse, R. M., Ioannidis, P., Kriventseva, E. v., and Zdobnov, E. M. (2015). BUSCO: Assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics*, *31*(19), 3210–3212.

Song, L., and Florea, L. (2015). Rcorrector: Efficient and accurate error correction for Illumina RNA-seq reads. *GigaScience*, *4*(1).

Valledor, L. and Weckwerth, W. (2014). An improved detergent-compatible gel-fractionation LC-LTQ-orbitrap-MS workflow for plant and microbial proteomics. *Methods in Molecular Biology*, 1072, 347–358.

Valledor, L., Escandón, M., Meijón, M., Nukarinen, E., Cañal, M. J., and 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.

Visser, E. A., Kampmann, T. P., Wegrzyn, J. L., and Naidoo, S. (2023). Multispecies comparison of host responses to Fusarium circinatum challenge in tropical pines show consistency in resistance mechanisms. *Plant Cell and Environment*, *46*(5), 1705–1725.

Zhang, S., Li, R., Zhang, L., Chen, S., Xie, M., Yang, L., Xia, Y., Foyer, C. H., Zhao, Z., and Lam, H. M. (2020). New insights into Arabidopsis transcriptome complexity revealed by direct sequencing of native RNAs. *Nucleic Acids Research*, *48*(14), 7700–7711.

Zhang, Y., Tian, H., Chen, D., Zhang, H., Sun, M., Chen, S., Qin, Z., Ding, Z., and Dai, S. (2023). Cysteine-rich receptor-like protein kinases: emerging regulators of plant stress responses. In *Trends in Plant Science* (Vol. 28, Issue 7, pp. 776–794).

Figure legends

Figure 1. Overview of Pra-GE-ATLAS. Pra-GE-ATLAS is a refined multi-omics platform compiling the largest transcriptomics and proteomics collections to date for *P. radiata*. Pra-GE-ATLAS provides user-friendly search functionalities and tools to explore and analyse processed tissue- and stress-related changes, as well as to extrapolate data from other species to this reference. Pra-GE-ATLAS DB is available at: https://rocesv.github.io/Pra-GE-ATLAS/.

Figure 2. Transcriptional module global description. A) Venn diagrams showing all intersections between gene expression (GE, upper) and alternative splicing (AS, lower) core sets (see **Methods**). Pan = genes/events that are expressed/alternatively spliced in the vast majority of samples; Stress = genes/events that are up/down regulated in stress experiments; Tissue = genes/events that are up/down regulated across tissues. **B)** Proportion of each type of AS event in each AS core set (see **Methods**). NR = non-regulated; Genome = background set constituted by events that passed the same coverage criteria and filters; IR = intron retention; ES = exon skipping; AltAD = alternative splice acceptor/donor sites; AS-Unknown = events that passed coverage criteria and filters without classification. Significant enrichment compared to genome background are marked with "*". **C)** Venn diagrams showing intersections between gene expression (GE, left) and alternative splicing (AS, right) genes for each core set. **D)** Percentage of intron retention (first), exon skipping (second) and alternative splice donor and acceptor sites (Alternative A/D, third) events belonging to the different AS core sets located out/in CDS regions. Among the latter category (in CDS regions), the percentage of events with potential effects in protein levels are indicated. Gen. = genome background; Not-CDS = outside CDS regions; PTC = sequence variation inside CDS regions introduce premature termination codons; Disrupt = sequence variation inside CDS regions force out of frame reading; Change = sequence variation inside CDS change CDS region sequence. Significant enrichment compared to genome background are marked with "*". **E)** Schematic representation of genomic regulatory features associated with each AS core sets for introns (first) and exons (second and third). Only features with statistical significant differences for each AS core set were represented. Arrows summarise which features show significant differences respect to Genome background and the direction of these differences (higher-red or lower-blue). "X" indicates no statistically significant

