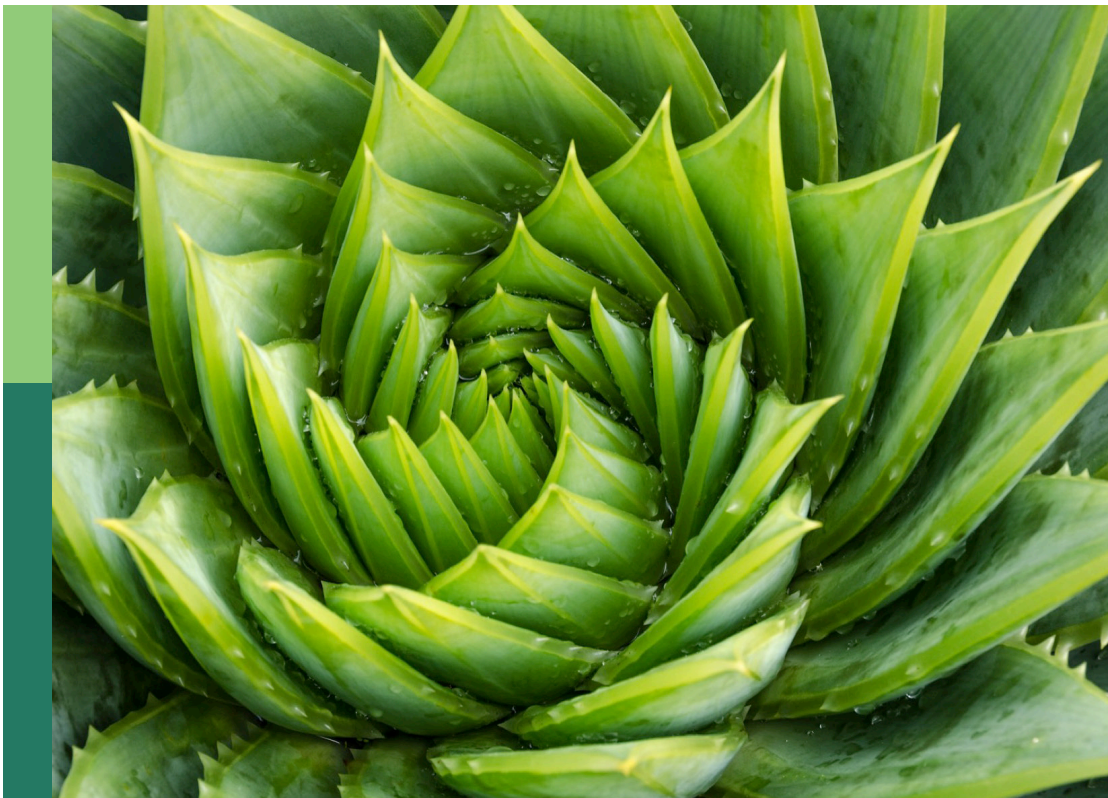
frontiers | Research Topics

Seed proteomics

Edited by
Dominique Job, Karine Gallardo, Pingfang Yang, Andrej Frolov,
Erwann Arc and Bing Bai

Published in
Frontiers in Plant Science



ISSN 1664-8714
ISBN 978-2-83251-203-6
DOI 10.3389/978-2-83251-203-6

About Frontiers

Frontiers is more than just an open access publisher of scholarly articles: it is a pioneering approach to the world of academia, radically improving the way scholarly research is managed. The grand vision of Frontiers is a world where all people have an equal opportunity to seek, share and generate knowledge. Frontiers provides immediate and permanent online open access to all its publications, but this alone is not enough to realize our grand goals.

Frontiers journal series

The Frontiers journal series is a multi-tier and interdisciplinary set of open-access, online journals, promising a paradigm shift from the current review, selection and dissemination processes in academic publishing. All Frontiers journals are driven by researchers for researchers; therefore, they constitute a service to the scholarly community. At the same time, the *Frontiers journal series* operates on a revolutionary invention, the tiered publishing system, initially addressing specific communities of scholars, and gradually climbing up to broader public understanding, thus serving the interests of the lay society, too.

Dedication to quality

Each Frontiers article is a landmark of the highest quality, thanks to genuinely collaborative interactions between authors and review editors, who include some of the world's best academicians. Research must be certified by peers before entering a stream of knowledge that may eventually reach the public - and shape society; therefore, Frontiers only applies the most rigorous and unbiased reviews. Frontiers revolutionizes research publishing by freely delivering the most outstanding research, evaluated with no bias from both the academic and social point of view. By applying the most advanced information technologies, Frontiers is catapulting scholarly publishing into a new generation.

What are Frontiers Research Topics?

Frontiers Research Topics are very popular trademarks of the *Frontiers journals series*: they are collections of at least ten articles, all centered on a particular subject. With their unique mix of varied contributions from Original Research to Review Articles, Frontiers Research Topics unify the most influential researchers, the latest key findings and historical advances in a hot research area.

Find out more on how to host your own Frontiers Research Topic or contribute to one as an author by contacting the Frontiers editorial office: frontiersin.org/about/contact

Seed proteomics

Topic editors

Dominique Job — UMR5240 Microbiologie, Adaptation et Pathogenie (MAP), France

Karine Gallardo— Institut National de recherche pour l'agriculture, l'alimentation et l'environnement (INRAE), France

Pingfang Yang — Hubei University, China

Andrej Frolov — Leipzig University, Germany

Erwann Arc — Institut für Botanik

Bing Bai — University of Copenhagen, Denmark

Citation

Job, D., Gallardo, K., Yang, P., Frolov, A., Arc, E., Bai, B., eds. (2023). *Seed proteomics*. Lausanne: Frontiers Media SA. doi: 10.3389/978-2-83251-203-6

Table of contents

frontiers in Plant Science

ORIGINAL RESEARCH
published: 30 August 2021
doi: 10.3389/fpls.2021.707184

Comparative Proteomics at the Critical Node of Vigor Loss in Wheat Seeds Differing in Storability

Xiuling Chen[1,2], Andreas Börner[3], Xia Xin[1], Manuela Nagel[3], Juanjuan He[1], Jisheng Li[2], Na Li[2], Xinxiong Lu[1] and Guangkun Yin[1]**

[1] National Crop Genebank, Institute of Crop Science, Chinese Academy of Agricultural Sciences, Beijing, China, [2] Applied Technology Research and Development Center for Sericulture and Special Local Products of Hebei Universities, Institute of Sericulture, Chengde Medical University, Chengde, China, [3] Department of Genebank, Leibniz Institute of Plant Genetics and Crop Plant Research, Gatersleben, Germany

The critical node (CN, 85% germination) of seed viability is an important threshold for seed regeneration decisions after long-term conservation. Dependent on the germplasm, the storage period until CN is reached varies and information on the divergence of the proteomic profiles is limited. Therefore, the study aims to identify key proteins and mechanisms relevant for a long plateau phase and a late CN during artificial seed aging of wheat. Seeds of the storage-tolerant genotype (ST) TRI 23248, and the storage-sensitive genotype (SS) TRI 10230 were exposed to artificial ageing (AA) and extracted embryos of imbibed seeds were analyzed using an iTRAQ-based proteomic technique. ST and SS required AA for 24 and 18 days to reach the CN, respectively. Fifty-seven and 165 differentially abundant proteins (DAPs) were observed in the control and aged groups, respectively. Interestingly, a higher activity in metabolic processes, protein synthesis, transcription, cell growth/division, and signal transduction were already found in imbibed embryos of control ST seeds. After AA, 132 and 64 DAPs were accumulated in imbibed embryos of both aged ST and SS seeds, respectively, which were mainly associated with cell defense, rescue, and metabolism. Moreover, 78 DAPs of ST appeared before CN and were mainly enriched in biological pathways related to the maintenance of redox and carbon homeostasis and they presented a stronger protein translation ability. In contrast, in SS, only 3 DAPs appeared before CN and were enriched only in the structural constituents of the cytoskeleton. In conclusion, a longer span of plateau phase might be obtained in seeds when proteins indicate an intense stress response before CN and include the effective maintenance of cellular homeostasis, and avoidance of excess accumulation of cytotoxic compounds. Although key proteins, inherent factors and the precise regulatory mechanisms need to be further investigated, the found proteins may also have functional potential roles during long-term seed conservation.

Keywords: wheat, seed longevity, long-term storage, differentially accumulated proteins, artificial aging

OPEN ACCESS

Edited by:
Bing Bai,
University of Copenhagen, Denmark

Reviewed by:
Christopher Edward West,
University of Leeds, United Kingdom
Julia Buitink,
INRA Centre Angers-Nantes Pays de la Loire, France

***Correspondence:**
Xinxiong Lu
luxinxiong@caas.cn
Guangkun Yin
yinguangkun@caas.cn

Specialty section:
This article was submitted to Plant Proteomics and Protein Structural Biology, a section of the journal Frontiers in Plant Science

Received: *17 May 2021*
Accepted: *10 August 2021*
Published: *30 August 2021*

Citation:
Chen X, Börner A, Xin X, Nagel M, He J, Li J, Li N, Lu X and Yin G (2021) Comparative Proteomics at the Critical Node of Vigor Loss in Wheat Seeds Differing in Storability. Front. Plant Sci. 12:707184. doi: 10.3389/fpls.2021.707184

Abbreviations: AA, accelerated aging treatment; CN, critical node; DAPs, differentially abundant proteins; GP, germination percentage; GLY I, lactoylglutathione lyase; isoAsp, abnormal isoaspartyl residues; LEA, late embryogenesis abundant protein; PIMT, protein L-isoaspartyl-O-methyltransferase; PER1, 1-Cys peroxiredoxin; ROS, reactive oxygen species; sHSPs, small heat shock proteins; ST, storage-tolerant; SS, storage-sensitive.

INTRODUCTION

The seed is an important genetic carrier of germplasm resources which form the basis for the earliest innovations and the future developments in agriculture. Most orthodox (desiccation tolerant) seeds are conserved under low temperature and low humidity conditions (Walters et al., 2005). Wheat (*Triticum aestivum* L.) is the second most important cereal after rice, and its safe conservation is vital for sustainable development in agriculture (Li et al., 2019). To date, more than 850,000 wheat germplasm resources (accessions) have been conserved at genebanks worldwide, of which 51,000 have been stored at −18°C in the National Genebank of China (unpublished). Though, storage conditions follow international standards (FAO, 2014), seeds deteriorate (Lu et al., 2005; Walters et al., 2005; Lee et al., 2019) which is accompanied by the accumulation of reactive oxygen species (ROS), disruption of cellular membranes, collapse of defense system, and oxidative damage to sugars, lipids, proteins, and nucleic acids (Rajjou et al., 2008; Waterworth et al., 2015; Yin et al., 2017; Nagel et al., 2019). During the loss of seed viability, a reverse S-shaped curve forms that comprises an initial plateau phase (PI), a rapidly decreasing phase (PII), and a slowly decreasing phase (PIII). The transition from P-I to P-II, especially when seed viability falls below 85% of initial germination is defined as the critical node (CN) which is an important threshold for regeneration decision in genebank management (FAO, 2014; Yin et al., 2016, 2017). The wide variation in falling below CN between and within species is based on the difference of seed longevity and affects seed storability (Nagel et al., 2009; Niedzielski et al., 2009; Lee et al., 2019).

Seed storability is a complex quantitative trait and defined as the ability of seeds to maintain alive during storage. It is affected by genetic and environmental factors during plant growth, seed maturation and post-harvest, and is reflected by the variations in the physiological and biochemical status and intrinsic genetic factors of seeds, such as the testa morphology, carbohydrates, lipids, and proteins (Clerkx et al., 2004; Nagel et al., 2009; Nguyen et al., 2015; He et al., 2016; Li et al., 2017). Genetic studies revealed quantitative trait loci (QTLs) for seed storability in bread but also durum wheat being distributed on many chromosomes indicating the complexity of the trait (Landjeva et al., 2010; Rehman Arif et al., 2012, 2017; Börner et al., 2018; Zuo et al., 2018; Rehman Arif and Börner, 2019, 2020). Recently, key proteins involved in the lipid peroxidation, energy metabolism, maintenance of redox homeostasis, scavenging of cytotoxic compounds, and repair of oxidative damage have been found to play vital role in the regulation of seed storability, and have been reported in many species including *Arabidopsis*, rice, tobacco, and chickpea (Verma et al., 2013; Rissel et al., 2014; Nisarga et al., 2017; Yazdanpanah et al., 2019). For example, NADP-malic enzyme 1, aldehyde dehydrogenase (ALDH), poly (ADPribose) polymerase (PARP), and protein L-isoaspartyl-*O*-methyltransferase (PIMT), are positively correlated with seed storability by possibly avoiding oxidative damage, maintain the genomic integrity and function of proteins (Verma et al., 2013; Rissel et al., 2014; Nisarga et al., 2017; Yazdanpanah et al., 2019). However, how these key proteins contribute to the higher storability is still unclear. In a previous study, we obtained two wheat accessions (TRI 23248 and TRI 10230) from the IPK Gatersleben genebank. Their storability was identified by comparing seed viability after storage for 10 years under the long-term storage (−18°C) and ambient (20°C) conditions. Seeds of TRI 23248 germinated at 84% and the genotype was characterized as storage-tolerant (ST), and seeds of TRI 10230 had a germination of 14% and the genotype classified as storage-sensitive (SS) (Chen et al., 2018). Gene expression and proteomic based two-dimensional polyacrylamide gel electrophoresis analysis of both genotypes showed that ST seeds imbibed for 24 h had higher abundance in some proteins related with antioxidant defense system, protein destination and storage, and others, leading to extended longevity during conservation (Chen et al., 2018). Although Yin et al. (2017) showed the abundance of proteins and carbonylation patterns at and around the CN for rice, the proteomic profiles and the potential regulatory mechanisms at the CN among are still ambiguous in wheat germplasms with apparent differences in seed storability. Therefore, the aim of the study is to elucidate molecular regulations around CN, an important threshold for genebank regeneration, and to identify key proteins relevant for long storability and possible prediction tools. Therefore, we regenerated the two wheat accessions TRI 23248 and TRI 10230 and applied artificial aging (AA) treatment to gain a seed germination around CN, at 90, 80, and 60%. Imbibed embryos were extracted from AA seeds and investigated using an iTRAQ-based quantitative proteomic technology to further explain the variation in protein expression regulation before CN.

MATERIALS AND METHODS

Plant Materials and Treatment

Two wheat accessions TRI 23248[1] and TRI 10230[2] obtained from the genebank of the Leibniz Institute of Plant Genetics and Crop Plant Research (IPK), Germany, were used. TRI 23248 (longer seed longevity), was termed the storage-tolerant genotype, while TRI 10230 (shorter seed longevity), was termed as the storage-sensitive genotype. Both accessions were regenerated in 2015. After harvest, the mature seeds were cleaned, dried to the same water content (7.2%) and stored at −20°C in aluminum foil bag until usage. AA treatment was performed by placing seeds in tightly closed desiccators with saturated sodium chloride solution at the bottom to obtain 75% relative humidity at 40°C for 36 days (Yin et al., 2016). Every 4 days, seeds were withdrawn from the treatment and immediately germinated to characterize sigmoidal shaped aging curve for each genotype.

Seed Germination Test

Seed germination was performed according to the protocol of the International Seeds Testing Association (2015). Triplicate with 50 seeds were germinated at 25°C in the dark. The germination percentage (GP) of the control group was 99%.

[1]https://doi.org/10.25642/IPK/GBIS/227191
[2]https://doi.org/10.25642/IPK/GBIS/10202

Based on GP, the viability curve of the seed was plotted to select appropriate nodes at P-I and P-II. The survival curve was fitted using Avrami equation [$\ln(N0/N) = (t/\Phi)^n$, $N0$ represented the average initial germination percentage, N was the average germination percentage for the time (t) in days, Φ and n were respectively the coefficient and exponential factor of Avrami equation] with OriginPro software (Walters et al., 2005), and then the calculated time for seed germination to decrease to 50% (P50) was obtained to illustrate the different declines between the two wheat accessions. Finally, seeds belonging to three aging groups (90, 80, and 60%) and two untreated control groups were imbibed in water at 25°C for 24 h. Subsequently, the seed embryos were extracted to further analyze the proteomic differences between the two wheat accessions.

Protein Preparation

Total protein was extracted as described by Chen et al. (2018). Briefly, 50 embryos were ground into a fine powder in liquid nitrogen, and suspended in 50 mM Tris-HCl buffer (pH 7.5) containing 0.8 M sucrose, 5 mM EDTA, and 65 mM dithiothreitol. The suspension was mixed with tris-saturated phenol and incubated in an ice bath for 30 min. After centrifugation at 10,000 *g* for 10 min, ammonium acetate/methanol (0.1 M) was used to precipitate the protein at –20°C. Finally, the protein pellet was washed thrice with chilled acetone, and vacuum-dried. Three biological replicates were used separately. The protein pellet was dissolved in hydration buffer containing 7 M urea, 2 M thiourea, 2% CHAPS, and 1% dithiothreitol. Subsequently, it was ultrasonicated for 15 s, the protein concentration was detected by the Bradford method (Bradford, 1976).

Protein Digestion

Proteins were digested using the filter-aided sample preparation method, as previously described by Wiśniewski et al. (2009). Briefly, the total protein samples (200 μg) diluted in hydration buffer were mixed with 50 mM Tris (2-carboxyethyl) phosphine (1:50, v/v) and incubated at 37°C for 2 h. Subsequently, 200 mM methyl methanethiosulfonate (1:100, v/v) in isopropanol was added to the mixture and incubated at room temperature in the dark for 30 min. The samples were transferred onto a 10 kDa filter (Sartorious, Göttingen, Germany) for ultrafiltration and centrifuged at 10,000 × *g* for 10 min, at 4°C. Subsequently, three wash steps were performed with 400 μL of ddH_2O. After each wash step, centrifugation was performed at 10,000 × *g* for 20 min. Finally, 85 μL of dissolution buffer (500 mM triethylammonium bicarbonate, pH 8.5) and 2 μg of trypsin (Applied Biosystems, Foster City, CA, United States) were added to each filter. The samples were incubated at 37°C overnight for 16∼18 h. The resulting peptides were collected as the filtrate and freeze-dried into a powder for protein labeling.

Protein Labeling

The peptide mixture was labeled with the iTRAQ reagent 8-plex multiplex kit (Applied Biosystems, United States), according to the manufacturer's instructions. In TRI 23248 and TRI 10230, the control and aging groups were separately labeled as (ST-99%)-113, (ST-90%)-114, (ST-80%)-115, (ST-60%)-116, (SS-99%)-117, (SS-90%)-118, (SS-80%)-119, and (SS-60%)-121. The labeled samples were incubated at room temperature for 2 h and the reaction was terminated by adding 100 μL of ddH_2O.The peptide mixtures were pooled, vacuum-frozen, and dried into powder. Protein labeling was performed in triplicate.

Peptide Separation Using SCX Chromatography

The dried peptide mixture was reconstituted and acidified with buffer A (10 mM KH_2PO_4 in 25% acetonitrile, pH 3.0) and loaded onto a 4.6 × 100 mm polysulfethyl column (5 μm, 200 Å, PolyLC, Inc., Columbia, MD, United States). The peptides were eluted at a flow rate of 1 mL min^{-1} with a gradient of 0∼5% buffer B (2 M KCl and 10 mM KH_2PO_4 in 25% ACN, pH 3.0) for 1 min, 5∼30% buffer B for 20 min, 30∼50% buffer B for 5 min, 50% buffer B for 5 min, 50∼100% buffer B for 5 min, and 100% buffer B for 10 min. The elution was monitored by measuring the absorbance at 214 nm of fractions collected every 1 min. Thirty-two fractions were finally combined into 16 pools and desalted on C18 Cartridges [Empore SPE Cartridges C18 (standard density), bed I.D. 7 mm, volume 3 mL, Sigma, St. Louis, MO, United States]. All fractions were dried and stored at –80°C until further analysis.

NanoRPLC-MS/MS Analysis

NanoRPLC-MS/MS experiments were performed on a Triple-TOF 5600 + system (AB SCIEX, Framingham, MA, United States) coupled to an Eksigent nanoLCUltra binary pump system (AB SCIEX). The peptide mixture was loaded onto a trap column (Eksigent, 100 × 20 mm) connected to an analytical column (Eksigent, 75 μm × 150 mm), in mobile phase A (2% acetonitrile, 0.1% formic acid), at a flow rate of 2 μL min^{-1}. Both the trap and analytical columns were filled with MAGIC C18 – AQ 5 μm 200 Å phase (MICHROM Bioresources, Inc., Auburn, CA, United States). The peptides were separated with mobile phase B (98% acetonitrile, 0.1% formic acid) at a linear gradient of 5 – 80%, for 90 min at a flow rate of 300 nL.min^{-1}, which was controlled by the Intelli Flow technology. The MS analysis was performed in the information-dependent mode. Mass spectrometry (MS) spectra were acquired across the mass range of 350∼1500 m/z in high-resolution mode (>30,000) using 250 ms as the accumulation time per spectrum. A maximum of 40 precursors per cycles was chosen for fragmentation from each MS spectrum, with a minimum accumulation time of 50 ms for each precursor, and dynamic exclusion for 20 s. Tandem mass spectra were recorded in high sensitivity mode (resolution > 15,000) with rolling collision energy. Data are available via ProteomeXchange with identifier PXD025860.

Protein Identification

The peptides were identified using ProteinPilot 4.5 software (AB SCIEX) with the Paragon database search algorithm and an integrated false discovery rate (FDR) analysis. Only unique peptide sequences were used as evidence for protein

identification. The data were run against the UniProt database. The user defined parameters as follows: (i) sample type, iTRAQ 8-plex (peptide labeled); (ii) cysteine alkylation, MMTS; (iii) digestion, trypsin; (iv) instrument, Triple-TOFTM 5600; (v) special factors, none; (vi) species, none; (vii) ID focus, biological modifications; (viii) database, Uniprot-*Triticum aestivum* 20170920.fasta; (ix) search effort, thorough; and (x) detected protein threshold, Unused ProtScore (conf) > 0.05 (10%). For determination of the false discovery rate (FDR), the data were searched against concatenated databases *in silico* for decoy sequences, by on-the-fly reversal, automatically by the software. Only proteins with a 1% global FDR from fit were used for further analysis.

Protein Quantification Analysis

Protein abundance was calculated as the sum of the peak areas of unique peptides (contribution > 0, confidence = 99%, and annotated as "auto"), followed by logarithmic transformation (Log10), and normalization using the median protein abundance for each label, to correct for systematic errors. Proteins present in two or three biological replicates were analyzed using the Limma ranked-product approach with the online software[3] described by Schwämmle et al. (2013). The ratio of protein abundance in the aged groups to that in the respective control groups, and between the two control groups, was used to evaluate the fold change (FC). An FDR-based estimate was used to obtain the set of significant proteins. Each protein was assigned a total of seven expression ratios with corresponding adjusted *p*-values. A moderate FC cut-off of < −1.5 was applied for lower abundance, whereas a cut off > 1.5 was applied for proteins with higher abundance. Proteins falling within these thresholds were considered as differentially abundant if the adjusted *p*-value was < 0.05 [differentially abundant proteins (DAPs)].

Bioinformatics Analysis

The identified proteins were annotated using BLAST 2 GO software (version 5.0 basic). DAPs were grouped according to their biological function using Gene Ontology (GO) terms[4], and mapped to reference authoritative pathways using the Kyoto Encyclopedia of Genes and Genomes (KEGG)[5], to determine the biologically active pathways. GO enrichment analysis was performed using g: Profiler with a filtering threshold of g: SCS less than 0.05, as described by Reimand et al. (2019). Dynamic changes in DAPs were analyzed using the Short Time-series Expression Miner software and clustering algorithm by setting 20 as the maximum number of model profiles, and 0.15 as the maximum unit change in model profiles, and 0.7 as the minimum correlation in the option for clustering profiles on the OmicShare tools platform, which is a free online platform for data analysis[6] (Ernst and Bar-Joseph, 2006).

[3]http://computproteomics.bmb.sdu.dk/Apps/LimmaRP/

[4]http://www.geneontology.org/

[5]http://www.genome.jp/kegg/

[6]http://www.omicshare.com/tools

Statistical Analyses

High-confidence results were obtained by using moderated *t*-test (Limma) with rank products based on well-defined null hypotheses. For the aging group, two paired tests were performed within eight samples by using ST-99%-113 and SS-99%-117 as controls. For the control group, some DAPs from the intersection of the two paired tests were selected as the variation between the two materials. Differences at the level of $p < 0.05$, 0.01, and 0.001 were considered as significant, which were respectively labeled as *, **, and ***, respectively.

RESULTS

Pattern of Seed Survival Curves and Number of DAPs Differ Between ST and SS Wheat Accessions After AA Treatment

Seed from ST and SS genotypes lost germination over an aging period of 36 days but slopes of survival curves were significantly different (**Figure 1**). Overall, a half-viability period (P50) was estimated to be 26.64 days and of 22.44 days for ST and SS seeds, respectively (**Figure 1**), confirming the previous results of seed storability after long-term cool and long-term ambient storage (Chen et al., 2018). ST seeds showed a longer plateau phase compared to SS seeds and were close to CN (85%) after 24 days (GP, 82.00% ± 8.00%). SS seeds approached CN after 18 days (GP, 80.67% ± 8.08%). In order to reveal differences in the proteome profile at the CN of ST and SS wheat accessions, imbibed embryos were extracted for comparative proteomics. Embryos were prepared from three stages around and close to CN and extracted from seeds imbibed for 24 h. For ST

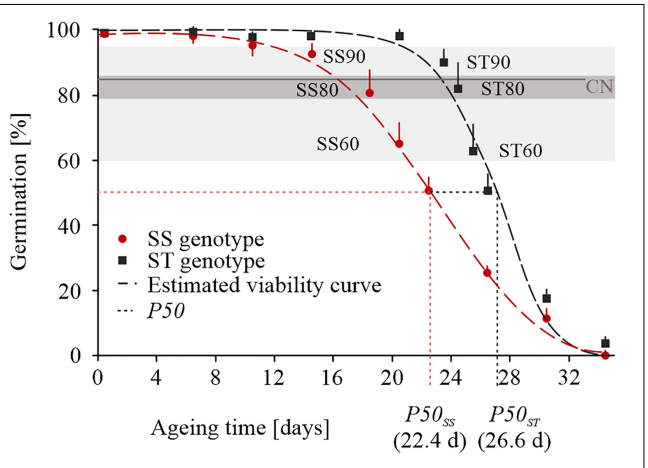


FIGURE 1 | Seed viability curves and P50 of the storage-tolerant (ST) wheat accession TRI 23248 and the storage-sensitive (SS) accession TRI 10230 following artificial aging treatment. Mean and standard error of three replicates is given. ST90, ST80, ST60, SS90, SS80, and SS60 represent stages (PI, PII, PIII) around and close to the critical node (CN) of seed germination (85%) that were used to proteomic analysis. The half-viability period (P50) is given for ST and SS seeds. Color code provided in the figure.

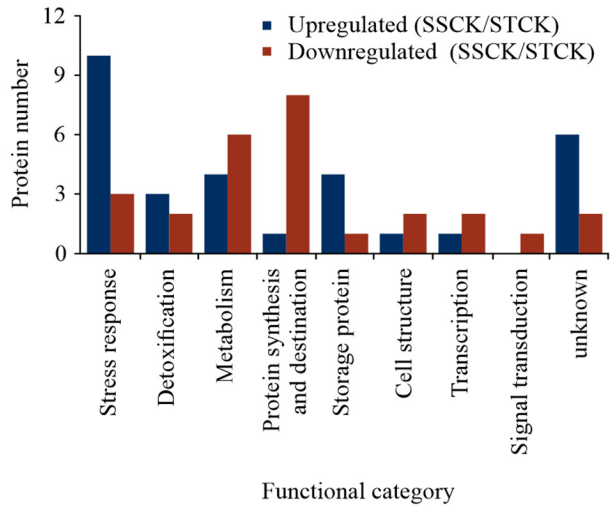


FIGURE 2 | Functional classification of identified differentially abundant proteins between the untreated control group of the storage-tolerant (ST) and of the storage-sensitive (SS) embryos with upregulation (blue) or downregulation (red) according to Bevan et al. (1998) and Zhang et al. (2016). SSCK/STCK represented the comparison of protein abundance between the control group of ST and SS seeds.

wheat, seeds of an untreated control were used and seeds aged for 23 days (GP, 90.00% ± 4.00% defined as ST90), for 24 days (GP, 82.00% ± 8.00% defined as ST80), for 25 days (GP, 62.67% ± 8.33% defined as ST60). For SS wheat, seeds from untreated control and seeds aged for 14 days (GP, 92.67% ± 3.06% defined as SS90), for 18 days (GP, 80.67% ± 8.08% defined as SS80), for 20 days (GP, 65.00% ± 5.00% defined as SS60) were used for embryo extraction.

A total of 2,118 proteins were identified after merging the data from all experimental groups (**Supplementary File 1**). By analyzing the embryos of imbibed SS and ST seeds, overall, 57 DAPs were identified in the untreated control groups (**Figure 2** and **Supplementary File 2**), and 165 DAPs were obtained as result of seed aging (**Figure 3** and **Supplementary File 3**) and are visualized in a heatmap (**Supplementary Figure 1**). In imbibed embryos of aged ST and SS seeds, 31 proteins were differentially accumulated in both accessions; 21 were upregulated and 10 were downregulated (**Figure 3C**). Besides, 101 DAPs (46 upregulated and 55 downregulated) occurred only in embryos of the aged ST seeds (**Figure 3C**), and 33 DAPs (11 upregulated and 22 downregulated) were found only in embryos of the aged SS seeds (**Figure 3C**).

In imbibed embryos of aged ST seeds, the abundance of in total 78 proteins changed when GP declined to ST90 (**Figure 3A**), while only 41 DAPs maintained the significant changes in all treated groups (PI to III). In addition, 10 DAPs were significant when GP approached CN at ST80, and 39 DAPs appeared only at ST60. However, in embryos of aged SS seeds (**Figure 3B**), only three DAPs occurred in all treated groups, 26 DAPs were presented when GP was close to CN at SS80 and 35 DAPs appeared at the SS60.

Functional Comparison of DAPs at the CN of Seed Viability From Two Wheat Accessions Differing in Storability

Gene Ontology (GO) annotations and category were performed according to Bevan et al. (1998) and Zhang et al. (2016). By contrast, these results showed that certain proteins associated with metabolism, protein synthesis and destination, transcription, cell structure, and signal transduction were more abundant in embryos of untreated imbibed ST seeds, while embryos of untreated imbibed SS seeds had a relatively higher abundance of proteins associated with the stress response and storage proteins indicating that embryos of ST and SS had already a different proteome after harvest (**Figure 2** and **Supplementary File 2**). After seed aging, DAPs associated with cell defense and rescue, and were most abundant and accounted for 40.15% in embryos obtained from aged ST seeds (**Figure 3D**) and 45.31% in embryos obtained from aged SS seeds (**Figure 3E**). DAPs associated for metabolism accounted for 20.45 and 17.19% in embryos obtained from aged ST and SS seeds, respectively. Interestingly, DAPs involved in protein synthesis and destination represented 9.85% in embryos from aged ST seeds; whereas, in aged SS seeds, the percentage of DAPs associated with this functionality was very low, only 3.13%. Moreover, embryos obtained from aged ST seeds were also significantly enriched in molecular functions related to antioxidant, oxidoreductase, and peroxidase activities, and mainly participated in certain biological pathways associated with stress response and metabolic processes (**Figure 4**). The DAPs in embryos obtained from aged SS seeds were only enriched in the structural constituents of the cytoskeleton (**Figure 4**). Concluding, the results indicated that imbibed embryos of ST seeds have a higher ability to perceive, defend, and repair various damages around CN, and ST seeds might be able to maintain this ability for a longer storage period compared to SS seeds.

Dynamic Changes of DAPs at the CN of Seed Viability From Two Wheat Accessions Differing in Storability

To gain deeper insight into the dynamic changes in proteins between imbibed embryos of SS and ST seeds during the response to AA treatment, we clustered the DAPs result from aging based on their abundance data using the STEM clustering algorithm, and the actual size based p-value protein enrichment was computed by hypothesis testing. Based significant thresholds for correlation statistics ($p < 0.05$) (Ernst and Bar-Joseph, 2006), the results showed that all the DAPs from imbibed embryos of aged ST and SS seeds were clustered into 13 profile models, of which 7 trends presented significant change including 141 DAPs were obtained (**Figure 5**, **Tables 1**–**4**, and **Supplementary File 4**).

Comparison between the two accessions showed that common 28 DAPs were significantly overrepresented in the profiles with apparent up-regulated or down-regulated changes (**Table 3**). Interestingly, the majority of ST DAPs including 10 late embryogenesis abundant (LEA) proteins, aldose reductase, topless-related protein 2 and others were enriched in cluster 1

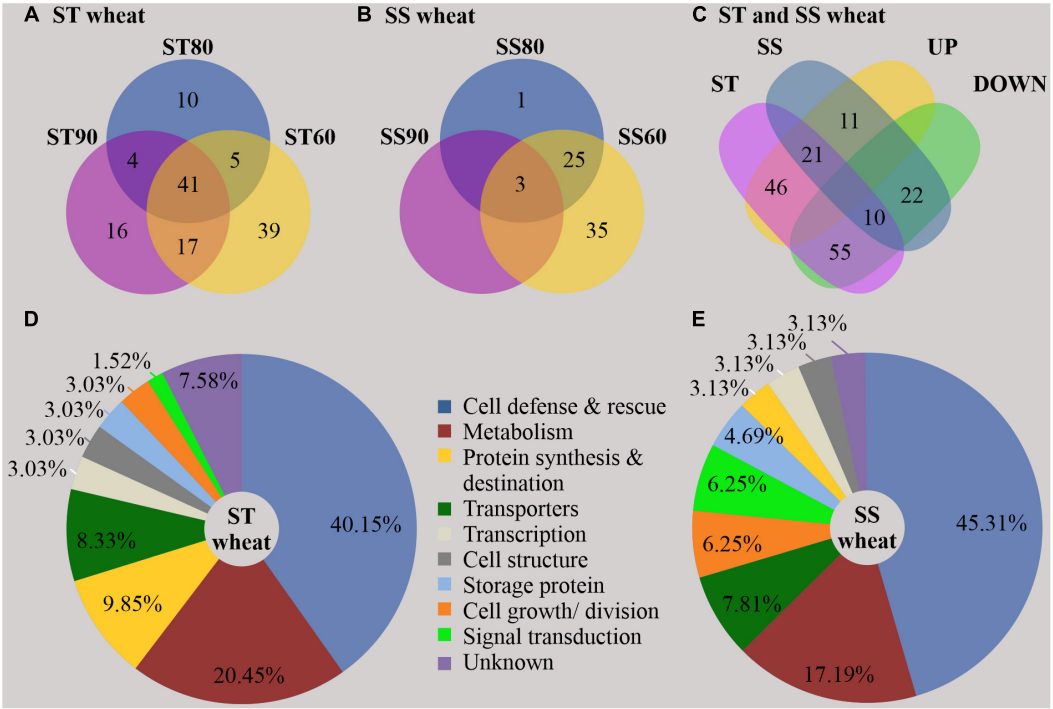


FIGURE 3 | Distribution and function of differentially abundant proteins (DAPs) resulting from artificial aging treatment of storage-tolerant (ST) and storage-sensitive (SS) seed. The Venn diagram shows the distribution of DAPs in ST **(A)** and SS **(B)** embryos around (SS90/ST90 and SS60/ST60) and close to the critical node (SS80/ST80 and their intersection) **(C)**. The Pie chart shows the functional classification according to Bevan et al. (1998) and Zhang et al. (2016). 132 DAPs were found in imbibed embryos of ST seeds **(D)**, and 64 DAPs in imbibed embryos of SS seeds **(E)**. % is given of the total number of proteins in ST and SS embryos, respectively, UP, upregulated DAPs; and DOWN, downregulated DAPs.

that displayed an increase at ST90 and maintained the level of abundance close to CN at ST80 and ST60. One LEA protein was enriched in cluster 2 that displayed a continuous increase during germination drop of ST seeds (**Figure 5A**). However, for embryos of aged SS seeds, 14 DAPs including 8 LEA proteins, 1 Probable protein phosphatase 2C 58, 1 Globulin-1 S allele and others had the same up-regulated profile (cluster 1), only 2 LEA proteins 1 aldose reductase and 1 topless-related protein 2 was enriched in cluster 3 that displayed an continuous increase at SS90 and close to CN at SS80, then maintained the level of abundance at SS60 (**Figure 5B**). Besides, 2 peroxidase-related proteins and 2 carbohydrate metabolism-related proteins were down-regulated at ST90, and then maintained the level of abundance close to CN at ST80 and ST60 in embryos of aged ST seeds (**Figure 5C**, cluster 4), but they displayed continuous a decrease in embryos of aged SS seeds (**Figure 5D**, cluster 5, 6, and 7). These proteins were mainly related to cell defense and rescue, carbohydrate metabolism and cell growth. On the whole, the change in DAPs clearly lagged in the SS accession compared to that in the ST accession (**Table 3**).

Comparison between the two accessions also showed that clear differences were displayed in the groups of DAPs that were specifically change in either embryos of ST or SS seeds. ST DAPs with significant up-regulation were significantly overrepresented in the cluster 1 and cluster 2 that were primarily assigned to "response to abiotic stimulus," of which 22 DAPs related to stress response and detoxification including dehydrin (FC, 1.61∼2.34), Em protein (FC, 2.40∼5.72), small heat shock proteins (sHSPs; FC, 1.62∼2.40), 1-Cys peroxiredoxin (PER1; FC, 1.61), lactoylglutathione lyase (FC, 1.50∼1.77), and protein-L-isoaspartate *O*-methyltransferase (PIMT; FC, 1.96), and 9 DAPs associated with energy and carbohydrate metabolism, protein synthesis, and transportation, such as 60S ribosomal protein (FC, 1.55∼2.01), glyceraldehyde-3-phosphate dehydrogenase (FC, 1.81), aldose reductase (FC, 1.96∼2.14), and non-specific lipid-transfer protein 4-like (FC, 1.89∼2.08) were enriched in cluster 1 (**Figure 5A** and **Table 1**). However, for aged SS seeds, only 2 LEA proteins were significantly overrepresented in the cluster 1, and only 9 DAPs related with stress response, carbohydrate metabolism and storage proteins, such as phosphoglucomutase and inositol-3-phosphate synthase, were enriched in cluster 3, which were primarily assigned to "anaerobic respiration" (**Figure 5B** and **Table 4**). These proteins may play an important role in maintaining higher germination after storage. In addition to that, for the groups of DAPs with significant down-regulation that were specifically change in ST or SS, 38 DAPs (38/47, 80.85%) of ST were mainly enriched in cluster 4 that displayed a decrease at ST90, and only nine were enriched in cluster 5, which were primarily assigned to "antioxidant activity, response to oxidative stress, and aromatic amino acid metabolic process" (**Figure 5C**

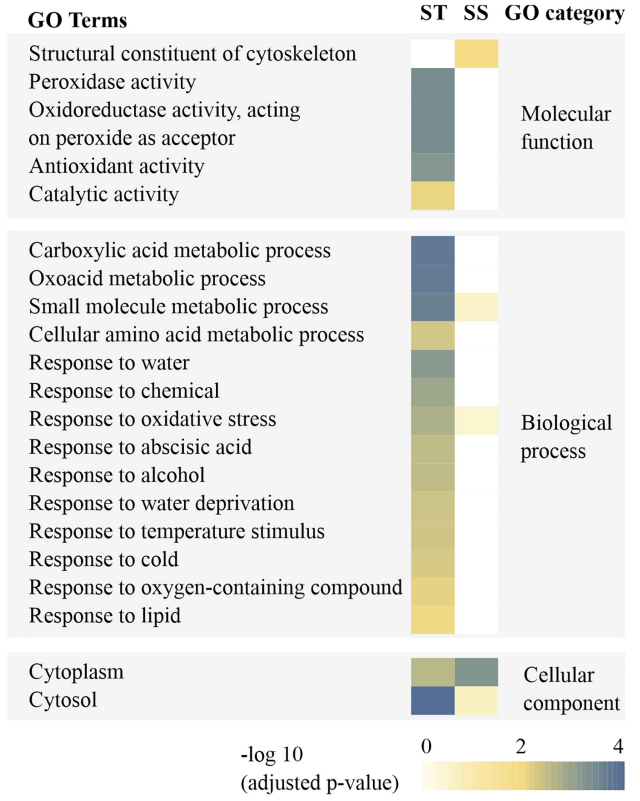


FIGURE 4 | Enrichment analysis for GO annotation of differentially abundant proteins (DAPs) resulting from imbibed embryos of aged storage-tolerant (ST) and storage-sensitive (SS) wheat accessions. GO categories were assigned as significant thresholds for correlation statistics ($p < 0.05$).

and **Table 2**); 13 DAPs of SS were enriched in cluster 4∼7, which were primarily assigned to "cytoskeleton" (**Figure 5D** and **Table 4**). The downregulated proteins mainly associated with metabolism, indicating that the inhibition of metabolic activity of imbibed embryos was not only a consequence of seed aging, but might be a strategy of avoiding rapid energy consumption for stored and *de novo* synthesis of substances to maintain germination ability after storage. These results indicate that seed aging results in a decreased the ability of maintain the normal abundance of proteins associated with antioxidant defense, whereas the response above or close to CN might endow seeds with other effective defensive mechanisms, that are more favorable for avoiding the rapid decline in seed viability and germination. Therefore, effective expression and regulation of proteins before CN plays an important role in maintaining the germination ability of ST seeds.

DISCUSSION

During seed conservation, changes in seed viability vary greatly among intraspecific germplasms (Nagel et al., 2009; Niedzielski et al., 2009; Lee et al., 2019). The safe conservation of germplasm resources is determined by the decline in seed viability from the original level to the CN. In a first study, to understand molecular changes during genebank storage, we had selected two wheat accessions differing in seed storability and investigated the proteome after long-term ambient and cold storage (Chen et al., 2018). In the present study, we aimed to elucidate the variations in the protein profiles around CN of seed viability by applying AA to the two regenerated wheat accessions. Overall, the variation found between the two accessions was mainly exhibited in the stress response, scavenging of cytotoxic compounds, and maintenance of redox and carbon homeostasis by multiple patterns of protein abundance.

Seeds With Longer Plateau Phase Show Stronger Stress Response Around CN

The interaction between non-reducing sugars and certain proteins, such as LEA proteins and small heat shock proteins sHSPs influence seed longevity (Chatelain et al., 2012; Kijak and Ratajczak, 2020). LEAs are hydrophilic and heat-soluble proteins that typically accumulate at high levels during seed maturation, and equip the seeds with desiccation tolerance (Kijak and Ratajczak, 2020). In the present study, we identified 25 LEA proteins showing significantly different expression patterns in imbibed embryos of aged seeds among two accessions differing in storability. Of these, dehydrins (Pfam 00257, DHN1, DHN4, and DHN6), group II LEA proteins were specifically upregulated in ST seeds. The protective activity of dehydrin is known to be related to a largely disordered structural state with conserved, short sequence motifs (K- and H-segments), which enable to enhance the stress tolerance of plants (Andersson et al., 2020; Li et al., 2020; Murvai et al., 2020; Nguyen et al., 2020). The downregulation of seed-expressed DHN can drastically shorten seed longevity in *A. thaliana* (Hundertmark et al., 2011) supporting a key role of some dehydrins for seed storability. Thereby, dehydrins may reduce the oxidative damage which accompanies seed aging by protecting DNA, lipids, and proteomes effectively activating redox enzymes and by facilitating autophagic degradation of cargo proteins (e.g., aquaporin). Thus, we speculate that ST seeds might accumulate abundant dehydrins at the maturation stage, so as to effectively protect from various damages during seed aging and imbibition, and enabling higher germination and consequently a longer plateau phase.

Other LEA proteins, such as the embryonic protein DC-8 and seed maturation protein PM41, attributed to group III (LEA_4, Pfam 02987), were significantly upregulated in long-term stored ST seeds (Liu et al., 2019). During seed aging, these proteins may protect against the inactivation of biomacromolecules by conformational changes to enable survival (Yu et al., 2019). The Em proteins, assigned to LEA_5 (group I, pfam 00477), had a very high correlation ($r^2 = 0.95$) with the acquisition of seed longevity (Chatelain et al., 2012) and showed also significant upregulation in aged ST seeds, enabling an effectively response to adverse environments (Liu et al., 2019). Thus, certain LEA proteins are essential for higher seed storability. However, to determine how these proteins contribute to maintaining the plateau phase of seeds, their precise regulatory networks need to be verified.