difference. Intron features (first) include (from top to bottom and left to right): length of the upstream (UP) exon, target intron, polypyrimidine tract (PT) and downstream (DO) exon; GC content of the upstream 5' splice region; number of introns; distance between branch point (BP) and 3' splice site (ss); score of the polypyrimidine tract; rank and/or position of the target intron. Exon features for exon skipping (second) include (from top to bottom and left to right): length of the upstream exon, upstream intron, upstream polypyrimidine tract, target exon, downstream intron, downstream exon and transcript; GC content of the target exon, 5' splice region and downstream exon; score of the upstream branch point, polypyrimidine tract, 5' splice region and downstream branch point; rank and/or position of the target exon. Exon features for alternative acceptor donor site (third) include (from top to bottom and left to right): length of the downstream exon and transcript; GC content of the upstream 5' splice region, target exon and downstream exon; score of the upstream and downstream branch points; rank and/or position of the target exon. **F)** Heatmaps depicting significant overrepresented Mercator functional categories (p-value adjusted < 0.1; -log10(p-adjusted)) and network modules-trait correlations (p-value adjusted < 0.05; pearson). Biosynthe = biosynthesis; hom = homeostasis; CHO = carbohydrate; met = metabolism; reg = regulation; org = organisation; PS = photosynthesis; resp = response; cellular resp = cellular respiration; transloc = translocation; mod = modification; dmg = damage.

Figure 3. Protein module global description. A) From left to right: heatmaps showing Mercator functional categories normalised enrichment scores (NES, first), significance (p-value adjusted < 0.1, second) and matrix layout (third) for all intersections of differential proteins between tissues. Letters in significance heatmap highlight for which tissue the functional term is significantly enriched. B = Bud; N = Needle. **B)** Summary of volcano analyses (see **Methods**) indicating top marker proteins for each differential contrast between tissues. **C)** Proteomic Age Index (PAI) corresponding high values to younger protein genes. Flat line test p-value < 0.05 highlight a significant evolutionary pattern. **D)** From left to right: heatmaps showing Mercator functional categories normalised enrichment scores (NES, first), significance (p-value adjusted < 0.1, second) and matrix layout (third) for all intersections of differential proteins between stress experiments. Letters in significance heatmap highlight for which condition the functional term is significantly enriched. H = Heat; U = UV; C = Control; R = Recovery. **E)** Summary of volcano analyses (see **Methods**) indicating top marker proteins for each differential contrast between stress conditions. **F)** Heatmap depicting significant network modules-trait correlations (p-value adjusted < 0.05; Pearson). T1-T4 correspond to low-very high stress intensities. FU = *Fusarium*. **G)** From left to right: heatmaps showing Mercator

functional categories normalised enrichment scores (NES, first) and significance (p-value adjusted < 0.1, second) for all network modules. "*" in significance heatmap highlights for which particular module the functional term is significant. Met = metabolism; hom = homeostasis; resp = response; cellular resp = cellular respiration; mod = modification; reg = regulation; org = organisation; PS = photosynthesis; act = action; dmg = damage.

Figure 4. Identification of the main transcriptional sources of variation in tissues and stresses. A) Percentage of explained variance (%) by each latent factor (LF) and regulatory layer (gene expression, GE; alternative splicing, AS) for ungrouped tissues framework. **B)** Scatter plot of latent factor 1 (x-axis) and latent factor 3 (y-axis) illustrating the variation described. Samples are coloured according to tissues. **C)** Table showing top absolute loading genes for latent factors 1 and 3. **D)** Heatmaps depicting significant (FDR < 0.1, -log10(FDR)) enriched Mercator functional terms (green), genes ages (purple) and family founder events ages (blue) for each regulatory layer. PS/F = gene/family-founder phylostratum. Lower phylostratum values correspond to genes with older origins. **E)** Experimental validation of tissues/age-induced AS events by RT-PCR. The primers used allow the amplification of multiple splice variants (see **Methods**). **F)** Percentage of explained variance (%) by each latent factor (LF) and regulatory layer (gene expression, GE; intron retention, IR; exon skipping, ES; alternative acceptor donor site, Alt; alternative splicing without classification, AS) for grouped stress framework. **G)** Scatter plots of latent factors 1, 2 and 3 illustrating the variation described. Colours denote stress treatments. Different figures denote genotypes. DO = *Dothistroma*; FU = *Fusarium*; HS = heat; PH = *Phytophthora*; dmg = damage. T1-T4 correspond to low-very high stress intensities. **H)** Table showing top absolute loading genes for latent factors 1, 2 and 3. **I)** Heatmaps depicting significant (FDR < 0.1, -log10(FDR)) enriched Mercator functional terms (green), genes ages (purple) and family founder events ages (blue) for gene expression and intron retention regulatory layers. PS1 = cellular organisms; PS2 = Eukaryota; PS3 = Viridiplantae; PS4 = Streptophyta; PS5 = Streptophytina; PS6 = Embryophyta; PS7 = Tracheophyta; PS8 = Euphyllophyta; PS9 = Spermatophyta; PS10 = Acrogymnospermae; PS11 = Pinidae; PS12 = Pinaceae; met = metabolism; CHO = carbohydrate; org = organisation; resp = response; reg = regulation; hom = homeostasis; mod = modification; transloc = translocation; PS = photosynthesis; biosynthe = biosynthesis.