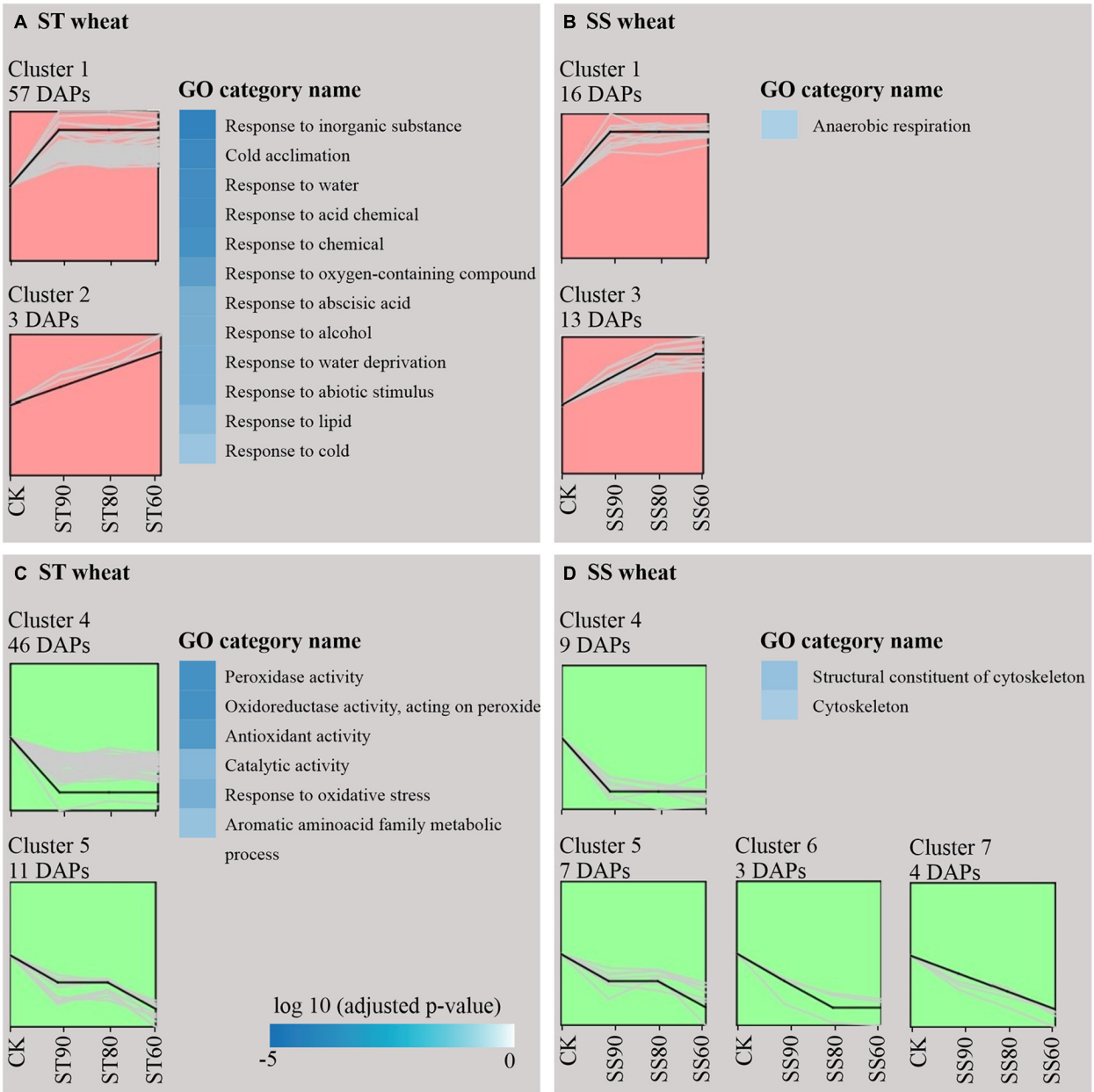


FIGURE 5 | Cluster analysis and GO categories for differentially abundant proteins (DAPs) around (SS90/ST90 and SS60/ST60) and close to the critical node (SS80/ST80) extracted from embryos of imbibed seeds. Up-accumulation of aged storage-tolerant (ST) seeds **(A)** and aged storage-sensitive (SS) seeds **(B)** and down-accumulation of ST seeds **(C)** and SS seeds **(D)**. GO category represent annotation in DAPs with significant trend of upregulation (red) or downregulation (green).

Beside LEA proteins, heat-stable polypeptides, such as sHSPs, may prevent protein aggregation during biotic and abiotic environmental stress (Waters, 2013) and correlated with the establishment of seed longevity (Chatelain et al., 2012). sHSPs with molecular weights ranging from 12 to 42 kDa function as ATP-independent molecular chaperones by methionine-rich amphipathic α-crystalline C-terminal domains (Waters, 2013). Here, we found that three sHSPs (two 16.9 kDa class I HSP1-like and one 23.2 kDa HSP) were upregulated in imbibed embryos of aged ST seeds, which may protect the seeds against ROS-mediated damage and maintain the germination and plateau phase. Indeed, during AA, the function as a chaperone at high temperature might be also considered and may to prevent irreversible thermal inactivation of some enzymes. In conclusion, these results imply that the higher abundance of LEA and sHSP determine the longer span of plateau phase in ST seed, thus, they may prolong span of seed longevity during AA and long-term conservation.

TABLE 1 | Differentially abundant proteins in embryos of ST seeds with significant trend in up-regulation during artificial aging at 40°C and 75% relative humidity.

Accession	Protein description	Function	Fold change			
			ST90/STCK	ST80/STCK	ST60/STCK	SSCK/STCK
Cell defense and rescue						
A0A0H4MAT1	Dehydrin 1	Stress response	1.73*	1.61	1.56	2.16
Q41579	Dehydrin 4	Stress response	2.34*	1.83	2.06*	1.96
W5GW81	Dehydrin 4	Stress response	1.67*	1.44	1.75*	1.63
W5ERW2	Dehydrin 6	Stress response	1.61*	1.27	1.62*	1.36
A0A1D5SU11	Em protein	Stress response	2.51**	2.40*	2.63**	1.60
A0A1D5WW07	Late embryogenesis abundant protein EMB564	Stress response	5.53*	5.72*	4.99*	2.44*
W5CL83	Em-like protein	Stress response	2.47*	2.24	3.43**	1.34
A0A1D6C183	16.9 kDa class I heat shock protein 1-like	Stress response	2.26*	1.77	2.05	1.48
A0A1D6CJL6	16.9 kDa class I heat shock protein 1-like	Stress response	1.83*	1.72*	1.85**	1.40
W5BBE4	23.2 kDa heat shock protein	Stress response	2.40**	2.13**	1.62*	3.94***
A0A1D6REH6	Embryonic protein DC-8	Stress response	2.26*	1.68	2.03*	1.57
E6Y2L2	Seed maturation protein PM41	Stress response	2.74*	2.05	1.87	1.96
W5EJM2	Seed maturation protein PM41	Stress response	1.56*	1.39	1.50*	1.49*
W4ZQK5	Small hydrophilic plant seed protein	Stress response	2.16*	1.69	1.79	1.09
A0A1D5XWV5	Late embryogenesis abundant protein D34	Stress response	1.85*	2.03*	2.18**	1.51
W5FLF3	Late embryogenesis abundant protein D34	Stress response	1.49	1.27	1.59*	1.51
A0A1D5ZRK3	Late embryogenesis abundant protein 3	Stress response	1.92	1.92	2.40*	0.71
A0A1D5URY0	LEA domain-containing protein	Stress response	1.73	1.75	1.95*	1.38
A0A1D6CKF4	Oleosin 16.4 kDa-like	Stress response	1.95**	1.85**	1.49*	1.69*
A0A1D5STX4	Predicted protein	Stress response	1.83***	1.85**	2.19***	2.11***
Q6W8Q2	1-Cys peroxiredoxin PER1	Detoxification	1.53	1.63	1.61*	1.21
A0A1D5ZD36	Glutathione-dependent formaldehyde activating enzyme	Detoxification	1.51*	1.41	1.46	1.29
A0A1D6CB85	Lactoylglutathione lyase	Detoxification	1.77*	1.44	1.50*	1.41
A0A1D5TZ33	Protein-L-isoaspartate *O*-methyltransferase	Cell rescue	1.54	1.56	1.96*	1.07
Metabolism						
W5A6D0	Aldose reductase	Carbohydrate	1.96*	2.14*	2.00*	1.47
W4ZUK2	Glucose and ribitol dehydrogenase homolog	Carbohydrate	1.36	1.44	1.72*	1,26
A0A1D5XI63	Ricin B-like lectin R40C1	Carbohydrate	2.05*	2.63**	1.72	1.58
C7C4 × 1	Glyceraldehyde-3-phosphate dehydrogenase	Energy	1.81*	1.32	1.57	1.22
Q1PBI3	Glucose-6-phosphate isomerase, cytosolic	Energy	1.69*	1.27	1.32	1.41
Protein synthesis and destination						
A0A1D5Z340	60S ribosomal protein L7a	Protein synthesis	2.01*	1.52	1.65	1.53
A0A1D6ASL7	60S ribosomal protein L6-like	Protein synthesis	1.26	1.55*	1.32	1.29
A0A1D5Z5S3	D-aminoacyl-tRNA deacylase	Protein synthesis	1.16	1.33	1.60*	1.25
Transporters						
A0A1D6CN29	Copper transport protein ATX1	Ions transport	1.68*	1.83*	1.58*	1.44
W5EC46	Non-specific lipid-transfer protein 4-like	Lipids transport	1.89*	2.08*	1.34	1.81
W5F7M9	Putative chaperone GrpE type 2	Protein transport	1.06	1.25	2.00*	1.33
Storage protein						
I6QQ39	Globulin-1 S allele	Nutrient reservoir	3.30*	2.07	2.8	1.11
A0A1D5ZZT8	Globulin-1 S allele-like	Nutrient reservoir	1.92*	1.85	2.20*	1.39
Unknown						
A0A1D5XIA7	Predicted protein	Unknown	1.56	1.61	1.89*	1.26
A0A1D6AI79	Predicted protein	Unknown	1.80**	1.41	1.97**	1.46
A0A1D6AZR7	Predicted protein	Unknown	2.43*	2.25	3.45*	2.02
W5CQP6	Predicted protein	Unknown	1.66*	1.73*	1.72*	1.61
W5BGG8	Uncharacterized protein LOC109781115	Unknown	1.68*	1.33	1.57	0.71

*ST90/STCK, ST80/STCK, and ST60/STCK respectively represented the comparison of protein abundance between the aging groups with germination percentage of ∼90, ∼80, ∼60% and the control group. *, **, *** represent that the significant fold change of protein abundance was at the level of $p < 0.05, 0.01, 0.001$, respectively.*

TABLE 2 | Differentially abundant proteins in embryos of ST seeds with significant trend in down-regulation during artificial aging at 40°C and 75% relative humidity.

Accession	Protein description	Function	Fold change			
			ST90/STCK	ST80/STCK	ST60/STCK	SSCK/STCK
Cell defense and rescue						
A0A1D5YB80	Heat shock 70 kDa protein, mitochondrial-like	Stress response	0.52**	0.55**	0.48***	0.47***
W5FEE5	Heat shock cognate 70 kDa protein 2-like	Stress response	0.36*	0.41*	0.48***	0.52
W5FA12	Em-like protein	Stress response	0.35*	0.40*	0.26**	0.42
W5H102	Soluble inorganic pyrophosphatase 6, chloroplastic	Stress response	0.73	0.69	0.39*	0.64
A0A1D5U1N2	GYF domain-containing protein	Stress response	0.69	0.56*	0.56*	0.79
A0A1D5V9V3	Peroxidase 2-like	Detoxification	0.45*	0.50*	0.40*	0.60
C3VQ52	Ascorbate peroxidase	Detoxification	0.59**	0.64*	0.50*	0.78
A0A1D5TUQ5	Cationic peroxidase SPC4-like	Detoxification	0.61	0.49*	0.64*	0.56
A0A1D5UU04	Cationic peroxidase SPC4-like	Detoxification	0.71	0.61*	0.49*	0.99
A0A1D6A4I6	Glutathione S-transferase zeta class	Detoxification	0.78	0.58*	0.61*	0.82
Q8GTB8	Glutathione S-transferase 3-like	Detoxification	0.40**	0.56*	0.58*	1.11
Metabolism						
A0A1D6C6R0	Indole-3-glycerol phosphate synthase, chloroplastic	Amino acid	0.61	0.56*	0.56*	0.85
Q8W430	Sucrose 1-fructosyltransferase	Carbohydrate	0.28*	0.26*	0.56*	0.32*
A0A1D5RYR3	UDP-glucuronic acid decarboxylase 2	Carbohydrate	0.61	0.58*	0.26*	0.65
A0A1D6CYP4	Probable galactinol–sucrose galactosyltransferase 1	Carbohydrate	0.46*	0.62	0.58*	0.69
A0A1D6RST6	Mannose-1-phosphate guanylyltransferase 1	Carbohydrate	0.64	0.49*	0.62	0.69
W5FDW8	UDP-glucose 6-dehydrogenase 4	Carbohydrate	0.3**	0.43*	0.49*	0.67
W5GW19	Predicted protein	Nucleotide	0.74	0.69	0.43*	0.72
A0A1D6S4A3	Deoxyuridine 5'-triphosphate nucleotidohydrolase	Nucleotide	0.68	0.55*	0.69	0.60
A0A1D5Z4H3	Lipoxygenase 2	Lipid	0.33**	0.41*	0.55*	0.59
A0A1D5RUI5	3-ketoacyl-CoA thiolase 2, peroxisomal	Lipid	0.35	0.32*	0.41*	0.46
A0A1D5TUT9	Anthocyanidin reductase	Secondary metabolite	0.57	0.52*	0.32*	0.72
A0A1D5UWN7	Phospho-2-dehydro-3-deoxyheptonate aldolase 2	Secondary metabolite	0.48*	0.44**	0.52*	0.56
A0A1D5Z1D2	Anthranilate synthase alpha subunit 1, chloroplastic	Secondary metabolite	0.77	0.77*	0.44**	0.94
A0A1D6RQI0	3-dehydroquinate dehydratase/shikimate dehydrogenase, chloroplastic-like isoform X1	Secondary metabolite	0.61***	0.69**	0.77*	0.57***
A0A1D5U6U1	Adenylylsulfate kinase	Secondary metabolite	0.58	0.39*	0.69**	0.65
Protein synthesis and destination						
A0A1D5YFC1	Phenylalanine–tRNA ligase alpha subunit, cytoplasmic	Protein synthesis	0.89	0.55*	0.69	0.91
W5AC26	Histidine–tRNA ligase, cytoplasmic	Protein synthesis	0.42*	0.67	0.55*	0.70
Q8L806	40S ribosomal protein S18	Protein synthesis	0.09***	0.12***	0.67	0.10***
A0A1D5RS63	Ribosomal L1 domain-containing protein 1-like	Protein synthesis	0.87	0.62*	0.12***	0.67
W5ATU2	Polyadenylate-binding protein RBP47	Protein synthesis	0.57*	0.43**	0.62*	0.60
A0A1D6BMF2	Protease 2	Protein degradation	0.46*	0.58	0.43**	0.73
W5EJA6	ATP-dependent Clp protease proteolytic subunit 5	Protein degradation	0.77	0.55*	0.58	0.92
W5ASV0	Oryzain alpha chain	Protein degradation	0.37**	0.43*	0.55*	0.67
Transcription						
A0A1D6CRD4	RNA-binding (RRM/RBD/RNP motifs) family protein	mRNA process	0.60*	0.72	0.43*	0.87
A0A1D5UZA9	Glycine-rich RNA-binding protein 4, mitochondrial-like	RNA bind	0.71	0.77	0.72	0.66
A0A1D6C5W5	RNA exonuclease 4	rRNA synthesis	0.95	0.56*	0.77	0.63
A0A1D6AYF2	KH domain-containing protein	mRNA synthesis	0.26*	0.33	0.56*	0.24*
Transporters						
A0A1D5YJB1	AP-2 complex subunit alpha-2-like	Protein transporter	0.58*	0.68	0.33	0.65
A0A1D5S4G8	Golgi apparatus membrane protein-like protein ECHIDNA	Others	0.66	0.47*	0.68	0.64
A0A1D5UDS7	Mitochondrial import inner membrane translocase subunit TIM44-2	Others	0.73	0.58*	0.47*	0.73
A0A1D5SL30	Mitochondrial dicarboxylate/tricarboxylate transporter DTC-like	Others	0.69	0.61*	0.58*	0.79

(Continued)

TABLE 2 | Continued

Accession	Protein description	Function	Fold change			
			ST90/STCK	ST80/STCK	ST60/STCK	SSCK/STCK
Cell structure						
A4K4Y1	Alpha tubulin1	Cytoskeleton	0.57	0.50*	0.61*	0.63
A0A1D5YN83	Annexin D5	Others	0.76	0.43*	0.50*	0.57
Cell growth/division						
A0A1D5ST90	DNA replication licensing factor MCM3	DNA replication	0.36**	0.40*	0.43*	0.57
W5FPI3	DNA replication licensing factor MCM7	DNA replication	0.28***	0.41**	0.40*	0.52
Unknown						
A0A1D5UJV6	Naked endosperm1	Unknown	0.65	0.59	0.41**	0.69

*ST90/STCK, ST80/STCK, and ST60/STCK respectively represented the comparison of protein abundance between the aging groups with germination percentage of ∼90, ∼80, ∼60% and the control group. *, **, *** Represent that the significant fold change of protein abundance was at the level of $p < 0.05, 0.01, 0.001$, respectively.*

The Maintenance of Redox Homeostasis Is a Key Aspect of Longer Plateau Phase

Excess accumulation of ROS leads to an inactivation of antioxidant defense systems and an imbalance of redox homeostasis and attack on cellular components or compartments, causing inactivation of antioxidant defense systems and ultimately aggravating the loss of seed viability (Hu et al., 2012; Xin et al., 2014; Yin et al., 2014, 2017). Downregulation of some peroxidase-related proteins damages the antioxidant system and disturbs redox homeostasis at an earlier stage, hence, effecting seed storability (Xin et al., 2014; Yin et al., 2014, 2017). In the present study, the abundance of 1-Cys peroxiredoxin (1-Cys Prx, also named PER1) was dramatically upregulated in imbibed embryos of aged ST seeds when GP declined below the CN. PER1 is considered as a seed-specific antioxidant with an important role during early water imbibition (Chen et al., 2020) and improves the resistance to stress from various environments (Chen et al., 2016). It detoxifies various ROS including H_2O_2, hydroxyl radicals, and alkyl hydroperoxides by cysteine residues (Chen et al., 2016). In general in wheat, the regeneration and antioxidant activity of PER1 is maintained by the reducing power of the NADPH-dependent thioredoxin reductase/thioredoxin system (Pulido et al., 2009). The expression of thioredoxin domain-containing protein was dramatically downregulated in aging SS seeds below the CN, whereas this expression remained unchanged in ST seeds. These results indicate that the ability to eliminate accumulated ROS is effectively maintained in ST seeds, although seed viability decreases to the below the CN. This ability was completely lost in SS seeds.

In organisms, various spontaneous age-related covalent modifications in proteins often cause fatal damage to their structure and function. Among the modification is the formation of abnormal isoaspartyl (isoAsp) residues, which is mainly caused by either isomerization of aspartyl residues or deamidation of asparaginyl residues (Ghosh et al., 2020). Many studies have shown that massive accumulation of isoAsp in combination with reduced PIMT activity occurs in naturally and artificially aged seeds, leading to a decrease in their germination (Verma et al., 2013). IsoAsp residues negatively affect the catalytic efficiency of antioxidant enzymes, particularly catalase and superoxide dismutase, and that of the PIMT-mediated protein repair system (Petla et al., 2016; Ghosh et al., 2020). PIMT repairs damages by restoring abnormal isoAsp residues to their normal aspartyl residues through an *S*-adenosyl-L-methionine-dependent methyl esterification reaction (Ghosh et al., 2020). The overexpression of PIMT in rice, chickpea, and *Arabidopsis* seeds contributed to improved seed traits, such as longevity and vigor (Ogé et al., 2008; Verma et al., 2013; Wei et al., 2015; Petla et al., 2016). In agreement with this, in our study, PIMT was significantly upregulated in imbibed embryos of aged ST seeds with GP below the CN. On the basis of these results, we speculated that PIMT might be a major endogenous factor to repair the damage, thereby restricting ROS accumulation and lipid peroxidation and protecting the antioxidant system in ST seeds.

In conclusion, higher abundance of proteins related to defense and repair systems in imbibed embryos of ST seeds could provide stronger potential in maintaining redox homeostasis to avoid over-accumulation of various damage, which endowing longer span of plateau phase for seeds with higher storability during conservation.

Maintaining the Carbon Homeostasis Might Be Essential for Longer Plateau Phase During AA

In the present study, sharp upregulation in glyceraldehyde-3-phosphate dehydrogenase involved in the transfer of glyceraldehyde-3-phosphate to 1, 3-bisphosphoglycerate, was only abundant above the CN in embryos of aged ST seeds (**Table 1**). Cytoplasmic glyceraldehyde-3-phosphate dehydrogenase can be induced under many stress conditions, such as anoxia and salt toxicity, and is involved in seed aging (Zhang et al., 2020). Cytoplasmic glucose-6-phosphate isomerase was also upregulated significantly only above the CN in embryos of aged ST seeds (**Table 1**). This enzyme catalyzes the inter-conversion of glycolysis intermediates (fructose-6-phosphate, Fru-6-P) and oxidative pentose phosphate pathway intermediates (glucose-6-phosphate, Glc-6-P) (Kunz et al., 2014). Cytoplasmic phosphoglucomutase was upregulated only below CN of imbibed embryos of aged SS seeds. Phosphoglucomutase

TABLE 3 | List of common differentially abundant proteins with significant trend in both in embryos of ST and SS seeds during artificial aging at 40°C and 75% relative humidity.

Accession	Protein description	Function	Fold change						
			ST90/STCK	ST80/STCK	ST60/STCK	SS90/SSCK	SS80/SSCK	SS60/SSCK	SSCK/STCK
Cell defense and rescue									
A0A1D5Z500	Late embryogenesis abundant protein D-34	Stress response	3.27***	3.08***	4.13***	1.24	2.06	1.89*	2.47**
W5FP47	Late embryogenesis abundant protein D-34	Stress response	1.92**	1.51	1.86**	1.25	1.60*	1.63*	1.47
A0A1D6CR43	11 kDa late embryogenesis abundant protein	Stress response	1.90*	2.11**	1.99**	2.00**	2.18***	2.10***	1.22
A0A1D5SU52	Late embryogenesis abundant protein B19.4	Stress response	4.60***	3.84***	4.92***	1.43	1.71*	1.85*	1.67
W5A4Z8	Late embryogenesis abundant protein B19.4	Stress response	3.40**	2.82*	3.32*	1.39	1.84*	1.88*	1.07
W5CUW2	Late embryogenesis abundant protein 14-A	Stress response	1.17	1.48	2.03*	1.40	2.55*	2.37*	0.90
A0A1D6BIR9	Seed maturation protein	Stress response	2.74	2.92	3.57*	1.25	2.56*	2.89*	1.93
A0A1D6CUR4	Seed maturation protein	Stress response	2.37**	1.93*	2.45**	1.74	2.65***	2.99***	1.48
A0A1D6SDR0	Seed maturation protein	Stress response	1.92*	1.75	2.50**	1.79*	2.99***	3.39***	1.31
Q53WS3	Early methionine-labeled polypeptide	Stress response	6.30***	6.38***	6.08***	1.14	1.73	1.65*	2.73*
Q9ZR70	Early methionine-labeled polypeptide	Stress response	2.15*	1.87*	2.41**	1.79	2.99**	3.67***	1.86
A0A1D5RSF5	Predicted protein	Stress response	2.72**	2.30*	3.12**	1.41	1.68	1.90*	1.08
W5FZJ8	DnaJ protein homolog	Stress response	0.48*	0.49*	0.53*	0.69	0.60*	0.55*	0.71
A0A1D6BJM1	Peroxidase 16	Detoxification	0.51**	0.57*	0.51**	0.83	0.68	0.62*	0.85
W5B6C4	Peroxidase 16	Detoxification	0.59	0.70	0.39*	0.62	0.53*	0.41**	0.94
W5EJK0	L-ascorbate peroxidase 1,	Detoxification	0.27**	0.31**	0.28**	0.56	0.41**	0.39***	0.53*
Metabolism									
A0A077RWS5	*S*-adenosylmethionine synthase 1	Amino acid	0.40**	0.44*	0.37**	0.74	0.66*	0.57*	0.57
W5AFS8	Aldose reductase	Carbohydrate	2.36**	2.30**	2.17**	2.14	2.24*	1.79*	1.49
A0A1D6BAP2	Alpha-amylase 1	Carbohydrate	0.22*	0.41	0.32	0.43	0.77	0.22*	0.50
W5FEY4	Predicted protein	Carbohydrate	0.27**	0.35**	0.30**	0.72	0.53	0.36*	0.59
A0A077RVB3	UDP-arabinopyranose mutase 1-like	Carbohydrate	0.60	0.66	0.46*	0.53	0.61*	0.78	0.72
Cell growth/division									
A0A1D5UQF4	Topless-related protein 2	Cell growth	4.47**	4.10**	3.70*	1.73	2.13*	2.20*	2.01
A0A1D5U093	Predicted protein	Cell growth	3.60*	3.11	3.20*	1.43	2.28*	2.25*	1.63
Signal transduction									
A0A1D5U8D4	Probable protein phosphatase 2C 58	Phosphatases	2.22	2.07	2.75*	1.22	1.94*	2.12*	1.44
Storage protein									
A0A1D5YEH0	Globulin-1 S allele	Nutrient reservoir	2.66**	2.23*	2.36**	1.26	1.49	1.72*	1.79
Transporters									
A0A1D6CVG5	ABC transporter F family member 1	ABC-type	0.47***	0.60**	0.50***	0.66*	0.60***	0.56***	0.86
Cell structure									
A0A1D5WND0	Tubulin beta chain	Cytoskeleton	0.29**	0.38**	0.49	0.57	0.46**	0.48**	0.65
Unknown									
A0A1D6BIB2	Hypothetical protein TRIUR3_06375	Unknown	2.59***	2.32**	2.19**	1.52	2.56***	2.49***	1.58

*ST90/STCK, ST80/STCK, ST60/STCK, SS90/SSCK, SS80/SSCK and SS60/SSCK, respectively, represented the comparison of protein abundance between the aging groups with germination percentage of ∼90, ∼80, ∼60% and the control group in ST and SS seeds. *, **, *** Represent that the fold change of protein abundance was at the level of $p < 0.05, 0.01, 0.001$, respectively.*

TABLE 4 | List of SS- specific differentially abundant proteins with significant trend in up- and down-regulation during artificial aging at 40°C and 75% relative humidity.

Accession	Protein description	Function	Fold change			
			SS90/SSCK	SS80/SSCK	SS60/SSCK	SSCK/STCK
Cell defense and rescue						
A0A1D5YGW1	Late embryogenesis abundant protein D-34	Stress response	1.55	1.94*	2.00*	1.43
A0A1D5YW17	Late embryogenesis abundant protein D-34	Stress response	1.49	1.85*	1.87*	1.21
A0A1D6RWC3	Late embryogenesis abundant protein D-34	Stress response	1.47	1.84*	2.38**	1.53
A0A1D5XF71	Late embryogenesis abundant protein D-34	Stress response	1.35	1.72	1.91*	1.24
A0A080YUM7	Calmodulin-binding protein 60 F-like isoform X2	Stress response	2.14	2.04	2.2*	0.53
Metabolism						
A0A1D5WYC0	Phosphoglucomutase, cytoplasmic	Carbohydrate	1.88	1.6	1.82*	0.83
W5E174	Inositol-3-phosphate synthase	Carbohydrate	2.68	1.89	2.03*	0.75
Storage protein						
W5EST8	Globulin-1 S allele	Nutrient reservoir	1.88	1.61*	1.82*	1.21
B7U6L5	Globulin-1 S allele	Nutrient reservoir	1.89	1.66	1.76*	4.13***
Signal transduction						
A0A1D5ZUF9	Ribosome-binding ATPase YchF	Mediators	1.58	1.6	1.78*	0.50
Transporters						
W5EMC7	Outer envelope pore protein 16-2, chloroplastic	Others	1.36	1.3	1.53*	1.15
Cell defense and rescue						
W5G9P1	DnaJ protein homolog 2-like isoform X2	Stress response	0.63	0.6*	0.48**	0.97
A0A1D5YRV8	Heat shock cognate 70 kDa protein 2-like	Stress response	0.66	0.62	0.43*	0.89
W5FIX9	Macrophage migration inhibitory factor homolog isoform X2	Stress response	0.73	0.77	0.59*	0.74
Metabolism						
A0A1D5XUF5	Outer envelope protein 64, chloroplastic	Amino acid	0.82	0.87	0.52*	1.20
A0A1D5U5C8	Probable pyridoxal 5′-phosphate synthase subunit PDX1.1	Secondary metabolite	0.74	0.67*	0.59*	0.67
Transporters						
Q9ZR95	Aquaporin TIP1-1	Ions transport	0.74	0.66	0.61*	1.26
W5GYF2	ADP, ATP carrier protein, mitochondrial	Others	0.61	0.63	0.59*	1.12
Signal transduction						
A0A1D6BU25	Profilin	Mediators	0.93	0.89	0.64*	0.98
A0A1D5SCB2	DNA-binding protein DDB_G0278111	Mediators	0.83	0.69	0.49*	0.97
Protein synthesis and destination						
Q7XYD5	60S acidic ribosomal protein P2B	Protein synthesis	0.88	0.65	0.58*	1.33
Transcription						
W5CRM8	Glycine-rich RNA-binding protein 4, mitochondrial	mRNA process	0.80	0.88	0.58*	0.91

*SS90/SSCK, SS80/SSCK, and SS60/SSCK, respectively, represented the comparison of protein abundance between the aging groups with germination percentage of ∼90, ∼80, ∼60% and the control group. *, **, *** Represent that the significant fold change of protein abundance was significant respectively at the level of $p < 0.05$, 0.01, 0.001, respectively.*

catalyzes the interconversion of Glc-6-P to Glc-1-P, which is mainly used in biosynthetic pathways (e.g., cell wall polysaccharide synthesis) together with Fru-6-P and Glc-6-P, which constitute the hexose-phosphate pool and represent an important node in primary metabolism (Kunz et al., 2014). Interestingly, this enzyme was significantly upregulated below the CN of viability in SS seeds. These results indicate that ST seeds have a strong ability to maintain highly efficient energy metabolism to assure sufficient energy resources for germination.

Many important enzymes correlated with formation of cell walls were significantly downregulated during aging. For example, below CN, common DAPs in both, in imbibed embryos from aged ST and SS seeds, were linked to the UDP-arabinopyranose mutase 1-like enzyme, important for cell wall establishment because it converts UDP-arabinopyranose to UDP-arabinofuranose (Saqib et al., 2019). More DAPs participating in the formation of arabinoxylans were downregulated in imbibed embryos from aged ST seeds. These include UDP-glucose 6-dehydrogenase 4 and UDP-glucuronic acid decarboxylase 2 that are involved in the synthesis of fructans and *myo*-inositol (Suzuki et al., 2004; Diedhiou et al., 2012). Subsequently, when GP was close to CN, mannose-1-phosphate guanylyltransferase 1 catalyzing the production of GDP-mannose, which is linked with certain secondary changes such as the cell wall composition, *N*-glycosylation, and ascorbic acid production, was also downregulated only in imbibed embryos of age ST seeds (Maeda and Dudareva, 2012). This implies that the establishment of the cell wall is seriously affected in embryos of aged SS seeds but also ST seeds. However, we assume that the difference

in carbon fluxes at the CN among embryos of ST and SS seeds caused the obvious divergence in the plateau phase of the viability curve by regulating carbohydrate metabolism to maintain carbon homeostasis.

Scavenge Cytotoxic Compounds May Cause Differences in Plateau Phase and Storability

Lipid oxidation (Wiebach et al., 2020) and peroxidation (Colville et al., 2012) are the major factor responsible for the loss of seed vigor. In our study, the downregulation of lipoxygenases, an important enzyme associated with the oxidation of polyunsaturated fatty acids, was observed only in imbibed embryos of aged ST seeds. This is consistent with the findings of previous studies in which higher suppression of lipoxygenase expression enhanced the viability and longevity (Song et al., 2007; Xu et al., 2015). In the early stage of seed aging, oxidative damage results in the massive accumulation of lipid peroxide-derived aldehydes and ketones, especially α, β-unsaturated carbonyls, which generate advanced glycation end products and advanced lipoxidation end products by aggregating with proteins and lipids (Nisarga et al., 2017; Fu et al., 2018). Furthermore, at a level below the CN, the 3-ketoacyl-CoA thiolase 2 (KAT), which participates in the last step of the β-oxidation cycle, was significantly downregulated in imbibed embryos of aged ST seeds. A previous study showed that these cytotoxic compounds seriously affected seed germination and seedling vigor by altering metabolic pathways (Nisarga et al., 2017).

Aldose reductase is a key enzyme involved in the accumulation of sorbitol by reducing aldose (e.g., glucose and glucose-6-phosphate) in the polyol pathway to enhance stress tolerance. It also participates in the detoxification of reactive carbonyl metabolites (e.g., acrolein) (Tari et al., 2010). In our study, we found that this enzyme was significantly upregulated in both ST and SS seeds; however, in ST seeds, aldose reductase always maintained a higher abundance at the three levels of around CN. Massive accumulation of methylglyoxal has also been demonstrated in aged rice seeds (Nisarga et al., 2017). This compound is formed via important metabolic pathways, such as the non-enzymatic breakdown of triose phosphate isomers during glycolysis, auto-oxidation of ketone bodies and sugars, Maillard reaction between reducing sugars and amino acids, and lipid peroxidation (Mostofa et al., 2018). Lactoylglutathione lyase (known as glyoxalase I, GLY I, EC 4.4.1.5) is involved in methylglyoxal detoxification by converting glutathione and methylglyoxal into *S*-D-lactoylglutathione (Mostofa et al., 2018). Moreover, there is a relationship between stress response and resistance in plants and the abundance of GLY I protein. In particular, tolerant genotypes demonstrate higher expression of GLY genes, proteins, and their activities (Hasanuzzaman et al., 2014; Yan et al., 2016). Similarly, only ST seeds with higher storability were significantly upregulated in the aging group, which gave the stronger ability in scavenging cytotoxic compounds to postpone the appearance of CN and rapid loss of seed viability, thereby maintaining a higher germination viability.

Based on the above results, the multiple responses in the abundance of protein related to scavenging of cytotoxic compounds provided effective protection, to obtain the longer span of plateau phase in aged ST seeds, thus prolonging span of seed longevity during conservation.

CONCLUSION

In conclusion, the present study compared the variations in the loss of seed viability and proteomic profiles of imbibed embryos of aged ST and SS seeds. The results of comparative proteomics performed around the CN showed that ST seeds with prolonged plateau phase during AA have a higher ability to prevent oxidative damage and excess accumulation of cytotoxic compounds, and to maintain redox and carbon homeostasis. Therefore, we conclude that ST seed have a higher capability to deal with the stresses occurring during long-term genebank storage. However, to identify potential inherent genetic factors and the fine regulatory mechanisms for prediction tool, further investigation is needed.

DATA AVAILABILITY STATEMENT

The original contributions presented in the study are publicly available. This data can be found here: ProteomeXchange with identifier PXD025860.

AUTHOR CONTRIBUTIONS

XC, GY, and XX designed the experiments and analyzed the data. XC and GY conducted the experiments. XX, JH, JL, and NL contributed materials, reagents, and analysis tools. XC, AB, MN, and GY wrote the manuscript. All authors contributed to the article and approved the submitted version.

FUNDING

This work was supported by the National Key R&D Program of China (2017YFE0105100) and Agricultural Science and Technology Innovation Program.

ACKNOWLEDGMENTS

We would like to thank Lu Zhuang from Plant Science Facility of the Institute of Botany, Chinese Academy of Sciences for her excellent analysis of MS/MS.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpls.2021.707184/full#supplementary-material

REFERENCES

Andersson, J. M., Pham, Q. D., Mateos, H., Eriksson, S., Harryson, P., and Sparr, E. (2020). The plant dehydrin LTI30 stabilizes lipid lamellar structures in varying hydration conditions. *J. Lipid Res.* 61, 1014–1024. doi: 10.1194/JLR.RA120000624

Bevan, M., Bancroft, I., Bent, E., Love, K., Goodman, H., Dean, C., et al. (1998). Analysis of 1.9 Mb of contiguous sequence from chromosome 4 of *Arabidopsis thaliana*. *Nature* 391, 485–488. doi: 10.1038/35140

Börner, A., Nagel, M., Agacka-Mołdoch, M., Gierke, P. U., Oberforster, M., Albrecht, T., et al. (2018). QTL analysis of falling number and seed longevity in wheat (*Triticum aestivum* L.). *J. Appl. Genet.* 59, 35–42. doi: 10.1007/s13353-017-0422-5

Bradford, M. (1976). A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding. *Anal. Biochem.* 72, 248–254. doi: 10.1006/abio.1976.9999

Chatelain, E., Hundertmark, M., Leprince, O., Le Gall, S., Satour, P., Deligny-Penninck, S., et al. (2012). Temporal profiling of the heat-stable proteome during late maturation of Medicago truncatula seeds identifies a restricted subset of late embryogenesis abundant proteins associated with longevity. *Plant Cell Environ.* 35, 1440–1455. doi: 10.1111/j.1365-3040.2012.02501.x

Chen, H., Ruan, J., Chu, P., Fu, W., Liang, Z., Li, Y., et al. (2020). AtPER1 enhances primary seed dormancy and reduces seed germination by suppressing the ABA catabolism and GA biosynthesis in *Arabidopsis* seeds. *Plant J.* 101, 310–323. doi: 10.1111/tpj.14542

Chen, H. H., Chu, P., Zhou, Y. L., Ding, Y., Li, Y., Liu, J., et al. (2016). Ectopic expression of NnPER1, a Nelumbo nucifera 1-cysteine peroxiredoxin antioxidant, enhances seed longevity and stress tolerance in *Arabidopsis*. *Plant J.* 88, 608–619. doi: 10.1111/tpj.13286

Chen, X., Yin, G., Börner, A., Xin, X., He, J., Nagel, M., et al. (2018). Comparative physiology and proteomics of two wheat genotypes differing in seed storage tolerance. *Plant Physiol. Biochem.* 130, 455–463. doi: 10.1016/j.plaphy.2018.07.022

Clerkx, E. J. M., Blankestijn-De Vries, H., Ruys, G. J., Groot, S. P. C., and Koornneef, M. (2004). Genetic differences in seed longevity of various *Arabidopsis* mutants. *Physiol. Plant.* 121, 448–461. doi: 10.1111/j.0031-9317.2004.00339.x

Colville, L., Bradley, E. L., Lloyd, A. S., Pritchard, H. W., Castle, L., and Kranner, I. (2012). Volatile fingerprints of seeds of four species indicate the involvement of alcoholic fermentation, lipid peroxidation, and Maillard reactions in seed deterioration during ageing and desiccation stress. *J. Exp. Bot.* 63, 6519–6530. doi: 10.1093/jxb/ers307

Diedhiou, C., Gaudet, D., Liang, Y., Sun, J., Lu, Z. X., Eudes, F., et al. (2012). Carbohydrate profiling in seeds and seedlings of transgenic triticale modified in the expression of sucrose:sucrose-1-fructosyltransferase (1-SST) and sucrose:fructan-6-fructosyltransferase (6-SFT). *J. Biosci. Bioeng.* 114, 371–378. doi: 10.1016/j.jbiosc.2012.05.008

Ernst, J., and Bar-Joseph, Z. (2006). STEM: a tool for the analysis of short time series gene expression data. *BMC Bioinformatics* 7:191. doi: 10.1186/1471-2105-7-191

FAO (2014). *Genebank Standards for Plant Genetic Resources for Food and Agriculture*. Rome: Food and Agriculture Organization of the United Nations.

Fu, S. Z., Yin, G. K., Xin, X., Wu, S. H., Wei, X. H., and Lu, X. X. (2018). Levels of crotonaldehyde and 4-hydroxy-(E)-2-nonenal and expression of genes encoding carbonyl-scavenging enzyme at critical node during rice seed aging. *Rice Sci.* 25, 152–160. doi: 10.1016/j.rsci.2018.04.003

Ghosh, S., Kamble, N. U., Verma, P., Salvi, P., Petla, B. P., Roy, S., et al. (2020). *Arabidopsis* protein L-ISOASPARTYL METHYLTRANSFERASE repairs isoaspartyl damage to antioxidant enzymes and increases heat and oxidative stress tolerance. *J. Biol. Chem.* 295, 783–799. doi: 10.1074/jbc.RA119.010779

Hasanuzzaman, M., Alam, M. M., Rahman, A., Hasanuzzaman, M., Nahar, K., and Fujita, M. (2014). Exogenous proline and glycine betaine mediated upregulation of antioxidant defense and glyoxalase systems provides better protection against salt-induced oxidative stress in two rice (*Oryza sativa* L.) varieties. *Biomed. Res. Int.* 2014:757219. doi: 10.1155/2014/757219

He, H., Willems, L. A. J., Batushansky, A., Fait, A., Hanson, J., Nijveen, H., et al. (2016). Effects of parental temperature and nitrate on seed performance are reflected by partly overlapping genetic and metabolic pathways. *Plant Cell Physiol.* 57, 473–487. doi: 10.1093/pcp/pcv207

Hu, D., Ma, G., Wang, Q., Yao, J., Wang, Y., Pritchard, H. W., et al. (2012). Spatial and temporal nature of reactive oxygen species production and programmed cell death in elm (*Ulmus pumila* L.) seeds during controlled deterioration. *Plant Cell Environ.* 35, 2045–2059. doi: 10.1111/j.1365-3040.2012.02535.x

Hundertmark, M., Buitink, J., Leprince, O., and Hincha, D. K. (2011). The reduction of seed-specific dehydrins reduces seed longevity in *Arabidopsis thaliana*. *Seed Sci. Res.* 21, 165–173. doi: 10.1017/S0960258511000079

Kijak, H., and Ratajczak, E. (2020). What do we know about the genetic basis of seed desiccation tolerance and longevity? *Int. J. Mol. Sci.* 21:3612. doi: 10.3390/ijms21103612

Kunz, H. H., Zamani-Nour, S., Häusler, R. E., Ludewig, K., Schroeder, J. I., Malinova, I., et al. (2014). Loss of cytosolic phosphoglucose isomerase affects carbohydrate metabolism in leaves and is essential for fertility of *Arabidopsis*. *Plant Physiol.* 166, 753–765. doi: 10.1104/pp.114.241091

Landjeva, S., Lohwasser, U., and Börner, A. (2010). Genetic mapping within the wheat D genome reveals QTL for germination, seed vigour and longevity, and early seedling growth. *Euphytica* 171, 129–143. doi: 10.1007/s10681-009-0016-3

Lee, J. S., Velasco-Punzalan, M., Pacleb, M., Valdez, R., Kretzschmar, T., McNally, K. L., et al. (2019). Variation in seed longevity among diverse Indica rice varieties. *Ann. Bot.* 124, 447–460. doi: 10.1093/aob/mcz093

Li, H., Zhou, Y., Xin, W., Wei, Y., Zhang, J., and Guo, L. (2019). Wheat breeding in northern China: achievements and technical advances. *Crop J.* 7, 718–729. doi: 10.1016/j.cj.2019.09.003

Li, T., Zhang, Y., Wang, D., Liu, Y., Dirk, L. M. A., Goodman, J., et al. (2017). Regulation of seed vigor by manipulation of raffinose family oligosaccharides in maize and *Arabidopsis thaliana*. *Mol. Plant* 10, 1540–1555. doi: 10.1016/j.molp.2017.10.014

Li, X., Liu, Q., Feng, H., Deng, J., Zhang, R., Wen, J., et al. (2020). Dehydrin MtCAS31 promotes autophagic degradation under drought stress. *Autophagy* 16, 862–877. doi: 10.1080/15548627.2019.1643656

Liu, H., Xing, M., Yang, W., Mu, X., Wang, X., Lu, F., et al. (2019). Genome-wide identification of and functional insights into the late embryogenesis abundant (LEA) gene family in bread wheat (*Triticum aestivum*). *Sci. Rep.* 9:13375. doi: 10.1038/s41598-019-49759-w

Lu, X. X., Chen, X. L., and Guo, Y. H. (2005). Seed germinability of 23 crop species after a decade of storage in the National Genebank of China. *Agric. Sci. China* 4, 408–412.