Figure 5. Characterisation of shared and unique sources of stress variation at protein level. Due to the high complexity of proteomics data four LFs were selected to perform in-depth characterisation. **A)** Percentage of explained variance (%) by each

latent factor (LF) for grouped all stresses total proteomes framework. FU = *Fusarium*; HS = heat. **B)** Scatter plots of latent factors 1, 2, 3 and 4 illustrating the variation described. Colours denote stress treatments. R = Recovery. T1-T4 correspond to low-very high stress intensities. **C)** Table showing top absolute loading proteins for latent factors 1, 2, 3 and 4. **D)** Heatmaps depicting significant (FDR < 0.1, -log10(FDR)) enriched Mercator functional terms (green), genes ages (purple) and family founder events ages (blue). PS/F = gene/family-founder phylostratum. Lower phylostratum values correspond to genes with older origins. **E)** Percentage of explained variance (%) by each latent factor (LF) for grouped abiotic stresses all proteomes framework. **F)** Scatter plots of latent factors 2, 6, 7 and 9 illustrating the variation described. Colours denote stress intensity. Figures denote subcellular location. Chloro = chloroplast. **G)** Table showing top absolute loading proteins for latent factors 2, 6, 7 and 9. **H)** Heatmaps depicting significant (FDR < 0.1, -log10(FDR)) enriched Mercator functional terms (green), genes ages (purple) and family founder events ages (blue). **I)** Percentage of explained variance (%) by each latent factor (LF) for grouped abiotic stresses chloroplast proteomes framework. LF1 was excluded because it only represented study batch effect. PopE = population E (non-stressed); PopT = population T (stressed) (see **supplementary fig. S2F**). **J)** Scatter plots of latent factors 2, 3, 5 and 6 illustrating the variation described. Colours denote stress intensity. Figures denote stress type. E = population E; T = population T. **K)** Table showing top absolute loading proteins for latent factors 2, 3, 5 and 6. **L)** Heatmaps depicting significant (FDR < 0.1, -log10(FDR)) enriched Mercator functional terms (green), genes ages (purple) and family founder events ages (blue). PS1 = cellular organisms; PS2 = Eukaryota; PS3 = Viridiplantae; PS4 = Streptophyta; PS5 = Streptophytina; PS6 = Embryophyta; PS7 = Tracheophyta; PS8 = Euphyllophyta; PS9 = Spermatophyta; PS10 = Acrogymnospermae; PS11 = Pinidae; PS12 = Pinaceae; met = metabolism; CHO = carbohydrate; org = organisation; resp = response; reg = regulation; hom = homeostasis; mod = modification; transloc = translocation; PS = photosynthesis; biosynthe = biosynthesis.