Maeda, H., and Dudareva, N. (2012). The shikimate pathway and aromatic amino acid biosynthesis in plants. *Annu. Rev. Plant Biol.* 63, 73–105. doi: 10.1146/annurev-arplant-042811-105439

Mostofa, M. G., Ghosh, A., Li, Z. G., Siddiqui, M. N., Fujita, M., and Tran, L. S. P. (2018). Methylglyoxal – a signaling molecule in plant abiotic stress responses. *Free Radic. Biol. Med.* 122, 96–109. doi: 10.1016/j.freeradbiomed.2018.03.009

Murvai, N., Kalmar, L., Szalaine Agoston, B., Szabo, B., Tantos, A., Csikos, G., et al. (2020). Interplay of structural disorder and short binding elements in the cellular chaperone function of plant dehydrin ERD14. *Cells* 9:1856. doi: 10.3390/cells9081856

Nagel, M., Seal, C. E., Colville, L., Rodenstein, A., Un, S., Richter, J., et al. (2019). Wheat seed ageing viewed through the cellular redox environment and changes in pH. *Free Radic. Res.* 53, 641–654. doi: 10.1080/10715762.2019.1620226

Nagel, M., Vogel, H., Landjeva, S., Buck-Sorlin, G., Lohwasser, U., Scholz, U., et al. (2009). Seed conservation in ex situ genebanks-genetic studies on longevity in barley. *Euphytica* 170, 5–14. doi: 10.1007/s10681-009-9975-7

Nguyen, P. N., Tossounian, M.-A., Kovacs, D. S., Thu, T. T., Stijlemans, B., Vertommen, D., et al. (2020). Dehydrin ERD14 activates glutathione transferase Phi9 in *Arabidopsis thaliana* under osmotic stress. *Biochim. Biophys. Acta Gen. Subj.* 1864:129506. doi: 10.1016/j.bbagen.2019.129506

Nguyen, T. P., Cueff, G., Hegedus, D. D., Rajjou, L., and Bentsink, L. (2015). A role for seed storage proteins in *Arabidopsis* seed longevity. *J. Exp. Bot.* 66, 6399–6413. doi: 10.1093/jxb/erv348

Niedzielski, M., Walters, C., Luczak, W., Hill, L. M., Wheeler, L. J., and Puchalski, J. (2009). Assessment of variation in seed longevity within rye, wheat and the intergeneric hybrid triticale. *Seed Sci. Res.* 19, 213–224. doi: 10.1017/S0960258509990110

Nisarga, K. N., Vemanna, R. S., Kodekallu Chandrashekar, B., Rao, H., Vennapusa, A. R., Narasimaha, A., et al. (2017). Aldo-ketoreductase 1 (AKR1) improves seed longevity in tobacco and rice by detoxifying reactive cytotoxic compounds generated during ageing. *Rice* 10:11. doi: 10.1186/s12284-017-0148-3

Ogé, L., Bourdais, G., Bove, J., Collet, B., Godin, B., Granier, F., et al. (2008). Protein repair L-Isoaspartyl methyltransferase l is involved in both seed longevity and germination vigor in *arabidopsis*. *Plant Cell* 20, 3022–3037. doi: 10.1105/tpc.108.058479

Petla, B. P., Kamble, N. U., Kumar, M., Verma, P., Ghosh, S., Singh, A., et al. (2016). Rice PROTEIN l-ISOASPARTYL METHYLTRANSFERASE isoforms differentially accumulate during seed maturation to restrict deleterious isoAsp and reactive oxygen species accumulation and are implicated in seed vigor and longevity. *New Phytol.* 211, 627–645. doi: 10.1111/nph.13923

Pulido, P., Cazalis, R., and Cejudo, F. J. (2009). An antioxidant redox system in the nucleus of wheat seed cells suffering oxidative stress. *Plant J.* 57, 132–145. doi: 10.1111/j.1365-313X.2008.03675.x

Rajjou, L., Lovigny, Y., Groot, S. P. C., Belghazi, M., Job, C., and Job, D. (2008). Proteome-wide characterization of seed aging in *Arabidopsis*: a comparison between artificial and natural aging protocols. *Plant Physiol.* 148, 620–641. doi: 10.1104/pp.108.123141

Rehman Arif, M. A., and Börner, A. (2019). Mapping of QTL associated with seed longevity in durum wheat (*Triticum durum* Desf.). *J. Appl. Genet.* 60, 33–36. doi: 10.1007/s13353-018-0477-y

Rehman Arif, M. A., and Börner, A. (2020). An SNP based GWAS analysis of seed longevity in wheat. *Cereal Res. Commun.* 48, 149–156. doi: 10.1007/s42976-020-00025-0

Rehman Arif, M. A., Nagel, M., Lohwasser, U., and Börner, A. (2017). Genetic architecture of seed longevity in bread wheat (*Triticum aestivum* L.). *J. Biosci.* 42, 81–89. doi: 10.1007/s12038-016-9661-6

Rehman Arif, M. A., Nagel, M., Neumann, K., Kobiljski, B., Lohwasser, U., and Börner, A. (2012). Genetic studies of seed longevity in hexaploid wheat using segregation and association mapping approaches. *Euphytica* 186, 1–13. doi: 10.1007/s10681-011-0471-5

Reimand, J., Isserlin, R., Voisin, V., Kucera, M., Tannus-Lopes, C., Rostamianfar, A., et al. (2019). Pathway enrichment analysis and visualization of omics data using g:Profiler, GSEA, Cytoscape and EnrichmentMap. *Nat. Protoc.* 14, 482–517. doi: 10.1038/s41596-018-0103-9

Rissel, D., Losch, J., and Peiter, E. (2014). The nuclear protein Poly(ADP-ribose) polymerase 3 (AtPARP3) is required for seed storability in *Arabidopsis thaliana*. *Plant Biol.* 16, 1058–1064. doi: 10.1111/plb.12167

Saqib, A., Scheller, H. V., Fredslund, F., and Welner, D. H. (2019). Molecular characteristics of plant UDP-arabinopyranose mutases. *Glycobiology* 29, 839–846. doi: 10.1093/glycob/cwz067

Schwämmle, V., León, I. R., and Jensen, O. N. (2013). Assessment and improvement of statistical tools for comparative proteomics analysis of sparse data sets with few experimental replicates. *J. Proteome Res.* 12, 3874–3883. doi: 10.1021/pr400045u

Song, M., Wu, Y., Zhang, Y., Liu, B. M., Jiang, J. Y., Xu, X., et al. (2007). Mutation of rice (*Oryza sativa* L.) LOX-1/2 near-isogenic lines with ion beam implantation and study of their storability. *Nucl. Instr. Methods Phys. Res. Sect. B Beam Interact. Mater. Atoms* 265, 495–500. doi: 10.1016/j.nimb.2007.09.035

Suzuki, K., Watanabe, K., Masumura, T., and Kitamura, S. (2004). Characterization of soluble and putative membrane-bound UDP-glucuronic acid decarboxylase (OsUXS) isoforms in rice. *Arch. Biochem. Biophys.* 431, 169–177. doi: 10.1016/j.abb.2004.08.029

Tari, I., Kiss, G., Deér, A. K., Csiszár, J., Erdei, L., Gallé, A., et al. (2010). Salicylic acid increased aldose reductase activity and sorbitol accumulation in tomato plants under salt stress. *Biol. Plant.* 54, 677–683. doi: 10.1007/s10535-010-0120-1

Verma, P., Kaur, H., Petla, B. P., Rao, V., Saxena, S. C., and Majee, M. (2013). PROTEIN L-ISOASPARTYL METHYLTRANSFERASE2 Is Differentially expressed in chickpea and enhances seed vigor and longevity by reducing abnormal isoaspartyl accumulation predominantly in seed nuclear proteins. *Plant Physiol.* 161, 1141–1157. doi: 10.1104/pp.112.206243

Walters, C., Walters, C., Wheeler, L. M., and Grotenhuis, J. M. (2005). Longevity of seeds stored in a genebank: species characteristics. *Seed Sci. Res.* 15, 1–20. doi: 10.1079/SSR2004195

Waters, E. R. (2013). The evolution, function, structure, and expression of the plant sHSPs. *J. Exp. Bot.* 64, 391–403. doi: 10.1093/jxb/ers355

Waterworth, W. M., Bray, C. M., and West, C. E. (2015). The importance of safeguarding genome integrity in germination and seed longevity. *J. Exp. Bot.* 66, 3549–3558. doi: 10.1093/jxb/erv080

Wei, Y., Xu, H., Diao, L., Zhu, Y., Xie, H., Cai, Q., et al. (2015). Protein repairl-isoaspartyl methyltransferase 1 (PIMT1) in rice improves seed longevity by preserving embryo vigor and viability. *Plant Mol. Biol.* 89, 475–492. doi: 10.1007/s11103-015-0383-1

Wiebach, J., Nagel, M., Börner, A., Altmann, T., Riewe, D. (2020). Age-dependent loss of seed viability is associated with increased lipid oxidation and hydrolysis. *Plant Cell Environ.* 43, 303–314. doi: 10.1111/pce.13651

Wiśniewski, J. R., Zougman, A., Nagaraj, N., and Mann, M. (2009). Universal sample preparation method for proteome analysis. *Nat. Methods* 6, 359–362. doi: 10.1038/nmeth.1322

Xin, X., Tian, Q., Yin, G., Chen, X., Zhang, J., Ng, S., et al. (2014). Reduced mitochondrial and ascorbate–glutathione activity after artificial ageing in soybean seed. *J. Plant Physiol.* 171, 140–147. doi: 10.1016/j.jplph.2013.09.016

Xu, H., Wei, Y., Zhu, Y., Lian, L., Xie, H., Cai, Q., et al. (2015). Antisense suppression of LOX3 gene expression in rice endosperm enhances seed longevity. *Plant Biotechnol. J.* 13, 526–539. doi: 10.1111/pbi.12277

Yan, G., Lv, X., Gao, G., Li, F., Li, J., Qiao, J., et al. (2016). Identification and characterization of a glyoxalase i gene in a rapeseed cultivar with seed thermotolerance. *Front. Plant Sci.* 7:150. doi: 10.3389/fpls.2016.00150

Yazdanpanah, F., Maurino, V. G., Mettler-Altmann, T., Buijs, G., Bailly, M., Karimi Jashni, M., et al. (2019). NADP-MALIC ENZYME 1 affects germination after seed storage in *Arabidopsis thaliana*. *Plant Cell Physiol.* 60, 318–328. doi: 10.1093/pcp/pcy213

Yin, G., Whelan, J., Wu, S., Zhou, J., Chen, B., Chen, X., et al. (2016). Comprehensive mitochondrial metabolic shift during the critical node of seed ageing in rice. *PLoS One* 11:e0148013. doi: 10.1371/journal.pone.0148013

Yin, G., Xin, X., Fu, S., An, M., Wu, S., Chen, X., et al. (2017). Proteomic and carbonylation profile analysis at the critical node of seed ageing in *Oryza sativa*. *Sci. Rep.* 7:40611. doi: 10.1038/srep40611

Yin, G., Xin, X., Song, C., Chen, X., Zhang, J., Wu, S., et al. (2014). Activity levels and expression of antioxidant enzymes in the ascorbate-glutathione cycle in artificially aged rice seed. *Plant Physiol. Biochem.* 80, 1–9. doi: 10.1016/j.plaphy.2014.03.006

Yu, Z., Wang, X., Tian, Y., Zhang, D., and Zhang, L. (2019). The functional analysis of a wheat group 3 late embryogenesis abundant protein in *Escherichia coli* and *Arabidopsis* under abiotic stresses. *Plant Signal. Behav.* 14:1667207. doi: 10.1080/15592324.2019.1667207

Zhang, L., Lei, D., Deng, X., Li, F., Ji, H., and Yang, S. (2020). Cytosolic glyceraldehyde-3-phosphate dehydrogenase 2/5/6 increase drought tolerance via stomatal movement and reactive oxygen species scavenging in wheat. *Plant Cell Environ.* 43, 836–853. doi: 10.1111/pce.13710

Zhang, Y. X., Xu, H. H., Liu, S. J., Li, N., Wang, W. Q., Møller, I. M., et al. (2016). Proteomic analysis reveals different involvement of embryo and endosperm proteins during aging of Yliangyou 2 hybrid rice seeds. *Front. Plant Sci.* 7:1394. doi: 10.3389/fpls.2016.01394

Zuo, J., Liu, J., Gao, F., Yin, G., Wang, Z., Chen, F., et al. (2018). Genome-wide linkage mapping reveals QTls for seed vigor-related traits under artificial aging in common wheat (*Triticum aestivum*). *Front. Plant Sci.* 9:1101. doi: 10.3389/fpls.2018.01101

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

frontiers in Plant Science

ORIGINAL RESEARCH
published: 15 September 2021
doi: 10.3389/fpls.2021.735719

Delayed Protein Changes During Seed Germination

Bing Bai[1†], Niels van der Horst[2], Jan H. Cordewener[3,4,5], Antoine H. P. America[3,4,5], Harm Nijveen[2] and Leónie Bentsink[1*]*

[1]Wageningen Seed Science Centre, Laboratory of Plant Physiology, Wageningen University, Wageningen, Netherlands, [2]Bioinformatics Group, Wageningen University, Wageningen, Netherlands, [3]BU Bioscience, Wageningen Plant Research, Wageningen, Netherlands, [4]Centre for BioSystems Genomics, Wageningen, Netherlands, [5]Netherlands Proteomics Centre, Utrecht, Netherlands

OPEN ACCESS

Edited by:
Wei Wang,
Henan Agricultural University, China

Reviewed by:
Ludovit Skultety,
Institute of Virology (SAS), Slovakia
Xiaojian Yin,
China Pharmaceutical University, China

***Correspondence:**
Bing Bai
bing.bai@bio.ku.dk
Leónie Bentsink
leonie.bentsink@wur.nl

†Present address:
Bing Bai,
Department of Biology, University of Copenhagen, Copenhagen, Denmark

Specialty section:
This article was submitted to Plant Proteomics and Protein Structural Biology, a section of the journal Frontiers in Plant Science

Received: *03 July 2021*
Accepted: *05 August 2021*
Published: *15 September 2021*

Citation:
Bai B, van der Horst N, Cordewener JH, America AHP, Nijveen H and Bentsink L (2021) Delayed Protein Changes During Seed Germination. Front. Plant Sci. 12:735719. doi: 10.3389/fpls.2021.735719

Over the past decade, ample transcriptome data have been generated at different stages during seed germination; however, far less is known about protein synthesis during this important physiological process. Generally, the correlation between transcript levels and protein abundance is low, which strongly limits the use of transcriptome data to accurately estimate protein expression. Polysomal profiling has emerged as a tool to identify mRNAs that are actively translated. The association of the mRNA to the polysome, also referred to as translatome, provides a proxy for mRNA translation. In this study, the correlation between the changes in total mRNA, polysome-associated mRNA, and protein levels across seed germination was investigated. The direct correlation between polysomal mRNA and protein abundance at a single time-point during seed germination is low. However, once the polysomal mRNA of a time-point is compared to the proteome of the next time-point, the correlation is much higher. 35% of the investigated proteome has delayed changes at the protein level. Genes have been classified based on their delayed protein changes, and specific motifs in these genes have been identified. Moreover, mRNA and protein stability and mRNA length have been found as important predictors for changes in protein abundance. In conclusion, polysome association and/or dissociation predicts future changes in protein abundance in germinating seeds.

Keywords: ribosome, proteome, translation delay, seed germination, translatome, *Arabidopsis* thaliana, polysome profiling

INTRODUCTION

The central dogma in molecular biology is that a gene is transcribed into a mRNA, which is further translated into a protein. Data produced by the high-throughput techniques, such as next-generation sequencing and mass spectrometry, have shown that there is a lack of correlation between mRNA levels and protein abundance levels (Chen et al., 2002; Nie et al., 2006; Haider and Pal, 2013). This lack of correlation might partly be hampered by the low throughput of protein analysis compared to the high-throughput transcriptome analysis. In addition, cellular processes, such as posttranscriptional and translational regulation as well as different half-lives for mRNAs and proteins, limit correlation between the mRNA and protein levels (Varshavsky, 1996; Grolleau et al., 2002; Becker et al., 2018). Multiple studies have shown

that protein abundance levels are better explained by ribosome-associated mRNA (translatome) measurements than by transcriptome data in a wide variety of organisms. Based on this, the translatome is proposed as a better proxy for predicting proteomic changes (Beyer et al., 2004; Lu et al., 2007; Smircich et al., 2015). In this study, germinating seeds were used as a system to identify whether indeed the translatome can be used to predict protein level changes.

Seed germination is an important trait that secures seedling establishment and thus the next generation of a plant. Plant seeds accumulate myriads of mRNAs during seed maturation, and these mRNAs are translated during early seed imbibition (Rajjou et al., 2004). This temporal separation of transcription and translation is similar to translational events during embryo development in other organisms (Kronja et al., 2014). Moreover, it has been shown that the translation of these seed stored mRNAs is essential for seed germination (Suzuki and Minamikawa, 1985; Beltran-Pena et al., 1995; Bai et al., 2020). Seed germination is a triphasic process triggered by the uptake of water, referred to as imbibition. Phase I involves water uptake during which *de novo* transcription is not essential (Rajjou et al., 2004). This phase ends when the water uptake reaches a plateau. In phase II, multiple essential cellular and biochemical processes, including protein synthesis, are activated. Phase III is the phase of *sensu stricto* germination ending with the embryonic root emergence (Bradford, 1990). Previously, we have shown substantial translational regulation of mRNAs at specific developmental stages during seed germination. Thousands of mRNAs showed distinct patterns of expression between the total mRNA and polysomal mRNA pool. This indicates that transcribed mRNAs are not always bound by ribosomes and directed for protein synthesis (Bai et al., 2017). Polysome profiling is tool to characterize the polysome-associated mRNAs (translatome; Mustroph et al., 2009, Pringle et al., 2019). The comparison of the polysomal associated mRNAs to the total mRNA on a genome-wide scale hints at the translational efficiency of these mRNAs (Branco-Price et al., 2005; Gamm et al., 2014; Bai et al., 2017, 2018). However, it is not known if polysomal mRNA changes can predict the protein abundance during the seed germination. Here, we combine our previous study in which we have identified polysome-associated mRNAs during seed germination with protein abundance obtained from the same samples. Therefore, disturbance by repeated sampling variation can be excluded. We reveal a global delay in protein-level changes and identified potential sequence features that might contribute to this delay in protein-level changes. Specific motifs identified on the mRNAs for which the protein-level changes are delayed suggest a role for RNA-binding proteins (RBPs) in mediating a delay in protein translation or degradation.

MATERIALS AND METHODS

Plant Material Sampling and Total Protein Extraction

Maturing *Arabidopsis thaliana* seeds (accession Columbia-0) were grown in three biological replicates with 50 plants each. Plants were grown on 4 × 4 cm Rockwool blocks irrigated with standard nutrient solution as (He et al., 2014) in a growth chamber at 20/18°C (day/night) under a 16-h day/8-h night photoperiod of artificial light (150 µmol m^{-2} s^{-1}) and 70% relative humidity. Seeds at specific physiological states during germination were harvested and frozen in liquid nitrogen, freeze-dried, and stored at −80°C until further analyzed. Protein extraction, purification, chromatography and mass spectrometry and database search for protein identification, and quantitative analysis are included in **Supplementary Methods**.

Data Retrieval and Statistical Analysis

The microarray data of total mRNA and polysomal mRNA were retrieved from the Gene Expression Omnibus repository[1] accession number GSE65780 (Bai et al., 2017). Log2-fold transformed microarray intensity data for the transcriptome/translatome and Log2 LFQ abundance signal for the proteomic data were used for the correlation analysis. Furthermore, sequence data [5'UTR, 3'UTR, and CDS (coding DNA sequence)] and predicted subcellular location were retrieved from the TAIR10 database from The Arabidopsis Information Resource (TAIR; Berardini et al., 2015). To identify any bias in the data, the total mRNA and polysomal mRNA Log2 intensity data were retrieved for the genes for which protein abundance is available from the current study. The expression for the subset which the protein data are available from current study is compared to the whole dataset, including all expressed genes on mRNA and polysomal mRNA level in the seeds (19,781 genes). The Wilcoxon Rank Sum test is used to test whether there is a significant difference between the groups. The predicted subcellular location was retrieved from the TAIR10 database, and the subcellular enrichment was calculated for the complete gene set and the subset for which protein expression was available. For correlation analysis, the Shapiro test was used to calculate whether the mRNA, polysomal mRNA, and protein abundance data are normally distributed. Spearmen correlation analysis was used to calculate the correlation coefficient. To calculate the differential expression between the time-points on polysome-associated mRNA and protein abundance level, a linear model based on empirical Bayes methods was used (Smyth, 2004) and applied by log2-fold change > 1 and p-value < 0.05 and adjusted by the false discovery rate calculated using the Benjamini and Hochberg (1995). The statistical analyses in this study were performed using R version 3.6.1[2] in RStudio.[3]

Subcellular Localization Classification and GO Enrichment Analysis

Subcellular localization data from TAIR10 were used to classify protein for their localization. The ratio of the protein from each subcellular location was calculated based on the number of proteins identified from each location in the proteomic dataset and the total number of proteins identified. ClusterProfiler (Yu et al., 2012) was used for GO enrichment analysis with

[1]http://www.ncbi.nlm.nih.gov/geo/
[2]https://www.r-project.org/
[3]https://rstudio.com/

total gene identified from proteomic dataset (1,438 genes) as background. GO terms were considered significant if the *p*-value < 0.05 adjusted for the false discovery rate calculated using the Benjamini and Hochberg (1995) method.

Mass Spectrometry and Proteomic Analysis

Raw lysate from the same sample extract as performed for total mRNA and polysomal mRNA analysis were used for total protein extraction (Bai et al., 2017). The detailed protein extraction, chromatography and mass spectrometry, and also the proteomic analysis are described in the **Supplementary Information**.

mRNA and Protein Sequence Analysis

mRNA and protein decay rata data were retrieved from Narsai et al. (2007) and Li et al. (2017). The distributions of sequence lengths of each compared group were evaluated separately for CDS, 5' untranslated region (5'UTR), 3'UTR, and full transcript as compared to the total identified protein coding genes from proteomic analysis as background. Given the non-normality of the distributions of values, a Wilcoxon signed-rank test was adopted for all statistical comparisons (median as the test statistic).

Motif Analysis

DNA motif analyses were performed using the MEME suite (Bailey et al., 2009), for full transcript, 5'UTR, CDS, and 3'UTR sequences, extracted from the TAIR10 database.[4] The minimum and maximum motif widths were set to five and 12, respectively. If a gene had multiple transcripts, only the TAIR10 representative splice form (first splicing form in the database) was used. Background dinucleotide frequencies were provided separately for each sequence type (including 5'UTR, CDS, cDNA, and 3'UTR). FIMO (Bailey et al., 2009) was used to test motif specificity, motifs with a FIMO adjusted *p*-value lower than 0.001 were considered as significant hits. Obtained motif counts from FIMO (Version 5.3.0) were used to compute the enrichment *p*-value for the identified genes using all mRNAs expressed in the experiment as background by means of a one-tailed Fisher's exact test, performed with a custom script and the R software package[5] where *p*-value < 0.05 was considered to be significant. FIMO is used for motif scanning to identify the gene transcripts at the region where the motif is identified. The identified motifs and RNA sequence from the regulated gene sets (including 5'UTR, CDS, cDNA, and 3'UTR) were used as input for the sequence scan. Value of $p < 0.0001$was used as cutoff to filter the matched sequences.

RESULTS AND DISCUSSION

Protein Expression Data Are Enriched for Highly Expressed Genes

Protein levels at five distinct physiological stages as Bai et al. (2017) during seed germination (dry, 6, 26, 48, and 72 hours after imbibition; HAI) were measured by Liquid Chromatography-Mass Spectrometry (LC-MS). Proteins were precipitated using chloroform/methanol instead of the canonical acetone precipitation. This resulted in a high protein recovery and an optimal solubility which is requested for the LC-MS analysis. An additional acetone precipitation gained only 5.8% of total protein; based on this result, it was decided to stick to the chloroform/methanol precipitation for further analyses (**Supplementary Table 2**, **Supplementary Information**). In total, 1,469 proteins were identified. To be marked as identified proteins had to be present in all three biological replicates in at least one of the five physiological stages (**Supplementary Table 1**). The number of identified proteins is much lower than the number of transcripts identified in transcriptome and polysome analyses that were performed on exact the same samples, which resulted in 19,781 expressed genes (Bai et al., 2017). Comparative analyses revealed that the proteins identified represent genes that are significantly higher expressed at both the mRNA and polysomal mRNA level compared to the total set of genes expressed in germinating seeds (Wilcoxon Rank Sum test, *p*-value < 0.0001; **Figure 1A**; **Supplementary Figure S1**). There was a bias in the predicted subcellular locations for the detected proteins. Genes localized in the chloroplast and cytoplasm were over-represented, while nucleus-localized genes were under-represented (**Figure 1B**). This bias is likely the result of the applied protein isolation method and the detection limit of the LC-MS. Proteins localized in the nucleus, such as transcription factors and signaling proteins, might therefore be excluded from these analyses. Abundant proteins dominate the LC-MS identification, limiting the detection of low abundant proteins. Moreover, in seeds, the most abundant proteins are located in the cytosol (Mamontova et al., 2019). Also, the protein isolation buffer may be more efficient for purifying hydrophilic proteins. Despite the possible bias in the proteome subset (1,438 proteins), these data were used to compare protein abundance, polysomal, and total mRNA levels. Comparing the different levels of gene regulation is currently largely unexplored in plants; however, this dataset on seed germination is unique in that it allows such a systematic comparison.

Polysomal mRNA and Protein Abundance Display a Delay in Protein-Level Changes

To investigate the correlation between transcriptome, translatome, and proteome expression, spearman correlation analyses were performed. A strong correlation was found between the total and polysomal mRNA in the proteome subset ($r = 0.96$); this correlation was significantly higher than that in the whole dataset (27,826 genes on the microarray; $r = 0.94$, Wilcoxon Rank Sum test, *p*-value < 0.001; **Figure 2A**). This is consistent with the high correlation between the transcriptome and translatome observed in mice tissue (Reid et al., 2017). In contrast, the correlation between mRNA and protein, and polysomal mRNA and protein is both relatively low ($r = 0.36$ and $r = 0.35$, respectively; **Figure 2A**). This low level of correlation may have different explanations: (1) It might be bias caused by the fact that only the high expressed

[4]http://www.arabidopsis.org/
[5]http://www.r-project.org/

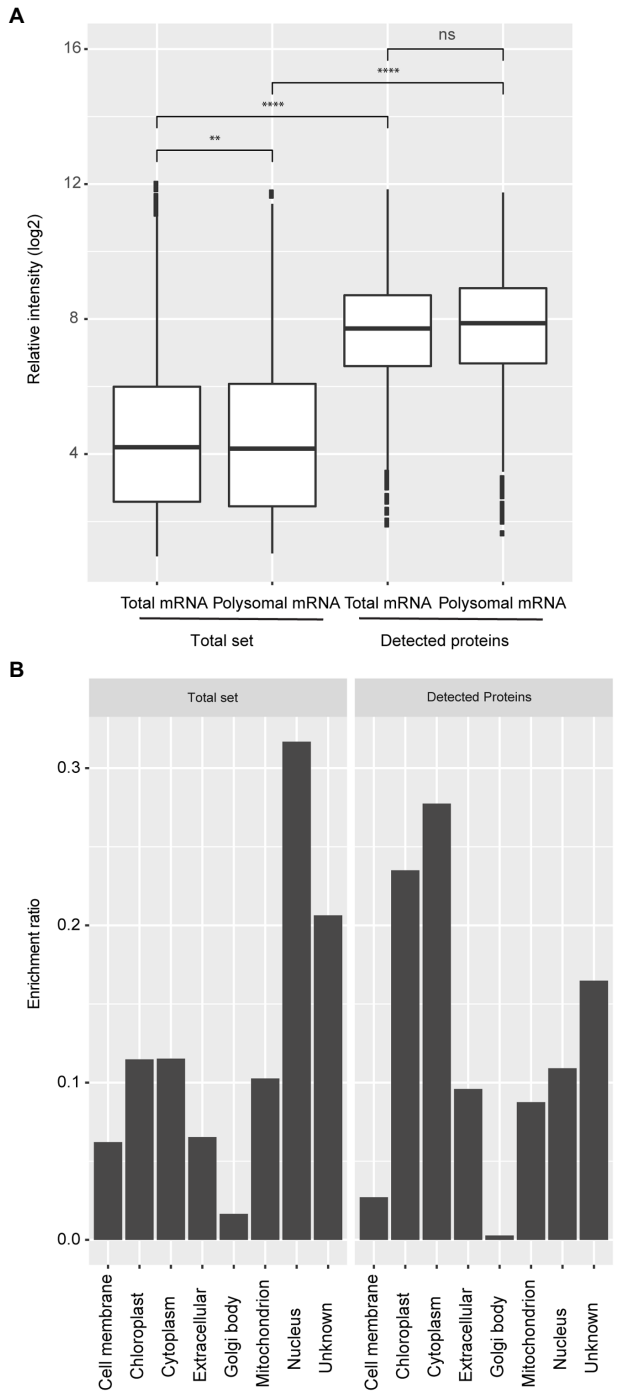


FIGURE 1 | Expression bias of detected protein coding genes and their predicted subcellular localization. **(A)** Boxplots representing the expression level of total and polysomal mRNA identified in the total set of genes that are expressed in seeds (Total set) and total and polysomal mRNA of the subset for which proteins have been identified (Detected proteins). The boxplots are based on the data from five time-points with three biological replicates (ns; $p > 0.05$; $^{**}p < 0.01$; $^{***}p < 0.001$; $^{****}p < 0.0001$; Wilcoxon rank-sum test). **(B)** Classification of subcellular locations of the genes for the whole set (Total set) and subset (Detected proteins). The ratio was calculated by dividing the number of genes/proteins from each subcellular location and the total number of genes/proteins identified from each dataset.

subset of the total number of genes that are expressed in seeds was taken into account, (2) there can be a delay in translation for translationally regulated proteins before translated proteins accumulate to a level that will be detected by proteomics, (3) it is also possible that translation in seeds occurs by single ribosomes (monosome) as well and not only by polysomes as is often thought (Bai et al., 2020), or (4) not all the polysome-associated mRNA are eventually translated into proteins due to, e.g., ribosome stalling (Rachter and Coller, 2015). Monosomes have been excluded in our analyses. For specific tissues, it has been shown that translation occurs by monosomes, *e.g.*, in the tiny synaptic compartment of rodent. mRNA is translated by monosomes rather than by polysomes, likely the result of the limited space in these neuron cells (Biever et al., 2020). Also, ribosomes can bind to mRNAs without translating, as was shown by the repression of translation initiation of the ribosome-associated *HAC1* mRNA due to lack of splicing in *Saccharomyces cerevisiae* (Kuhn et al., 2001). The correlation between the polysome occupancy (as proxy for translational efficiency) and protein abundance is also low (**Figure 2A**), revealing an uncoupling between polysome association and protein abundance during seed germination. This contrasts with the study by Beyer et al. (2004). These authors showed a slightly higher correlation between translational efficiency (represented by ribosome density) and total protein abundance than between total mRNA and total protein abundance. This discrepancy is likely caused by the translational status of the two different systems. In yeast, translation is sustained in a constantly active state, while in seed, translation is activated during germination after being completely inactive in the dry stage (Bai et al., 2017).

Possible delays in protein-level changes were investigated by comparing the total mRNA, the polysomal mRNA, and the total protein abundance at each germination time-point to its preceding time-point (**Figure 2B**). Differential expression analysis at total mRNA and polysomal mRNA level identified substantially more changes during the early time-points (dry, 6 and 26 HAI) compared to the later time-points (**Figure 2B**; **Supplementary Table 3**). This pattern was reversed at the protein level, where only a low number of differentially abundant proteins were identified in the early time-points and later time-points showed larger changes. Two tests were performed to investigate whether the low correlation between the polysome and the proteome could be the result of a delay of changes in protein level. First, the total mRNA and polysome mRNA at a time-point were compared to the protein abundance at the next time-point (i.e., per mRNA dry seeds vs. protein 6 HAI). This resulted in a significantly enhanced correlation (p-value < 0.001, from average 0.35–0.41; **Figure 3A**). Then, the time-point at which the gene expression changed for the first time at the polysomal level (both up and down) was compared to its protein abundance at all time-points. Similar significant correlation enhancement was identified (p-value < 0.001, from average 0.33–0.40). Both tests revealed that the up- and downregulation of (polysomal) mRNAs preceded the change at the protein abundance level, indicating a delay in either protein accumulation/synthesis or degradation. Since we do not know the reason for the delay, we refer to "delay in protein level changes" this article. The correlations between

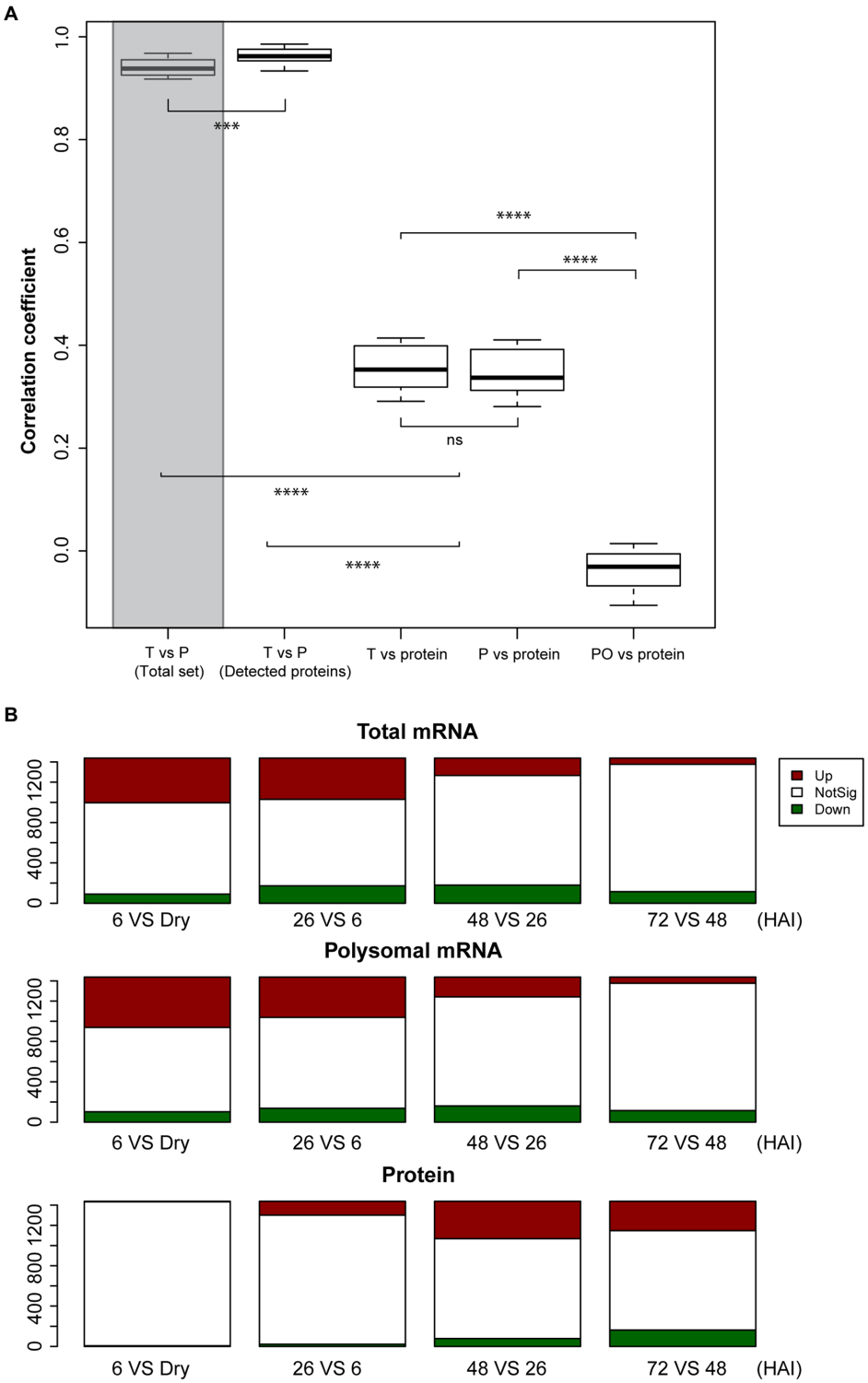


FIGURE 2 | Lack of correlation between total mRNA, polysomal mRNA, and protein abundance during seed germination. **(A)** Distribution of Spearman correlation coefficients (five time-points and three biological replicates) between total mRNA (T) and polysomal mRNA (P) for the total set of identified mRNAs (Total set; shaded bar) and the mRNAs with identified protein (Detected proteins), total mRNA and protein, polysomal mRNA, and protein as well as the polysomal occupancy (PO) and protein (ns; $p > 0.05$; $^{**}p < 0.01$; $^{***}p < 0.001$; $^{****}p < 0.0001$; Wilcoxon rank-sum test). **(B)** Differential expressed genes in total mRNA, polysomal mRNA, and protein during seed germination following hours after imbibition (HAI). The genes upregulated (Up, red) and downregulated (Down, green) are displayed in comparison with the non-significant (NotSig, white) changed genes/proteins.

changes at the total mRNA and protein level and the polysomal mRNA and the protein levels were similar (**Figure 3A**) and indicate that the delay occurs from the polysome association to the protein synthesis.

Unique Functions of the Delayed Response Genes

To further characterize the delayed response genes, two gene sets were created: one containing all Delayed Response Upregulated Genes (DUGs) and the other containing the Delayed Response Downregulated Genes (DDGs), 425 and 93 genes in each set, respectively (**Figure 3C**; **Supplementary Table 4**). Two public data sets were used to validate the delayed protein changes in the DUG set; these include a high-density transcriptome time-course and neosynthesized protein during seed germination (Dekkers et al., 2013; Galland et al., 2014). Six genes that were detected as *de novo* synthesized proteins were also identified in the current DUG set, and thus, their RNA profile and neosynthesized protein synthesis dynamics could be compared. The peak of protein synthesis for these six delayed genes lagged behind the peak of their corresponding mRNA signal during germination (**Supplementary Figure S2**). This development-dependent delay is far beyond the sequential time interval during mRNA transcription and protein translation, which normally occurs within seconds or minutes (Boehlke and Friesen, 1975; Ingolia et al., 2011). Therefore, the delay between changes at the polysomal mRNA level and actual changes in protein abundance suggests a temporal regulation between polysome association and protein synthesis/degradation. Such a phenomenon is not new to seeds. Previously, it has been reported that there is a delay between the transcription and translation of some Late Embryogenesis Abundant proteins (Galau et al., 1987, 1991; Espelund et al., 1992; Chatelain et al., 2012; Verdier et al., 2013). It has been suggested that this posttranscriptional regulation during seed maturation is linked to the plasticity of seed tissues to respond to fluctuations in the environment and as such contribute to a better seed storability (Verdier et al., 2013). Such a delay of translation also occurs during seed germination. The GO enrichment analysis showed that the DUG set was enriched for translation, ribosome biogenesis, photosynthesis, and plastid organization (FDR < 0.05; **Supplementary Table 5**), genes that are essential for protein synthesis to support seed germination and photomorphogenesis during seedling establishment. The DDG set was enriched for the GO terms "response to abscisic acid" and "oxidation-reduction process" (FDR < 0.05; **Supplementary Table 5**) and includes seed storage proteins, NAD(P)-binding proteins, glutathione transferases, and peroxygenases. The delayed reduction of these stress-related proteins would allow seeds to respond to changes in the environment, i.e., oxidation and maintain their plasticity when adverse conditions are met before germination (Nguyen et al., 2015; Sano et al., 2016). During seed germination, the most evident delay in protein translation occurred from 0 (dry) to 6 and from 6 to 26 HAI (**Figure 3C**), corresponding to seed germination Phases I and II (Bewley, 1997). From dry to six HAI, the translation of genes related to translation and ribosome biogenesis were delayed. This is consistent with the polysome

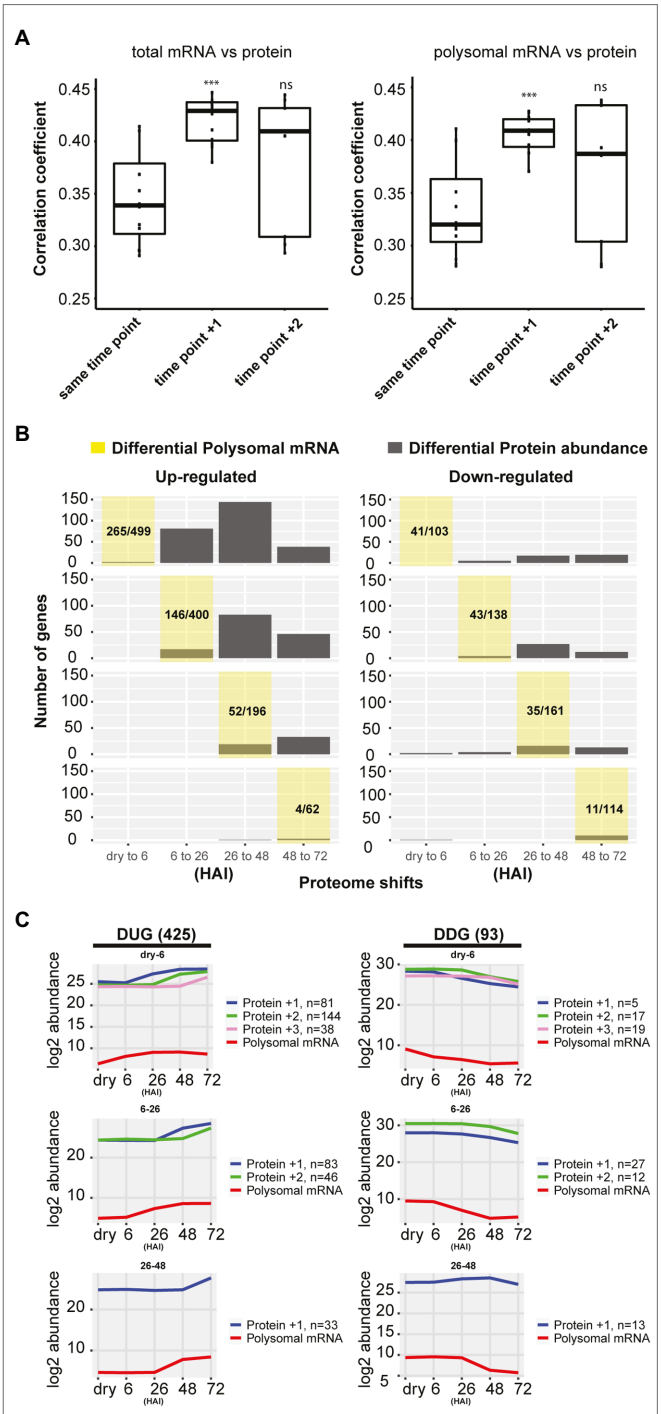


FIGURE 3 | Delay in protein synthesis during seed germination. **(A)** Correlation analysis between the total mRNA, polysomal mRNA level, and the protein abundance of the next time-points. Time-points +1 and +2 for the polysomal mRNA are used for the correlation with the protein abundance (***$p < 0.001$; Wilcoxon rank-sum test). **(B)** Delay in protein abundance changes compared to changes in polysome-associated mRNAs comparing two consecutive hours after imbibition (HAI) time-points. The yellow shaded area represents the germination time-point under investigation and indicates when the polysomal mRNAs are for the first time up-/downregulated. The corresponding distribution of protein changes of these genes changed in polysomal mRNA at the yellow shaded time-point is indicated in grey. The proportions of up- or downregulated

(continued)

FIGURE 3 | polysomal mRNA for the detected proteins (sum of grey bars) compared to the total number of differentially regulated polysomal mRNA at each time-point are shown. **(C)** The average abundance profiles of the polysome-associated mRNA and proteins of the delayed response genes at different HAI during seed germination. At the left, the Delayed Response Upregulated Genes (DUGs) are indicated, and at the right, the Delayed Response Downregulated Genes (DDGs). The polysome mRNA is plotted at each consecutive time-point, and the corresponding protein is plotted according to the first time they are up-/downregulated indicated by +1/+2/+3 shift for the detected time-point. The gene/protein changes in each category are shown.

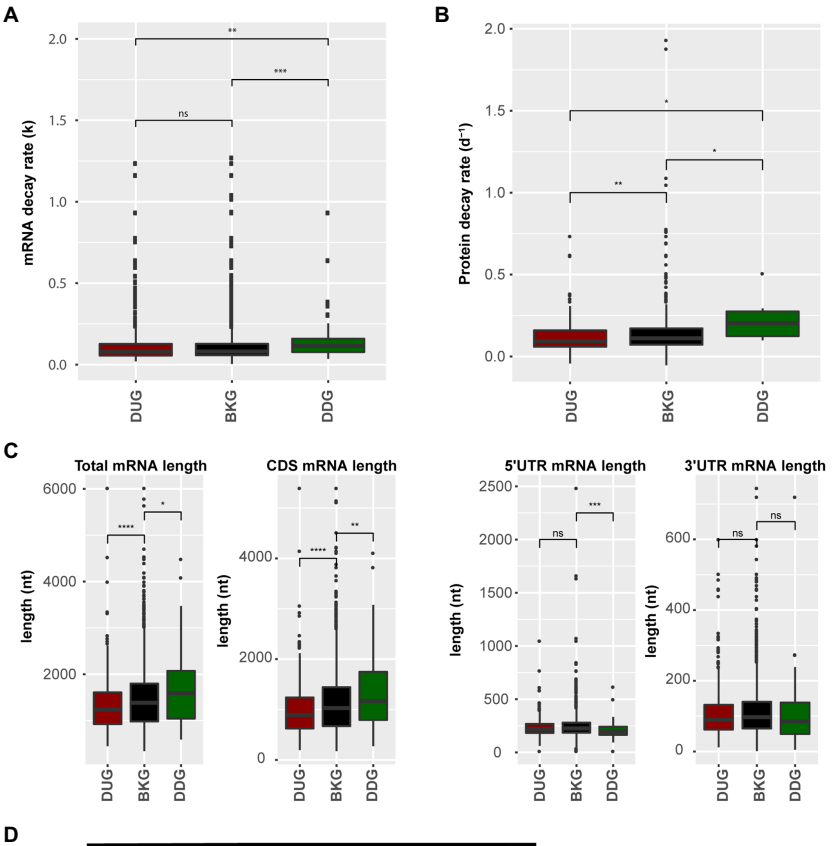
profile during early seed germination, which indicates that ribosome biogenesis only starts after six HAI (Bai et al., 2017). Moreover, among the delayed genes from 6 to 48 HAI were those related to photosynthesis (**Supplementary Table 5**), indicating that although the mRNA is associated with polysomes, protein synthesis does not occur until the embryo needs to become autotrophic during seedling establishment.