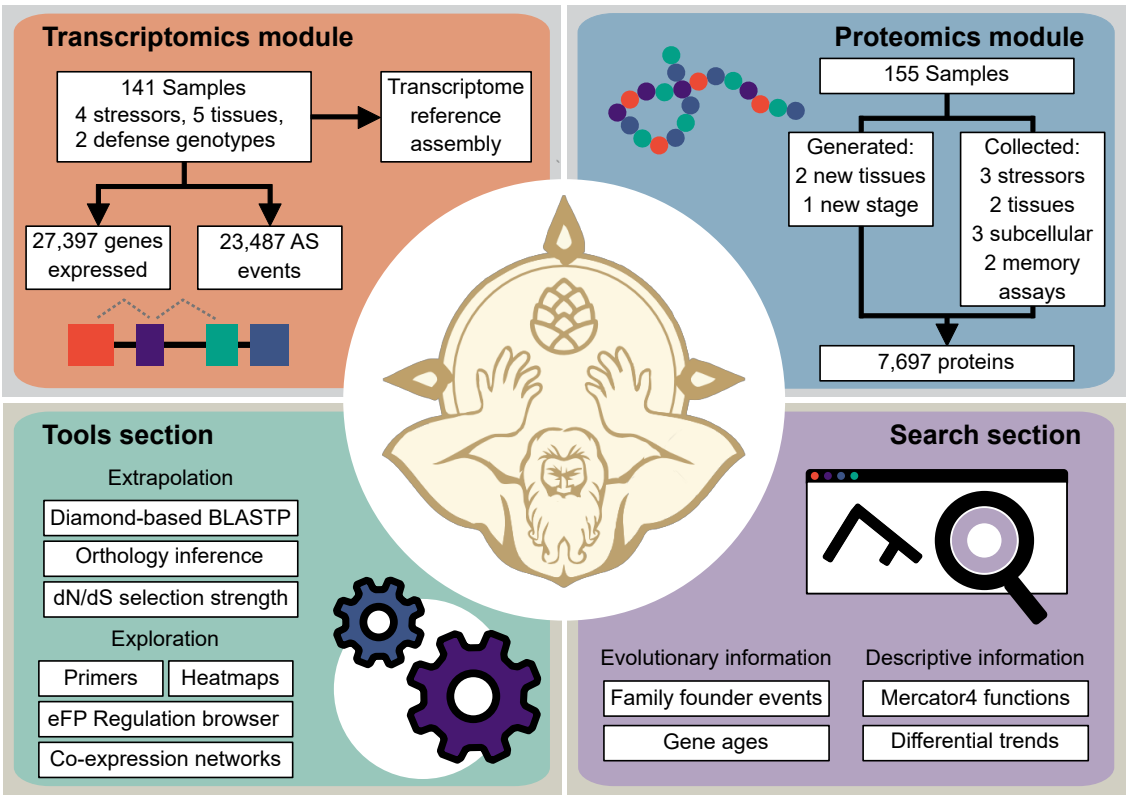

Transcriptomics module
141 Samples
4 stressors, 5 tissues,
2 defense genotypes
Transcriptome
reference
assembly
27,397 genes
expressed
23,487 AS
events
Proteomics module
155 Samples
Generated:
2 new tissues
1 new stage
Collected:
3 stressors
2 tissues
3 subcellular
2 memory
assays
7,697 proteins
Tools section
Extrapolation
Diamond-based BLASTP
Orthology inference
dN/dS selection strength
Exploration
Primers
Heatmaps
eFP Regulation browser
Co-expression networks
Search section
Evolutionary information
Family founder events
Gene ages
Descriptive information
Mercator4 functions
Differential trends

Figure 1

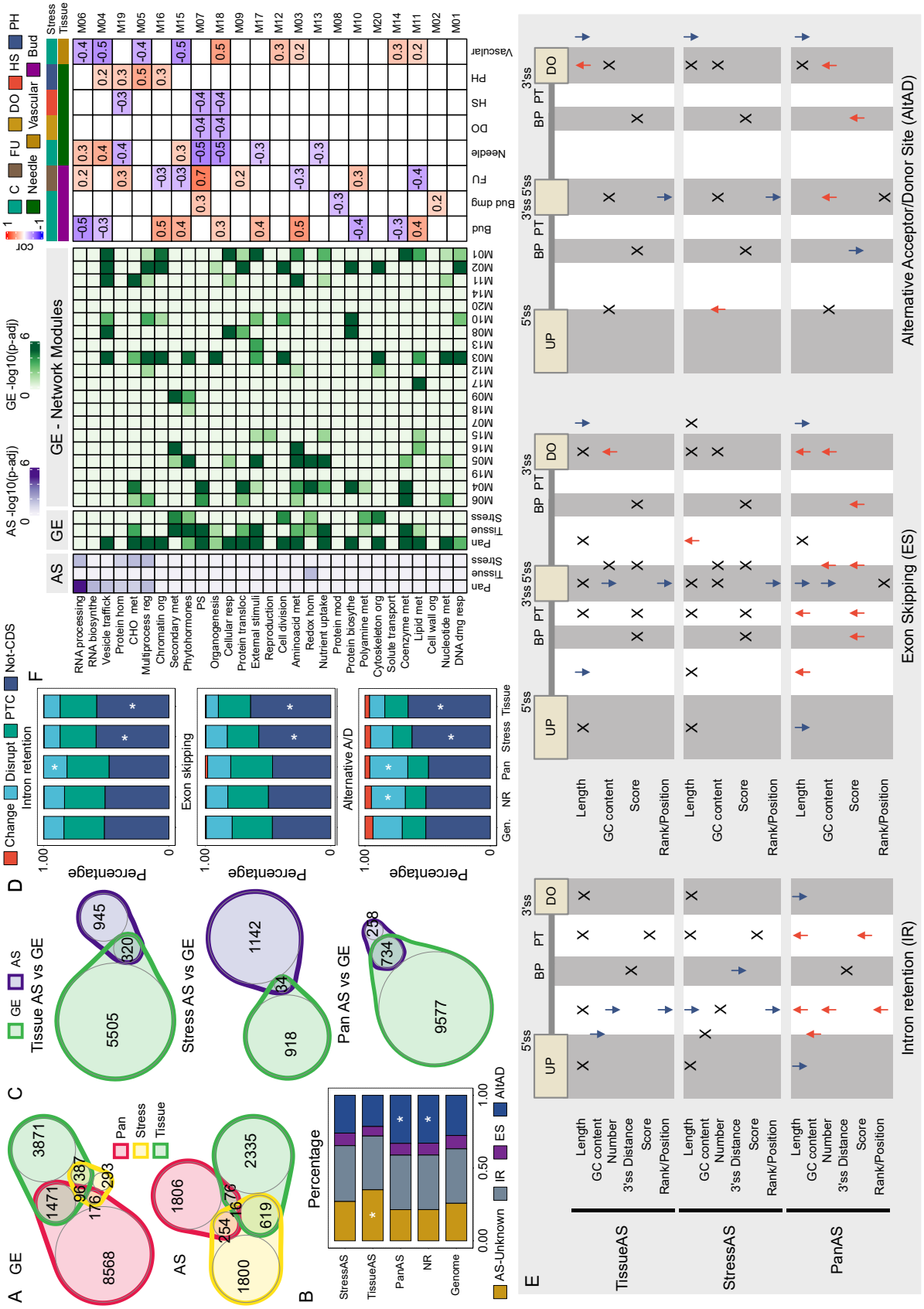
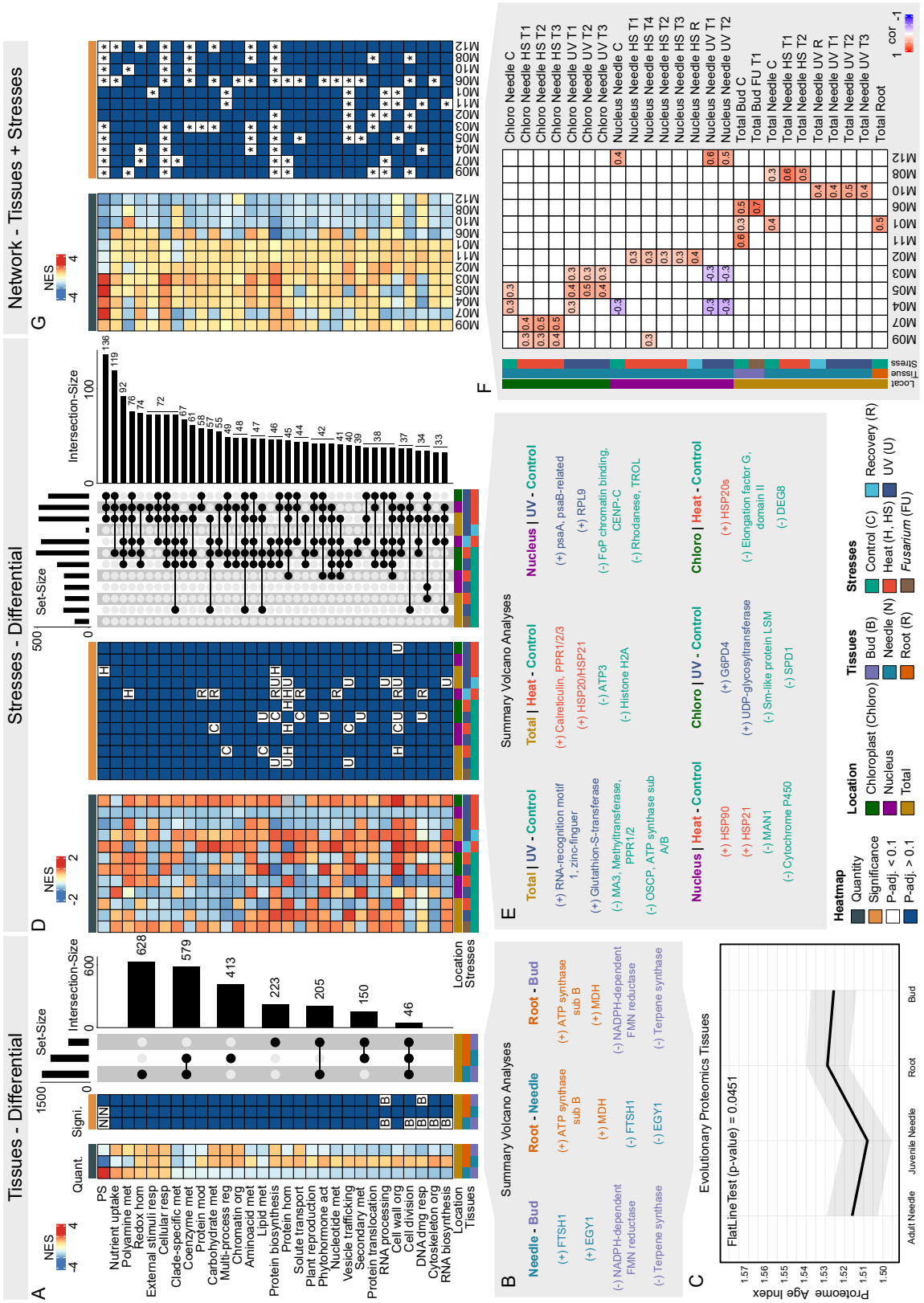