Delaying translation might be an evolutionary advantage. The translation is energy consuming; therefore, the delay of the protein translation could increase the chance of survival in case the conditions do not allow the plant to fulfil its life cycle. Germination is an all or nothing event, and the decision to germination can only be made once. The delay of translation might refer to a checkpoint that needs to be passed for germination to occur. Until they have germinated, seeds are resilient to environmental stress and can even regain their desiccation tolerance and rehydrate and germinate when conditions allow this (Maia et al., 2011).

mRNA Sequence Features and Protein Stability Wire the Delay in Protein-Level Changes

Genes with a delayed response in protein abundance are likely under translational control. The control of mRNA translation involves the interaction between mRNA, ribosome, and associated translation factors, such as translation initiation, elongation, and termination factors. This interaction is substantially influenced by mRNA sequence features, mRNA stability, and the folding rate of the synthesized peptides (Chatelain et al., 2012; Radhakrishnan and Green, 2016; Farias-Rico et al., 2018; Waudby et al., 2019). By quantifying the mRNA decay rate and protein decay rate *in silico* for both the DUG and DDG dataset compared to the background control (Narsai et al., 2007; Li et al., 2017), it was identified that mRNA and protein stability correlated with the delay in protein changes. The mRNA stability, based on both the mRNA decay rate and mRNA half-life times, of the DDG set was significantly lower when compared to the background set (**Figure 4A**; **Supplementary Figure S3**). Similarly, the protein decay rate of the DDG set was significantly higher when considering the total set of delay genes (**Figure 4B**). A closer look at the time-points revealed that mRNAs high in decay rate were mainly identified from 26 to 48 HAI time-point when germination occurs (**Supplementary Figures S4A–C**; **Supplementary Tables 6** and **7**). In contrast, the DUG set mainly contained genes with the lower protein decay rate (0–26 HAI; **Supplementary Figure S4C**). Interestingly, the DUG set contains relatively short sequences, especially the set of genes that are expressed during early germination (0–6 HAI; **Supplementary Figure S4D**). The DDG set contains mainly longer genes at the same time-point (0–6 HAI; Figure 4C, **Supplementary Figure S4C**). These analyses suggest a role for length of the coding region (CDS), mRNA, and protein stability on the delay of the changes in protein abundance.

RNA-binding proteins are known to regulate translation by binding to specific motifs at different positions in the mRNA. To investigate a possible role for RBPs in the delay of protein changes, motif enrichment analyses were performed on the 5'UTR, 3'UTR and CDS of mRNAs in both DUG and DDG. 11 motifs were found in different regions (p-value <0.05), including one motif in the 5' UTR, five motifs in the 3'UTR, three motifs in the CDS of DUG set, one motif in the 5' UTR, and one motif in the CDS of DDG set (**Figure 4D**). Motif UCUUCUUC, identified in both the 5'"UTR and CDS of the DUG set, is a known binding site of the Glycine-rich RNA-binding protein 7 (GRB7, AT2G21660; Zhang and Mount, 2009; Meyer et al., 2017). Motif GAAGAAGA, identified from the 3'UTR of the same set, was previously identified as intronic splicing regulatory element recognized by Serine/arginine-rich splicing factor 45 (SR45, AT1G16610) and Serine/arginine-rich SC35-like splicing factor 33 (SCL33, AT1G55310), two exonic splicing enhancer with RNA binding capacity (Pertea et al., 2007; Thomas et al., 2012; Xing et al., 2015). These three RBPs are identified in the ribosome-associated mRNA fractions in seeds (Bai et al., 2020), indicating their potential role for translational regulation. The disruption of GRB7 and SR45 results in defects during seed germination and seedling growth in stress conditions, including abscisic acid hypersensitivity and retarded germination and seedling growth as a result of salt and dehydration (Cao et al., 2006; Carvalho et al., 2010). Therefore, GRB7 and SR45 are likely involved in determining the plasticity of seed germination. The regulation of mRNA translation by these RBPs likely contributes to the delay in the protein synthesis. Such a delay could reduce the rate of translation initiation and elongation, which is important for protein co-translational folding thereby ensuring the correct functionality of the encoded proteins (Zhao et al., 2019). Potential RBPs targets were identified by scanning for the sequence motif. In total, 18 mRNAs were identified to contain the GAAGAAGA motif at their 3'UTR. 54 and 66 mRNAs were identified to contain the UCUUCUUC motif in their 5'UTR and CDS, respectively (**Supplementary Table 8**), among which 20 and 24 were identified as GRP7 binding targets previously (Cao et al., 2006). The GAAGAAGA motif containing mRNAs was enriched for GO category of response to cold, while the UCUUCUUC motif containing mRNAs was enriched for GO categories, such as salt response, translation, and ribosome biogenesis (**Supplementary Table 9**). Therefore, RBPs might mediate the timing of protein translation during seed germination. Moreover, translation that is on hold or delayed might result in the formation of the so-called stress granules (SGs) or processing bodies. Earlier, we have identified SG



D

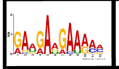
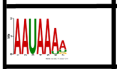
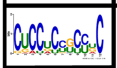
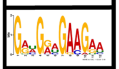
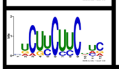
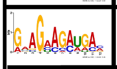
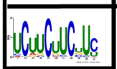
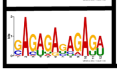
Motif logo	Discovery region	P-value	FoldEnrich
	3'UTR delay up	1.02e-09	1.35
	3'UTR delay down	6.87e-08	1.69
	CDS delay up	4.35e-39	1.46
		3.54e-32	1.34
		9.72e-31	1.38
	CDS delay down	6.98e-04	1.16
	5'UTR delay up	1.35e-12	1.50
		1.26e-13	1.51
		3.19e-07	1.55
		2.93e-11	1.45
		3.27e-04	1.30

FIGURE 4 | mRNA and protein sequence features for the protein delayed genes. **(A)** mRNA decay rate of the proteins that are delayed in abundance. **(B)** Protein decay rate for the delayed response genes. **(C)** mRNA sequence length including total mRNA length, length of coding sequence, 5'UTR, and 3'UTR for the delayed response genes (ns; $p > 0.05$; $^{*}p < 0.05$; $^{**}p < 0.01$; $^{***}p < 0.001$; $^{****}p < 0.0001$; Wilcoxon rank-sum test). **(D)** Sequence motifs identified in the delayed response genes. Motif logo, the discovered sequence region (5'UTR, 3'UTR, and CDS), the *p*-value (Fisher test), and the fold enrichment compared to the background set (total identified protein coding genes) are shown for motifs identified in the DUG and the DDG sets.

and processing bodies proteins in association with the ribosome in seeds, which suggests that during seed germination these cellular foci are also formed (Bai et al., 2020).

CONCLUSION

In this study, we have compared protein-level changes during seed germination with transcriptome and polysome changes. This unique data set allows correlation analyses between the three levels of regulation (transcriptome, translatome, and proteome) over a series of time-points. We identified a high correlation between total mRNA and polysomal mRNA levels, but only a low correlation between polysomal mRNA and protein abundance during seed germination. The low correlation is likely caused by a delay in protein translation and degradation. Multi-level correlation studies have not been performed before due to the lack of consistent sampling schemes on multiple "omic" levels. Whether the identified delay on protein level compared with polysomal mRNA level is unique for germinating seeds remains elusive since similar types of analyses have not yet been reported for other systems.

Seeds are key for plant survival, and thus, it is not surprising that the germination of seeds is strongly regulated. Seed germination can be halted at several moments and even after the embryonic root (radicle) has protruded its surrounding layers, seeds can survive desiccation (Maia et al., 2011). Earlier, we suggested a role for RBPs in the control of translation in seeds (Sajeev et al., 2019); the current study revealed some possible candidates based on the identified sequence motifs and germination phenotypes of RBPs that may recognize these motifs. In depth, studies identifying the mRNA, RBP, and ribosome interactome will allow a further investigation of the regulatory mechanisms underlying the translational control of seed germination.

DATA AVAILABILITY STATEMENT

The original contributions presented in the study are publicly available. These data can be found at ProteomeXchange with following accession: PXD027345.

AUTHOR CONTRIBUTIONS

BB and LB designed the experiment. BB, AA, and JC performed the proteomic analysis. BB, NH, and HN analyzed microarray and proteomic data. BB, NH, HN, and LB wrote the paper. All authors contributed to the article and approved the submitted version.

FUNDING

This work was supported by The Netherlands Organization for Scientific Research and Technology Hotel Program (Grant number: 435002002) from The Netherlands Organization for Health Research and Development.

ACKNOWLEDGMENTS

We thank Wenying Huo for technical assistance in data validation.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpls.2021.735719/full#supplementary-material

REFERENCES

Bai, B., Novak, O., Ljung, K., Hanson, J., and Bentsink, L. (2018). Combined transcriptome and translatome analyses reveal a role for tryptophan-dependent auxin biosynthesis in the control of DOG1-dependent seed dormancy. *New Phytol.* 217, 1077–1085. doi: 10.1111/nph.14885

Bai, B., Peviani, A., van der Horst, S., Gamm, M., Snel, B., Bentsink, L., et al. (2017). Extensive translational regulation during seed germination revealed by polysomal profiling. *New Phytol.* 214, 233–244. doi: 10.1111/nph.14355

Bai, B., van der Horst, S., Cordewener, J. H. G., America, T., Hanson, J., and Bentsink, L. (2020). Seed-stored mRNAs that are specifically associated to Monosomes are translationally regulated during germination. *Plant Physiol.* 182, 378–392. doi: 10.1104/pp.19.00644

Bailey, T. L., Boden, M., Buske, F. A., Frith, M., Grant, C. E., Clementi, L., et al. (2009). MEME SUITE: tools for motif discovery and searching. *Nucleic Acids Res.* 37, W202–W208. doi: 10.1093/nar/gkp335

Becker, K., Bluhm, A., Casas-Vila, N., Dinges, N., Dejung, M., Sayols, S., et al. (2018). Quantifying post-transcriptional regulation in the development of Drosophila melanogaster. *Nat. Commun.* 9:4970. doi: 10.1038/s41467-018-07455-9

Beltran-Pena, E., Ortiz-Lopez, A., and Sanchez de Jimenez, E. (1995). Synthesis of ribosomal proteins from stored mRNAs early in seed germination. *Plant Mol. Biol.* 28, 327–336. doi: 10.1007/BF00020251

Benjamini, Y., and Hochberg, Y. (1995). Controlling the false discovery rate - a practical and powerful approach to multiple testing. *J Roy Stat Soc B Met* 57, 289–300. doi: 10.1111/j.2517-6161.1995.tb02031.x

Berardini, T. Z., Reiser, L., Li, D., Mezheritsky, Y., Muller, R., Strait, E., et al. (2015). The Arabidopsis information resource: making and mining the "gold standard" annotated reference plant genome. *Genesis* 53, 474–485. doi: 10.1002/dvg.22877

Bewley, J. D. (1997). Seed germination and dormancy. *Plant Cell* 9, 1055–1066. doi: 10.1105/tpc.9.7.1055

Beyer, A., Hollunder, J., Nasheuer, H. P., and Wilhelm, T. (2004). Post-transcriptional expression regulation in the yeast Saccharomyces cerevisiae on a genomic scale. *Mol. Cell. Proteomics* 3, 1083–1092. doi: 10.1074/mcp.M400099-MCP200

Biever, A., Glock, C., Tushev, G., Ciirdaeva, E., Dalmay, T., Langer, J. D., et al. (2020). Monosomes actively translate synaptic mRNAs in neuronal processes. *Science* 367:eaay4991. doi: 10.1126/science.aay4991

Bradford, K. J. (1990). A water relations analysis of seed germination rates. *Plant Physiol.* 94, 840–849. doi: 10.1104/pp.94.2.840

Branco-Price, C., Kawaguchi, R., Ferreira, R. B., and Bailey-Serres, J. (2005). Genome-wide analysis of transcript abundance and translation in Arabidopsis seedlings subjected to oxygen deprivation. *Ann. Bot.* 96, 647–660. doi: 10.1093/aob/mci217

Boehlke, K. W., and Friesen, J. D. (1975). Cellular content of ribonucleic acid and protein in Saccharomyces cerevisiae as a function of exponential growth rate: calculation of the apparent peptide chain elongation rate. *J. Bacteriol.* 121, 429–433. doi: 10.1128/jb.121.2.429-433.1975

Cao, S., Jiang, L., Song, S., Jing, R., and Xu, G. (2006). AtGRP7 is involved in the regulation of abscisic acid and stress responses in Arabidopsis. *Cell. Mol. Biol. Lett.* 11, 526–535. doi: 10.2478/s11658-006-0042-2

Carvalho, R. F., Carvalho, S. D., and Duque, P. (2010). The plant-specific SR45 protein negatively regulates glucose and ABA signaling during early seedling development in Arabidopsis. *Plant Physiol.* 154, 772–783. doi: 10.1104/pp.110.155523

Chatelain, E., Hundertmark, M., Leprince, O., Le Gall, S., Satour, P., Deligny-Penninck, S., et al. (2012). Temporal profiling of the heat-stable proteome during late maturation of Medicago truncatula seeds identifies a restricted subset of late embryogenesis abundant proteins associated with longevity. *Plant Cell Environ.* 35, 1440–1455. doi: 10.1111/j.1365-3040.2012.02501.x

Chen, G., Gharib, T. G., Huang, C. C., Taylor, J. M., Misek, D. E., Kardia, S. L., et al. (2002). Discordant protein and mRNA expression in lung adenocarcinomas. *Mol. Cell. Proteomics* 1, 304–313. doi: 10.1074/mcp.M200008-MCP200

Dekkers, B. J. W., Pearce, S., van Bolderen-Veldkamp, R. P., Marshall, A., Widera, P., Gilbert, J., et al. (2013). Transcriptional dynamics of two seed compartments with opposing roles in Arabidopsis seed germination. *Plant Physiol.* 163, 205–215. doi: 10.1104/pp.113.223511

Espelund, M., Saeboe-Larssen, S., Hughes, D. W., Galau, G. A., Larsen, F., and Jakobsen, K. S. (1992). Late embryogenesis-abundant genes encoding proteins with different numbers of hydrophilic repeats are regulated differentially by abscisic acid and osmotic stress. *Plant J.* 2, 241–252. doi: 10.1046/j.1365-313X.1992.t01-46-00999.x

Farias-Rico, J. A., Ruud Selin, F., Myronidi, I., Fruhauf, M., and von Heijne, G. (2018). Effects of protein size, thermodynamic stability, and net charge on cotranslational folding on the ribosome. *Proc. Natl. Acad. Sci. U. S. A.* 115, E9280–E9287. doi: 10.1073/pnas.1812756115

Galau, G. A., Bijaisoradat, N., and Hughes, D. W. (1987). Accumulation kinetics of cotton late embryogenesis-abundant mRNAs and storage protein mRNAs: coordinate regulation during embryogenesis and the role of abscisic acid. *Dev. Biol.* 123, 198–212. doi: 10.1016/0012-1606(87)90442-8

Galau, G. A., Jakobsen, K. S., and Hughes, D. W. (1991). The controls of late dicot embryogenesis and early germination. *Physiol. Plant.* 81, 280–288. doi: 10.1111/j.1399-3054.1991.tb02142.x

Galland, M., Huguet, R., Arc, E., Cueff, G., Job, D., and Rajjou, L. (2014). Dynamic proteomics emphasizes the importance of selective mRNA translation and protein turnover during Arabidopsis seed germination. *Mol. Cell. Proteomics* 13, 252–268. doi: 10.1074/mcp.M113.032227

Gamm, M., Peviani, A., Honsel, A., Snel, B., Smeekens, S., and Hanson, J. (2014). Increased sucrose levels mediate selective mRNA translation in Arabidopsis. *BMC Plant Biol.* 14:306. doi: 10.1186/s12870-014-0306-3

Grolleau, A., Bowman, J., Pradet-Balade, B., Puravs, E., Hanash, S., Garcia-Sanz, J. A., et al. (2002). Global and specific translational control by rapamycin in T cells uncovered by microarrays and proteomics. *J. Biol. Chem.* 277, 22175–22184. doi: 10.1074/jbc.M202014200

Haider, S., and Pal, R. (2013). Integrated analysis of transcriptomic and proteomic data. *Curr. Genomics* 14, 91–110. doi: 10.2174/1389202911314020003

He, H. Z., Vidigal, D. D., Snoek, L. B., Schnabel, S., Nijveen, H., Hilhorst, H., et al. (2014). Interaction between parental environment and genotype affects plant and seed performance in Arabidopsis. *J. Exp. Bot.* 65, 6603–6615. doi: 10.1093/jxb/eru378

Ingolia, N. T., Lareau, L. F., and Weissman, J. S. (2011). Ribosome profiling of mouse embryonic stem cells reveals the complexity and dynamics of mammalian proteomes. *Cell* 147, 789–802. doi: 10.1016/j.cell.2011.10.002

Kronja, I., Yuan, B., Eichhorn, S. W., Dzeyk, K., Krijgsveld, J., Bartel, D. P., et al. (2014). Widespread changes in the posttranscriptional landscape at the drosophila oocyte-to-embryo transition. *Cell Rep.* 7, 1495–1508. doi: 10.1016/j.celrep.2014.05.002

Kuhn, K. M., DeRisi, J. L., Brown, P. O., and Sarnow, P. (2001). Global and specific translational regulation in the genomic response of Saccharomyces cerevisiae to a rapid transfer from a fermentable to a nonfermentable carbon source. *Mol. Cell. Biol.* 21, 916–927. doi: 10.1128/MCB.21.3.916-927.2001

Li, L., Nelson, C. J., Trosch, J., Castleden, I., Huang, S., and Millar, A. H. (2017). Protein degradation rate in Arabidopsis thaliana leaf growth and development. *Plant Cell* 29, 207–228. doi: 10.1105/tpc.16.00768

Lu, P., Vogel, C., Wang, R., Yao, X., and Marcotte, E. M. (2007). Absolute protein expression profiling estimates the relative contributions of transcriptional and translational regulation. *Nat. Biotechnol.* 25, 117–124. doi: 10.1038/nbt1270

Maia, J., Dekkers, B. J., Provart, N. J., Ligterink, W., and Hilhorst, H. W. (2011). The re-establishment of desiccation tolerance in germinated Arabidopsis thaliana seeds and its associated transcriptome. *PLoS One* 6:e29123. doi: 10.1371/journal.pone.0029123

Mamontova, T., Afonin, A. M., Ihling, C., Soboleva, A., Lukasheva, E., Sulima, A. S., et al. (2019). Profiling of seed proteome in pea (Pisum sativum L.) lines characterized with high and low responsivity to combined inoculation with nodule bacteria and Arbuscular Mycorrhizal fungi. *Molecules* 24:1603. doi: 10.3390/molecules24081603

Meyer, K., Koster, T., Nolte, C., Weinholdt, C., Lewinski, M., Grosse, I., et al. (2017). Adaptation of iCLIP to plants determines the binding landscape of the clock-regulated RNA-binding protein AtGRP7. *Genome Biol.* 18:204. doi: 10.1186/s13059-017-1332-x

Mustroph, A., Juntawong, P., and Bailey-Serres, J. (2009). Isolation of plant polysomal mRNA by differential centrifugation and ribosome immunopurification methods. *Methods Mol. Biol.* 553, 109–126. doi: 10.1007/978-1-60327-563-7_6

Narsai, R., Howell, K. A., Millar, A. H., O'Toole, N., Small, I., and Whelan, J. (2007). Genome-wide analysis of mRNA decay rates and their determinants in Arabidopsis thaliana. *Plant Cell* 19, 3418–3436. doi: 10.1105/tpc.107.055046

Nguyen, T. P., Cueff, G., Hegedus, D. D., Rajjou, L., and Bentsink, L. (2015). A role for seed storage proteins in Arabidopsis seed longevity. *J. Exp. Bot.* 66, 6399–6413. doi: 10.1093/jxb/erv348

Nie, L., Wu, G., Brockman, F. J., and Zhang, W. (2006). Integrated analysis of transcriptomic and proteomic data of Desulfovibrio vulgaris: zero-inflated Poisson regression models to predict abundance of undetected proteins. *Bioinformatics* 22, 1641–1647. doi: 10.1093/bioinformatics/btl134

Pertea, M., Mount, S. M., and Salzberg, S. L. (2007). A computational survey of candidate exonic splicing enhancer motifs in the model plant Arabidopsis thaliana. *BMC bioinformatics* 8:159. doi: 10.1186/1471-2105-8-159

Pringle, E. S., McCormick, C., and Cheng, Z. (2019). Polysome profiling analysis of mRNA and associated proteins engaged in translation. *Curr. Protoc. Mol. Biol.* 125:e79. doi: 10.1002/cpmb.79

Rachter, J. D., and Coller, J. (2015). Pausing on polyribosomes: make way for elongation in translational control. *Cell* 163, 292–300. doi: 10.1016/j.cell.2015.09.041

Radhakrishnan, A., and Green, R. (2016). Connections underlying translation and mRNA stability. *J. Mol. Biol.* 428, 3558–3564. doi: 10.1016/j.jmb.2016.05.025

Rajjou, L., Gallardo, K., Debeaujon, I., Vandekerckhove, J., Job, C., and Job, D. (2004). The effect of alpha-amanitin on the Arabidopsis seed proteome highlights the distinct roles of stored and neosynthesized mRNAs during germination. *Plant Physiol.* 134, 1598–1613. doi: 10.1104/pp.103.036293

Reid, D. W., Xu, D., Chen, P., Yang, H., and Sun, L. (2017). Integrative analyses of translatome and transcriptome reveal important translational controls in brown and white adipose regulated by microRNAs. *Sci. Rep.* 7:5681. doi: 10.1038/s41598-017-06077-3

Sajeev, N., Bai, B., and Bentsink, L. (2019). Seeds: a unique system to study translational regulation. *Trends Plant Sci.* 24, 487–495. doi: 10.1016/j.tplants.2019.03.011

Sano, N., Rajjou, L., North, H. M., Debeaujon, I., Marion-Poll, A., and Seo, M. (2016). Staying alive: molecular aspects of seed longevity. *Plant Cell Physiol.* 57, 660–674. doi: 10.1093/pcp/pcv186

Smircich, P., Eastman, G., Bispo, S., Duhagon, M. A., Guerra-Slompo, E. P., Garat, B., et al. (2015). Ribosome profiling reveals translation control as a key mechanism generating differential gene expression in Trypanosoma cruzi. *BMC Genomics* 16:443. doi: 10.1186/s12864-015-1563-8

Smyth, G. K. (2004). Linear models and empirical bayes methods for assessing differential expression in microarray experiments. *Stat. Appl. Genet. Mol. Biol.* 3:Article3. doi: 10.2202/1544-6115.1027

Suzuki, Y., and Minamikawa, T. (1985). On the role of stored mRNA in srotein synthesis in embryonic axes of germinating vigna unguiculata seeds. *Plant Physiol.* 79, 327–331. doi: 10.1104/pp.79.2.327

Thomas, J., Palusa, S. G., Prasad, K. V., Ali, G. S., Surabhi, G. K., Ben-Hur, A., et al. (2012). Identification of an intronic splicing regulatory element involved in auto-regulation of alternative splicing of SCL33 pre-mRNA. *Plant J.* 72, 935–946. doi: 10.1111/tpj.12004

Varshavsky, A. (1996). The N-end rule: functions, mysteries, uses. *Proc. Natl. Acad. Sci. U. S. A.* 93, 12142–12149. doi: 10.1073/pnas.93.22.12142

Verdier, J., Lalanne, D., Pelletier, S., Torres-Jerez, I., Righetti, K., Bandyopadhyay, K., et al. (2013). A regulatory network-based approach dissects late maturation processes related to the acquisition of desiccation tolerance and longevity of Medicago truncatula seeds. *Plant Physiol.* 163, 757–774. doi: 10.1104/pp.113.222380

Waudby, C. A., Dobson, C. M., and Christodoulou, J. (2019). Nature and regulation of protein folding on the ribosome. *Trends Biochem. Sci.* 44, 914–926. doi: 10.1016/j.tibs.2019.06.008

Xing, D., Wang, Y., Hamilton, M., Ben-Hur, A., and Reddy, A. S. (2015). Transcriptome-wide identification of RNA targets of Arabidopsis SERINE/ARGININE-RICH45 uncovers the unexpected roles of This RNA binding protein in RNA processing. *Plant Cell* 27, 3294–3308. doi: 10.1105/tpc.15.00641

Yu, G., Wang, L. G., Han, Y., and He, Q. Y. (2012). clusterProfiler: an R package for comparing biological themes among gene clusters. *OMICS* 16, 284–287. doi: 10.1089/omi.2011.0118

Zhang, X. N., and Mount, S. M. (2009). Two alternatively spliced isoforms of the Arabidopsis SR45 protein have distinct roles during normal plant development. *Plant Physiol.* 150, 1450–1458. doi: 10.1104/pp.109.138180

Zhao, J., Qin, B., Nikolay, R., Spahn, C. M. T., and Zhang, G. (2019). Translatomics: the global view of translation. *Int. J. Mol. Sci.* 20:212. doi: 10.3390/ijms20235988

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

frontiers in Plant Science

ORIGINAL RESEARCH
published: 25 October 2021
doi: 10.3389/fpls.2021.748487

Label-Free Quantitative Proteome Analysis Reveals the Underlying Mechanisms of Grain Nuclear Proteins Involved in Wheat Water-Deficit Response

Tingting Li[1†], Dong Zhu[1†], Zhisheng Han[1†], Junwei Zhang[1], Ming Zhang[2] and Yueming Yan[1*]*

[1] Beijing Key Laboratory of Plant Gene Resources and Biotechnology for Carbon Reduction and Environmental Improvement, College of Life Sciences, Capital Normal University, Beijing, China, [2] College of Agricultural and Biological Engineering (College of Tree Peony), Heze University, Heze, China

OPEN ACCESS

Edited by:
Pingfang Yang,
Hubei University, China

Reviewed by:
Haiying Li,
Heilongjiang University, China
Sun-Hee Woo,
Chungbuk National University, South Korea
Mehanathan Muthamilarasan,
University of Hyderabad, India

***Correspondence:**
Ming Zhang
zhangming5152007@163.com
Yueming Yan
yanym@cnu.edu.cn

† These authors have contributed equally to this work

Specialty section:
This article was submitted to Plant Proteomics and Protein Structural Biology, a section of the journal Frontiers in Plant Science

Received: *28 July 2021*
Accepted: *05 October 2021*
Published: *25 October 2021*

Citation:
Li T, Zhu D, Han Z, Zhang J, Zhang M and Yan Y (2021) Label-Free Quantitative Proteome Analysis Reveals the Underlying Mechanisms of Grain Nuclear Proteins Involved in Wheat Water-Deficit Response. Front. Plant Sci. 12:748487. doi: 10.3389/fpls.2021.748487

In this study, we performed the first nuclear proteome analysis of wheat developing grains under water deficit by using a label-free based quantitative proteomic approach. In total, we identified 625 unique proteins as differentially accumulated proteins (DAPs), of which 398 DAPs were predicted to be localized in nucleus. Under water deficit, 146 DAPs were up-regulated and mainly involved in the stress response and oxidation-reduction process, while 252 were down-regulated and mainly participated in translation, the cellular amino metabolic process, and the oxidation-reduction process. The *cis*-acting elements analysis of the key nuclear DAPs encoding genes demonstrated that most of these genes contained the same *cis*-acting elements in the promoter region, mainly including ABRE involved in abscisic acid response, antioxidant response element, MYB responsive to drought regulation and MYC responsive to early drought. The *cis*-acting elements related to environmental stress and hormones response were relatively abundant. The transcription expression profiling of the nuclear up-regulated DAPs encoding genes under different organs, developmental stages and abiotic stresses was further detected by RNA-seq and Real-time quantitative polymerase chain reaction, and more than 50% of these genes showed consistency between transcription and translation expression. Finally, we proposed a putative synergistic responsive network of wheat nuclear proteome to water deficit, revealing the underlying mechanisms of wheat grain nuclear proteome in response to water deficit.

Keywords: wheat, developing grains, water deficit, nuclear proteome, label-free quantitation

INTRODUCTION

As an allohexaploid species and the product of intensive breeding and long-term artificial domestication, wheat (*Triticum aestivum* L., $2n = 6x = 42$, AABBDD) is the second leading cereal crop widely cultivated in the world. Wheat grain flour can be used to make various special foods such as breads, noodles, cakes, and biscuits (Shewry, 2009). The mature wheat grains

contain a variety of nutrients including starch and storage protein that are rapidly synthesized and accumulated after flowering (Guo et al., 2012; Liu et al., 2012). Their compositions and properties play important roles in wheat yield and quality formation. In particular, gliadins, and glutenins are the major storage protein of endosperm and confer dough viscoelasticity that is closely related to bread making quality (Shewry, 2009; Yan et al., 2009; Xie et al., 2010).

Wheat often encounters various biotic and abiotic stresses such as drought, high temperature, and insect pests during growth and development. Drought stress that has occurred at wheat grain developing stages severely affects protein and starch biosynthesis, leading to a significant reduction in grain weight and yield (Dupont and Altenbach, 2003; Duan et al., 2020). In the face of drought and other adverse stress events, plants have developed various mechanisms for survival. The signal transduction is a key mechanism for perceiving and transmitting water stress signals so that adaptive responses are activated (Zhu, 2002). Drought also induces the formation of plant hormones such as abscisic acid (ABA), jasmonate and ethylene to promote the production of reactive oxygen species (ROS). Plant signal transduction in response to drought stress includes an ABA-dependent osmotic stress signal transduction, MAPKKK signal cascade transduction pathway and an ROS scavenging response pathway. In addition, plants can rid themselves of excessive toxic substances such as ROS caused by drought, as well as protect their membrane system by activating the enzymatic and non-enzymatic scavenging systems (Pandey et al., 2008; Ge et al., 2012). With the recent development of wheat genomics, proteomics has made great progress in revealing plant synergistic response mechanisms to drought stress. To date, proteomic responses to drought stress of wheat's different organs have been widely studied, including seedling roots and leaves (Cheng et al., 2015; Hao et al., 2015), grain development (Ge et al., 2012; Deng et al., 2018; Duan et al., 2020), flag leaves (Deng et al., 2018; Zhu et al., 2020), and glume and awn (Deng et al., 2019). Various protein post-translational modifications also participate in regulating drought responses such as phosphorylation and acetylation modifications (Zhang et al., 2014a,b; Zhu et al., 2018). In recent years, subcellular proteomics has been developed to decipher different organelle proteins involved in abiotic stress defense such as endoplasmic reticulum, plasma membrane, chloroplast, and mitochondria (Wang and Komatsu, 2016; Zhang et al., 2021; Zhu et al., 2021).

Considerable transcriptome and proteome work showed that wheat developing grains contained a large number of genes and proteins that are involved in biotic and abiotic stress responses (Gao et al., 2009; Ge et al., 2012; Yu et al., 2016). The nucleus is a highly ordered subcellular organization structure of core area of a cell. It contains biological genetic information that regulates gene expression and stress responses (Pandey et al., 2008). At present, studies on nuclear proteome involved in plant growth and stress response have been performed in some plant species, including Arabidopsis (Palm et al., 2016), rice (Choudhary et al., 2009), soybean (Yin and Komatsu, 2015; Wang and Komatsu, 2016), chickpea (Pandey et al., 2008; Kumar et al., 2014), and *Medicago truncatula* (Repetto et al., 2008). In wheat, only a few studies on nuclear proteome have been reported so far. Bonnot et al. (2015) used nano liquid chromatography-tandem mass spectrometry to identify 387 nuclear protein spots in the developing wheat grains. The researchers found that some proteins were involved in regulating transcription such as HMG1/2-like protein and histone deacetylase HDAC2 with abundant expression in the phase between cellularization to grain-filling. Bancel et al. (2015) analyzed the nuclear proteome of two wheat varieties at two developmental stages by mass spectrometry. The study found some differently expressed proteins between species and developmental stages, and many these proteins may be involved in regulating protein synthesis in grains. However, wheat nuclear proteome responses to abiotic stress during grain development remain unknown.

Over the past several decades, there have been great developments in mass spectrometry technology. As a major breakthrough, label-free techniques enable protein quantification relative to the levels of corresponding peptides in a sample and greatly improve the sensitivity for protein detection, especially for the detection of low abundance proteins (Bantscheff et al., 2007). In this study, we performed the first label-free quantitative nuclear proteome analysis to uncover the underlying drought response and defense mechanisms of wheat nuclear proteins in developing wheat grains. Our results revealed a synergistic response network of wheat nuclear proteins under field water-deficit conditions, thereby providing new insights into the mechanisms of plant abiotic stress response.

MATERIALS AND METHODS

Plant Material, Field Trial and Water-Deficit Treatment

Plant material used in this study was Chinese elite winter wheat cultivar Zhongmai 175. This cultivar has many advantages such as better adaptation, high yield and superior noodle quality widely cultivated in the north wheat production area of China in recent years (Li et al., 2016). Field trial was conducted at the experimental station of China Agricultural University, Beijing during 2019–2020 growing season, including a control group with well-watered irrigation and a water-deficit treatment group without irrigation after flowering. The traditional flood irrigation with 75 mm at both jointing and flowering stages was set as well-watered irrigation while rain-fed region without artificial water irrigation was used as no irrigation group. Three biological replicates were performed for each treatment (each plot was 4 m × 9 m with row spaced at 0.16 m). To minimize the interference from adjacent plots, 1-m-wide zone between plots was set as an unirrigated zone. According to the previous study, the stage of 9–18 day post-anthesis (DPA) is the critical period of starch synthesis and accumulation during grain filling in response to water deficit (Yang et al., 2004). Thus, the developing grain samples from 15 DPA were collected for later analysis in this study.

Isolation and Purification of Cell Nucleus From Wheat Developing Grains

Isolation and purification of grain cell nuclear were based on Bancel et al. (2015) with modifications. The developing grain samples about 2 *g* from 15 DPA were grounded into powder with liquid nitrogen. The powder was thoroughly mixed in 8 mL extraction buffer with 20 mM HEPES-KOH (pH 7.0), 5 mM $MgCl_2$, 10 mM 2-mercaptoethanol, 0.5 mM PMSF, and 0.1% (v/v) phosphatase inhibitor (Sigma-Aldrich). After filtering through a two-layer Miracloth filter (Calbiochem, pore size 25 μm) for removing cell debris, the homogenate was transferred into a 15 mL centrifuge tube and more extraction buffer was added with a final volume of 12 mL containing 0.5% (v/v) Triton X-100 for lysing membranes, and then incubated at 4°C for 15 min. After centrifugation at 1,000 *g* and 4°C for 10 min, the supernatant was collected, and the precipitation was washed four times with extraction buffer. Meanwhile, the supernatant was successively collected in five-time extractions and, respectively, marked as sample S1–S5.

The crude extracts of cell nucleus were resuspended in 4 mL extraction buffer and placed on the top of a stepwise Percoll gradient (4 mL 30% Percoll solution and 4 mL 80% Percoll solution). After centrifuging at 930 *g* and 4°C for 30 min, the nucleus floating at the interface of 30 and 80% liquid layers were collected and washed twice with 4 mL extraction buffer. The nuclear pellets were resuspended in PBS buffer (1.47 mM KH_2PO_4, 4.3 mM $Na_2HPO_4{\cdot}2H_2O$, 2.7 mM KCl, and 137 mM NaCl), and an aliquot was stained in 0.1 μg/mL Hoechst 33,258 fluorescent dye solution for 20 min. Finally, the nuclear signals were detected by confocal microscope (Leica TCS SP5, Germany) to evaluate nuclear integrity and enrichment flux.

Nuclear Protein Extraction

The isolated nuclear samples were resuspended in 500 μL of TRI reagent (Sigma-Aldrich) and incubated at room temperature for 5 min. Then, 50 μL of 1-bromo-3-chloropropane was added to the mixture and vortexed for 15 s. After incubation at room temperature for 15 min and centrifugation at 12,000 *g* and 4°C for 15 min, RNA located on the upper phase was removed. To precipitate DNA, 150 μL of 95% (v/v) ethanol was added, and incubated at room temperature for 3 min. The mixture was centrifuged at 2,000 *g* and 4°C for 5 min and the supernatant was collected. Then, 750 μL of 2-propanol was added and incubated at room temperature for 10 min. After centrifugation at 12,000 *g* and 4°C for 10 min, the protein pellets were collected and washed three times with 95% (v/v) ethanol. Freeze drying was performed and then the protein samples were kept at -80°C for later use.

Trypsin Digestion

Dithiothreitol was added to dried nuclear proteins with a final concentration of 5 mM and incubated at 56°C for 30 min. After that, proper amount of iodoacetamide was added to the protein mix with a final concentration of 11 mM, and further incubated at room temperature for 15 min in darkness. Subsequently, the protein samples were diluted by adding 100 mM NH_4HCO_3 to a urea concentration less than 2 M and digested with trypsin buffer at a final enzyme/protein ratio of 1:50 (w/w) at 37°C overnight. Then, second digestion was performed with trypsin buffer at a final enzyme/protein ratio of 1:100 (w/w) for 4 h.

UPLC Fractionation

The digested peptides were fractionated through reversed-phase high performance liquid chromatogram by using a chromatographic column (Agilent 300 Extend C18) with an internal diameter of 4.6 mm and a particle size of 5 μm. The peptides were eluted with gradient acetonitrile (8–32%; $NaHCO_3$ buffer, pH 9) and collected for later analysis.

NanoLC-MS/MS Analysis

The grain nuclear proteome identification was performed by using EASY-nLC 1000 (Thermo/Finnigan, Sanjose, CA, United States) according to the recent report (Zhang et al., 2021).

Sequence Database Search and Data Analysis

The resulting MS/MS data were searched using Maxquant search engine (v.1.5.2.8) against Uniprot *T. aestivum* (taxonomy: 4565 143 146 sequences). Trypsin/P was specified as cleavage enzyme allowing up to 2 missing cleavages. Mass precursor tolerance was set to 5 ppm, and MS/MS tolerance was set to 0.02 Da. False positive rate was adjusted to <1%, and minimum score for peptides was set to >40. In quantitative analysis, the quantitative values of each sample were obtained via LFQ intensity in three replicates. The ratio of the mean LFQ intensity between the two samples represents the protein fold-change value. LFQ intensity was taken as log2 transform, and then used to calculate the significant *p*-value of differential abundance between two samples. The *p*-value was calculated by using two-tailed test when the protein was quantified three times in the two compared samples. And, the available quantitative proteins were regarded as differentially accumulated proteins (DAPs) with at least two-fold change (Student's *t*-test, $p < 0.05$) based on Bostanci et al. (2018).

Western Blotting

The extracted nuclear proteins (N) and the five supernatant samples (S1–S5) collected during the washing process of nuclear precipitation were separated by SDS-PAGE gel, and then transferred to PVDF membrane (GE, United States) using a wet transblot system (Bio-Rad, United States). The PVDF membrane was incubated in blocking buffering for 1 h at room temperature, and then further incubated overnight at 4°C with polyclonal antibody at recommend dilution of application example. After washing three times with PBST buffer, the PVDF membrane was incubated with horseradish peroxidase-conjugated goat anti rabbit IgG at 1:5000 dilution for 1 h at room temperature. The signal was visualized by using ECL plus Western blotting detection kit (GE Healthcare Bio-Sciences AB, Uppsala, Sweden) based on the manufacturer's instructions.

Subcellular Localization Prediction and Verification of Nuclear Proteins

The subcellular localization of the identified DAPs were predicted according to WoLF PSORT[1], CELLO[2], Plant-mPLoc[3], and UniProtKB[4]. To verify the predicted results of subcellular localization, the full-length coding sequences of the selected DAPs encoding genes were amplified by PCR and ligated into pSAT1-GFP-N (Pe3449) vector containing green fluorescent protein (GFP) gene. The recombinant plasmid DNA and pSAT1-GFP-N (Pe3449) empty control were transformed into wheat protoplasts based on Zou et al. (2020). After incubating overnight at 23°C in the dark, the GFP signal was observed by using a confocal laser scanning microscope (Leica TCS SP5, Germany).

Functional Annotation

Gene Ontology (GO) annotation of the nuclear proteins was conducted by using AgBase version 2.00, including molecular function, cellular components, and biological functions.

Detection of the *Cis*-Acting Elements in the Genes Encoding the Nuclear Differentially Accumulated Proteins

The gene ID of DAP was reunified into IWGSC gene ID. The *cis*-acting elements in in the 1,500 bp promoter region of the nuclear protein encoding gene were detected by PlantCARE[5]. Genome database (IWGSC RefSeqv1.1) was used to search the sequences of these nuclear proteins encoding genes with a coverage rate of 94% from GRAMENE[6].

Transcriptional Expression Analysis by RNA-seq

The transcriptional expression profiling of the nuclear DAPs encoding genes was analyzed using the publicly available bread wheat RNA-seq database, which included expression data from different grain developmental stages, and various stress treatments. RNA-seq data for the nuclear DAPs encoding genes in wheat grains were derived from expVIP[7].

RNA Extraction and Real-Time Quantitative Polymerase Chain Reaction

To detect the dynamic transcriptional expression changes of the key nuclear DAPs encoding genes under water-deficit condition, total RNA from three grain developmental stages (10, 15, and 20 DPA) was extracted using TRIZOL Reagent according to the manufacturer's instructions. Genomic DNA removal and the reverse transcription reactions were executed by using Prime ScriptRT reagent Kit with gDNA Eraser (TaKaRa, Japan). Gene-specific primers were designed using the online tool Primer 3 Plus[8]. Ubiquitin was used as the reference for normalization. Real-time quantitative polymerase chain reaction (qRT-PCR) was conducted using CFX96 Real-Time PCRD detection System (Bio-Rad) according to Yu et al. (2016). All data were analyzed with CFX Manager Software (Bio-Rad).

RESULTS

Nuclear Isolation and Quality Assessment of the Grain Nuclear Proteins

To evaluate the integrity of the isolated grain cell nucleus, we used Hoechst staining to detect the morphology of the nucleus extracted from wheat developing grains at 15 DPA (**Supplementary Figure 1A**). The results showed that the average diameter of the nucleus was approximately 5–10 μm, and most were complete. This means that the isolated cell nucleus had better integrity. However, when observed under a bright field environment (**Supplementary Figure 1B**), some substances like starch granules and cell debris were stained with fluorescent dye. This could lead the nuclear proteins to become contaminated.

TRI Reagent was used to extract grain nuclear proteins from the enriched nuclear components. To estimate the efficiency of nuclear protein extraction, five supernatant samples (S1–S5) collected during the washing process of nuclear precipitation were separated by SDS-PAGE. Coomassie blue staining showed that no significant amount of proteins in S3–S5 was present, indicating that samples S1 and S2 can extract most of the nuclear proteins. As a nuclear organelle marker protein, the core histone was detectable in the nuclear proteome fractions (**Figure 1A**). Western blotting allowed us to further examine the purity of nuclear protein fractions by using specific antibodies against marker proteins from different organelles (**Figures 1B–F**), including histone H3 (H3) from cell nuclei, photosystem II reaction center protein D1 (PsbA) from chloroplasts, plant alternative oxidase 1 and 2 (AOX1/2) from mitochondria, cytoplasm UDP-glucose pyrophosphorylase from cytoplasm, and plasma membrane H^+-ATPase (H^+-ATPase) from plasma membrane. We detected the extremely strong nuclear specific protein H3 band and the weak chloroplast specific protein PsbA band in the nuclear protein fractions of wheat developing grains while the remaining antibodies did not show clear protein bands. These results indicated that the extracted grain cell nuclear proteins had a high purity without clear cross-contamination.

Nuclear Proteome Response to Water Deficit

Label-free quantitative nuclear proteomic analysis of wheat developing grains under control and water-deficit treatment groups identified 193,026 peptides (**Supplementary Table 1**), corresponding to 5,850 unique proteins with a high confidence (**Supplementary Table 2**). Among them, we determined 625 DAPs with at least two-fold differences under water deficit

[1]http://www.genscript.com/wolf-psort.html
[2]http://cello.life.nctu.edu.tw/
[3]http://www.csbio.sjtu.edu.cn/bioinf/plant-multi/
[4]https://www.uniprot.org/
[5]http://bioinformatics.psb.ugent.be/webtools/plantcare/html/
[6]http://ensembl.gramene.org/
[7]http://www.wheat-expression.com/
[8]www.bioinformatics.nl/cgi-bin/primer3plus/primer3plus.cgi

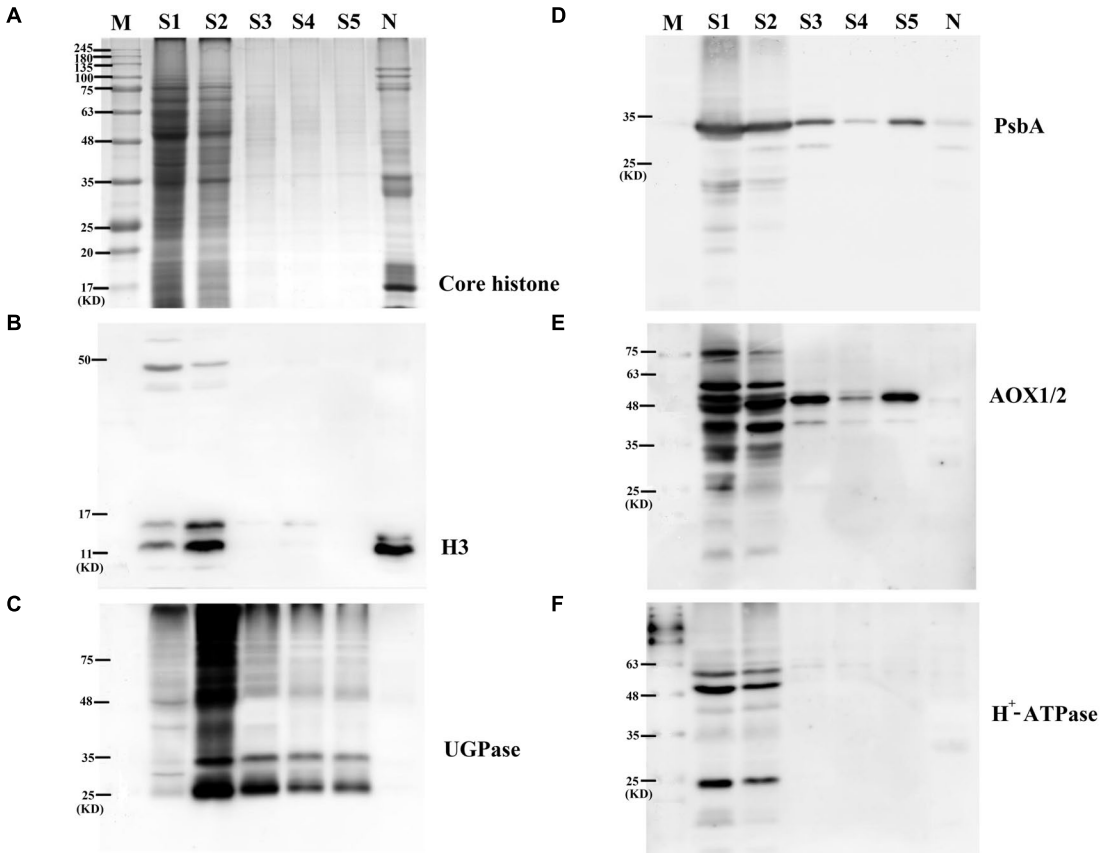


FIGURE 1 | Enrichment and purity assessment of nuclear proteome from wheat developing grains. **(A)** Analytical 1-D electrophoresis profile (12.5% SDS-PAGE, CBB staining) of protein content in different stages of nuclei enrichment progress. **(B)** Immunoblot analysis with anti-histone H3. **(C)** Immunoblot analysis with anti-vacuolar UGP. **(D)** Immunoblot analysis anti-photosystem PsbA. **(E)** Immunoblot analysis with anti-AOX. **(F)** Immunoblot analysis with anti-H^+-ATPase. S1–S5, five supernatant samples (S1–S5) collected during the washing process of nuclear precipitation; N, extracted nuclei proteins.

(**Supplementary Table 3**). Based on the subcellular localization prediction by the four databases of UniProt KB, CELLO, WoLF PSORT, and Plant-mPLoc, a total of 398 DAPs accounting for 63.78% were localized in nucleus (**Figure 2A** and **Supplementary Table 3**). We predicted that the remaining DAPs would localize in different organelles, mainly including chloroplast (13.62%), cytoplasm (6.25%), plasma membrane (5.13%), extracellular (3.69%), and vacuole (2.72%). These results are generally consistent with previously reports on wheat grain nuclear proteome analysis (Bancel et al., 2015; Bonnot et al., 2015).

To verify the prediction results of the subcellular localization, we selected six DAPs predicted in the cell nucleus for further subcellular localization assay through protein transient expression in wheat protoplast, including WD-40 repeat-containing protein (FVE), transcription factor TCP20 (TCP20), mitogen-activated protein kinase (MAPK), low temperature-responsive RNA-binding protein (LTR) and transcription factors YABBY6 (YABBY6) and WRKY transcription factor 19 isoform X1 (WRKY19). We cloned the encoding sequences of these DAPs encoding genes using specific PCR primers (**Supplementary Table 4**), and then recombined them into pSAT1-GFP-N (Pe3449) vector carrying the GFP encoding gene. We transformed the vectors connected with the target genes into wheat protoplast for transient expression, and then observed them through a confocal microscope. The results showed that strong signals of these six DAPs occurred only in the nucleus (**Figure 2B**), confirming that they were all localized in the cell nucleus. These results were consistent with the website localization predictions (**Supplementary Table 3**).

Among 398 nucleus-subcellular DAPs, the abundance of 146 and 252 proteins under water deficit condition were up-regulated and down-regulated, respectively, (**Figures 3A,B**). To further understand the characteristics and potential roles of these DAPs in response to water deficit, we performed analyses of the GO functional classification according to their biological process, molecular function and cellular component annotation (**Figures 3C,D** and **Supplementary Table 5**). Based on our molecular function analysis, the up-regulated DAPs were mainly associated with DNA binding, histidine phosphotransfer kinase activity, while the down-regulated DAPs were mainly involved in RNA binding, DNA binding, nucleic acid binding and structure constituent of ribosome. The results of biological process analysis showed that the up-regulated DAPs were involved in nucleosome assembly, protein phosphorylation, and the phosphorelay signal

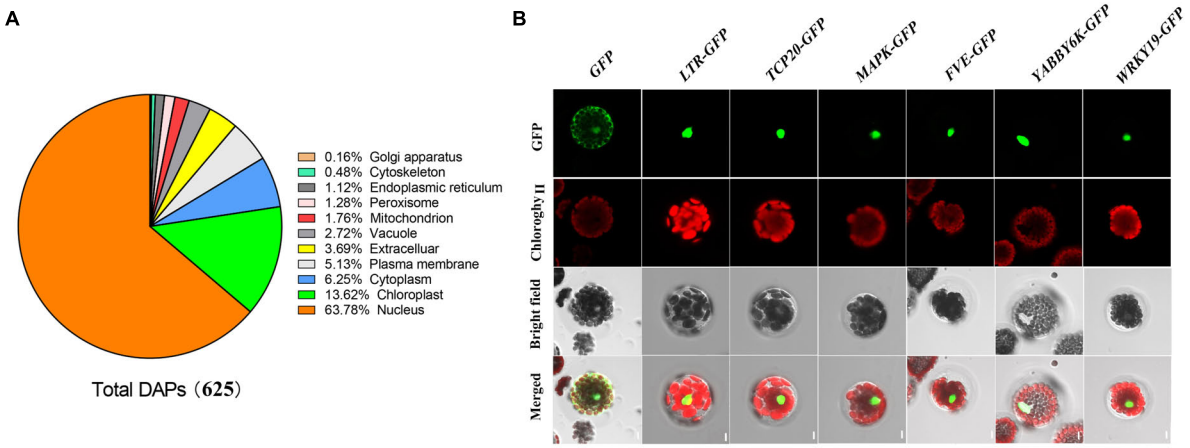


FIGURE 2 | Subcellular localization of differentially accumulated proteins identified under water deficit. **(A)** Subcellular localization of differentially accumulated proteins in wheat developing grains under drought stress. **(B)** Subcellular localization of six reprehensive proteins by wheat protoplast transformation. GFP, GFP fluorescence; Chlorophyll, chlorophyll autofluorescence; Bright field, the field of bright light; Merged, merged GFP fluorescence, chlorophyll autofluorescence, and the field of bright light. LTR, low temperature-responsive RNA-binding protein; TCP20, TCP20 transcription factors; MAPK, mitogen-activated protein kinase; FVE,WD-40 repeat-containing protein; YABBY6, transcription factor YABBY6; and WRKY, transcription factor WRKY.

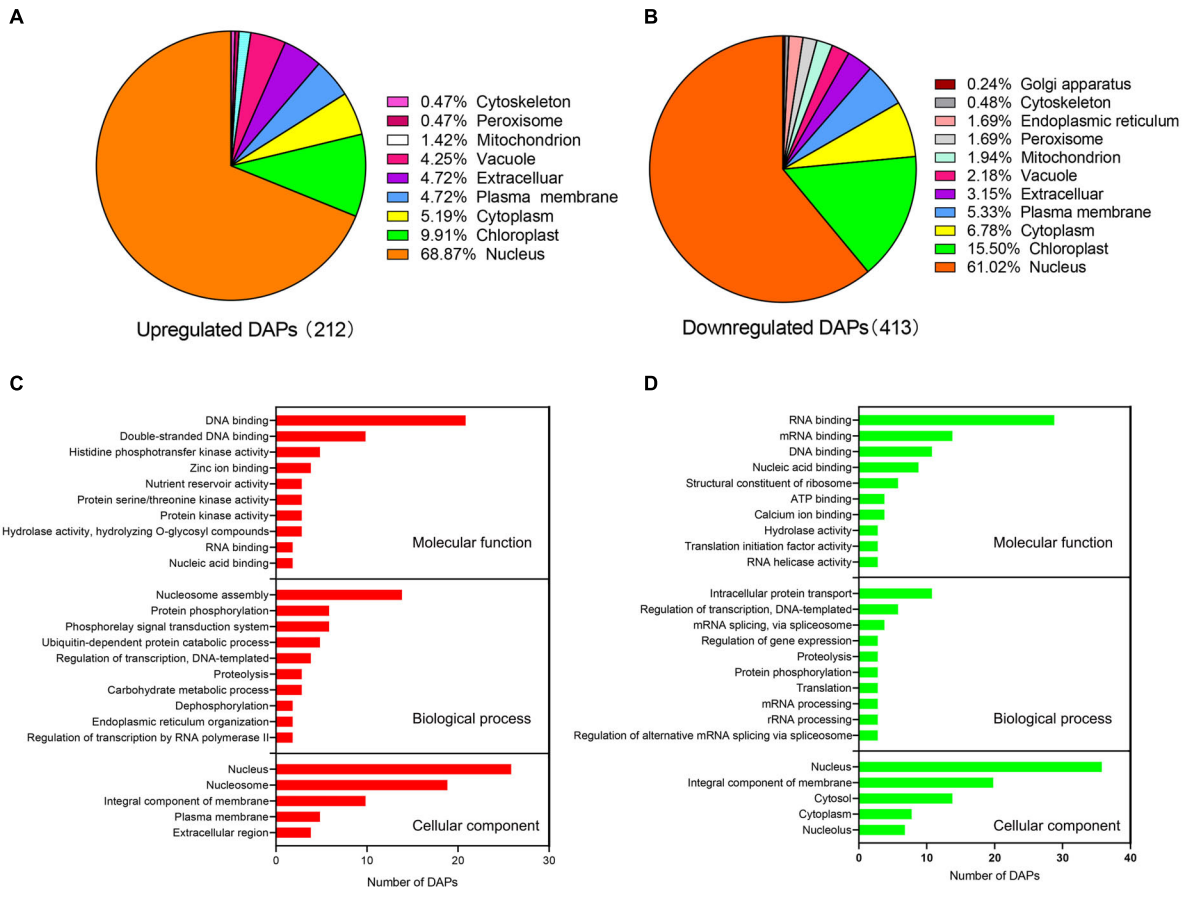


FIGURE 3 | Subcellular localization and GO annotation of differentially accumulated proteins (DAPs) identified from wheat nucleus under water deficit. **(A)** Subcellular localization of significantly up-regulated nuclear proteins. **(B)** Subcellular localization of significantly down-regulated nuclear proteins. **(C)** GO annotation of significantly up-regulated proteins. **(D)** GO annotation of significantly down-regulated proteins.

transduction system. The down-regulated DAPs were associated with intracellular protein transport, regulation of transcription, mRNA splicing and regulating gene expression. In terms of the cellular component analysis, the up-regulated DAPs were mainly distributed in nucleus and nucleosome, while the down-regulated DAPs were mainly distributed in nucleus and integral component of membrane (**Figures 3C,D**).

Main *Cis*-Acting Elements Analysis of 22 Water-Deficit Responsive Nuclear Proteins Encoding Genes

We analyzed the main *cis*-acting elements in the 1,500 bp promoter region of 22 significantly up-regulated nuclear proteins encoding genes that showed a significant up-regulation in response to water deficit. In total, we identified 25 major *cis*-acting elements among these genes, which could be divided into five categories: metabolism related elements, environmental stress-related elements, hormones response elements, promoter related elements, and development related elements (**Supplementary Table 6**). Among them, *cis*-acting elements related to environmental stress and hormones response were particularly abundant. The environmental stress-related elements included LTR, antioxidant response element (ARE), MYC, MYB, WUN-motif, GC-motif, TC-rich repeats, and MBS. MYB in response to drought regulation and MYC in response to early drought were the most abundant. ARE and MBS (MYB binding site involved in drought induction) occurred frequently in some DAPs encoding genes such as PHD finger protein, TCP20 and E3 ubiquitin-protein ligase. The widely distributed *cis*-acting elements related to hormone response mainly included ABRE involved in ABA response and TGACG-motif involved in MeJA response. Additionally, some *cis*-acting elements related to transcription such as TATA-box and TGACG-box were also abundant.

Transcriptional Expression Profiling by RNA-seq

Using RNA-seq, we analyzed the transcriptional expression profiling of the 22 water-deficit responsive nuclear proteins encoding genes in different organs and growth periods as well as in response to various abiotic stresses. The publicly available RNA-seq database of bread wheat (var. Chinese Spring) allowed us to collect the relative expression levels of these genes. All the nuclear protein Uniprot ID was converted to wheat genetic ID by integrating it into GRAMENE, and then we submitted the gene ID to the bread wheat Chinese Spring public RNA-seq database to obtain the relative expression levels of 22 nuclear proteins encoding genes (**Supplementary Table 7**) and constructed the heat maps (**Figure 4**).

The expression profiling in different organs (root, leaf/shoot, and grain) showed that most DAPs encoding genes had a high expression level in root and leaf/shoot, and some genes had organ or growth stage expression preference (**Figure 4A**). In particular, the PHD finger protein and auxin response factor (ARF) encoding genes expressed in all organs. The DNA polymerases epsilon and sucrose 1-fructosyltransferase (1-SST) encoding genes showed an expression preference in roots, and five nuclear proteins encoding genes presented an expression preference in leaves/shoots, including *serine/arginine-rich-splicing factor SR34 isoform X3* (*SR34*), *ARF*, *CBL-interacting protein kinase 11* (*CIPK11*), *RING-box protein* (*RBX1*), and *mediator of RNA polymerase II transcription subunit 4* (*MED4*). In the grain, the PHD finger protein, E3 ubiquitin-protein ligase, and MAPK encoding genes showed a relatively high expression level at 10 or 15 DPA. In addition, some genes had relatively high expression in the root at the tilling stage, such as *E3 ubiquitin-protein ligase*, *DNA polymerases epsilon*, and *1-SST*. Others showed a relatively high expression level in the leaves/shoots at the milk grain stage, such as *TCP20*, *ARF*, and *CIPK11*.

Under abiotic stresses (water, drought, heat, and cold), most genes displayed distinct expression patterns (**Figure 4B**). The expression of ubiquitin-conjugating enzyme E2, TCP20, HMG1/2-like protein (HMG1/2), LRR-RLK, DNA topoisomerase, histone H2AX and B3 domain-containing protein encoding genes were up-regulated under water stress, while the expression of ARF and ubiquitin carboxy-terminal hydrolase encoding genes were up-regulated under cold treatment. In addition, some stress-responsive proteins encoding genes showed significant upregulation under multiple

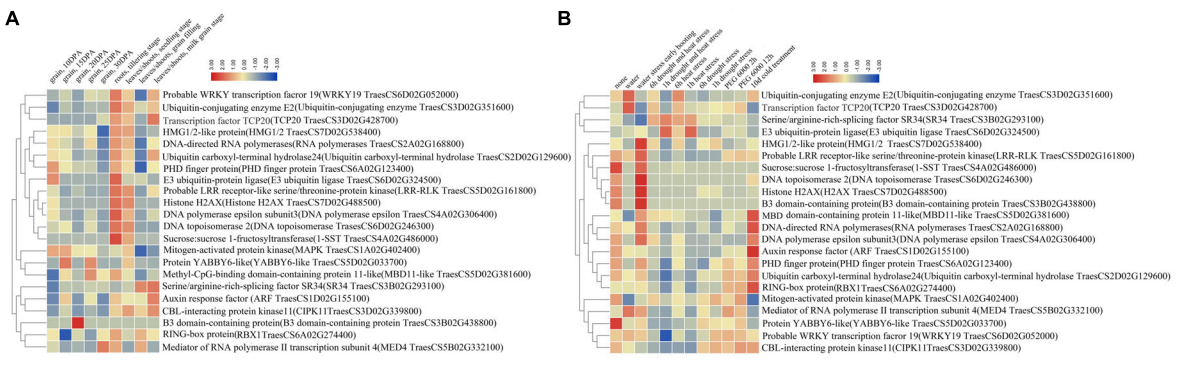


FIGURE 4 | The relative expression profiles of the 38 up-regulated nuclear differentially accumulated proteins (DAPs) encoding genes. **(A)** In different development period of wheat different organs. **(B)** Different abiotic stresses. Red means high expression level, green means low expression level.

abiotic stresses, including SR34 and E3 ubiquitin-protein ligase under heat and drought stresses, MED4 encoding gene under water and PEG6000 stresses, MBD11, RNA polymerases and DNA polymerases epsilon encoding genes under water and cold stresses.

Dynamic Transcriptional Expression Analysis by Real-Time Quantitative Polymerase Chain Reaction

We selected eight DAPs encoding genes closely related to wheat drought stress response to further detect their dynamic transcriptional expression patterns during grain development under water deficit by qRT-PCR. These DAPs encoding genes were significantly up-regulated or down-regulated under water-deficit conditions via preceding transcriptional analyses, including *RBX1*, *ARF*, *SR34*, *MBD11*, *MED4*, *CIPK11*, *putative E3 ubiquitin-protein ligase SINA-like 6* (*UPL*), and *TCP20*. **Supplementary Table 8** lists the gene-specific primers used for qRT-PCR analysis and **Figure 5** shows the results.

In general, eight nuclear DAPs encoding genes showed an up-down or down regulation expression during grain development under water-deficit treatment. Among them, five genes (*ARF*, *SR34*, *MBD11*, *MED4,* and *CIPK11*) were up-regulated at certain stages in response to water deficit while three (*RBX1*, *UPL,* and *TCP20*) were down-regulated. This was consistent with the results by RNA-seq (**Figure 4**). Compared to the protein levels, the transcriptional expression of four genes (*ARF, MBD11, SR34,* and *CIPK11*) showed a high consistency. However, the remaining four genes displayed a poor consistency between their transcription and translation expression as the previous report (Deng et al., 2018).

DISCUSSION

As sessile organisms, higher plants often face various environmental challenges during their growth and development. In the long-term evolutionary process, higher plants have formed a series of defense systems to cope with variable extreme environments, including environmental stress responses and gene expression regulation (Boyer, 1982). Transcription factors play a vital role in regulating gene expression, and many are associated with various abiotic stress responses (Zhu et al., 2015). In this study, we identified some transcription factor proteins with significant expression changes in response to water deficit, including TCP, NAC, and WRKY transcription factors (**Supplementary Table 3**).

T protein are plant-specific transcription factors characterized by the TCP domain, a motif encompassing a non-canonical basic-helix-loop-helix structure (Mar and Pilar, 2010). TCPs not only play an important role in controlling plant architecture (Aguilar-Martínez et al., 2007) and pollen development (Takeda et al., 2006), but they are also involved in responding to multiple abiotic stresses (Almeida et al., 2017; Liu et al., 2020). The expression level of *PeTCP10* in moso bamboo (*Phyllostachys edulis*) was induced by drought and ABA, and its overexpression significantly enhanced the tolerance of Arabidopsis and rice to drought stress (Liu et al., 2020). In Rice, *PCF2* was demonstrated to associate with salinity tolerance by directly targeting and positively regulating NHX1 that encodes a tonoplast-localized K^+-Na^+/H^+

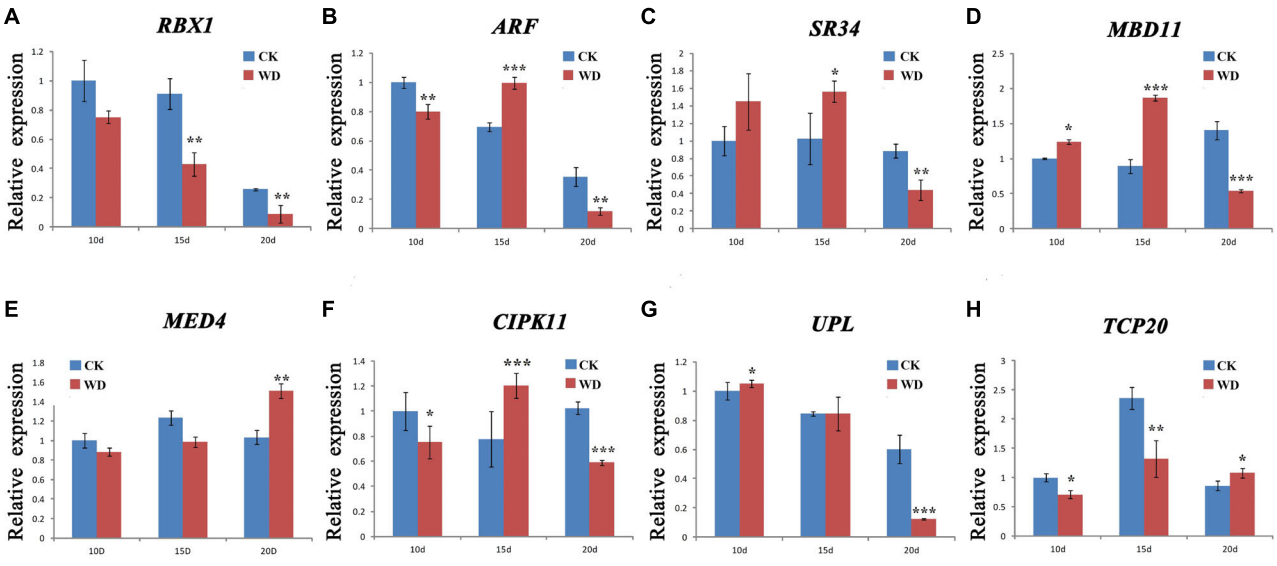


FIGURE 5 | qRT-PCR analysis of eight representative differentially accumulated proteins (DAPs) encoding genes in nuclear-enriched fraction of Zhongmai 175 grain under water deficit. **(A)** RBX1, E3 ubiquitin-protein ligase RBX1; **(B)** ARF, Auxin response factor; **(C)** SR34, Serine/arginine-rich-splicing factor SR34 isoform X3; **(D)** MBD11, Methyl-CpG-binding domain-containing protein 11; **(E)** MED4, Mediator of RNA polymerase II transcription subunit 4; **(F)** CIPK11, CBL-interacting protein kinase 11; **(G)** UPL,putative E3 ubiquitin-protein ligase SINA-like 6; and **(H)** TCP20, transcription factor TCP20. Statistically significant differences were calculated based on an independent Student's *t*-tests: $^*p < 0.05$; $^{**}p < 0.01$; and $^{***}p < 0.001$. CK, control; WD, water deficit.

antiporter protein (Almeida et al., 2017). In addition, TCP21 and PCF6 were reported as targets of microRNA319 and positively regulate cold-stress tolerance (Wang et al., 2014). In the current study, we found that transcription factor TCP20 was significantly up-regulated in wheat developing grains under water-deficit conditions (**Supplementary Table 3**) and the peak expression level of the *TCP20* gene occurred at the grain milk stage (**Figure 4A**), indicating that TCP20 could play important roles in water-deficit response during wheat grain development.

The NAC proteins are one of the largest transcription factor family only discovered in plants (Mao et al., 2012). They play important roles in responding to various environmental stresses (Zhu et al., 2015). In Arabidopsis, the ectopic expression of *ANAC055*, *ANAC019*, or *ANAC072/RD26* was associated with enhanced tolerance to both salt and drought stresses (Tran et al., 2004). As reported, *ANAC096* functions as a positive regulator of drought stress response and cooperates with the ABF2 and ABF4 to activate the transcription of ABA-inducible genes *RD29A* in response to dehydration and osmotic stresses (Xu et al., 2015). In rice, drought and salt tolerance significantly increase according to overexpressing of *OsNAC10*, *OsNAC063*, *OsNAC5*, or *SANC1/OsNAC9* (Jeong et al., 2010, 2013; Fang et al., 2015; Shen et al., 2017). *SNAC3* confer heat and drought tolerance by modulating ROS homeostasis through enhancing ROS scavenging genes' expression, including *R45* (RbohF, LOC_Os08g35210) and *R54* (Prx IIE2, LOC_Os02g09940; Fang et al., 2015). In *Pyrus betulifolia*, *beNAC1* improved stress tolerance by interacting with PbeDREB1 and PbeDREB2A to enhance the mRNA levels of some stress-associated genes (Jin et al., 2017). To date, an increasing number of NAC transcription factors in wheat were found to play crucial roles in a variety of abiotic stress responses, including *TaNAC2*, *TaNAC67*, *TaNAC29,* and *TaSNAC4* (Mao et al., 2012, 2014; Xu et al., 2015; Mei et al., 2021). In this study, quantitative grain nuclear proteome analysis identified three NAC domain-containing proteins, of which A0A3B6LX64 was significantly induced by water stress (**Supplementary Table 3**). This NAC domain-containing protein may be a new member of wheat NAC family involved in drought response.

WRKY transcription factor 19 (WRKY19, A0A3B6Q9R1) was also identified in the grain nuclear proteome (**Supplementary Table 3**). Plant WRKY transcription factors are downstream proteins of MAPK cascades, which are induced by ROS (H_2O_2) and in turn play an essential role in plant drought responses by activating transcription factors (Taj et al., 2011). In rice, *OsWRKY30* can be activated by MAPKs via phosphorylation, and its overexpression dramatically enhances drought tolerance (Shen et al., 2012). WRKY transcription factors are also key components in ABA signaling. *AtWRKY63* in Arabidopsis, activated by ABI5, activated downstream stress-inducible genes such as *RD29A* and *COR47* (Rushton et al., 2012). Our results showed that the protein accumulation level of WRKY19 (A0A3B6Q9R1) increased 2.244 times under water deficit (**Supplementary Table 3**) and the identified *cis*-acting elements in the promoter region of the *WRKY19* gene were predominantly associated with the stress- and hormone-responsive elements, including MYB, MYC, MBS, and ABRE and TGACG-motif (**Supplementary Table 6**). Moreover, the accumulation level of MAPK protein (A0A3B5Y6A9) was also significantly up-regulated under water-deficit condition (**Supplementary Table 3**). We speculated that *WRKY19* could be activated by MAPK protein, and then participated in water-deficit response via controlling downstream stress-related genes during wheat developing grains.

Most abiotic stresses can induce the changes of Ca^{2+} concentration, which are generally accepted as a secondary messenger to transduce the cellular responses to extracellular stimuli (Hofer and Brown, 2003; Kolukisaoglu et al., 2004). In plants, many Ca^{2+}-sensing protein kinases are involved in the stress responses, including calcium-dependent protein kinases (Chandran et al., 2006) and suc non-fermentation-related kinases (SnRKs; Nozawa et al., 2001). CBL protein-interacting protein kinases (CIPKs) are members of the SnRK3 subfamily. In maize, overexpression of *ZmCIPK21* significantly enhances plant salt tolerance by decreasing accumulation of Na^+ and allowing retention of relative high K^+ (Chen et al., 2014). In rice, overexpressing *OsCIPK03, OsCIPK12*, and *OsCIPK15* showed a higher tolerance to cold, drought and salt stress, respectively. Meanwhile, researchers have found a higher content of proline and soluble sugar in *OsCIPK03*- and *OsCIPK12*-overexpressing transgenic plants (Xiang et al., 2007). Recent study has found that *TaCIPK27* dramatically enhanced drought tolerance of Arabidopsis by regulating ABA-related genes' expression such as *ABI1*, *RD29B* and drought-related genes such as *DREB1B* and *DREB2A* (Wang et al., 2018). In this study, the abundance of CIPK11 (A0A3B6H0R7) protein was increased by 2.015 times and its transcription level was also significantly up-regulated at 15 DPA under water deficit (**Figure 5**). These results indicate that CIPK11 may contribute to water-deficit response by regulating the expression of downstream stress-related genes during wheat grain development.

Histone acetylation is a dynamic process that affects chromatin structure and regulates gene expression. This process is regulated by two types of enzymes: histone acetyltransferases and histone deacetylases (HDAs). In plants, HDACs could be classified into four different subfamilies: reduced potassium dependency protein 3 (RPD3), histone deacetylase 1 protein (HDA1), HD2-like proteins (HD2) and the silent information regulator protein 2 (SIR2; Lusser et al., 1997; Ruijter et al., 2003). Previous studies reported that Arabidopsis HD2C, a HD2-type HDA, was involved in the ABA and salt-stress response by interacting with HDA6 (RPD3-type HDA) and modulating the expression of stress-responsive genes such as *ABI1*, *ABI2,* and *AtERF4* (Luo et al., 2012). The transgenic Arabidopsis with *BdHD1* overexpression displayed a hypersensitive phenotype to ABA and exhibited better survival under drought conditions (Song et al., 2019). These results indicated that HDAs were closely associated with response to abiotic stress. Recently, lysine acetylproteome profiling under wheat water deficit found that many acetylated proteins are involved in grain development and starch biosynthesis (Zhu et al., 2018). We also found that the histone deacetylase complex subunit SAP18 (A0A3B6PGM4) was up-regulated under water deficit (**Supplementary Table 3**), which may positively regulate drought responses by activating

the expression of stress-responsive genes via regulating histone acetylation levels during grain development.

The ubiquitin-proteasome system is mechanistically conserved in eukaryotes consisting of an intricate collection of enzymes and enzyme complexes that conjugate ubiquitin to target proteins and facilitate the degradation of ubiquitinated proteins (Vierstra, 2009). Target proteins are ubiquitinated via an ATP-dependent reaction cascade that generally involves the sequential action of E1 ubiquitin-activating enzyme, E2 ubiquitin-conjugating enzyme, and E3 ubiquitin ligase. In plants, both the E2 ubiquitin-conjugating enzyme and the E3 ubiquitin ligase are pivotal in responding to drought stress (Lyzenga and Stone, 2012; Shu and Yang, 2017). In Arabidopsis, overexpression of E2 ubiquitin-conjugating enzyme, soybean *GmUBC2*, peanut *AhUBC2* and mung bean *VrUBC1* significantly enhanced the drought tolerance through modulating abiotic stress-responsive gene expression (Zhou et al., 2010; Wan et al., 2011; Chung et al., 2013). In addition, SDIR1 (SALT- AND DROUGHT-INDUCED RING FINGER 1), a RING type E3 ligase, serves as a positive regulator of ABA signaling by promoting the transcription of *ABI3* and *ABI5* genes and the overexpression of *SDIR1* consistently enhanced plant drought tolerance (Zhang et al., 2007). Furthermore, another RING type E3 ligase AIRP1 (ABA-insensitive RING protein 1), also functions as a positive regulator in the ABA dependent drought responses, and *AIRP1*-overexpresing transgenic plants showed a drought tolerant phenotype while the *atairp1* mutant was highly susceptible to water-deficit condition (Ryu et al., 2010). In this study, we found that ubiquitin-conjugating enzyme E2 (A0A3B6PIG4) and E3 ubiquitin-protein ligase RBX1 (W5GNW2) were, respectively, increased by 2.011 and 2.459 times, indicating that the process of ubiquitination may play important roles in wheat grain responding to drought stress during development.

Based on the results of this study and previous reports, we propose a putative pathway of wheat grain nuclear subproteome to respond to water deficit (**Figure 6**). Under water deficit, ABA and Ca^{2+} signal induces the expression of transcription factors such as *TCP20*, *WRKY*, *NAC,* and *ARF*, thereby activating the expression of stress-related genes to cope with water-deficit stress. A MAPK cascade signaling system and CBL-CIPK signaling pathway of ABA-dependent endow plants with resistance to water deficits. In addition, some proteins involved in the process of chromatin remodeling and ubiquitin-proteasome system were also activated to adapt or resist water deficit during wheat grain development.

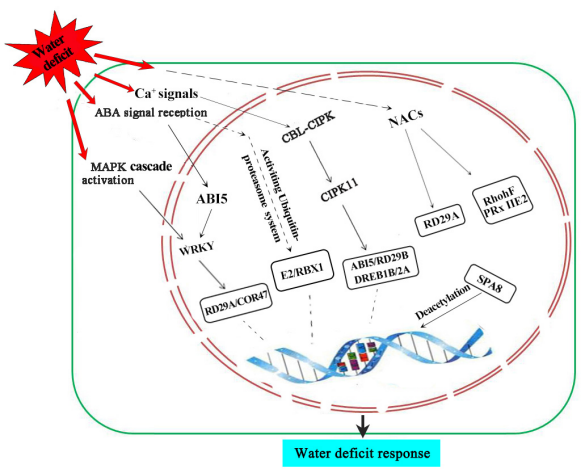


FIGURE 6 | A putative synergistic responsive network of the wheat grain proteome to water-deficit stress. SnRK2, suc non-fermentation-related kinases; CIPK11, CBL protein-interacting protein kinase 11.

CONCLUSION

Label-free quantitative proteomic analysis identified 398 water-deficit responsive DAPs in the cell nucleus subproteome of wheat developing grains. Among them, 146 up-regulated DAPs were mainly involved in stress response and oxidation-reduction process while 252 down-regulated DAPs mainly participated in translation, the cellular amino metabolic process, and the oxidation-reduction process. In particular, the promoter region of 22 up-regulated DAPs encoding genes under water deficit contained abundant *cis*-acting elements related to environmental stresses and hormones response such as ABRE, ARE, MYB, and MYC. We detected the transcriptional expression profiling of these DAPs encoding genes in different organs, developmental stages and abiotic stresses by using RNA-seq and qRT-PCR. We proposed a putative metabolic pathway of wheat grain nuclear subproteome in response to water deficit, thereby providing new evidence from a subcellular proteome level for further understanding the molecular mechanisms of plant drought stress response.

DATA AVAILABILITY STATEMENT

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found below: http://www.proteomexchange.org/, PXD021315.

AUTHOR CONTRIBUTIONS

TL, DZ, and ZH performed most of the experiments, data analysis, and wrote the manuscript. JZ formed part of the experiments and data collection. MZ and YY designed and supervised the experiments. All authors contributed to the article and approved the submitted version.

FUNDING

This research was financially supported by grants from the National Natural Science Foundation of China (31971931 and 31771773). All the mass spectrometry data were deposited in the Proteome X change Consortium

(http://proteomecentral.proteomexchange.org) via the PRIDE partner repository. The dataset identifier number is PXD021315. The English in this document has been checked by professional editors, native speakers of English. For a certificate, please see:https://submit.proofreadingmanuscripts.com//uploads/papers/1627356573641.pdf.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpls.2021.748487/full#supplementary-material

Supplementary Figure 1 | Morphological observation of the nucleus extracted from wheat developing grains by using Hoechst staining.

Supplementary Table 1 | The list of peptides identified from extracted nucleus of wheat developing grains under water deficit.

Supplementary Table 2 | The list of unique proteins identified from extracted nucleus of wheat developing grains.

Supplementary Table 3 | Subcellular localization analysis of differentially accumulated proteins (DAPs) identified from extracted nucleus of wheat developing grains under water deficit.

Supplementary Table 4 | List of primers used for gene cloning and vector constructions.

Supplementary Table 5 | GO annotation of nucleus localized differentially accumulated proteins (DAPs) identified from extracted nucleus of wheat developing grains under water deficit.

Supplementary Table 6 | The *cis*-acting elements in the 1,500 bp upstream promoter regions of 22 significantly up-regulated nuclear proteins encoding genes under water deficit.

Supplementary Table 7 | Transcription expression profile of 22 water deficit responsive nuclear proteins encoding genes in different tissues and developing stages and abiotic stresses based on RNA-seq data.

Supplementary Table 8 | Sequence of specific primers for qRT-PCR analysis.

REFERENCES

Aguilar-Martínez, J. A., Poza-Carrión, C., and Cubas, P. (2007). *Arabidopsis* BRANCHED1 acts as an integrator of branching signals within axillary buds. *Plant Cell* 19, 458–472. doi: 10.1105/tpc.106.048934

Almeida, D. M., Gregorio, G. B., Oliveira, M. M., and Saibo, N. J. (2017). Five novel transcription factors as potential regulators of *OsNHX1* gene expression in a salt tolerant rice genotype. *Plant Mol. Biol.* 93, 61–77. doi: 10.1007/s11103-016-0547-7

Bancel, E., Bonnot, T., Davanture, M., Branlard, G., Zivy, M., and Martre, P. (2015). Proteomic approach to identify nuclear proteins in wheat grain. *J. Proteome Res.* 14, 4432–4439. doi: 10.1021/acs.jproteome.5b00446

Bantscheff, M., Schirle, M., Sweetman, G., Risk, J., and Kuster, B. (2007). Quantitative mass spectrometry in proteomics: a critical review. *Anal. Bioanal. Chem.* 389, 1017–1031.

Bonnot, T., Bancel, E., Chambon, C., Boudet, J., Branlard, G., and Martre, P. (2015). Changes in the nuclear proteome of developing wheat (*Triticum aestivum* L.) grain. *Front. Plant Sci.* 6:905. doi: 10.3389/fpls.2015.00905

Bostanci, N., Selevsek, N., and Wolski, W. (2018). Targeted proteomics guided by label-free quantitative proteome analysis in saliva reveal transition signatures from health to periodontal disease. *Mol. Cell. Proteomics* 17, 1392–1409. doi: 10.1074/mcp.ra118.000718

Boyer, J. S. (1982). Plant productivity and environment. *Science* 218, 443–448. doi: 10.1126/science.218.4571.443

Chandran, V., Stollar, E., Lindorff-Larsen, K., Harper, J. F., Chazin, W. K., Dobson, C. M., et al. (2006). Structure of the regulatory apparatus of a calcium-dependent protein kinase (CDPK): a novel mode of calmodulin-target recognition. *J. Mol. Biol.* 357, 400–410. doi: 10.1016/j.jmb.2005.11.093

Chen, X. J., Huang, Q. S., Zhang, F., Wang, B., Wang, J. H., and Zheng, J. (2014). *ZmCIPK21*, a maize CBL-Interacting Kinase, enhances salt stress tolerance in *Arabidopsis thaliana*. *Int. J. Mol. Sci.* 15, 14819–14834. doi: 10.3390/ijms150814819

Cheng, Z., Dong, K., Ge, P., Bian, Y., Dong, L., Deng, X., et al. (2015). Identification of leaf proteins differentially accumulated between wheat cultivars distinct in their levels of drought tolerance. *PLoS One* 10:e0125302. doi: 10.1371/journal.pone.0125302

Choudhary, M. K., Basu, D., Datta, A., Chakraborty, N., and Chakraborty, S. (2009). Dehydration-responsive nuclear proteome of rice (*Oryza sativa* L.) illustrates protein network, novel regulators of cellular adaptation, and evolutionary perspective. *Mol. Cell. Proteomics* 8, 1579–1598. doi: 10.1074/mcp.m800601-mcp200

Chung, E., Cho, C. W., So, H. A., Kang, J. S., Chung, Y. S., and Lee, J. H. (2013). Overexpression of *VrUBC1*, a mung bean E2 ubiquitin-conjugating enzyme, enhances osmotic stress tolerance in *Arabidopsis*. *PLoS One* 8:e66056. doi: 10.1371/journal.pone.0066056

Deng, X., Liu, Y., Xu, X., Liu, D., Zhu, G., Yan, X., et al. (2018). Comparative proteome analysis of wheat flag leaves and developing grains under water deficit. *Front. Plant Sci.* 9:425. doi: 10.3389/fpls.2018.00425

Deng, X., Zhen, S., Liu, D., Liu, Y., Li, M., Liu, N., et al. (2019). Integrated proteome analyses of wheat glume and awn reveal central drought response proteins under water deficit conditions. *J. Plant Physiol.* 232, 270–283. doi: 10.1016/j.jplph.2018.11.011

Duan, W., Zhu, G., Zhu, D., and Yan, Y. (2020). Dynamic proteome changes of wheat developing grains in response to water deficit and high-nitrogen fertilizer conditions. *Plant Physiol. Biochem.* 156, 471–483. doi: 10.1016/j.plaphy.2020.08.022

Dupont, F. M., and Altenbach, S. B. (2003). Molecular and biochemical impacts of environmental factors on wheat grain development and protein synthesis. *J. Cereal Sci.* 38, 133–146. doi: 10.1016/s0733-5210(03)00030-4

Fang, Y., Liao, K., Du, H., Xu, Y., Song, H., Li, X., et al. (2015). A stress-responsive NAC transcription factor *SNAC3* confers heat and drought tolerance through modulation of reactive oxygen species in rice. *J. Exp. Bot.* 66, 6803–6817. doi: 10.1093/jxb/erv386

Gao, L., Wang, A., Li, X., Dong, K., Wang, K., Appels, R., et al. (2009). Wheat quality related differential expressions of albumins and globulins revealed by two-dimensional difference gel electrophoresis (2-D DIGE). *J. Proteomics* 73, 279–296. doi: 10.1016/j.jprot.2009.09.014

Ge, P., Ma, C., Wang, S., Gao, L., Li, X., Guo, G., et al. (2012). Comparative proteomic analysis of grain development in two spring wheat varieties under drought stress. *Anal. Bioanal. Chem.* 402, 1297–1313. doi: 10.1007/s00216-011-5532-z

Guo, G., Lv, D., Yan, X., Subburaj, S., Ge, P., Li, X., et al. (2012). Proteome characterization of developing grains in bread wheat cultivars (*Triticum aestivum* L.). *BMC Plant Biol.* 12:147. doi: 10.1186/1471-2229-12-147

Hao, P., Zhu, J., Gu, A., Lv, D., Ge, P., Chen, G., et al. (2015). An integrative proteome analysis of different seedling organs in tolerant and sensitive wheat cultivars under drought stress and recovery. *Proteomics* 15, 1544–1563. doi: 10.1002/pmic.201400179

Hofer, A. M., and Brown, E. M. (2003). Extracellular calcium sensing and signalling. *Nat. Rev. Mol. Cell Biol.* 4, 530–538. doi: 10.1038/nrm1154

Jeong, J. S., Kim, Y. S., Baek, K. H., Jung, H., Ha, S. H., Kim, M., et al. (2010). Root-specific expression of *OsNAC10* improves drought tolerance and grain yield in rice under field drought conditions. *Plant Physiol.* 153, 185–197. doi: 10.1104/pp.110.154773

Jeong, J. S., Kim, Y. S., Redillas, M., Jang, G., Jung, H., Bang, S. W., et al. (2013). *OsNAC5* overexpression enlarges root diameter in rice plants leading to enhanced drought tolerance and increased grain yield in the field. *Plant Biotechnol. J.* 11, 101–114. doi: 10.1111/pbi.12011

Jin, C., Li, K., Xu, X., Zhang, H., Chen, H., Chen, Y., et al. (2017). A novel NAC transcription factor, *PbeNAC1*, of *Pyrus betulifolia* confers cold and drought

tolerance via interacting with *PbeDREBs* and activating the expression of stress-responsive genes. *Front. Plant Sci.* 8:1049. doi: 10.3389/fpls.2017.01049

Kolukisaoglu, U., Weinl, S., Blazevic, D., Batistic, O., and Kudla, J. (2004). Calcium sensors and their interacting protein kinases: genomics of the *Arabidopsis* and rice CBL-CIPK signaling networks. *Plant Physiol.* 134, 43–58. doi: 10.1104/pp.103.033068

Kumar, R., Kumar, A., Subba, P., Gayali, S., Barua, P., Chakraborty, S., et al. (2014). Nuclear phosphoproteome of developing chickpea seedlings (*Cicer arietinum* L.) and protein-kinase interaction network. *J. Proteomics* 105, 58–73. doi: 10.1016/j.jprot.2014.04.002

Li, F., Xu, X., Xu, Y., He, Z., and Wang, Z. (2016). Effect of high nitrogen on yield related traits and nitrogen utilization efficiency in Zhongmai 175 and Jingdong 17. *Acta Agron. Sin.* 42, 1853–1863. doi: 10.3724/sp.j.1006.2016.01853

Liu, H. L., Gao, Y. M., Wu, M., Shi, Y. N., Wang, H., Wu, L., et al. (2020). *TCP10*, a TCP transcription factor in moso bamboo (*Phyllostachys edulis*), confers drought tolerance to transgenic plants. *Environ. Exp. Bot.* 172:104002. doi: 10.1016/j.envexpbot.2020.104002

Liu, W., Zhang, Y., Gao, X., Wang, K., Wang, S., Zhang, Y., et al. (2012). Comparative proteome analysis of glutenin synthesis and accumulation in developing grains between superior and poor quality bread wheat cultivars. *J. Sci. Food Agric.* 92, 106–115. doi: 10.1002/jsfa.4548

Luo, M., Wang, Y., Liu, X., Yang, S., and Wu, K. (2012). HD2C interacts with HDA6 and is involved in ABA and salt stress response in *Arabidopsis*. *J. Exp. Bot.* 63, 3297–3306. doi: 10.1093/jxb/ers059

Lusser, A., Kölle, D., and Loidl, P. (1997). Histone acetylation: lessons from the plant kingdom. *Trends Plant Sci.* 6, 59–65. doi: 10.1016/s1360-1385(00)01839-2

Lyzenga, W. J., and Stone, S. L. (2012). Abiotic stress tolerance mediated by protein ubiquitination. *J. Exp. Bot.* 63, 599–616. doi: 10.1093/jxb/err310

Mao, X., Chen, S., Li, A., Zhai, C., and Jing, R. (2014). Novel NAC transcription factor *TaNAC67* confers enhanced multi-abiotic stress tolerances in *Arabidopsis*. *PLoS One* 9:e84359. doi: 10.1371/journal.pone.0084359

Mao, X. G., Zhang, H. Y., Qiang, X. Y., Li, A., Zhao, G. Y., and Jing, R. L. (2012). *TaNAC2*, a NAC-type wheat transcription factor conferring enhanced multiple abiotic stress tolerances in *Arabidopsis*. *J. Exp. Bot.* 63, 2933–2946.