A
GE
AS
3871
1471
96
387
176
293
8568
Pan
Stress
Tissue
1806
254
16
76
2335
619
1800
B
Percentage
StressAS
TissueAS
PanAS
NR
Genome
0.00
0.50
1.00
AS-Unknown
IR
ES
AltAD
C
GE
AS
Tissue AS vs GE
945
320
5505
Stress AS vs GE
1142
84
918
Pan AS vs GE
298
734
9577
D
Change
Disrupt
PTC
Not-CDS
Intron retention
Exon skipping
Alternative A/D
Percentage
1.00
0
Gen.
NR
Pan
Stress
Tissue
F
AS
GE
GE - Network Modules
AS -log10(p-adj)
GE -log10(p-adj)
cor
1
-1
C
FU
DO
HS
PH
Needle
Vascular
Bud
Stress
Tissue
RNA processing
RNA biosynthe
Vesicle traffick
Protein hom
CHO met
Multiprocess reg
Chromatin org
Secondary met
Phytohormones
PS
Organogenesis
Cellular resp
Protein transloc
External stimuli
Reproduction
Cell division
Aminoacid met
Redox hom
Nutrient uptake
Protein mod
Protein biosythe
Polyamine met
Cytoskeleton org
Solute transport
Coenzyme met
Lipid met
Cell wall org
Nucleotide met
DNA dmg resp
Pan
Tissue
Stress
Bud
Bud dmg
FU
Needle
DO
HS
PH
Vascular
M06
M04
M19
M05
M16
M15
M07
M18
M09
M17
M12
M03
M13
M08
M10
M20
M14
M11
M02
M01
E
TissueAS
StressAS
PanAS
Length
GC content
Number
3'ss Distance
Score
Rank/Position
5'ss
BP
PT
3'ss
UP
DO
Intron retention (IR)
Exon Skipping (ES)
Alternative Acceptor/Donor Site (AltAD)