Mar, M. T., and Pilar, C. (2010). TCP genes: a family snapshot ten years later. *Trends Plant Sci.* 15, 31–39. doi: 10.1016/j.tplants.2009.11.003

Mei, F., Chen, B., Li, F., Zhang, Y., and Mao, H. (2021). Overexpression of the wheat NAC transcription factor *TaSNAC4-3A* gene confers drought tolerance in transgenic *Arabidopsis*. *Plant Physiol. Biochem.* 160, 37–50. doi: 10.1016/j.plaphy.2021.01.004

Nozawa, A., Koizumi, N., and Sano, H. (2001). An *Arabidopsis* SNF1-related protein kinase, AtSR1, interacts with a calcium-binding protein, AtCBL2, of which transcripts respond to light. *Plant Cell Physiol.* 42, 976–981. doi: 10.1093/pcp/pce126

Palm, D., Simm, S., Darm, K., Weis, B. L., Ruprecht, M., Schleiff, E., et al. (2016). Proteome distribution between nucleoplasm and nucleolus and its relation to ribosome biogenesis in *Arabidopsis thaliana*. *RNA Biol.* 13, 441–454. doi: 10.1080/15476286.2016.1154252

Pandey, A., Chakraborty, S., Datta, A., and Chakraborty, N. (2008). Proteomics approach to identify dehydration responsive nuclear proteins from chickpea (*Cicerarietinum* L.). *Mol. Cell. Proteomics* 7, 88–107. doi: 10.1074/mcp.m700314-mcp200

Repetto, O., Rogniaux, H., Firnhaber, C., Zuber, H., Küster, H., Larré, C., et al. (2008). Exploring the nuclear proteome of *Medicagotruncatula* at the switch towards seed filling. *Plant J.* 56, 398–410.

Ruijter, A. J., Gennip, A. H., Caron, H. N., Kemp, S., and Kuilenburg, A. B. (2003). Histone deacetylases (HDACs): characterization of the classical HDAC family. *Biochem. J.* 370, 737–749. doi: 10.1042/bj20021321

Rushton, D. L., Tripathi, P., Rabara, R. C., Lin, J., Ringler, P., Boken, A. K., et al. (2012). WRKY transcription factors: key components in abscisic acid signalling. *Plant Biotechnol. J.* 10, 2–11. doi: 10.1111/j.1467-7652.2011.00634.x

Ryu, M. Y., Cho, S. K., and Kim, W. T. (2010). The *Arabidopsis* C3H2C3-Type RING E3 ubiquitin ligase atairp1 is a positive regulator of an abscisic acid-dependent response to drought stress. *Plant Physiol.* 154, 1983–1997. doi: 10.1104/pp.110.164749

Shen, H., Liu, C., Zhang, Y., Meng, X., Zhou, X., Chu, C., et al. (2012). OsWRKY30 is activated by MAP kinases to confer drought tolerance in rice. *Plant Mol. Biol.* 80, 241–253. doi: 10.1007/s11103-012-9941-y

Shen, J., Lv, B., Luo, L., He, J., Cao, C., Xi, D., et al. (2017). Corrigendum: the NAC-type transcription factor *OsNAC2* regulates ABA-dependent genes and abiotic stress tolerance in rice. *Sci. Rep.* 7:46890.

Shewry, P. R. (2009). Wheat. *J. Exp. Bot.* 60, 1537–1553.

Shu, K., and Yang, W. (2017). E3 ubiquitin ligases: ubiquitous actors in plant development and abiotic stress responses. *Plant Cell Physiol.* 58, 1461–1476. doi: 10.1093/pcp/pcx071

Song, J. P., Henry, H. A. L., and Tian, L. N. (2019). Brachypodium histone deacetylase BdHD1 positively regulates ABA and drought stress responses. *Plant Sci.* 283, 355–365. doi: 10.1016/j.plantsci.2019.03.009

Taj, G., Agarwal, P., Grant, M., and Kumar, A. (2011). Co-expression and in-silico interaction studies for inter-linking the activation of *MAPK3* and *LOX* genes during pathogenesis of *Alternaria brassicae* in *Brassica juncea*. *J Oilseeds Res.* 2, 13–20.

Takeda, T., Amano, K., Ohto, M.-A., Nakamura, K., Sato, S., Kato, T., et al. (2006). RNA interference of the *Arabidopsis* putative transcription factor *TCP16* gene results in abortion of early pollen development. *Plant Mol. Biol.* 61, 165–177. doi: 10.1007/s11103-006-6265-9

Tran, L. S., Nakashima, K., Sakuma, Y., Simpson, S. D., Fujita, Y., Maruyama, K., et al. (2004). Isolation and functional analysis of *Arabidopsis* stress-inducible NAC transcription factors that bind to a drought-responsive cis-element in the early responsive to dehydration stress 1 promoter. *Plant Cell* 16, 2481–2498. doi: 10.1105/tpc.104.022699

Vierstra, R. D. (2009). The ubiquitin-26S proteasome system at the nexus of plant biology. *Nat. Rev. Mol. Cell Biol.* 10, 385–397. doi: 10.1038/nrm2688

Wan, X. R., Mo, A. Q., Liu, S., Yang, L. X., and Li, L. (2011). Constitutive expression of a peanut ubiquitin-conjugating enzyme gene in *Arabidopsis* confers improved water-stress tolerance through regulation of stress-responsive gene expression. *J. Biosci. Bioeng.* 111, 478–484. doi: 10.1016/j.jbiosc.2010.11.021

Wang, S., Sun, X., Hoshino, Y., Yu, Y., Jia, B., Sun, Z., et al. (2014). MicroRNA319 positively regulates cold tolerance by targeting *OsPCF6* and *OsTCP21* in rice (*Oryza sativa* L.). *PLoS One* 9:e91357. doi: 10.1371/journal.pone.0091357

Wang, X., and Komatsu, S. (2016). Plant subcellular proteomics: application for exploring optimal cell function in soybean. *J. Proteomics* 143, 45–56. doi: 10.1016/j.jprot.2016.01.011

Wang, Y., Li, T., John, S. J., Chen, M., Chang, J., Yang, G., et al. (2018). A CBL-interacting protein kinase TaCIPK27 confers drought tolerance and exogenous ABA sensitivity in transgenic *Arabidopsis*. *Plant Physiol. Biochem.* 123, 103–113. doi: 10.1016/j.plaphy.2017.11.019

Xiang, Y., Huang, Y., and Xiong, L. (2007). Characterization of stress-responsive *CIPK* genes in rice for stress tolerance improvement. *Plant Physiol.* 144, 1416–1428. doi: 10.1104/pp.107.101295

Xie, Z., Wang, C., Wang, K., Wang, S., Li, X., Zhang, Z., et al. (2010). Molecular characterization of the celiac disease epitope domains in α-gliadin genes in *Aegilops tauschii* and hexaploid wheats (*Triticum aestivum* L.). *Theor. Appl. Genet.* 121, 1239–1251. doi: 10.1007/s00122-010-1384-8

Xu, Z., Gongbuzhaxi, Wang, C., Xue, F., Zhang, H., and Ji, W. (2015). Wheat NAC transcription factor *TaNAC29* is involved in response to salt stress. *Plant Physiol. Biochem.* 96, 356–363. doi: 10.1016/j.plaphy.2015.08.013

Yan, Y. M., Jiang, Y., An, X. L., Pei, Y. H., Li, X. H., Zhang, Y. Z., et al. (2009). Cloning, expression and functional analysis of HMW glutenin subunit *1By8* gene from Italy pasta wheat (*Triticum turgidum* L. ssp. durum). *J. Cereal Sci.* 50, 398–406. doi: 10.1016/j.jcs.2009.08.004

Yang, J. C., Zhang, J. H., Wang, Z. Q., Xu, G. W., and Zhu, Q. S. (2004). Activities of key enzymes in sucrose-to-starch conversion in wheat grains subjected to water deficit during grain filling. *Plant Physiol.* 135, 1621–1629.

Yin, X., and Komatsu, S. (2015). Quantitative proteomics of nuclear phosphoproteins in the root tip of soybean during the initial stages of flooding stress. *J. Proteomics* 119, 183–195. doi: 10.1016/j.jprot.2015.02.004

Yu, Y., Zhu, D., Ma, C. Y., Cao, H., Wang, Y., Xu, Y., et al. (2016). Transcriptome analysis reveals key differentially expressed genes involved in wheat grain development. *Crop J* 4, 92–106. doi: 10.1016/j.cj.2016.01.006

Zhang, J. W., Liu, D. M., Zhu, D., Liu, N. N., and Yan, Y. M. (2021). Endoplasmic reticulum subproteome analysis reveals underlying defense mechanisms of wheat seedling leaves under salt stress. *Int. J. Mol. Sci.* 22:4840. doi: 10.3390/ijms22094840

Zhang, M., Lv, D., Ge, P., Bian, Y., Chen, G., Zhu, G., et al. (2014a). Phosphoproteome analysis reveals new drought response and defense mechanisms of seedling leaves in bread wheat (*Triticum aestivum* L.). *J. Proteomics* 109, 290–308. doi: 10.1016/j.jprot.2014.07.010

Zhang, M., Ma, C. Y., Lv, D. W., Zhen, S. M., Li, X. H., and Yan, Y. M. (2014b). Comparative phosphoproteome analysis of the developing grains in bread wheat (*Triticum aestivum* L.) under well-watered and water deficit conditions. *J. Proteome Res.* 13, 4281–4297.

Zhang, Y., Yang, C. W., Li, Y., Zheng, N. Y., Chen, H., Zhao, Q. Z., et al. (2007). SDIR1 is a RING finger E3 ligase that positively regulates stress responsive abscisic acid signaling in *Arabidopsis*. *Plant Cell* 19, 1912–1929. doi: 10.1105/tpc.106.048488

Zhou, G. A., Chang, R. Z., and Qiu, L. J. (2010). Overexpression of soybean ubiquitin-conjugating enzyme gene *GmUBC2* confers enhanced drought and salt tolerance through modulating abiotic stress-responsive gene expression in *Arabidopsis*. *Plant Mol. Biol.* 72, 357–367. doi: 10.1007/s11103-009-9575-x

Zhu, D., Luo, F., Zou, R., Liu, J. X., and Yan, Y. M. (2021). Integrated physiological and chloroplast proteome analysis of wheat seedling leaves under salt and osmotic stresses. *J. Proteomics* 234, 104097. doi: 10.1016/j.jprot.2020.104097

Zhu, D., Zhu, G., Zhang, Z., Wang, Z., Yan, X., and Yan, Y. (2020). Effects of independent and combined water-deficit and high-nitrogen treatments on flag leaf proteomes during wheat grain development. *Int. J. Mol. Sci.* 21:2098. doi: 10.3390/ijms21062098

Zhu, G. R., Chen, G. X., Zhu, J. T., Zhu, Y., Lu, X. B., Li, X. H., et al. (2015). Molecular characterization and expression profiling of NAC transcription factors in *Brachypodium distachyon* L. *PLoS One* 10:e0139794. doi: 10.1371/journal.pone.0139794

Zhu, G. R., Yan, X., Zhu, D., Deng, X., Wu, J. S., Xia, J., et al. (2018). Lysine acetylproteome profiling under water deficit reveals key acetylated proteins involved in wheat grain development and starch biosynthesis. *J. Proteomics* 185, 8–24. doi: 10.1016/j.jprot.2018.06.019

Zhu, J. K. (2002). Salt and drought stress signal transduction in plants. *Annu. Rev. Plant Biol.* 53, 247–273. doi: 10.1146/annurev.arplant.53.091401.143329

Zou, R., Wu, J., Wang, R., and Yan, Y. (2020). Grain proteomic analysis reveals central stress responsive proteins caused by wheat-*Haynaldiavillosa* 6VS/6AL translocation. *J. Integr. Agric.* 19, 2–16.

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

frontiers in Plant Science

ORIGINAL RESEARCH
published: 17 December 2021
doi: 10.3389/fpls.2021.792057

TMT-Based Quantitative Proteomic Analysis Reveals the Physiological Regulatory Networks of Embryo Dehydration Protection in Lotus (*Nelumbo nucifera*)

Di Zhang[1]*, *Tao Liu*[1], *Jiangyuan Sheng*[1], *Shan Lv*[1] and *Li Ren*[2]*

[1] School of Design, Shanghai Jiao Tong University, Shanghai, China, [2] Institute for Agri-Food Standards and Testing Technology, Shanghai Academy of Agricultural Sciences, Shanghai, China

Lotus is an aquatic plant that is sensitive to water loss, but its seeds are longevous after seed embryo dehydration and maturation. The great difference between the responses of vegetative organs and seeds to dehydration is related to the special protective mechanism in embryos. In this study, tandem mass tags (TMT)-labeled proteomics and parallel reaction monitoring (PRM) technologies were used to obtain novel insights into the physiological regulatory networks during lotus seed dehydration process. Totally, 60,266 secondary spectra and 32,093 unique peptides were detected. A total of 5,477 proteins and 815 differentially expressed proteins (DEPs) were identified based on TMT data. Of these, 582 DEPs were continuously downregulated and 228 proteins were significantly up-regulated during the whole dehydration process. Bioinformatics and protein-protein interaction network analyses indicated that carbohydrate metabolism (including glycolysis/gluconeogenesis, galactose, starch and sucrose metabolism, pentose phosphate pathway, and cell wall organization), protein processing in ER, DNA repair, and antioxidative events had positive responses to lotus embryo dehydration. On the contrary, energy metabolism (metabolic pathway, photosynthesis, pyruvate metabolism, fatty acid biosynthesis) and secondary metabolism (terpenoid backbone, steroid, flavonoid biosynthesis) gradually become static status during lotus embryo water loss and maturation. Furthermore, non-enzymatic antioxidants and pentose phosphate pathway play major roles in antioxidant protection during dehydration process in lotus embryo. Abscisic acid (ABA) signaling and the accumulation of oligosaccharides, late embryogenesis abundant proteins, and heat shock proteins may be the key factors to ensure the continuous dehydration and storage tolerance of lotus seed embryo. Stress physiology detection showed that H_2O_2 was the main reactive oxygen species (ROS) component inducing oxidative stress damage, and glutathione and vitamin E acted as the major antioxidant to maintain the REDOX balance of lotus embryo during the dehydration process. These results provide new insights to reveal the physiological regulatory networks of the protective mechanism of embryo dehydration in lotus.

Keywords: ***Nelumbo nucifera*, seed embryo, dehydration, desiccation tolerance, abiotic stress, TMT proteomics, parallel reaction monitoring**

OPEN ACCESS

Edited by:
Pingfang Yang,
Hubei University, China

Reviewed by:
Jinwei Suo,
Zhejiang Agriculture and Forestry University, China
Arindam Ghatak,
University of Vienna, Austria

***Correspondence:**
Di Zhang
zhangdi2013@sjtu.edu.cn
Li Ren
renliaqx@163.com

Specialty section:
This article was submitted to Plant Proteomics and Protein Structural Biology, a section of the journal Frontiers in Plant Science

Received: *09 October 2021*
Accepted: *19 November 2021*
Published: *17 December 2021*

Citation:
Zhang D, Liu T, Sheng J, Lv S and Ren L (2021) TMT-Based Quantitative Proteomic Analysis Reveals the Physiological Regulatory Networks of Embryo Dehydration Protection in Lotus (Nelumbo nucifera). Front. Plant Sci. 12:792057. doi: 10.3389/fpls.2021.792057

INTRODUCTION

Lotus (*Nelumbo nucifera* Gaertn.), an ancient aquatic eudicot, belongs to the family Nelumbonaceae, which comprises only two species, *N. nucifera* and *N. lutea*, named Chinese lotus and American lotus, respectively (Shen-Miller, 2002; Shen-Miller et al., 2013). Lotus is among the top ten famous flowers in China, and it has been cultivated in China as vegetables and medicinal plants for over 7,000 years (Guo, 2009). Lotus has adapted to aquatic environments and plays an important role in wetland preservation and ecological restoration worldwide (Ming et al., 2013). The seed plays an important role in plant life cycle. It stores genetic information and nutrients to guarantee the reproduction of the next generation (Wang et al., 2016). Lotus seeds are black, hard, and ovoid. In its late immature stages, the seed is covered in a soft green husk containing a moist and soft endosperm and the developing embryo. When the seed reaches maturity, the husk turns dark brown and hardens, and both the endosperm and embryo become considerably dry (Moro et al., 2015). Lotus seeds have the characteristics of longevity and extremely durable storage, known as the "millennium ancient lotus seed." At the beginning of the twentieth century, some ancient lotus seeds were excavated from peat layer in Liaoning province of China, which can still germinate and grow normally. According to ^{14}C isotope detection, these seeds had a lifespan of more than 1,300 years, being one of the longest-lived seeds (Shen-Miller, 2002). The great difference between the responses of vegetative organs and seeds to dehydration is related to the special protective mechanism in embryos.

Seed longevity varies greatly between different plants, which can be as short as a few hours or as long as thousands of years (Ming et al., 2013). The difference in seed life span depends on genetic and environmental factors. Seed coat and pericarp have sealing function and play an important role in protecting seed vigor; however, lotus seeds removed pericarp can maintain 100% germination rate and develop into normal seedlings after 100°C treatment for 24 h (Huang et al., 2003), suggesting that the special anti-stress and protection mechanism of lotus embryo is an important internal cause to ensure the longevity of lotus seed. Lotus seeds mainly accumulate starch, which accounts for about 60% of its total dry weight. It also accumulates about 8% protein in immature seeds and as high as 24% in the mature desiccated seeds (Wang et al., 2016). Metabolomic and proteomic profiles revealed a highly significant metabolic switch at 15 days after pollination (DAP) going through a transition of metabolically highly active tissue to the preparation of storage tissue (Wang et al., 2016). Seed life span is a complex biological process from quantitative change to qualitative change. Seed formation is classically divided into two major phases: early embryogenesis during which the pattern and morphology of the embryo are established, and late embryogenesis during which seeds accumulate storage products, acquire desiccation tolerance, and finally fall in dormancy (Delseny et al., 2001). A series of complex and multifaceted responses are involved in desiccation tolerance in plants, including the structural or component alteration of cell wall, organelles, or organs, induction of the repair system, removal of free radicals, and accumulation of macromolecules. During seed desiccation, the accumulation of macromolecules such as oligosaccharides and proteins greatly increases cytoplasmic viscosity and usually causes the formation of bioglasses. Thus, bioglasses have been suggested to provide intracellular protection against the denaturation of large molecules to stabilize plasma membranes (Burke, 1986; Shih et al., 2008). On the molecular level, a number of mechanisms influencing seed survival in the dry state have been discovered in many plants. These mechanisms include the synthesis of protective molecules, such as non-reducing sugars including raffinose, stachyose, and cyclitols (Hoekstra et al., 2001; Verdier et al., 2013), late embryogenesis abundant (LEA) proteins (Hundertmark et al., 2011; Chatelain et al., 2012), heat shock proteins (HSPs) (Prieto-Dapena et al., 2006), various other stress proteins (Sugliani et al., 2009), and a set of antioxidant including glutathione (Kranner et al., 2006), tocopherols (Sattler et al., 2004), and flavonoids (Debeaujon et al., 2000). Interestingly, the phytohormone abscisic acid (ABA) is a central regulator of plant development and responses to environmental stresses. ABA controls seed developmental processes, including accumulation of food reserves, the acquisition of dormancy, and desiccation tolerance (Dekkers et al., 2015).

The genome of lotus has been sequenced and annotated recently (Ming et al., 2013; Wang et al., 2013), which has provided ample information for genomics and proteomics research of this aquatic species. Previous studies on lotus seed were mainly focused on the identification of its nutritional constituents, medicinal components, and biosynthesis regulation (Mukherjee et al., 2010; Wang et al., 2016). However, there are few reports on the systematic identification and characterization of dehydration tolerance and dehydration stress-related proteins in lotus seeds. Dehydration tolerance research has diverse applications, such as improving drought tolerance of crop species, optimizing germplasm resources, and stabilizing biomolecules and eukaryotic cells. In this study, TMT-labeled proteomics and PRM technologies were used to obtain novel insights into the biological network that regulates lotus seed dehydration tolerance and dehydration stress and laid a theoretical foundation for revealing the mystery of lotus seed longevity.

MATERIALS AND METHODS

Plant Materials

Lotus (*N. nucifera* "TaiKong 36") seeds were obtained from a cultivation pond under natural conditions in Shanghai Jiao Tong University, Shanghai, during the summer of 2018. Lotus plants are manually pollinated every morning and recorded the time from June to July. The developing seeds were collected at 15, 18, 21, 24, 27, 30, 40, and 50 DAP. The seed coat and cotyledon tissue were removed carefully by sharp shears, and 0.2 g of lotus embryos were collected and stored in a centrifugal tube. The collected lotus embryo samples were immediately frozen in liquid nitrogen and stored at −80°C for further proteomic analysis and physiological measurements.

Detection of Physiological Indices

For relative water content (RWC) detection, the embryo at different developmental stages were weighted for fresh weight (FW) and then dried at 70°C for constant weight (DW). The RWC was calculated using the following formula: RWC = (FW–DW)/FW * 100%. Five biological replicates per seed developmental stage were detected. A DDS-307 conductivity meter (INESA, Shanghai, China) was used to determine the relative electrical conductivity (REC) as described by Yang et al. (2019). Soluble sugar and soluble protein content were used with anthracene ketone-sulfuric acid and Coomassie bright blue (G250) method to detect the absorbance at 620 and 595 nm, respectively. Catalase (CAT), superoxide dismutase (SOD) and peroxidase (POD) activities, superoxide anion (O_2^-) inhibition activity, H_2O_2 content, hydroxyl radical (OH·) inhibition activity, ascorbic acid (AsA), glutathione (GSH), tocopherol (vitamin E, V_E), and malondialdehyde (MDA) contents were detected using biological assay kits (Nanjing Jiancheng Bioengineering Institute, China) following the manufacturer's instructions. ABA concentration was measured using a phytohormone ELISA Kit (Shanghai Enzyme-Linked Biotechnology Co., Ltd.) following the manufacturer's protocol. Three biological replicates were detected.

Protein Extraction

The protein extraction was performed according to Bo's et al. (2020) method with minor modifications. Frozen lotus embryos were mixed the steel beads and ground at the power of 60 Hz for 2 min. The samples were then supplemented with 1 ml of extraction buffer and mixed with Tris-phenol buffer for 30 min at 4°C. Next, the mixtures were centrifuged at 7,100 *g* for 10 min at 4°C and the phenol supernatants were collected. The supernatants were added to 5 times the volumes of 0.1 M cold ammonium acetate-methanol buffer and precipitated at −20°C overnight. After precipitation, the samples were centrifuged at 12,000 *g* for 10 min and the precipitation was collected. Then, the precipitation was washed by cold methanol and centrifuged at 12,000 *g* for 10 min at 4°C and the precipitation was collected. Methanol was then replaced by acetone and the wash stem was repeated twice to remove methanol contamination. Afterward, the samples were centrifuged at 12,000 *g* for 10 min at 4°C to collect precipitation and dried at room temperature for 3 min and then dissolved in lysis buffer for 3 h. Finally, the samples were centrifuged at 12,000 *g* for 10 min to collect supernatants. The supernatants were centrifuged again to remove precipitation completely. Protein concentration was determined by a BCA assay kit (Sangon Biotech, Shanghai, China) and aliquoted to store at −80°C.

Protein Digestion and TMT Labeling

Protein digestion was performed according to the filter-aided sample preparation (FASP) method (Wiśniewski et al., 2009). Briefly, the proteins were digested with trypsin (enzymes/substrate ratio 1:50) overnight after reductively alkylation with 10 mM dithiothreitol (DTT) (1 h, 60°C) and 50 mM iodoacetamide (IAA) (40 min, room temperature). The digested peptides were lyophilized and resolved in 200 mM triethylamine buffer (TEAB). Finally, the lyophilized samples were labeled using the 10 plex TMT reagent kit (Thermo scientific, USA) following the manufacturer's instructions. Briefly, the tryptic peptides (100 μg each) were labeled with TMT 10-plex with 126-tag (21d-1), 127N-tag (21d-2), 127C-tag (21d-3), 128N-tag (27d-1), 128C-tag (27d-2), 129N-tag (27d-3), 129C-tag (40d-1), 130N-tag (40d-2), 130C-tag (40d-3), and 131-tag (MIX). Proteomics platform was provided by Shanghai Luming Biotech. Co., Ltd. (Shanghai, China).

Mass Spectrometry Analyses

Mass Spectrometry (MS) detection was carried out as previously described (Gong et al., 2021). Reversed phase (RP) separation was performed on a 1,100 HPLC System (Agilent) using an Agilent Zorbax Extend RP column (5 μm, 150 × 2.1 mm). Mobile phases A (2% acetonitrile) and B (98% acetonitrile) were used for RP gradient. The solvent gradient was set as follows: 0–8 min, 98% A; 8.00–8.01 min, 98–95% A; 8.01–48 min, 95–75% A; 60–60.01 min, 60–10% A; 60.01–70 min, 10% A; 70–70.01 min, 10–98% A; 70.01–75 min, 98% A. Tryptic peptides were separated at a fluent flow rate of 300 μl/min and monitored at 210 and 280 nm. Dried samples were harvested from 8 to 50 min and elution buffer was collected every minute. The separated peptides were lyophilized for MS detection.

Mass spectrometry analyses were performed by a Q-Exactive mass spectrometer (Thermo, USA) equipped with a Nanospray Flex source (Thermo, USA). Samples were loaded and separated by a C18 column (15 cm × 75 μm) on an EASY-nLC^{TM} 1,200 system (Thermo, USA). The flow rate was 300 nl/min and the linear gradient was 60 min. Afterward, intact peptides were detected in the Orbitrap at a resolution of 70,000 with an automatic gain control (AGC) target of 1e6. The 10 most abundant precursor ions from the survey scan (300–1,600 *m/z*) were selected for higher-energy collisional dissociation (HCD) fragmentation at a normalized collision energy of 32. The resulting fragments were analyzed with the Orbitrap at a resolution of 35,000 with an AGC target of 2e5, maximum inject time of 80 ms, and dynamic exclusion duration of 30.0 s.

Protein Identification and Bioinformatics Analyses

The raw MS/MS data were processed using Proteome Discoverer (Version 2.2) software and tandem mass spectra were searched against the Uniprot-proteome_UP000189703-*Nelumbo nucifera* (Sacred lotus) (Strain cv. China Antique) database. Peptides with FDR <1% are considered to be effective. Credible proteins are selected based on the criteria of unique peptide ≥1 and Score Sequest >0, and the blank value is removed.

T-test was performed on three biological replicates in each group to calculate the statistical significance of all identified proteins between the comparison groups, and the average of the three biological replicates was used as the final protein expression rate. In this study, differentially expressed proteins (DEPs) were defined by the fold change criteria (FC) ≥ 1.5 or ≤0.67 ($p < 0.05$). Expression pattern clustering of all identified DEPs was performed using Genesis 1.8.1 software. Gene ontology (GO) function and Kyoto encyclopedia of genes

and genomes (KEGG) pathway enrichment annotation were analyzed using the Omicsbean platform (www.omicsbean.cn). The protein interaction regulatory network was constructed by using STRING website (https://string-db.org).

PRM Analysis

Parallel Reaction Monitoring (PRM) assay was performed as previously described with minor modifications (Bo et al., 2020; Gong et al., 2021). Eighteen key candidate DEPs were selected from the TMT data. PRM was used to quantify the expression, which was performed by Shanghai Ouyi Biological Co., Ltd. (Shanghai, China). The protein exaction was the same as the above procedure. The digested peptides were submitted to the PRM analysis. Three replicates for each sample were evaluated. The signature peptides of the target proteins were defined by the TMT dataset, and only the unique peptide sequence was selected for PRM analysis. In brief, the tryptic peptides were mixed in the solvent A and eluted in a reversed phase analytical column using the gradient solvent B (8–25% over 60 min, 25–45% over 20 min, 45–100% over 1 min, and 100% over the last 10 min) at a rate of 300 nl/min. The peptides measured were analyzed with a Q Exactive HF Orbitrap mass spectrometer (ThermoFisher Scientific, Waltham, MA, USA). The full MS scans (350–16,500 m/z) were obtained at a resolution of 1,200,000 using AGC of 3e6 and a highest injection time (MIT) of 100 ms. A data independent protocol (one MS scan followed by 20 MS/MS scans) was used for the MS/MS scans with the following parameters: resolution, 30,000; NEC, 27; AGC, 2E5; MIT, 80 ms; and insulation window, 1.4 m/z. The PRM data was analyzed using Skyline 3.6. The results were quantified for every peptide, and the DEPs detected were screened and compared with the MS data derived from the TMT.

Statistical Analysis

All experiments were repeated at least three times, and statistical analysis was performed using SAS 9.1.3 software and one-way ANOVA followed by least significant difference (LSD) multiple range tests ($p < 0.05$). The correlation analysis of oxidative physiological parameters under stresses was performed using SAS 9.1.3 software. $p < 0.05$ was considered significant.

RESULTS

Development and Dehydration Characteristics of Lotus Seed Embryos

The whole process of the development and maturation of lotus seeds needs about 40 d, during which the morphological and color characteristics of embryo and pericarp changed clearly (**Figure 1A**). In the first stage, the embryo mainly undergoes embryogenesis and development. The embryo reached morphological maturity during 18–21 DAP, and embryo and pericarp appeared bright green color, the fresh weight of embryo reached the maximum, and dry matter accumulated rapidly during this stage (**Figure 1**). In the second stage, lotus seeds enter the dehydrated mature process from 21 to 50 DAP, the embryo and pericarp gradually shrunk in shape, and the color gradually became dark brown (**Figure 1A**). The RWC (dropped from 73.7 to 21.0%) and fresh weight (dropped from 119.3 to 56.6 mg) of the embryo decreased dramatically, and dry matter accumulated slowly during 21–27 DAP (**Figure 1B**). The soluble sugar and soluble protein contents increased significantly during the rapid dehydration stage (18–27 DAP) (**Figure 1C**). Then, the RWC, fresh weight, dry weight, and soluble sugar and proteins of lotus embryos enter the steady phase during 30–50 DAP (**Figures 1B,C**). These results indicated that the lotus embryo reached morphological maturity, rapid dehydration, and dehydration maturity at 21, 27, and 40 DAP, respectively. Accordingly, embryos from these three time points were used for subsequent proteomic studies.

TMT Data and DEP Identification

Proteome analysis of fresh lotus embryos at the three dehydration stages (21, 27, and 40 DAP) was performed by TMT labeling technology. Totally, 60,266 secondary spectra and 32,093 unique peptides were detected. After removing repetitive proteins, 5,477 proteins were identified based on TMT data. The statistical information of identified proteins is shown in **Supplementary Table 1**, including the molecular weight, peptide sequence length, distribution of peptide number, and protein quantification. PCA analysis was performed to evaluate the test samples, and the results showed that the different groups of lotus embryos were well-distinguished, and three repeated samples of each group were aggregated together. The first two components of PCA accounted for more than 70 % of the total variance, separating the total proteins from the different groups (**Figure 2A**). A total of 815 protein abundance changed significantly between the three test groups. Of these, 198 (32%) and 198 (26%) proteins increased in abundance, while 420 (68%) and 558 (74%) proteins decreased in abundance in the 27 DAP and 40 DAP groups, respectively (**Figure 2B**). Only 30 proteins abundance changed significantly between 27 DAP and 40 DAP groups (**Figure 2C**). Quantitative and functional annotation information of all DEPs is shown in **Supplementary Table 2**.

Bioinformatics Analysis of the DEGs

The 815 DEPs identified from lotus embryos during seed dehydration maturity were conducted using hierarchical clustering and protein functional classification analysis (**Figure 3**). All DEPs were classified into 5 main expression profiles. Cluster A (contains 16 proteins) and cluster B (566 proteins) contained DEPs that were continuously downregulated, and cluster C (39 proteins) and cluster E (185 proteins) were gradually up-regulated during the seed dehydration process. Cluster D only contains 4 proteins that had a specific up-regulated expression at the rapid dehydration stage (27 DAP), including dehydrin RAB18 (A0A1U7ZRR9), digalactosyldiacylglycerol synthase (A0A1U7ZHL0), protein TORNADO (A0A1U7ZLU9), and an uncharacterized protein (A0A1U8B9X3). Protein functional classification analysis showed that 815 DEPs encoded proteins that were mainly associated with photosynthesis, lipid metabolism, amino acid

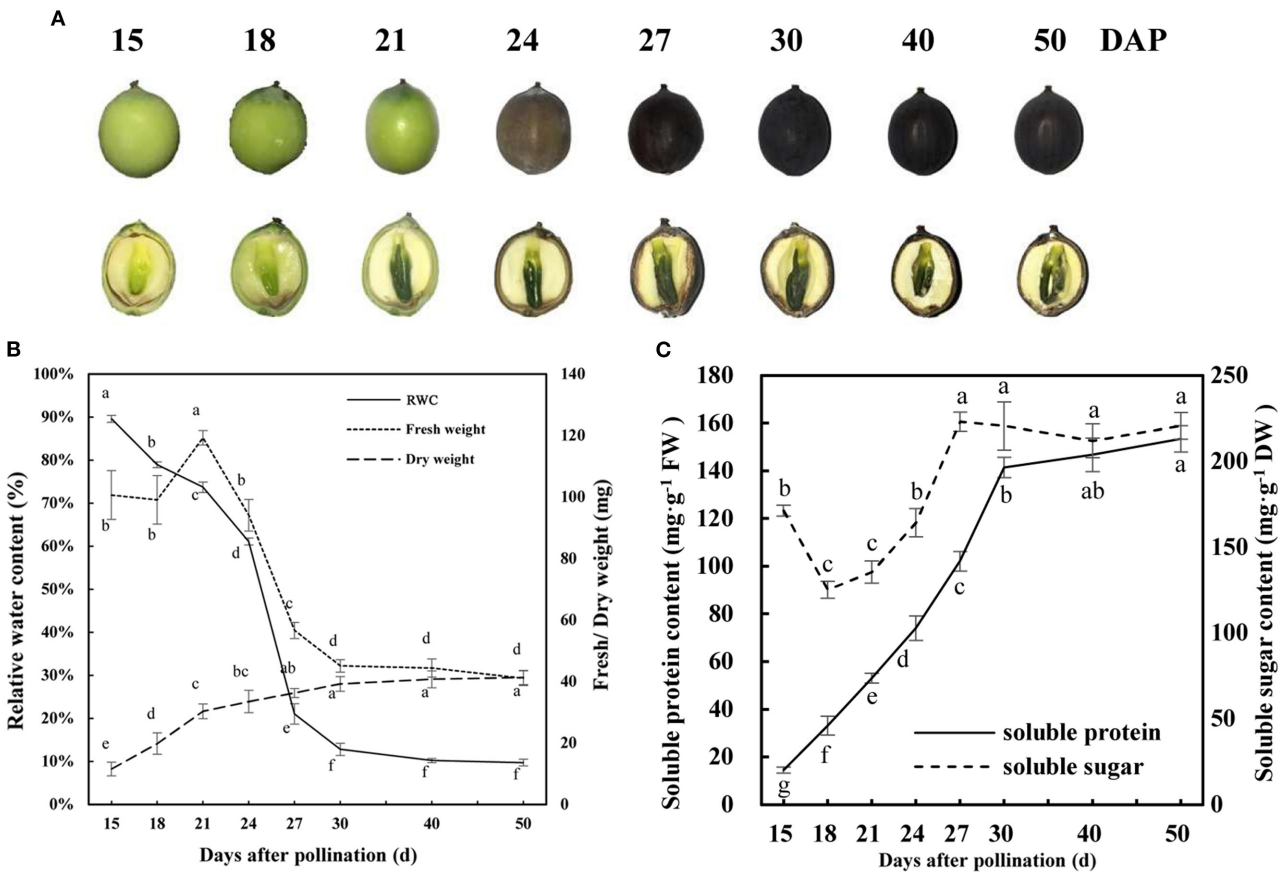


FIGURE 1 | Development and dehydration characteristics of lotus seed embryos. **(A)** Morphological traits of developing lotus seed; **(B)** Fresh/dry weight and relative water content of developing lotus seed embryo; **(C)** Soluble protein and sugar contents of developing lotus seed embryo. Lower-case letters indicate that the indexes have significant differences between different developmental stages; $p < 0.05$.

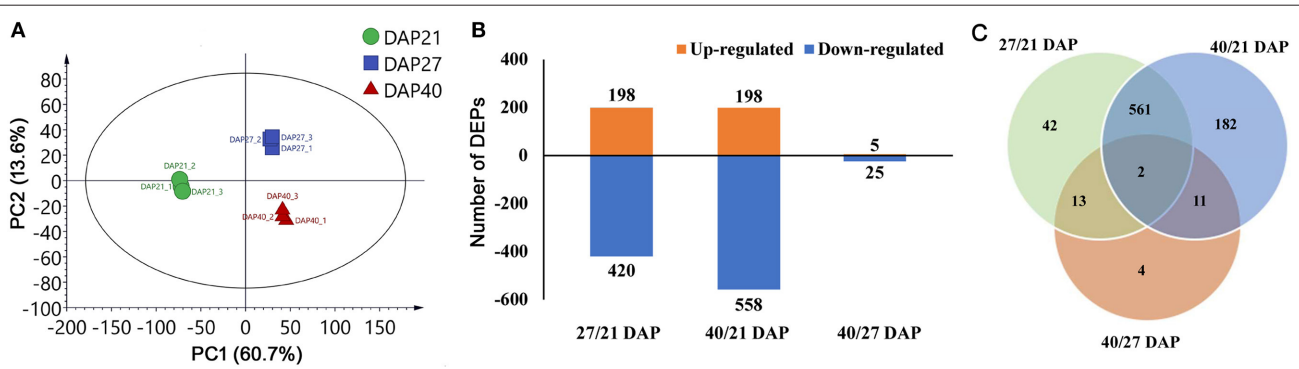


FIGURE 2 | Statistic of differentially expressed proteins (DEPs) of lotus embryos during seed dehydration maturity. **(A)** Principal components analysis of each sample; **(B)** Statistics of up and down-regulated DEPs; **(C)** Veen diagram analysis of DEPs among different groups.

metabolism, hormone metabolism, RNA, protein processing, cell wall, secondary metabolism, stress, and development (**Figure 3**).

Pathway and GO annotation and significant enrichment methods were used for understanding the seed embryo dehydration events and associated downstream biological processes regulated by the DEPs. According to the results of clustering analysis, the DEPs were divided into up-regulated expression groups (clusters C and E) and downregulated expression groups (clusters A and B), which contained 224 and 582 proteins, respectively. A total of 224 up-regulated DEPs

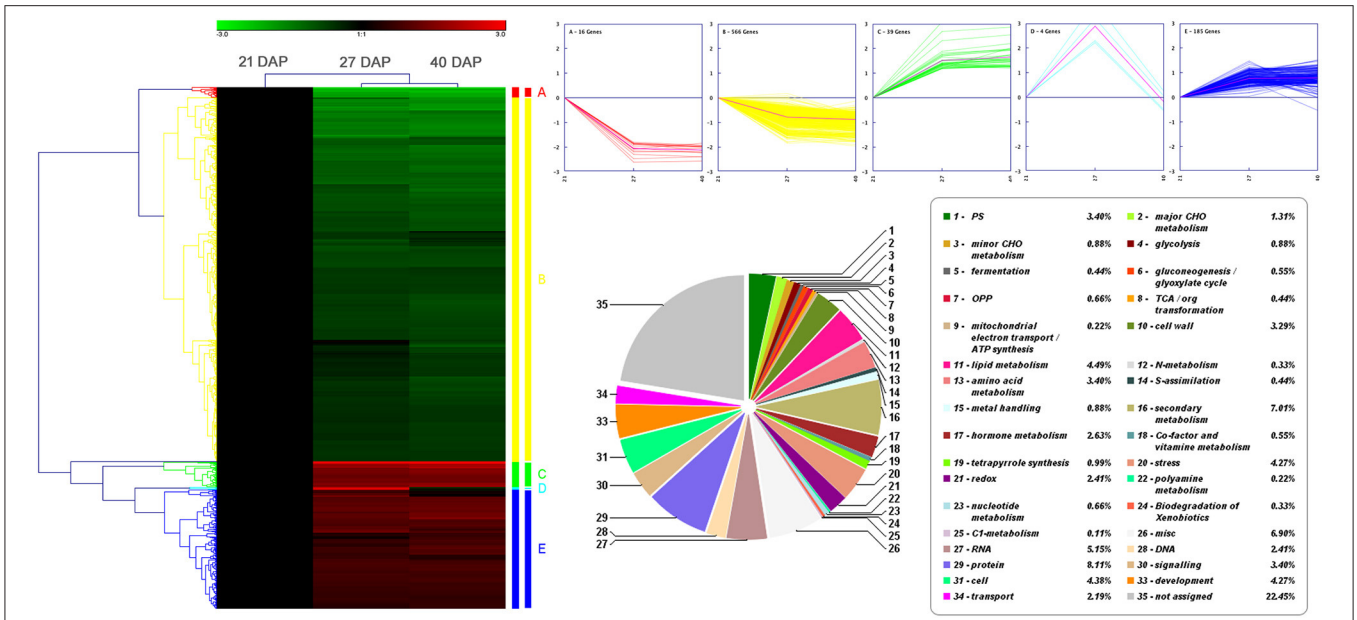


FIGURE 3 | Hierarchical clustering and protein functional classification analysis of 815 DEPs in lotus embryos during seed dehydration maturity.

were annotated to GO terms (**Supplementary Figure 1A**), and the main categories in biological process (BP) include response to stress (74, 37.2%), response to chemical (64, 32.2%), and response to abiotic stimulus (59, 29.6%); the prominent cell component (CC) categories were cell part (152, 76.4%) and intracellular part (135, 67.8%); and the most significant categories in molecular function (MF) were carbon-carbon lyase activity (6, 3.0%) and sulfur compound binding (4, 2.0%). A total of 582 downregulated DEPs were performed with GO annotation (**Supplementary Figure 1B**), and the main categories in BP were single-organism process (395, 75.7%) and single-organism cellular process (339, 64.8%); the prominent CC categories were cell part (461, 88.1%), intracellular part (422, 80.7%) and chloroplast; and the most significant categories in MF were protein binding (143, 27.3%) and nucleoside binding (72, 13.8%). Additionally, AgriGO significant enrichment analysis showed that up-regulated DEPs were significantly enriched in "carbohydrate metabolism process," "catalytic activity," and "RNA banding" GO terms, and downregulated DEPs were significantly enriched in "amino acid and derivative metabolic process," "carbohydrate metabolism process," "lipid metabolic process," "generation of precursor metabolites and energy," "photosynthesis," "cytoskeleton," "Golgi apparatus," "endoplasmic reticulum," and "peroxisome" GO terms (**Supplementary Figure 2**). KEGG pathway annotation analysis showed that the 815 DEPs were significantly enriched in "pyruvate metabolism," "glycolysis/gluconeogenesis," "pentose phosphate pathway," "fatty acid biosynthesis and metabolism," "biosynthesis of amino acids," "amino sugar and nucleotide sugar metabolism," "glutathione metabolism," "terpenoid backbone biosynthesis," "steroid biosynthesis," "flavonoid biosynthesis," "photosynthesis," "biosynthesis of secondary metabolites," and "metabolic pathway" (**Supplementary Figure 3**). Furthermore, the up-regulated DEPs were mainly enriched in "galactose metabolism," "pentose phosphate pathway," "glycolysis/gluconeogenesis," "protein processing in endoplasmic reticulum," and "glutathione metabolism;" and the downregulated DEPs were significantly enriched in "metabolism pathway," "biosynthesis of secondary metabolites," "pyruvate metabolism," "photosynthesis," and "fatty acid biosynthesis and metabolism" (**Supplementary Figure 4**). The above results suggested that environmental stimuli, stress response, energy metabolism, and protein processing in ER-related biological process and pathway are vigorous during lotus embryo continuous dehydration and maturation. Nevertheless, the cell metabolism, biosynthesis of secondary metabolites, photosynthesis, amino acids biosynthesis, and fatty acid metabolism are distinctly decreased or inhibited in this water loss process.