Figure 2



990

991 Figure 3

992

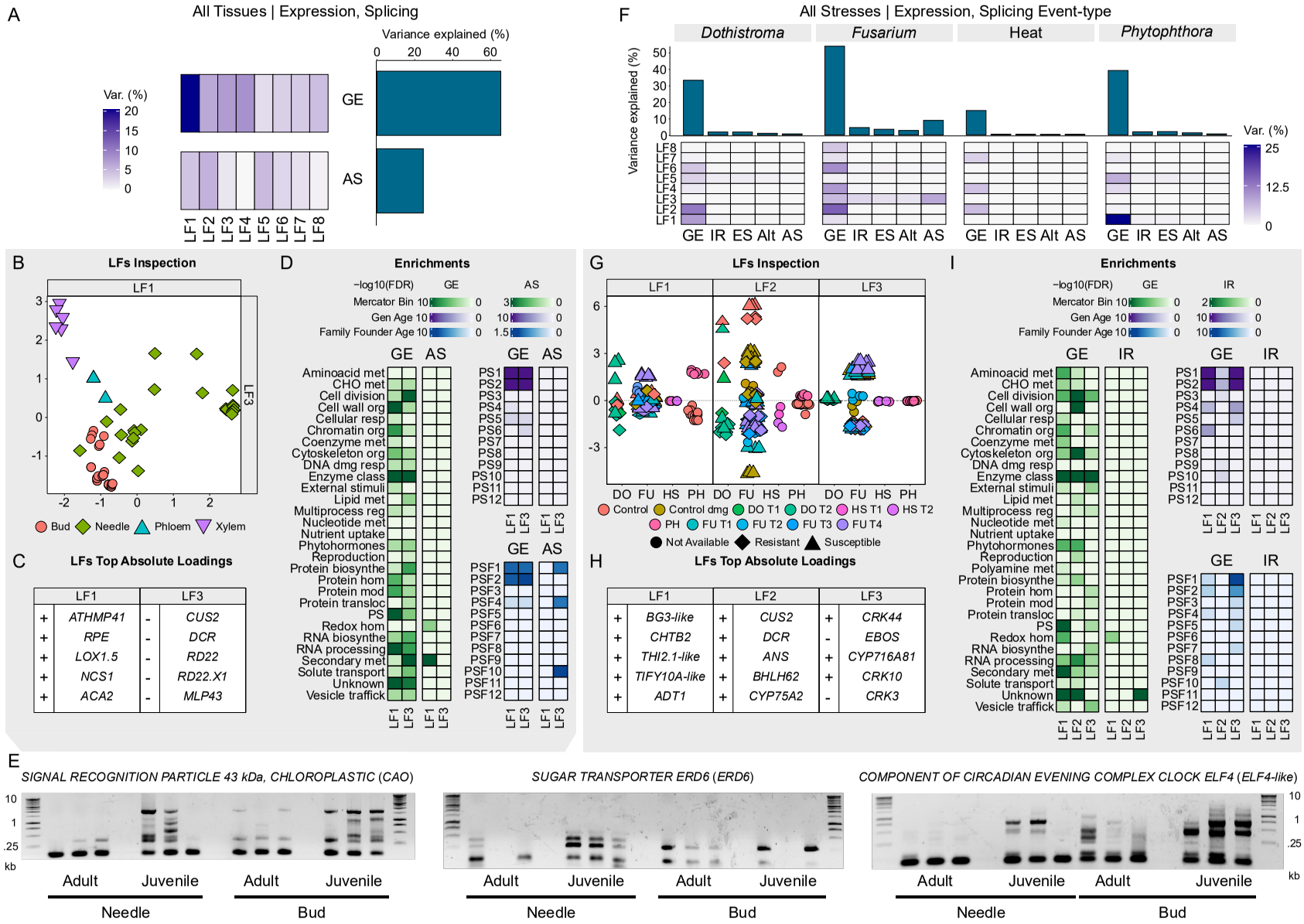

A
All Tissues | Expression, Splicing
Variance explained (%)
Var. (%)
GE
AS
LF1 LF2 LF3 LF4 LF5 LF6 LF7 LF8
F
All Stresses | Expression, Splicing Event-type
Dothistroma
Fusarium
Heat
Phytophthora
Variance explained (%)
GE IR ES Alt AS
B
LFs Inspection
Bud Needle Phloem Xylem
C
LFs Top Absolute Loadings
LF1 LF3
+ ATHMP41 - CUS2
+ RPE - DCR
+ LOX1.5 - RD22
+ NCS1 - RD22.X1
+ ACA2 - MLP43
D
Enrichments
−log10(FDR) GE AS
Mercator Bin 10
Gen Age 10
Family Founder Age 10
Aminoacid met
CHO met
Cell division
Cell wall org
Cellular resp
Chromatin org
Coenzyme met
Cytoskeleton org
DNA dmg resp
Enzyme class
External stimuli
Lipid met
Multiprocess reg
Nucleotide met
Nutrient uptake
Phytohormones
Reproduction
Protein biosynthe
Protein hom
Protein mod
Protein transloc
PS
Redox hom
RNA biosynthe
RNA processing
Secondary met
Solute transport
Unknown
Vesicle traffick
PS1 PS2 PS3 PS4 PS5 PS6 PS7 PS8 PS9 PS10 PS11 PS12
PSF1 PSF2 PSF3 PSF4 PSF5 PSF6 PSF7 PSF8 PSF9 PSF10 PSF11 PSF12
E
SIGNAL RECOGNITION PARTICLE 43 kDa, CHLOROPLASTIC (CAO)
SUGAR TRANSPORTER ERD6 (ERD6)
COMPONENT OF CIRCADIAN EVENING COMPLEX CLOCK ELF4 (ELF4-like)
kb
Adult Juvenile Adult Juvenile
Needle Bud
G
LFs Inspection
DO FU HS PH
Control Control dmg DO T1 DO T2 HS T1 HS T2
PH FU T1 FU T2 FU T3 FU T4
Not Available Resistant Susceptible
H
LFs Top Absolute Loadings
LF1 LF2 LF3
+ BG3-like + CUS2 + CRK44
+ CHTB2 + DCR - EBOS
+ THI2.1-like + ANS + CYP716A81
+ TIFY10A-like + BHLH62 + CRK10
+ ADT1 + CYP75A2 - CRK3