Protein-Protein Interaction Networks Analysis of DEPs

Protein-protein interaction (PPI) networks analysis were performed based on the DEPs during different seed dehydration stage, i.e., 618 DEPs from rapid dehydration stage (27/21 DAP) and 30 DEPs from terminal dehydration stage (40/27 DAP). In the rapid dehydration stage, up-regulated DEPs interact with each other and participate in "galactose metabolism," "glycolysis/gluconeogenesis," "protein processing in ER," "starch and sucrose metabolism," and "base excision repair" GO/KEGG terms, and dehydrin Rab18-like (A0A1U7ZRR9), STAR-1 Family DOT2 (A0A1U8A3P7), and digalactosyldiacylglycerol synthase (A0A1U7ZHL0) have the highest fold change of protein abundance in this network (**Figure 4**). PPI network

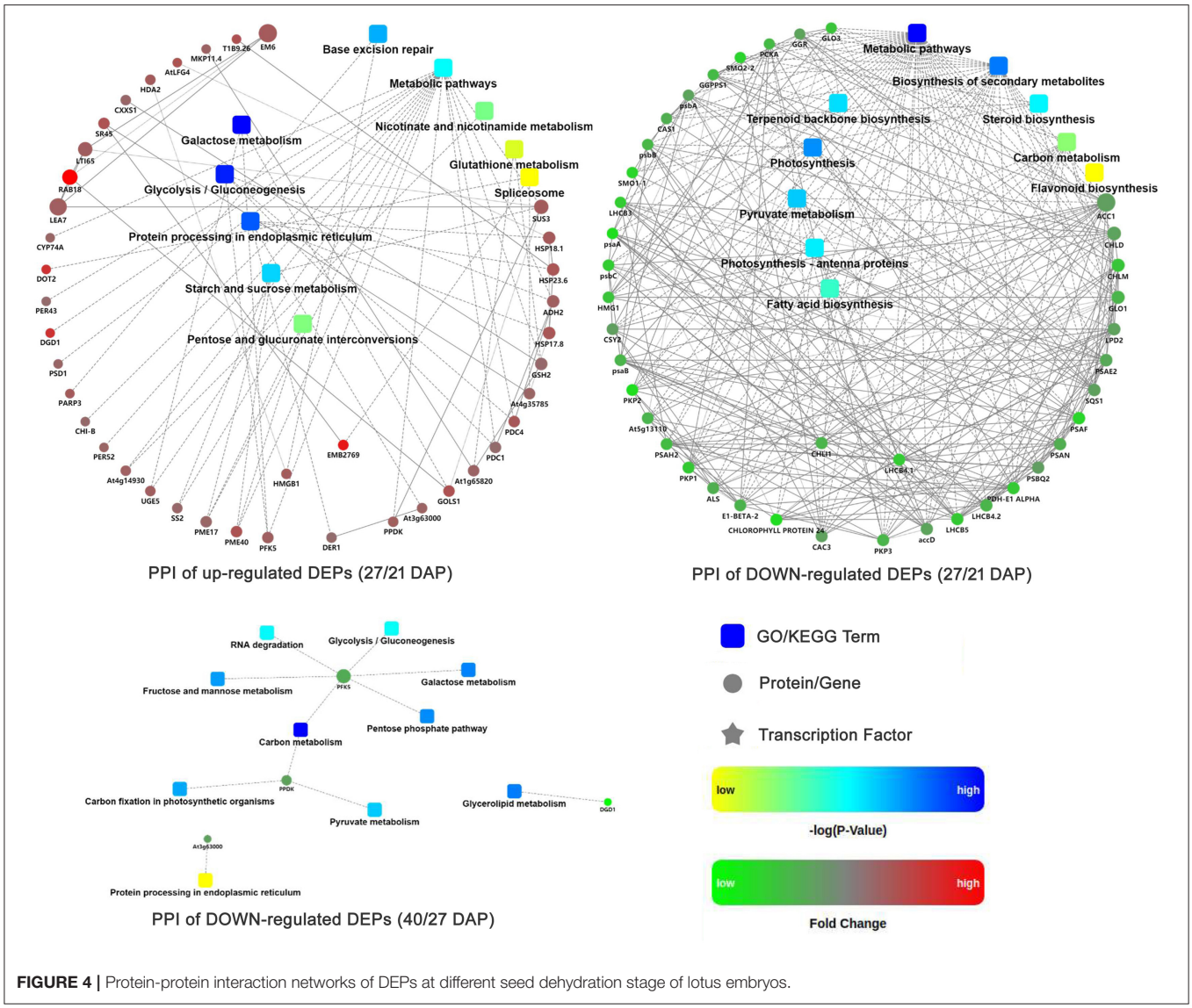


FIGURE 4 | Protein-protein interaction networks of DEPs at different seed dehydration stage of lotus embryos.

of downregulated DEPs (27/21 DAP) mainly include "metabolic pathway," "biosynthesis of secondary metabolites, "photosynthesis," "pyruvate metabolism," "terpenoid backbone biosynthesis," and "fatty acid biosynthesis" (**Figure 4**). In the terminal dehydration stage (40/27 DAP), only downregulated DEPs can construct PPI networks, and annotated proteins were involved in "carbon metabolism," "galactose metabolism," "pyruvate metabolism," "glycerolipid metabolism," and "protein processing in ER." Furthermore, according to the DEP Venn diagram of 27/21 DAP and 40/21 DAP, 563 (69.4%) proteins were overlapped in lotus embryos at 27 DAP and 40 DAP, while 55 (6.8%) and 193 (23.8%) DEPs were unique at 27 DAP and 40 DAP, respectively (**Figure 5A**). PPI networks calculated by the STRING database showed that 27 DAP-specific DEPs were involved in the metabolism pathway (**Figure 5B**) while 193 specific DEPs of the 40 DAP group were related to 6 biological processes including fatty acid metabolism, oxidation-reductive process, response to ABA, carbohydrate metabolism, protein modification and processing, and protein transport (**Figure 5C**). The overlapped DEPs were mainly involved in 12 biological processes, such as cell wall organization, carbohydrate metabolism, RNA binding, carbohydrate and fatty acid metabolism, and antioxidant activities (**Figure 5D**). These PPI networks results indicated that carbohydrate metabolism (including glycolysis/gluconeogenesis, galactose, starch and sucrose metabolism, pentose phosphate pathway, and cell wall organization), protein processing in ER, DNA repair, and antioxidative events have a positive response to lotus embryo dehydration and maturation, and ABA signal and dehydrin Rab18 protein may play an important role in this process. On the contrary, the energy metabolism (metabolic pathway, photosynthesis, pyruvate metabolism, and fatty acid biosynthesis) and secondary metabolism (including terpenoid backbone, steroid, and flavonoid biosynthesis) gradually become

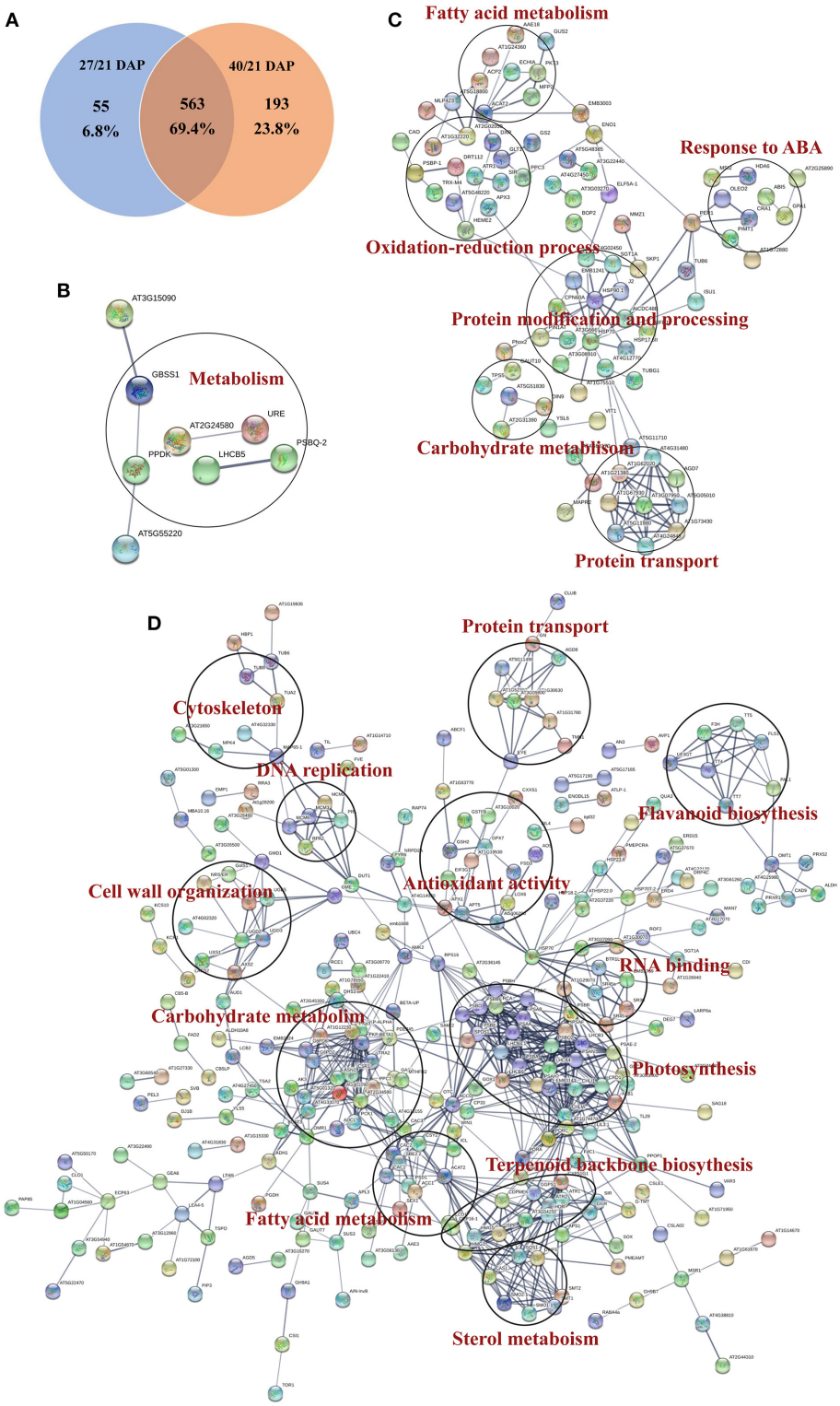


FIGURE 5 | Functional protein association networks analysis of DEGs from tandem mass tags (TMT) proteomics data. **(A)** Venn diagram; **(B)** Interaction network of common DEPs; **(C)** Interaction network of DEPs in 27 days after pollination (DAP); **(D)** Interaction network of specific DEPs in 40 DAP.

TABLE 1 | Comparison of the quantitation results between tandem mass tags (TMT) and parallel reaction monitoring (PRM).

Accession	Description	TMT result (FC)		PRM result (FC)	
		27/21	40/21	27/21	40/21
A0A1U8AFZ9	Pyruvate decarboxylase 2	2.286	2.416	4.480	6.323
A0A1U7ZDJ8	Glucose-6-Phosphate 1-Dehydrogenase	1.517	1.553	1.733	2.727
A0A1U8BJ10	Phosphoenolpyruvate carboxylase	0.650	0.548	0.382	0.455
A0A1U8BAL0	Citrate synthase	0.652	0.576	0.343	0.489
A0A1U8APR5	Sucrose synthase	2.069	2.302	1.783	5.099
A0A1U8B0W4	Acetyl-CoA carboxylase 1-like	0.652	0.636	0.218	0.345
A0A1U8A0H7	Glutathione S-transferase-like	1.553	1.570	ns	1.750
A0A1U8Q237	L-Ascorbate peroxidase 2	0.202	0.188	0.006	0.005
A0A1U7ZHC7	Thioredoxin-like protein CXXS1	1.639	1.908	1.883	3.239
A0A1U7ZLU2	Allene oxide synthase 1	1.669	1.684	1.756	4.629
A0A1U7ZBZ6	Protein-L-Isoaspartate O- methyltransferase	ns	1.508	ns	2.376
A0A1U7ZRR9	Dehydrin Rab18-like	14.646	ns	2.044	1.761
A0A1U8ACK5	Late embryogenesis abundant protein EMB564	2.173	2.241	2.435	3.244
A0A1U8BJB0	Late embryogenesis abundant protein D-34-like	2.230	2.304	3.856	5.512
A0A1U7ZRG9	Late embryogenesis abundant protein D-34	2.203	2.338	3.658	4.603
A0A1U7YX92	Cysteine proteinase inhibitor	2.215	2.383	3.246	4.488
A0A1U8Q4V5	Histone deacetylase 2	2.110	1.858	3.030	4.213
A0A1U7ZYD0	DNA damage repair protein	1.730	1.676	1.956	3.014

"ns" means there was no significant difference in protein abundance changes, and gray background indicates the results of PRM and TMT were different.

static status during lotus embryo dehydration and maturation. Based on the above bioinformatics analysis, a total of 92 key DEPs were selected and shown in **Supplementary Table 3**. Those proteins were involved in carbohydrate and energy metabolism, redox homeostasis, stress/defense, protein modification and degradation, response to ABA signaling, and DNA repair.

PRM Validation

To validate the TMT proteomic data, 18 candidate DEPs that were closely related to the dehydration protection event of lotus seed embryo were selected for PRM quantitative detection (**Supplementary Figure 5**). PRM identified 18 candidate DEPs which include pyruvate decarboxylase, glucose-6-phosphate dehydrogenase, phosphoenolpyruvate carboxylase, citrate synthase, sucrose synthase, acetyl-CoA carboxylase 1-like, glutathione S-transferase-like, L-ascorbate peroxidase, thioredoxin-like, allene oxide synthase, protein-L-isoaspartate O-methyltransferase, dehydrin Rab18-like, LEA EMB564, LEA D34-like, LEA D34, cysteine proteinase inhibitor, histone deacetylase, and DNA damage repair protein. These protein expression patterns showed similar trends between PRM and TMT quantitation results. However, the fold change value of protein abundance detected by PRM was greater than that of TMT data (**Table 1**). This result largely supports the reliability of the TMT data.

Stress Physiological Indicators Validation

Based on the bioinformatics analysis of TMT proteomics, stress physiological indicators were examined to verify protective mechanism of dehydration in lotus seed embryos. The change of REC and MDA contents was opposite to that of RWC as their levels increased rapidly during rapid dehydration stage (**Figures 6A,B**). ABA content increased slightly before the rapid dehydration and dehydration maturity stages (**Figure 6C**). The changing trend of the above indicators suggested that ABA as an upstream signal is involved in inducing dehydration process, and serious membrane lipid peroxidation and plasma membrane damage occurred during rapid dehydration stage in lotus embryo. O_2^- and OH· inhibition activities were significantly decreased and H_2O_2 levels were increased 7-fold during 18–27 DAP (**Figures 6D–F**), which indicated that excessive ROS components are produced during embryo dehydration process. Additionally, CAT activities were significantly increased at the early stage of dehydration and then continuously decreased. SOD and POD activities showed gradually declines in response to embryo water loss (**Figures 6G–I**). Meanwhile, GSH and V_E contents increased 7.3 and 4.5 times, respectively, from 15 to 27 DAP, and then their levels tend to stabilize. However, AsA, as a water-soluble antioxidant, was continuously decreased during embryo dehydration process (**Figures 6J–L**). Furthermore, the correlation analysis of stress physiological indicators showed that REC had a significant negative correlation to MDA ($p < 0.01$), POD ($p < 0.05$), H_2O_2 ($p < 0.01$), GSH ($p < 0.01$), and V_E ($p < 0.05$). REC and MDA content had a significant positive correlation to H_2O_2, GSH, and V_E (**Supplementary Table 3**). H_2O_2 levels showed a significant positive correlation to GSH and V_E, and had a negative correlation to AsA and O_2^- contents and OH· inhibition activities (**Supplementary Table 3**). These results indicated that H_2O_2 was the main ROS component inducing

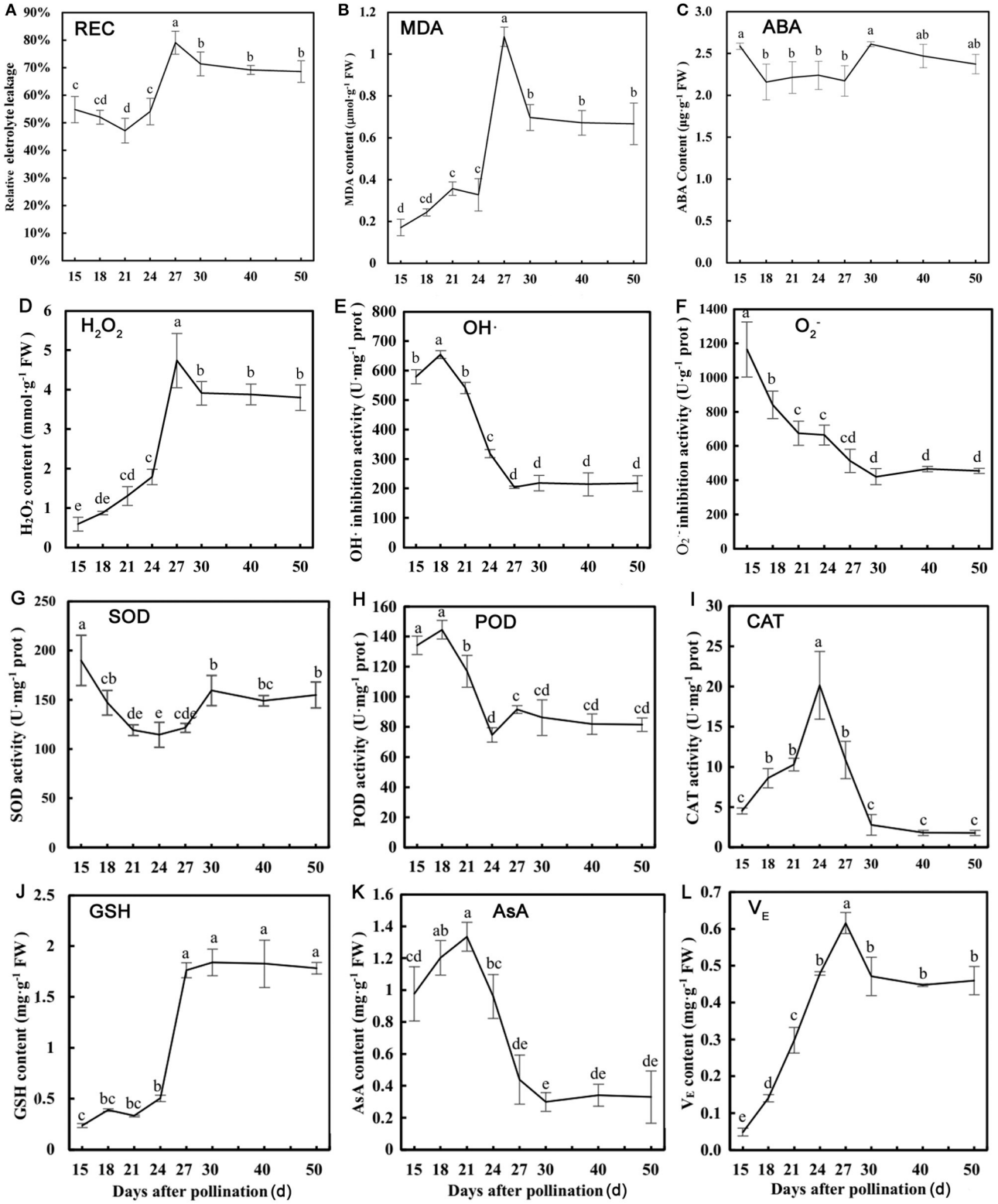


FIGURE 6 | Physiological parameters detection of lotus embryos at different seed dehydration stages. **(A)** Relative electrical conductivity (REC); **(B)** malondialdehyde (MDA) content; **(C)** abscisic acid (ABA) content; **(D)** H_2O_2 content; **(E)** hydroxyl radical (OH·) inhibition activity; **(F)** O_2^- inhibition activity; **(G)** superoxide dismutase (SOD) activity; **(H)** peroxidase (POD) activity; **(I)** catalase (CAT) activity; **(J)** glutathione (GSH) content; **(K)** ascorbic acid (AsA) content; **(L)** tocopherol (V_E) content. Lower-case letters represent the significant differences ($p < 0.05$, least significant difference test) between different dehydration stages of seed embryos.

oxidative stress damage, and GSH and V_E acts as the major antioxidant to maintain the REDOX balance of lotus embryo during the dehydration process.

DISCUSSION

Carbohydrate and Energy Metabolism During Lotus Seed Dehydration and Maturity

Glycolysis, tricarboxylic acid (TCA) cycle, and pentose phosphate pathway (PPP) are the three main pathways of carbohydrate and energy metabolism in plants. Glycolysis is a cytoplasmic pathway that breaks down glucose into two pyruvates and generates energy such as ATP and NADH. In aerobic conditions, pyruvate, through oxidative decarboxylation reaction, enters the TCA cycle that intermediates energy metabolism and participates in carbohydrates, amino acid, and fatty acids metabolism (Zhang et al., 2013). The PPP is different from the first two carbohydrate metabolism pathways, since it is no consumption and production of ATP during the reaction. The proportion of the PPP is relatively low in plant cells in normal environments, but increase when plants are under stress conditions, and the reduced state of glutathione and the redox balance of plant cells could be maintained by PPP (Kruger and von Schaewen, 2003; Scheibe, 2004; Scharte et al., 2009). The proteomics study of rice grains during development found that proteins involved in glycolysis and TCA cycle accumulated at a high level at the end of grain maturity (Lee and Koh, 2011). In wheat seeds, glycolysis and starch synthesis was significantly enhanced at the late stage of grain filling, which provided much energy for grain maturity (Zhang N. et al., 2015; Zhang et al., 2021). In this study, 16 DEPs in lotus embryo involved in the glycolysis pathway, of which 7 proteins (fructokinase, pyruvate decarboxylase, fructose-bisphosphate aldolase, etc.) were up-regulated and 9 proteins (isocitrate lyase, phosphoenolpyruvate carboxylase, phosphoenolpyruvate carboxykinase, etc.) were downregulated. However, 5 proteins involved in TCA cycle were all significantly downregulated (**Supplementary Figure 6**), and two significantly up-regulated proteins were found to participate in PPP, including glucose-6-phosphate dehydrogenase (G6PDH, A0A1U7ZDJ8) and 6-phosphate gluconolactonease (A0A1U7ZD76). G6PDH, as the key rate-limiting enzyme in PPP, can promote the regeneration of reduced glutathione and improve the cellular antioxidant capacity. Therefore, the carbohydrate and energy metabolism of lotus embryos decrease with dehydration and maturity, and PPP plays an important role in maintaining the redox homeostasis and alleviating stress damage of embryo cells in this process. Additionally, raffinose family oligosaccharides (RFOs) are a type of functional oligosaccharides that is unique to plants. Some studies have shown that RFOs usually accumulate during the late stages of seed development and are closely related to seed vigor, dehydration tolerance, and storage tolerance (Leprince et al., 2017). In this study, soluble sugar contents increased gradually during the rapid dehydration stage, and 4 DEPs related to RFOs synthesis pathway involved in raffinose and stachyose biosynthesis, including UDP-glucose 4-epimerase (A0A1U8APY8, A0A1U8Q8U3), inositol 3-α-galactosyltransferase (A0A1U8B9X9), sucrose synthetase (A0A1U8APR5), and UDP-glucose-6-dehydrogenase (A0A1U8A842). Those enzymes synergistically promote the synthesis of RFOs. Therefore, some oligosaccharides may accumulate in lotus embryo to resist dehydration osmotic stress.

Stress Response and Redox Homeostasis Status During Lotus Seed Dehydration Maturity

Cells usually suffer severe stress damage, lipid peroxidation, and macromolecules denaturation due to extreme dehydration during seed maturation and development. Biological macromolecules, including plasma membrane, enzymes, and nucleic acids, may be attacked by excessive ROS, which is the main factor causing seed aging and death. ROS is mainly eliminated by enzymatic or non-enzymatic antioxidants. In this study, stress physiological indicators indicated that H_2O_2 was the main ROS component accumulated in lotus embryos, and GSH and V_E act as the major antioxidants to resist oxidative stress during the seed dehydration process. Several antioxidant enzymes (SOD, APX, and GPX) were significantly downregulated during dehydration of lotus embryos. However, glutathione synthase (GS; A0A1U7ZUJ5) and two glutathione S-transferase (GST; A0A1U8A0H7, A0A1U7YZC7) and TRX (A0A1U7ZHC7) were significantly up-regulated in this process (**Supplementary Figure 5**). GS can promote glutathione production, and GST can catalyze GSH to chelate metal ions or combine with hydrophobic and electrophilic substrates to degrade the harmful substances of cells, helping cells resist oxidative damage (Zhang D. et al., 2015). TRX contains disulfide bonds with redox activity, which could regulate the redox state of cells through the reversible conversion of thiol-disulfide bonds. GST is positively related to the antioxidant capacity of plants and has been found in *Nicotiana tabacum*, rice, and *Arabidopsis* for its ability to resist oxidative stress (Gallé et al., 2009; Rezaei et al., 2013). It is speculated that GST may catalyze the chelation of GSH with metal ions to prevent metal ions from combining with ROS to turn into highly toxic hydroxyl radicals. The GSH pathway involving GS and GST actively responds to oxidative stress during the dehydration of lotus embryos. The above results further suggested that non-enzymatic antioxidants play a fundamental role in stress damage protection during water loss process in lotus embryos.

When undergoing desiccation, cells have to cope with significant changes in turgor pressure, which lead to cell shrinkage and mechanical damages (Scoffoni et al., 2014). ABA is found to be likely important in the acquisition of dehydration tolerance (DT) and drought tolerance during seed maturation and development (Umezawa et al., 2010; Hauser et al., 2011). ABA Insensitive (ABI) protein is a positive transcription factor in the ABA signaling pathway, which is involved in plant ABA signaling-mediated seed development or abiotic stress process (Gutierrez et al., 2007). In *Arabidopsis*, ABI5 can

regulate dormancy and germination of seeds (Lopez-Molina et al., 2001; Maia et al., 2014). ABI5 is the hub of the seed survival gene regulatory network under drought conditions in *Medicago truncatula* (Verdier et al., 2013). The function of ABI5 in the process of seed maturity was subsequently revealed in *Pisum sativum*, suggesting that ABI5, which is related to the synthesis and accumulation of raffinose family RFOs and LEA protein, is an important regulator of maturity development and longevity in legume seeds (Zinsmeister et al., 2016). In *Physcomitrella patens*, knockouts of the *ABI3* gene that regulate ABA signaling were shown to affect DT acquisition of the seed (Khandelwal et al., 2010). Some *abi3-5* mutant seeds failed to acquire DT and displayed a reduced longevity (Sugliani et al., 2009), while two *Mtabi3* mutants in *Medicago* led to a lower expression of LEA proteins, abnormal chlorophyll degradation in cotyledons, and defective in chromatin compaction and nuclear size reduction (Delahaie et al., 2013; Delmas et al., 2013; van Zanten et al., 2014). In lotus seed embryos, ABA contents were increased at 15 and 30 DAP, and three ABI5 proteins (A0A1U8B646, A0A1U7ZTI7, A0A1U7YYQ4) were significantly differently expressed at the late dehydration stage, indicating that ABA signaling can induce embryo dehydration and regulate downstream biological events in lotus. During seed drying, the accumulation of macromolecules such as oligosaccharides, LEA, or HSPs proteins greatly increases cytoplasmic viscosity and usually causes the formation of bioglasses. Thus, bioglasses have been suggested to provide intracellular protection against the denaturation of large molecules to stabilize plasma membranes. The major seed proteins involved in bioglasses formation should be LEA proteins (Shih et al., 2008). The LEA proteins are a family of protective proteins that accumulate at the late stages of seed development. On the one hand, it can take advantage of its highly hydrophilic characteristics to prevent excessive loss of water from the cell, thereby stabilizing the membrane structure and renaturing unfolded proteins. On the other hand, it can reduce oxidative stress damage to cells by chelating metal ions (Tunnacliffe and Wise, 2007; Yang et al., 2019). Some studies have been proved that the LEA family could maintain the vitality of seed embryos and improve cell resistance during storage in maize, alfalfa, *Arabidopsis*, and wheat (Kalemba and Pukacka, 2008; Rajjou et al., 2008; Wu et al., 2011; Chatelain et al., 2012). *ATEM1* and *ATEM6* (LEA family) knockout mutations of *Arabidopsis* display premature seed dehydration and maturation (Manfre et al., 2006), and recent studies have demonstrated that seed longevity is reduced when the seed-specific dehydrins were decreased in *A. thaliana* (Rajjou et al., 2008; Hundertmark et al., 2011). In rice, LEA proteins and their mRNAs accumulate to high concentrations during seed development and maturation (Sano et al., 2013), and glutathione-related proteins, DNA-damage-repair proteins and LEA protein might be correlated with seed storability (Gao et al., 2016). Furthermore, HSPs, as molecular chaperones, are mainly involved in the processing and transportation of nascent peptides, and also, in the repair and degradation of damaged proteins under stress. HSP accumulated at the late stage of maturation in many seed embryos, such as *Arabidopsis*, sunflower, and pea (Coca et al., 1994; DeRocher and Vierling, 1994; Wehmeyer et al., 1996), and had a close relationship with plant seed resistance and storage tolerance (Downs et al., 1999; Guo et al., 2007). In this study, 3 ABI5-like, 8 LEA proteins (significantly up-regulated), and 17 HSPs (8 up-regulated and 9 down-regulated) showed obvious expression changes during dehydration process in lotus embryos, in particular, RAB18 (response to ABA) dehydrin was the most abundant protein among all DEPs. These findings suggested that ABA signaling and the accumulation of LEA and HSPs proteins may be the key factors to ensure the continuous dehydration and storage tolerance of lotus seed embryos.

DEPs Involved in Protein Modification and Degradation

Histone acetylation modification is a key mechanism of plant gene transcriptional regulation and is closely related to gene expression. Histone deacetylase (HDAC) could participate in the remodeling of chromatin structure, playing an important role in the epigenetic regulation of plant genes (Zhang and Zhong, 2014). Numerous studies have reported that HDAC could regulate the maturation and development of seeds or abiotic stress in plants (Wu et al., 2000; Zhou et al., 2004; Gao et al., 2015). The *AtHD2C*-transgenic *Arabidopsis* are less sensitive to ABA, high salt, and drought stresses (To et al., 2011). In rice, overexpression of *SRT1* gene could enhance the ability of rice to resist oxidative stress (Zhong et al., 2013). In our study, we identified three types of HDAC family members (A0A1U7YW34, A0A1U8Q4V5, A0A1U8BLQ1), and their expressions were significantly up-regulated during maturation drying, indicating that HDAC positively regulating the maturation and development of lotus embryos. The up-regulated HDAC may respond to the ABA signaling pathway and enhance the tolerance of lotus embryo cells during continuous dehydration through regulation of gene transcription. Plants usually remove some damaged or useless organelles and proteins through autophagy (on type of programed cell death, PCD) to maintain the normal physiological activities when growing under abiotic stress conditions. However, excessive cellular autophagy may lead to serious protein degradation and cell death, causing severe damage to cells (Iakimova et al., 2005). Cysteine proteinase (CP) and subtilisin-like protease (SBT), as important proteolytic enzymes related to plant PCD events, are involved in the removal of damaged cells and the hydrolysis of most storage proteins in plants (Hatsugai et al., 2009; Tran et al., 2014). Cysteine proteinase inhibitor (CPI) can prevent the PCD events, such as hydrolysis of storage proteins or excessive autophagy, by inhibiting CP activity (Massonneau et al., 2005). In *Zea mays*, CPI was accumulated in mature seed embryos, which regulates the turnover of storage proteins during seed development (Wu et al., 2011). In our study, remarkably, CPI was significantly up-regulated, while CP and SBT were downregulated in lotus embryos. The expression changes of the above proteins suggested that PCD events are involved in lotus embryo morphogenesis, but PCD and storage proteolysis events are weakened during seed dehydration and maturity process, which beneficial to the nutrient accumulation and storage in embryos.

DEPs Involved in DNA Repair

DNA damage, such as nucleobase modification and DNA strand breaking, usually occurs when seeds suffered from dehydration or some adverse environmental conditions exacerbating seed deterioration and aging (Cheah and Osborne, 1978). DNA repair enzymes could effectively repair damaged DNA and maintain normal physiological activities of cells. A previous study found that DNA repair proteins are downregulated in rice embryos which are not resistant to storage, indicating that these proteins have a certain relationship with the storage tolerance of seed embryos (Manfre et al., 2006). High mobility group proteins (HMG) are widely involved in gene expression and regulation, including DNA replication, transcription, recombination, and DNA repair process (Kalyna et al., 2003; Grasser et al., 2007). In addition, the accumulation of histones was also found in *Brachypodium distachyon* embryos during seed maturation (Wolny et al., 2017). These proteins play an important role in DNA repair by rearranging chromatin structure to reduce ROS damage during seed maturation drying (Kalemba and Pukacka, 2008). In this study, we found that a number of proteins related to DNA repair were significantly up-regulated, including DNA damage repair proteins (A0A1U7ZYD0), HMG proteins (A0A1U7ZKY2, A0A1U8ATJ2, A0A1U7Z4F5), and histones (A0A1U7Z5K0, A0A1U8AWH0), indicating that DNA regulation and repair mechanism play a positive role in maintaining normal physiological activities and cell defense during lotus embryos dehydration and maturity process.

CONCLUSIONS

In this study, a TMT-based quantitative proteomic approach was used to reveal differential protein profiling during the lotus embryo continuous dehydration and maturity process. Bioinformatics analysis showed that environmental stimuli, stress responses, and protein processing related biological processes are vigorous, and carbohydrate metabolism, protein processing in ER, DNA repair, and antioxidative events have a positive response to lotus embryo dehydration. Furthermore, H_2O_2 was the main ROS component inducing oxidative stress damage, and GSH and V_E acted as the major antioxidant to maintain the REDOX balance of lotus embryo. ABA signal and the accumulation of oligosaccharides, LEA, and HSPs may be the key factors to ensure the continuous dehydration and storage tolerance of lotus seed embryo. Future studies focusing on characterizing the biological significance of LEA, HSPs, and oligosaccharide synthesis-related key proteins will be highly valuable in designing molecular breeding or engineering programs for enhancing plant tolerance to dehydration.

DATA AVAILABILITY STATEMENT

The original contributions presented in the study are publicly available. This data can be found here: ProteomeXchange via the PRIDE database, with the accession number PXD029187.

AUTHOR CONTRIBUTIONS

DZ conceived the original research plans and designed the experiments. TL, JS, and SL collected and analyzed the data. DZ and LR wrote the manuscript. All authors have read and approved the final manuscript.

FUNDING

This work was sponsored by the National Natural Science Foundation of China (Grant Nos. 31971705 and 31870686) and Natural Science Foundation of Shanghai (Grant No. 21ZR1434200).

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpls.2021.792057/full#supplementary-material

Supplementary Figure 1 | Gene Oncology (GO) terms annotation of all differentially expressed proteins (DEPs) in lotus embryos during seed dehydration maturity. **(A)** GO functional annotation of up-regulated DEPs; **(B)** GO functional annotation of down-regulated DEPs.

Supplementary Figure 2 | AgriGO significant enrichment functional annotation of DEPs from lotus embryo during seed dehydration maturity.

Supplementary Figure 3 | KEGG pathway term significant enrichment analysis of all DEPs in lotus embryos during seed dehydration maturity.

Supplementary Figure 4 | KEGG pathway significant enrichment analysis of DEPs from lotus embryo during seed dehydration and maturity. **(A)** Significant enrichment analysis of up-regulated DEPs; **(B)** significant enrichment analysis of down-regulated DEPs.

Supplementary Figure 5 | Parallel reaction monitoring quantitative verification of eighteen proteins from tandem mass tags (TMT) proteomics data. Asterisks indicate statistically significant differences (***$p < 0.001$; **$p < 0.01$; *$p < 0.05$) between lotus embryos at different dehydration stages.

Supplementary Figure 6 | Carbohydrates, energy, and antioxidation system regulation model of key DEPs from lotus embryo during seed dehydration maturity.

Supplementary Table 1 | Detailed quantitative information of identified proteins from lotus embryo during seed dehydration maturity.

Supplementary Table 2 | Detailed quantitative information of screened differentially expressed proteins from lotus embryo during seed dehydration maturity. The orange background represents up-regulated DEPs and the green background represents downregulated DEPs.

Supplementary Table 3 | Functional classification of key DEPs from lotus embryo during dehydration maturity.

Supplementary Table 4 | Correlation analysis of stress physiological indices of lotus embryo during dehydration maturity.

REFERENCES

Bo, C., Geng, X., Zhang, J., Sai, L., Zhang, Y., Yu, G., et al. (2020). Comparative proteomic analysis of silica-induced pulmonary fibrosis in rats based on tandem mass tag (TMT) quantitation technology. *PLoS ONE* 15:e0241310. doi: 10.1371/journal.pone.0241310

Burke, M. (1986). *The Glassy State and Survival of Anhydrous Biological Systems*. Ithaca, NY: Cornell University Press.

Chatelain, E., Hundertmark, M., Leprince, O., Gall, S. L., Satour, P., Deligny-penninck, S., et al. (2012). Temporal profiling of the heat-stable proteome during late maturation of *Medicago truncatula* seeds identifies a restricted subset of late embryogenesis abundant proteins associated with longevity. *Plant Cell Environ.* 35, 1440–1455. doi: 10.1111/j.1365-3040.2012.02501.x

Cheah, K. S. E., and Osborne, D. J. (1978). DNA lesions occur with loss of viability in embryos of ageing rye seed. *Nature* 272, 593–599. doi: 10.1038/272593a0

Coca, M. A., Almoguera, C., and Jordano, J. (1994). Expression of sunflower low-molecular-weight heat-shock proteins during embryogenesis and persistence after germination: localization and possible functional implications. *Plant Mol. Biol.* 25, 479–492. doi: 10.1007/BF00043876

Debeaujon, I., Léon-Kloosterziel, K. M., and Koornneef, M. (2000). Influence of the testa on seed dormancy, germination, and longevity in *Arabidopsis*. *Plant Physiol.* 122, 403–414. doi: 10.1104/pp.122.2.403

Dekkers, B. J. W., Costa, M. C. D., Maia, J., Bentsink, L., Ligterink, W., and Hilhorst, H. W. M. (2015). Acquisition and loss of desiccation tolerance in seeds: From experimental model to biological relevance. *Planta* 241, 563–577. doi: 10.1007/s00425-014-2240-x

Delahaie, J., Hundertmark, M., Bove, J., Leprince, O., Rogniaux, H., and Buitink, J. (2013). LEA polypeptide profiling of recalcitrant and orthodox legume seeds reveals ABI3-regulated LEA protein abundance linked to desiccation tolerance. *J. Exp. Bot.* 64, 4559–4573. doi: 10.1093/jxb/ert274

Delmas, F., Sankaranarayanan, S., Deb, S., Widdup, E., Bournonville, C., Bollier, N., et al. (2013). ABI3 controls embryo degreening through Mendel's locus. *Proc. Natl. Acad. Sci. U.S.A.* 110:E3888. doi: 10.1073/pnas.1308114110

Delseny, M., Bies-Etheve, N., Carles, C., Hull, G., Vicient, C., Raynal, M., et al. (2001). Late Embryogenesis Abundant (LEA) protein gene regulation during *Arabidopsis* seed maturation. *J. Plant Physiol.* 158, 419–427. doi: 10.1078/0176-1617-00353

DeRocher, A. E., and Vierling, E. (1994). Developmental control of small heat shock protein expression during pea seed maturation. *Plant J.* 5, 93–102. doi: 10.1046/j.1365-313X.1994.5010093.x

Downs, C. A., Coleman, J. S., and Heckathorn, S. A. (1999). The chloroplast 22-Ku heat-shock protein: a lumenal protein that associates with the oxygen evolving complex and protects photosystem II during heat stress. *J. Plant Physiol.* 155, 477–487. doi: 10.1016/S0176-1617(99)80042-X

Gallé, Á., Csiszár, J., Secenji, M., Guóth, A., Cseuz, L., Tari, I., et al. (2009). Glutathione transferase activity and expression patterns during grain filling in flag leaves of wheat genotypes differing in drought tolerance: response to water deficit. *J. Plant Physiol.* 166, 1878–1891. doi: 10.1016/j.jplph.2009.05.016

Gao, J., Fu, H., Zhou, X., Chen, Z., Luo, Y., Cui, B., et al. (2016). Comparative proteomic analysis of seed embryo proteins associated with seed storability in rice (*Oryza sativa* L.) during natural aging. *Plant Physiol. Bioch.* 103, 31–44. doi: 10.1016/j.plaphy.2016.02.026

Gao, M., Li, X., Huang, J., Gropp, G. M., Gjetvaj, B., Lindsay, D. L., et al. (2015). SCARECROW-LIKE15 interacts with HISTONE DEACETYLASE19 and is essential for repressing the seed maturation programme. *Nat. Commun.* 6:7243. doi: 10.1038/ncomms8243

Gong, M., Zhang, H., Wu, D., Zhang, Z., Zhang, J., Bao, D., et al. (2021). Key metabolism pathways and regulatory mechanisms of high polysaccharide yielding in *Hericium erinaceus*. *BMC Genomics* 22:160. doi: 10.1186/s12864-021-07480-x

Grasser, K. D., Launholt, D., and Grasser, M. (2007). High mobility group proteins of the plant HMGB family: dynamic chromatin modulators. *Genet. Resour. Crop Ev.* 1769, 346–357. doi: 10.1016/j.bbaexp.2006.12.004

Guo, H. (2009). Cultivation of lotus (*Nelumbo nucifera* Gaertn. ssp. *nucifera*) and its utilization in China. *Genet. Resour. Crop Ev.* 56, 323–330. doi: 10.1007/s10722-008-9366-2

Guo, S., Zhou, H., Zhang, X., Li, X., and Meng, Q. (2007). Overexpression of *CaHSP26* in transgenic tobacco alleviates photoinhibition of PSII and PSI during chilling stress under low irradiance. *J. Plant Physiol.* 164, 126–136. doi: 10.1016/j.jplph.2006.01.004

Gutierrez, L., van Wuytswinkel, O., Castelain, M., and Bellini, C. (2007). Combined networks regulating seed maturation. *Trends Plant Sci.* 12, 294–300. doi: 10.1016/j.tplants.2007.06.003

Hatsugai, N., Iwasaki, S., Tamura, K., Kondo, M., Kentaro, F., Kimi, O., et al. (2009). A novel membrane fusion-mediated plant immunity against bacterial pathogens. *Gene. Dev.* 23, 2496–2506. doi: 10.1101/gad.1825209

Hauser, F., Waadt, R., and Schroeder, J. I. (2011). Evolution of abscisic acid synthesis and signaling mechanisms. *Curr. Biol.* 21, R346–R355. doi: 10.1016/j.cub.2011.03.015

Hoekstra, F. A., Golovina, E. A., and Buitink, J. (2001). Mechanisms of plant desiccation tolerance. *Trends Plant Sci.* 6, 431–438. doi: 10.1016/S1360-1385(01)02052-0

Huang, S., Tang, X., Zhang, L., and Rui, F. (2003). Thermotolerance and activity of antioxidative enzymes in lotus seeds. *J. Plant Physiol. Mol. Biol.* 5, 421–424. (in Chinese).