I
Enrichments
−log10(FDR) GE IR
Mercator Bin 10
Gen Age 10
Family Founder Age 10
Aminoacid met
CHO met
Cell division
Cell wall org
Cellular resp
Chromatin org
Coenzyme met
Cytoskeleton org
DNA dmg resp
Enzyme class
External stimuli
Lipid met
Multiprocess reg
Nucleotide met
Nutrient uptake
Phytohormones
Reproduction
Polyamine met
Protein biosynthe
Protein hom
Protein mod
Protein transloc
PS
Redox hom
RNA biosynthe
RNA processing
Secondary met
Solute transport
Unknown
Vesicle traffick
PS1 PS2 PS3 PS4 PS5 PS6 PS7 PS8 PS9 PS10 PS11 PS12
PSF1 PSF2 PSF3 PSF4 PSF5 PSF6 PSF7 PSF8 PSF9 PSF10 PSF11 PSF12

Figure 4

A
All Stresses | Total Proteomes
Variance explained (%)
FU
HS
UV
LF4
LF3
LF2
LF1
Var. (%)
B
LFs Inspection
C
LFs Top Absolute Loadings
D
Enrichments
−log10(FDR)
Mercator Bin
Gen Age
Family Founder Age
Aminoacid met
CHO met
Cell division
Cell wall org
Cellular resp
Chromatin org
Coenzyme met
Cytoskeleton org
Enzyme class
External stimuli
Lipid met
Multiprocess reg
Nucleotide met
Nutrient uptake
Phytohormones
Reproduction
Protein biosynthe
Protein hom
Protein mod
Protein transloc
PS
Redox hom
RNA biosynthe
RNA processing
Secondary met
Solute transport
Unknown
Vesicle traffick
PS1
PS2
PS3
PS4
PS5
PS6
PSF1
PSF2
PSF3
PSF4
LF1
LF2
LF3
LF4
E
Abiotic Stresses | All Proteomes
HS
UV
LF1
LF2
LF3
LF4
LF5
LF6
LF7
LF8
LF9
LF10
LF11
F
LFs Inspection
G
LFs Top Absolute Loadings
H
Enrichments
PS9
PS12
PSF5
PSF6
PSF12
LF9
I
Abiotic Stresses | Chloroplast Proteomes
PopE
PopT
J
LFs Inspection
K
LFs Top Absolute Loadings
L
Enrichments

Figure 5