Hundertmark, M., Buitink, J., Leprince, O., and Hincha, D. K. (2011). The reduction of seed-specific dehydrins reduces seed longevity in *Arabidopsis thaliana*. *Seed Sci. Res.* 21, 165–173. doi: 10.1017/S0960258511000079

Iakimova, E., Kapchina-Toteva, V., de Jong, A., Atanassov, A., and Woltering, E. (2005). Involvement of ethylene, oxidative stress and lipid-derived signals in cadmium-induced programmed cell death in tomato suspension cells. *BMC Plant Biol.* 5:S19. doi: 10.1186/1471-2229-5-S1-S19

Kalemba, E. M., and Pukacka, S. (2008). Changes in late embryogenesis abundant proteins and a small heat shock protein during storage of beech (*Fagus sylvatica* L.) seeds. *Environ. Exp. Bot.* 63, 274–280. doi: 10.1016/j.envexpbot.2007.12.011

Kalyna, M., Lopato, S., and Barta, A. (2003). Ectopic expression of atRSZ33 reveals its function in splicing and causes pleiotropic changes in development. *Mol. Biol. Cell* 14, 3565–3577. doi: 10.1091/mbc.e03-02-0109

Khandelwal, A., Cho, S. H., Marella, H., Sakata, Y., Perroud, P. F., Pan, A., et al. (2010). Role of ABA and ABI3 in desiccation tolerance. *Science* 327:546. doi: 10.1126/science.1183672

Kranner, I., Birtić, S., Anderson, K. M., and Pritchard, H. W. (2006). Glutathione half-cell reduction potential: a universal stress marker and modulator of programmed cell death? *Free Radical Bio. Med.* 40, 2155–2165. doi: 10.1016/j.freeradbiomed.2006.02.013

Kruger, N. J., and von Schaewen, A. (2003). The oxidative pentose phosphate pathway: structure and organisation. *Curr. Opin. Plant Biol.* 6, 236–246. doi: 10.1016/S1369-5266(03)00039-6

Lee, J., and Koh, H. (2011). A label-free quantitative shotgun proteomics analysis of rice grain development. *Proteome Sci.* 9:61. doi: 10.1186/1477-5956-9-61

Leprince, O., Pellizzaro, A., Berriri, S., and Buitink, J. (2017). Late seed maturation: drying without dying. *J. Exp. Bot.* 68, 827–841. doi: 10.1093/jxb/erw363

Lopez-Molina, L., Mongrand, S., and Chua, N. (2001). A postgermination developmental arrest checkpoint is mediated by abscisic acid and requires the ABI5 transcription factor in *Arabidopsis*. *Proc. Natl. Acad. Sci. U.S.A.* 98:4782. doi: 10.1073/pnas.081594298

Maia, J., Dekkers, B. J. W., Dolle, M. J., Ligterink, W., and Hilhorst, H. W. M. (2014). Abscisic acid (ABA) sensitivity regulates desiccation tolerance in germinated *Arabidopsis* seeds. *New Phytol.* 203, 81–93. doi: 10.1111/nph.12785

Manfre, A. J., Lanni, L. M., and Marcotte, W. R. (2006). The *Arabidopsis* group 1 late embryogenesis abundant protein ATEM6 is required for normal seed development. *Plant Physiol.* 140, 140–149. doi: 10.1104/pp.105.072967

Massonneau, A., Condamine, P., Wisniewski, J., Zivy, M., and Rogowsky, P. M. (2005). Maize cystatins respond to developmental cues, cold stress and drought. *BBA Gene Struct. Expr.* 1729, 186–199. doi: 10.1016/j.bbaexp.2005.05.004

Ming, R., VanBuren, R., Liu, Y., Yang, M., Han, Y., Li, L., et al. (2013). Genome of the long-living sacred lotus (*Nelumbo nucifera* Gaertn.). *Genome Biol.* 14:R41. doi: 10.1186/gb-2013-14-5-r41

Moro, C. F., Fukao, Y., Shibato, J., Rakwal, R., Agrawal, G. K., Shioda, S., et al. (2015). Immature seed endosperm and embryo proteomics of the Lotus (*Nelumbo Nucifera* Gaertn.) by one-dimensional gel-based tandem mass spectrometry and a comparison with the mature endosperm proteome. *Proteomes* 3, 184–235. doi: 10.3390/proteomes3030184

Mukherjee, D., Khatua, T. N., Venkatesh, P., Saha, B. P., and Mukherjee, P. K. (2010). Immunomodulatory potential of rhizome and seed

extracts of *Nelumbo nucifera* Gaertn. *J. Ethnopharmacol.* 128, 490–494. doi: 10.1016/j.jep.2010.01.015

Prieto-Dapena, P., Castaño, R., Almoguera, C., and Jordano, J. (2006). Improved resistance to controlled deterioration in transgenic seeds. *Plant Physiol.* 142, 1102–1112. doi: 10.1104/pp.106.087817

Rajjou, L., Lovigny, Y., Groot, S. P. C., Belghazi, M., Job, C., and Job, D. (2008). Proteome-Wide characterization of seed aging in *Arabidopsis*: a comparison between artificial and natural aging protocols. *Plant Physiol.* 148, 620–641. doi: 10.1104/pp.108.123141

Rezaei, M. K., Shobbar, Z., Shahbazi, M., Abedini, R., and Zare, S. (2013). Glutathione S-transferase (GST) family in barley: Identification of members, enzyme activity, and gene expression pattern. *J. Plant Physiol.* 170, 1277–1284. doi: 10.1016/j.jplph.2013.04.005

Sano, N., Masaki, S., Tanabata, T., Yamada, T., Hirasawa, T., and Kanekatsu, M. (2013). Proteomic analysis of stress-related proteins in rice seeds during the desiccation phase of grain filling. *Plant Biotechnol. Nar.* 30, 147–156. doi: 10.5511/plantbiotechnology.13.0207a

Sattler, S. E., Gilliland, L. U., Magallanes-Lundback, M., Pollard, M., and DellaPenna, D. (2004). Vitamin E is essential for seed longevity and for preventing lipid peroxidation during germination. *Plant Cell* 16, 1419–1432. doi: 10.1105/tpc.021360

Scharte, J., Schön, H., Tjaden, Z., Weis, E., and von Schaewen, A. (2009). Isoenzyme replacement of glucose-6-phosphate dehydrogenase in the cytosol improves stress tolerance in plants. *Proc. Natl. Acad. Sci. U.S.A.*106:8061. doi: 10.1073/pnas.0812902106

Scheibe, R. (2004). Malate valves to balance cellular energy supply. *Physiol. Plant.* 120, 21–26. doi: 10.1111/j.0031-9317.2004.0222.x

Scoffoni, C., Vuong, C., Diep, S., Cochard, H., and Sack, L. (2014). Leaf shrinkage with dehydration: coordination with hydraulic vulnerability and drought tolerance. *Plant Physiol.* 164, 1772–1788. doi: 10.1104/pp.113.221424

Shen-Miller, J. (2002). Sacred lotus, the long-living fruits of *China Antique*. *Seed Sci. Res.* 12, 131–143. doi: 10.1079/SSR2002112

Shen-Miller, J., Lindner, P., Xie, Y., Villa, S., Wooding, K., Clarke, S. G., et al. (2013). Thermal-stable proteins of fruit of long-living sacred lotus *Nelumbo nucifera* Gaertn var. China antique. *Trop. Plant Biol.* 6, 69–84. doi: 10.1007/s12042-013-9124-2

Shih, M., Hoekstra, F. A., and Hsing, Y. C. (2008). *Late Embryogenesis Abundant Proteins*. Pittsburgh, PA: Academic Press. doi: 10.1016/S0065-2296(08)00404-7

Sugliani, M., Rajjou, L., Clerkx, E. J. M., Koornneef, M., and Soppe, W. J. J. (2009). Natural modifiers of seed longevity in the *Arabidopsis* mutants *abscisic acid insensitive3-5* (*abi3-5*) and *leafy cotyledon1-3* (*lec1-3*). *New Phytol.* 184, 898–908. doi: 10.1111/j.1469-8137.2009.03023.x

To, T. K., Kim, J., Matsui, A., Kurihara, Y., Morosawa, T., Ishida, J., et al. (2011). *Arabidopsis* HDA6 regulates locus-directed heterochromatin silencing in cooperation with MET1. *PLoS Genet.* 7:e1002055. doi: 10.1371/journal.pgen.1002055

Tran, V., Weier, D., Radchuk, R., Thiel, J., and Radchuk, V. (2014). Caspase-like activities accompany programmed cell death events in developing barley grains. *PLoS ONE* 9:e109426. doi: 10.1371/journal.pone.0109426

Tunnacliffe, A., and Wise, M. J. (2007). The continuing conundrum of the LEA proteins. *Naturwissenschaften* 94, 791–812. doi: 10.1007/s00114-007-0254-y

Umezawa, T., Nakashima, K., Miyakawa, T., Kuromori, T., Tanokura, M., Shinozaki, K., et al. (2010). Molecular basis of the core regulatory network in ABA responses: sensing, signaling and transport. *Plant Cell Physiol.* 51, 1821–1839. doi: 10.1093/pcp/pcq156

van Zanten, M., Zöll, C., Wang, Z., Philipp, C., Carles, A., Li, Y., et al. (2014). HISTONE DEACETYLASE 9 represses seedling traits in *Arabidopsis thaliana* dry seeds. *Plant J.* 80, 475–488. doi: 10.1111/tpj.12646

Verdier, J., Lalanne, D., Pelletier, S., Torres-Jerez, I., Righetti, K., Bandyopadhyay, K., et al. (2013). A regulatory network-based approach dissects late maturation processes related to the acquisition of desiccation tolerance and longevity of *Medicago truncatula* seeds. *Plant Physiol.* 163, 757–774. doi: 10.1104/pp.113.222380

Wang, L., Fu, J., Li, M., Fragner, L., Weckwerth, W., and Yang, P. (2016). Metabolomic and proteomic profiles reveal the dynamics of primary metabolism during seed development of Lotus (*Nelumbo nucifera*). *Front. Plant Sci.* 7:750. doi: 10.3389/fpls.2016.00750

Wang, Y., Fan, G., Liu, Y., Sun, F., Shi, C., Liu, X., et al. (2013). The sacred lotus genome provides insights into the evolution of flowering plants. *Plant J.* 76, 557–567. doi: 10.1111/tpj.12313

Wehmeyer, N., Hernandez, L. D., Finkelstein, R. R., and Vierling, E. (1996). Synthesis of small heat-shock proteins is part of the developmental program of late seed maturation. *Plant Physiol.* 112, 747–757. doi: 10.1104/pp.112.2.747

Wiśniewski, J. R., Zougman, A., Nagaraj, N., and Mann, M. (2009). Universal sample preparation method for proteome analysis. *Nat. Methods* 6, 359–362. doi: 10.1038/nmeth.1322

Wolny, E., Braszewska-Zalewska, A., Kroczek, D., and Hasterok, R. (2017). Histone H3 and H4 acetylation patterns are more dynamic than those of DNA methylation in *Brachypodium distachyon* embryos during seed maturation and germination. *Protoplasma* 254, 2045–2052. doi: 10.1007/s00709-017-1088-x

Wu, K., Tian, L., Malik, K., Brown, D., and Miki, B. (2000). Functional analysis of HD2 histone deacetylase homologues in *Arabidopsis thaliana*. *Plant J.* 22, 19–27. doi: 10.1046/j.1365-313x.2000.00711.x

Wu, X., Liu, H., Wang, W., Chen, S., Hu, X., and Li, C. (2011). Proteomic analysis of seed viability in maize. *Acta Physiol. Plant.* 33, 181–191. doi: 10.1007/s11738-010-0536-4

Yang, Z., Sheng, J., Lv, K., Ren, L., and Zhang, D. (2019). Y_2SK_2 and SK_3 type dehydrins from *Agapanthus praecox* can improve plant stress tolerance and act as multifunctional protectants. *Plant Sci.* 284, 143–160. doi: 10.1016/j.plantsci.2019.03.012

Zhang, D., Ren, L., Chen, G., Zhang, J., Reed, B. M., and Shen, X. (2015). ROS-induced oxidative stress and apoptosis-like event directly affect the cell viability of cryopreserved embryogenic callus in *Agapanthus praecox*. *Plant Cell Rep.* 34, 1499–1513. doi: 10.1007/s00299-015-1802-0

Zhang, D., Ren, L., Yue, J., Wang, L., Zhuo, L., and Shen, X. (2013). A comprehensive analysis of flowering transition in *Agapanthus praecox* ssp. orientalis Leighton by using transcriptomic and proteomic techniques. *J. Proteomics* 80, 1–25. doi: 10.1016/j.jprot.2012.12.028

Zhang, J., and Zhong, Q. (2014). Histone deacetylase inhibitors and cell death. *Cell. Mol. Life Sci.* 71, 3885–3901. doi: 10.1007/s00018-014-1656-6

Zhang, N., Chen, F., Huo, W., and Cui, D. (2015). Proteomic analysis of middle and late stages of bread wheat (*Triticum aestivum* L.) grain development. *Front. Plant Sci.* 6:735. doi: 10.3389/fpls.2015.00735

Zhang, S., Ghatak, A., Bazargani, M. M., Bajaj, P., Varshney, R. K., Chaturvedi, P., et al. (2021). Spatial distribution of proteins and metabolites in developing wheat grain and their differential regulatory response during the grain filling process. *Plant J.* 107, 669–687. doi: 10.1111/tpj.15410

Zhong, X., Zhang, H., Zhao, Y., Sun, Q., Hu, Y., Peng, H., et al. (2013). The rice NAD^+-dependent histone deacetylase OsSRT1 targets preferentially to stress- and metabolism-related genes and transposable elements. *PLoS ONE* 8:e66807. doi: 10.1371/journal.pone.0066807

Zhou, C., Labbe, H., Sridha, S., Wang, L., Tian, L., Latoszek-Green, M., et al. (2004). Expression and function of HD2-type histone deacetylases in *Arabidopsis* development. *Plant J.* 38, 715–724. doi: 10.1111/j.1365-313X.2004.02083.x

Zinsmeister, J., Lalanne, D., Terrasson, E., Chatelain, E., Vandecasteele, C., Vu, B. L., et al. (2016). ABI5 is a regulator of seed maturation and longevity in legumes. *Plant Cell* 28, 2735–2754. doi: 10.1105/tpc.16.00470

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

frontiers in Plant Science

ORIGINAL RESEARCH
published: 10 February 2022
doi: 10.3389/fpls.2022.824439

Carrot (*Daucus carota* L.) Seed Germination Was Promoted by Hydro-Electro Hybrid Priming Through Regulating the Accumulation of Proteins Involved in Carbohydrate and Protein Metabolism

*Shuo Zhao[1], Hao Zou[2], Yingjie Jia[3], Xueqin Pan[3] and Danfeng Huang[1,3]**

[1] School of Agriculture and Biology, Shanghai Jiao Tong University, Shanghai, China, [2] School of Mechanical Engineering, Institute of Refrigeration and Cryogenics, Shanghai Jiao Tong University, Shanghai, China, [3] Shanghai Vegetable Research Institute, Shanghai, China

OPEN ACCESS

Edited by:
Andrej Frolov,
Leipzig University, Germany

Reviewed by:
Jinwei Suo,
Zhejiang Agriculture and Forestry University, China
Łukasz Wojtyla,
Adam Mickiewicz University, Poland

***Correspondence:**
Danfeng Huang
hdf@sjtu.edu.cn

Specialty section:
This article was submitted to Plant Proteomics and Protein Structural Biology, a section of the journal Frontiers in Plant Science

Received: *29 November 2021*
Accepted: *20 January 2022*
Published: *10 February 2022*

Citation:
Zhao S, Zou H, Jia Y, Pan X and Huang D (2022) Carrot (Daucus carota L.) Seed Germination Was Promoted by Hydro-Electro Hybrid Priming Through Regulating the Accumulation of Proteins Involved in Carbohydrate and Protein Metabolism. Front. Plant Sci. 13:824439. doi: 10.3389/fpls.2022.824439

Asynchronized and non-uniform seed germination is causing obstacles to the large-scale cultivation of carrot (*Daucus carota* L.). In the present study, the combination of high voltage electrostatic field treatment (EF) with hydropriming (HYD), namely hydro-electro hybrid priming (HEHP), significantly improved all germination indicators of carrot seeds, and the promoting effect was superior to that of the HYD treatment. A tandem mass tags (TMT)-based proteomic analysis identified 4,936 proteins from the seeds, and the maximum number of differentially abundant proteins (DAPs) appeared between CK and HEHP. KEGG analysis revealed that the upregulated DAPs were mainly enriched in the pathways related to protein synthesis and degradation such as "ribosome" and "proteasome," while the downregulated DAPs were mainly enriched in photosynthesis-related pathways. Furthermore, the maximum DAPs were annotated in carbohydrate metabolism. Some proteins identified as key enzymes of the glyoxylate cycle, the tricarboxylate cycle, glycolysis and the pentose phosphate pathway showed enhanced abundance in priming treatments. The activities of several key enzymes involved in carbohydrate metabolism were also enhanced by the priming treatments, especially the HEHP treatment. Real-time quantitative PCR (qRT-PCR) analysis revealed that the effect of priming is mainly reflected before sowing. In conclusion, the optimal effect of HEHP is to regulate the synthesis and degradation of proteins in seeds to meet the requirements of germination and initiate the utilization of seed storage reserves and respiratory metabolism. The present work expanded the understanding of the response mechanism of carrot seed germination to priming and the biological effects of high voltage electrostatic field.

Keywords: hydro-electro hybrid priming, carrot, seed germination, TMT-based proteomic analysis, protein synthesis and degradation, carbohydrate metabolism

INTRODUCTION

Rapid and uniform germination is a crucial factor in determining field performance and crop yield, especially under the increasing uncertainty caused by global climatic change (Marcos-Filho, 2015; Finch-Savage and Bassel, 2016). Asynchronized and non-uniform seed germination is causing obstacles to the large-scale cultivation of carrot (*Daucus carota* L.). The mechanized precision sowing widely used in carrot cultivation also puts forward higher requirements for the germination performance of carrot seeds. Our group has developed a novel seed priming technology named hydro-electro hybrid priming (HEHP) by combining high voltage electrostatic field treatment (EF) and hydropriming (HYD). Moreover, we applied this method to carrot seeds for the first time, which expanded the model of the presowing treatment of carrot seeds.

Seed germination is essential to determine the connection between two successive plant generations, which is generally composed of three phases: rapid imbibition (phase I), metabolism reactivation (phase II) and radicle emergence (phase III) (Finch-Savage and Leubner-Metzger, 2006; Dekkers et al., 2013). Phase II is the critical phase of germination for extensively activated gene expression and metabolic activity, during which imbibed seeds either complete the germination process or remain dormant (Carrera et al., 2007; Han et al., 2017). Seed priming tends to advance some metabolic processes related to germination and avoid radicle extension by controlling the hydration state of seeds in phase II (Zhao et al., 2018; Srivastava et al., 2021). After desiccation, the seeds can still be stored for a certain period of time for commercial transportation and sales (Gallardo et al., 2001; Fabrissin et al., 2021). Seeds after priming usually show excellent germination performance by reducing the imbibition time (Espanany et al., 2016; Zhao et al., 2020; Chakma et al., 2021). As a non-chemical method, HYD is mainly carried out by soaking seeds with distilled water or controlling the water supply under optimal temperature conditions, which is environmentally safe and low cost, and the effect is remarkable (Yan, 2016; Nakao et al., 2020), but the seeds are easily infected by harmful microorganisms in the process of long-term priming (Singh et al., 2015). By applying EF treatment on the basis of HYD, HEHP achieves the goal of reducing the priming time, and this novel priming method has been successfully applied to onion seeds (Zhao et al., 2018).

At the stage of seed germination and seedling establishment, proteins are directly responsible for regulating biochemical reactions and metabolic processes in seeds (Li et al., 2019). The synthesis of specific proteins is necessary for the smooth progress of seed germination. For example, amanitine (an RNA polymerase II inhibitor) did not inhibit seed germination, while cycloheximide (a protein synthesis inhibitor) prevented the extension of the radicle (Rajjou et al., 2004; Sano et al., 2012). Studies on *Arabidopsis thaliana* have shown that more than 10,000 kinds of mRNA are stored in dry seeds (Nakabayashi et al., 2005). Within a few hours after seed imbibition, seeds can use these mRNAs and polysomal protein-synthesizing complexes appear in seeds for protein synthesis (Bewley, 1997). Meanwhile, the amino acids needed for *de novo* protein synthesis after germination are realized through the degradation of storage proteins. Various types of 20S proteasomes related to proteolysis increased abundance during germination (Rajjou et al., 2006; Wang et al., 2012). In addition, increased accumulation of proteins related to glycolysis, tricarboxylate cycle (TCA cycle) and pentose phosphate pathway were found in the germination process (Sano et al., 2012; Gu et al., 2016). The early stage of seed imbibition experienced a short anaerobic period. When the production of mitochondrial ATP is limited by hypoxia, the glycolysis pathway is dominant, while the pentose phosphate pathway is dominant after mitochondria become active (Nonogaki et al., 2010), and thus the accumulation of proteins involved in these pathways is required during early seed germination.

Proteomic techniques can effectively identify the characteristics of protein composition, which is an effective way to explore the molecular physiological basis of seed priming (Macovei et al., 2017; Marthandan et al., 2020). Numerous studies have used proteomic techniques to further uncover the molecular mechanisms of seed priming such as durum wheat (Fercha et al., 2013), maize (Araujo et al., 2021), rapeseed (Kubala et al., 2015) and alfalfa (Yacoubi et al., 2013), which focused on key proteins involved in reserve mobilization, protein metabolism, cell cycle and antioxidant protection through the analysis of the differentially abundant proteins (DAPs). However, as a representative of Apiaceae crops, the seeds of carrot come from cremocarp, so it is worth exploring whether the response of carrot seeds to priming is also related to these processes. Nevertheless, few studies have investigated the biological mechanism of seed priming on carrot seed germination and the proteome response of carrot seeds mediated by seed priming remains unknown.

In the present study, the tandem mass tags (TMT) isobaric labeling approach was applied to quantitatively analyze the proteome of carrot seeds, and through the analysis of DAPs, we focused our attention on the proteins involved in protein and carbohydrate metabolism. In conjunction with physiological determination, we explored the initiation effect of HEHP treatment on carrot seed germination and compared it with that of HYD treatment.

Abbreviations: DAPs, differentially abundant proteins; EF, high voltage electrostatic field treatment; FDR, false discovery rate; FC, fold change; FPA, fructose-bisphosphate aldolase; FW, fresh weight; GI, germination index; GPDH, glucose-6-phosphate dehydrogenase; HEHP, hydro-electro hybrid priming; HYD, hydropriming; ICL, isocitrate lyase; ME, malic enzyme; MS, malate synthase; PFK, 6-phosphofructokinase; RL, radicle length; SAM, S-adenosylmethionine synthase; SDH, succinate dehydrogenase subunit; SS, sucrose synthase; TCA cycle, tricarboxylate cycle; TMT, tandem mass tags; VI, vigor index.

MATERIALS AND METHODS

Materials and Seed Treatment Protocols

Carrot seeds (*Daucus carota* L. Naaisi) were harvested in 2019 and purchased from Shanghai Wells Seeds Co., Ltd (Shanghai,

China). Seeds were stored under proper conditions (5°C and 50% relative humidity).

For the HYD and HEHP treatments, the carrot seeds were first soaked in distilled water at 20°C for 6 h. Then for HEHP, the positive and negative electrodes of a BM-201 electrostatic field generator (made in Jiangsu, China) were connected to two 10 cm × 10 cm copper plates placed 1 cm apart; the seeds were placed on the lower plate (cathode) and exposed to a 2 kV/cm EF for 90 s after absorbing the surface moisture with absorbent paper. The EF parameters were widely selected by pre-experiments. Then HYD and HEHP were imbibed in a climate chamber (QHX-300BSH-III, made in Shanghai, China) set at 22°C at 98% humidity for 48 h in the dark, and desiccated at 25°C in a drum wind dryer until the initial weight was achieved. The HEHP treatment procedures are shown in **Figure 1**. The control (CK) was not treated in any way. Seed germination experiments and proteomic analysis were conducted on the seed samples from the three treatments (CK, HYD and HEHP).

Seed Germination Conditions

Fifty carrot seeds were sown in a Petri dish covered with 3 layers of sterile filter paper and moistened with 4 ml distilled water in a climate chamber (QHX-300BSH-III, made in Shanghai, China) set at 20°C and 80% relative humidity. The photoperiod was set to 14/10 h of light/dark obtained by using TL lights (TLD 30W/865 cool daylight, PHILIPS, made in Tailand) set at 40 μmol/(m^2·s). Distilled water (0.5 ml) was added to the Petri dish every 24 h after sowing. A seed was considered germinated when the radicle protruded through the seed coat. The number of germinated seeds and germination percentage (%) every 24 h were recorded after sowing until fifth day after sowing. Average radicle length (RL, mm) and fresh weight per 10 randomly selected seedlings (FW, g) were determined on the fifth day. The germination index (GI) and vigor index (VI) were calculated by the following formula. A total of 18 samples (3 treatments × 6 biological replications) were used for germination test.

$$GI = \sum Gt/Dt$$

$$VI = GI \times FW$$

Dt is the number of days for counting the number of germinated seeds, and Gt is the number of germinated seeds counted corresponding to the number of days.

Total Protein Extraction

The seed samples used for proteomic analysis were dry seeds with different treatments. The samples of carrot seeds (0.5 g per sample, approximately 200 seeds) were ground into a fine powder in liquid nitrogen. Then the powder was suspended in lysis buffer (8 M urea + 1% SDS, including protease inhibitor). The mixture was allowed to settle at 4°C for 30 min, and processed by ultrasound for 2 min at 0°C. After centrifugation at 12,000g at 4°C for 30 min, protein quantification was performed using BCA protein assay kit (Pierce, Thermo, United States). A total of 9 samples (3 treatments × 3 biological replications) were used for analysis.

Protein Digestion and Tandem Mass Tags Labeling

Protein digestion was performed according to the standard procedure and the resulting peptide mixture was labeled using 10-plex TMT reagent (Thermo Fisher, Art.No.90111) according to the manufacturer's instructions. 100 μg protein taken from each sample was mixed with 100 μL of the lysate. TCEP (10 mM) was added and then it was stored at 37°C for 60 min. Then iodoacetamide (40 mM) was added and stored in dark at room temperature for 40 min. Precooled acetone was added and precipitated at −20°C for 4 h, then centrifuged at 10,000 g for 20 min, and took the precipitate. The sample was dissolved with 50 mmol/L riethylammonium bicarbonate (TEAB), then trypsin

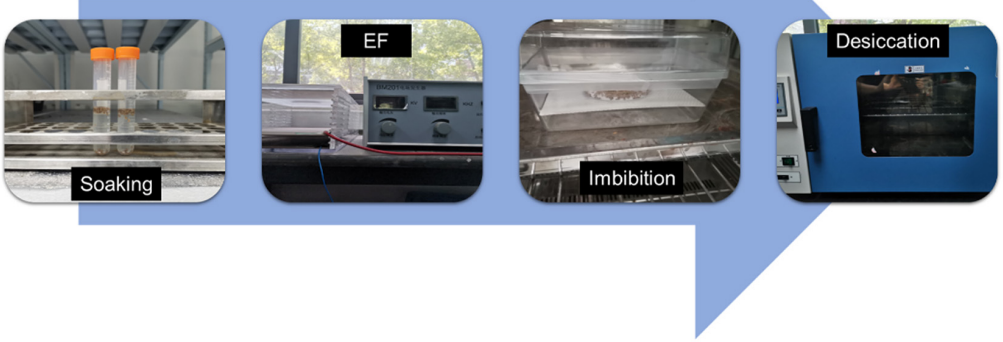


FIGURE 1 | Processing protocols of the HEHP treatment. EF: high voltage electrostatic field treatment.

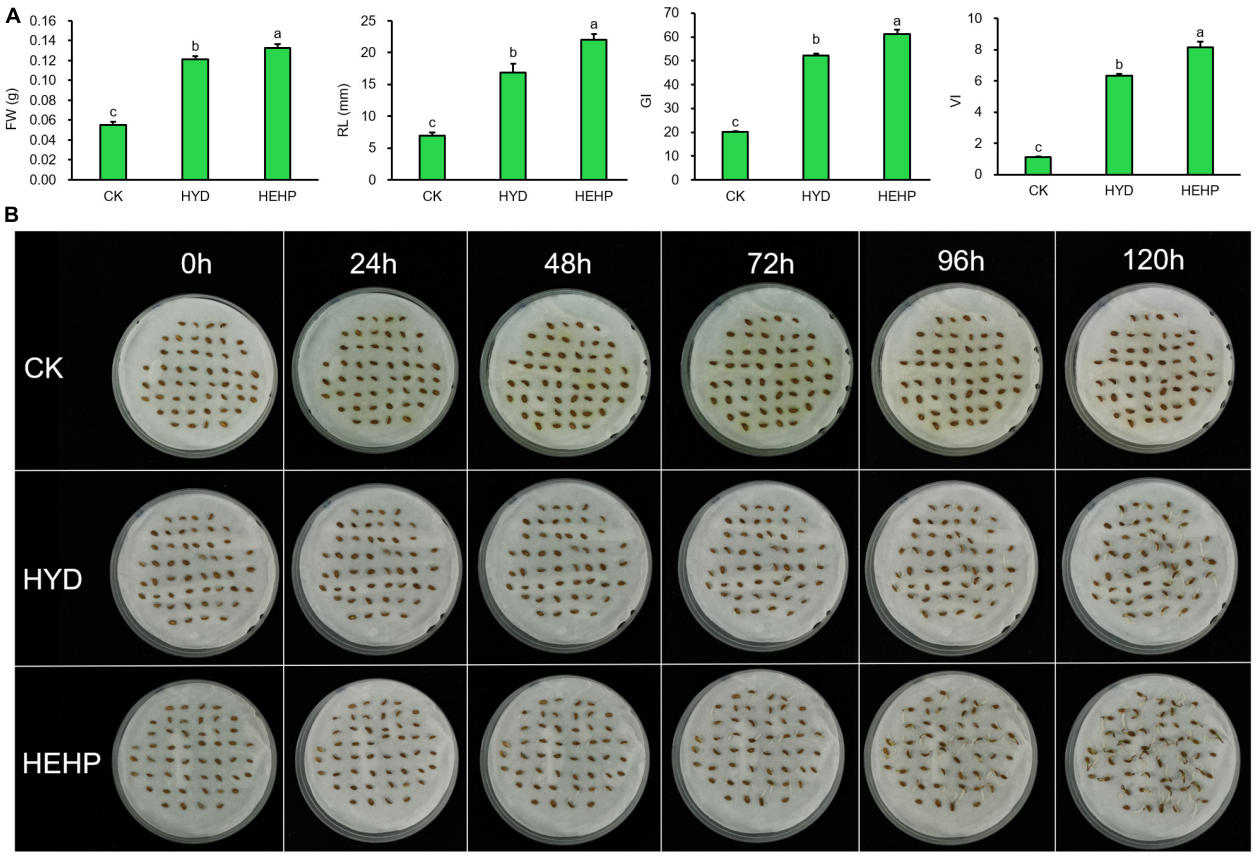


FIGURE 2 | Effects of different treatments on carrot seed germination. CK, the control; HYD, hydro-priming treatment; HEHP, hydro-electro hybrid priming treatment. **(A)** Effects of different treatments on carrot seed germination characteristics. **(B)** Photograph of seed germination experiment every 24 h. Values represent the means ± SE from six biological replications. The different lowercase letters above the bars indicate significant differences in 95% probability level ($p < 0.05$, Duncan test was performed after ANOVA analysis).

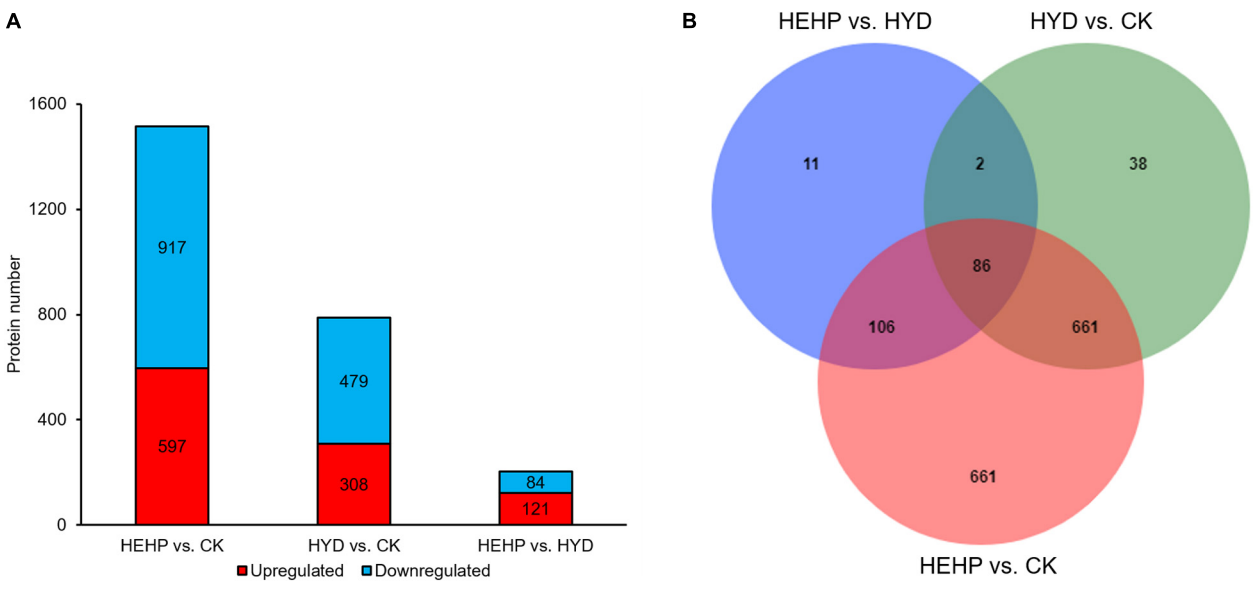


FIGURE 3 | Statistics of DAPs. **(A)** The number of DAPs among the three groups. **(B)** Venn diagram showing specific and overlapping DAPs among the three groups.

was added according to the mass ratio of 1: 50 and enzymolysis at 37°C overnight.

Tandem Mass Tags reagent was taken out at −20°C and was restored to room temperature. Then acetonitrile was added. After vortex centrifugation, TMT reagent was added per 100 μg polypeptide and incubated at room temperature for 2 h. Hydroxylamine was added and reacted at room temperature for 15 min. The same amount of labeled product was mixed in a tube and vacuum-dried.

High Performance Liquid Chromatography Separation

The pooled samples were fractionated by ACQUITY Ultra Performance liquid chromatography (Waters, United States) with an ACQUITY UPLC BEH C18 Column (1.7 μm, 2.1 mm × 150 mm, Waters) to increase proteomic depth. Briefly, peptides were first separated with a gradient of elution (Phase B: 5 mM ammonium hydroxide solution containing 80% acetonitrile, pH 10) over 48 min at a flow rate of 200 μl/min. A total of twenty fractions were collected according to peak shape and time, combined into ten fractions, and concentrated by vacuum centrifugation.

LC-MS/MS Analysis

Labeled peptides were analyzed by online nanoflow liquid chromatography tandem mass spectrometry performed on a 9RKFSG2_NCS-3500R system (Thermo, United States) connected to a Q Exactive Plus quadrupole orbitrap mass spectrometer (Thermo, United States) through a nanoelectrospray ion source. The C18 chromatographic column (75 μm × 25 cm, Thermo, United States) as equilibrated with solvent A (A:2% formic acid with 0.1% formic acid) and solvent B (B: 80% ACN with 0.1% formic acid). The peptides were eluted using the gradient (0–4 min, 0–5% B; 4–66 min, 5–23% B; 66–80 min, 23–29% B; 80–89 min,29–38% B; 89–91 min, 38–48% B; 91–92 min, 48–100% B; 92–105 min, 100% B; 105–106 min, 100–0% B) at a flow rate of 300 nL/min. The Q Exactive Plus was operated in the data-dependent acquisition mode to automatically switch between full scan MS and MS/MS acquisition. The survey of full scan MS spectra (m/z 350–1,300) was acquired in the Orbitrap. Then the top 20 most intense precursor ions were selected for secondary fragmentation, and the dynamic exclusion time was 18 s.

Protein Identification, Quantification and Bioinformatics Analysis

The raw data files were analyzed using Proteome Discoverer against the carrot database.[1] The false discovery rate (FDR) of peptide identification during library search is set to FDR ≤ 0.01. The protein contains at least one specific peptide.

The screening criteria of differentially abundant proteins (DAPs) were as follows: P value < 0.05 and fold change (FC) > 1.5 were identified as upregulated DAPs, P value < 0.05 and FC < 0.67 were identified as downregulated DAPs. Annotation of DAPs was performed using Gene Ontology (GO)[2,3] and Kyoto Encyclopedia of Genes and Genomes (KEGG).[4] DAPs were also used for KEGG enrichment analysis.

Determination of Key Enzyme Activities

The samples used for enzyme activity determination were conducted on the seeds (0.2 g for each sample, approximately 80 dry seeds or 50 imbibed seeds) at S0 (before sowing) and S20 (during germination, 20 h imbibed after sowing). All samples were frozen in liquid nitrogen and stored in a refrigerator at −80°C. The activity of malate synthase (MS, EC 2.3.3.9) was determined by assay kits (Michy Biomedical Technology Co., Ltd., Jiangsu, China) based on the principle that malate synthase catalyzes acetyl-CoA and glyoxylate to produce malate and CoA and transforms colorless DTNB into yellow TNB, with characteristic absorbance at 412 nm. The activity of sucrose synthase (SS, breakdown direction, EC 2.4.1.13) was determined by assay kits (Comin Biotechnology Co., Ltd., Jiangsu, China) based on the principle that sucrose synthase catalyzes the production of free fructose and UDPG from sucrose and UDP and the content of reducing sugar was determined by the 3,5-dinitrosalicylic acid method to reflect the enzyme activity. The activities of 6-phosphofructokinase (PFK, EC 2.7.1.40) and glucose-6-phosphate dehydrogenase (GPDH, EC 1.1.1.49) were determined by assay kits (Comin Biotechnology Co., Ltd., Jiangsu, China) and the enzyme activities were calculated by monitoring the rate of decrease in the absorbance value at 340 nm which is the position of the NADH absorption peak. A total of 18 samples (3 treatments × 2 sampling time points × 3 biological replications) were used for the analysis of each enzyme.

Real-Time Quantitative PCR Analysis

The samples used for real-time quantitative PCR (qRT-PCR) analysis were also conducted on the seeds (0.2 g for each sample, approximately 80 dry seeds or 50 imbibed seeds) at S0 and S20. Total RNA was extracted from the seed samples using Plant RNA Purification Reagent according to the manufacturer's instructions (Invitrogen, Waltham, MA, United States), and genomic DNA was removed using DNase I (TaKaRa, Shiga, Japan). The primers were designed with Primer 5.0 (**Supplementary Table 1**). The inner reference used the carrot gene *actin-7* (LOC108202619). Reverse transcription amplification was performed with a Goldenstar RT6 cDNA synthesis kit ver 2 (Tsingke Biotechnology Co., Ltd., Shanghai, China). qRT-PCR was performed with 2 × T5 Fast qRT-PCR Mix (SYBR Green I) (Tsingke Biotech, Beijing, China) on a CFX Connect Real-Time PCR Detection System (Bio-Rad, Hercules, CA, United States). The qRT-PCR amplification parameters were as follows: 95°C for 1 min; then 40 cycles of 95°C for 15 s, 60°C for 15 s, and 72°C for 30 min; followed by 1 cycle of 95°C for 5 s, 60°C for 1 min, and 50°C for 30 s. The relative quantitative results were calculated by the $2^{-\Delta\Delta Ct}$ method according to the Ct values. Three biological replications were performed for each reaction.

[1]https://www.uniprot.org/proteomes/UP000077755

[2]http://www.blast2go.com/b2ghome

[3]http://geneontology.org/

[4]http://www.genome.jp/kegg/

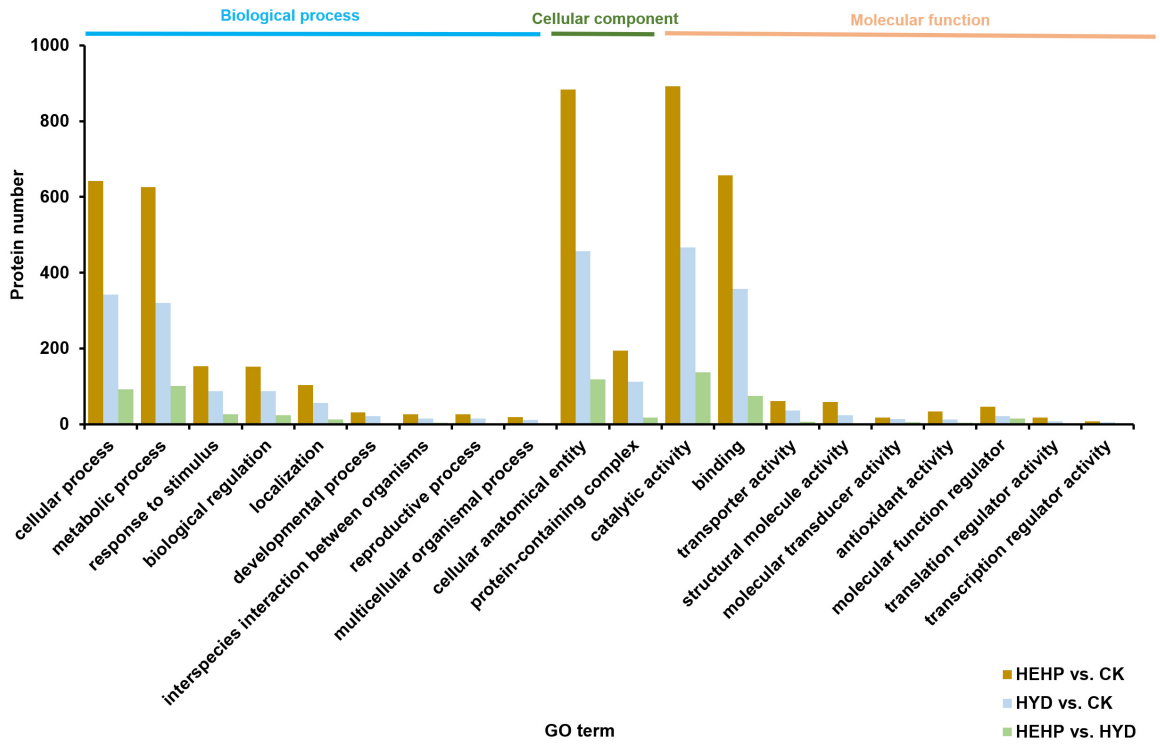


FIGURE 4 | Gene ontology (GO) annotation of DAPs in HEHP vs. CK, HYD vs. CK and HEHP vs. HYD.

Statistical Analysis

Statistical analysis of the germination indicators, enzyme activities and gene expression data were carried out based on the analysis of variance (ANOVA) test with IBM SPSS Statistics 19. Duncan's multiple range test was used to compare the mean values in different treatments.

RESULTS

Effect of Priming Treatments on Carrot Seed Germination

As shown in **Figure 2**, carrot seed germination was greatly improved by priming treatments. HYD and HEHP resulted in a significant difference in the germination indicators. The FW, RL, GI and VI of HEHP were increased by 142, 218, 206, and 641% compared with those of CK, respectively (**Figure 2A**). Moreover, all the germination indicators of HEHP were significantly superior to those of HYD, indicating that HEHP was able to accelerate germination more adorably than single HYD. Germination was observed at 24 h after sowing in the HEHP and HYD treatments, while CK germinated at 72 h after sowing (**Figure 2B**).

Identification and Quality Control of Identified Proteins

A total of 29,732 peptides and 4,936 proteins were identified in carrot seeds by MS and database searches. Among the identified proteins, proteins with molecular weights of 21–41 kDa accounted for 33%, followed by proteins with molecular weights of 41–61 and 1–21 kDa (**Supplementary Figure 1A**). The coverage range with the most protein quantity distribution was 20∼40, 1∼5, and 10∼20, accounting for 24, 21, and 20%, respectively (**Supplementary Figure 1B**). The number of proteins decreased with the increase in the number of peptides covering the protein (**Supplementary Figure 1C**), and the length of the peptide was generally between 7 and 13 (**Supplementary Figure 1D**). Principal component analysis (PCA) showed that the repeatability between the samples was high and the proteome profiles of the HEHP and HYD treatments differed from those of the control seeds, but there was a similarity between the HEHP and HYD treatments (**Supplementary Figure 1E**).

Statistics of Abundant Proteins

As shown in **Figure 3A** and **Supplementary Table 2**, with the screening threshold of FC $>$ 1.5 or $<$ 0.67 and P value $<$ 0.05, the maximum number of DAPs appeared between CK and HEHP (1514 DAPs including 597 upregulated and 917 downregulated), while the minimum number of DAPs appeared between HYD and HEHP (205 DAPs including 121 upregulated and 84 downregulated), indicating that the HEHP treatment can induce the differential expression of more proteins on the basis of the HYD treatment. Between CK and HYD, there were 787 DAPs, including 308 upregulated and 479 downregulated DAPs. The Venn diagram showed that 86 DAPs appeared simultaneously among

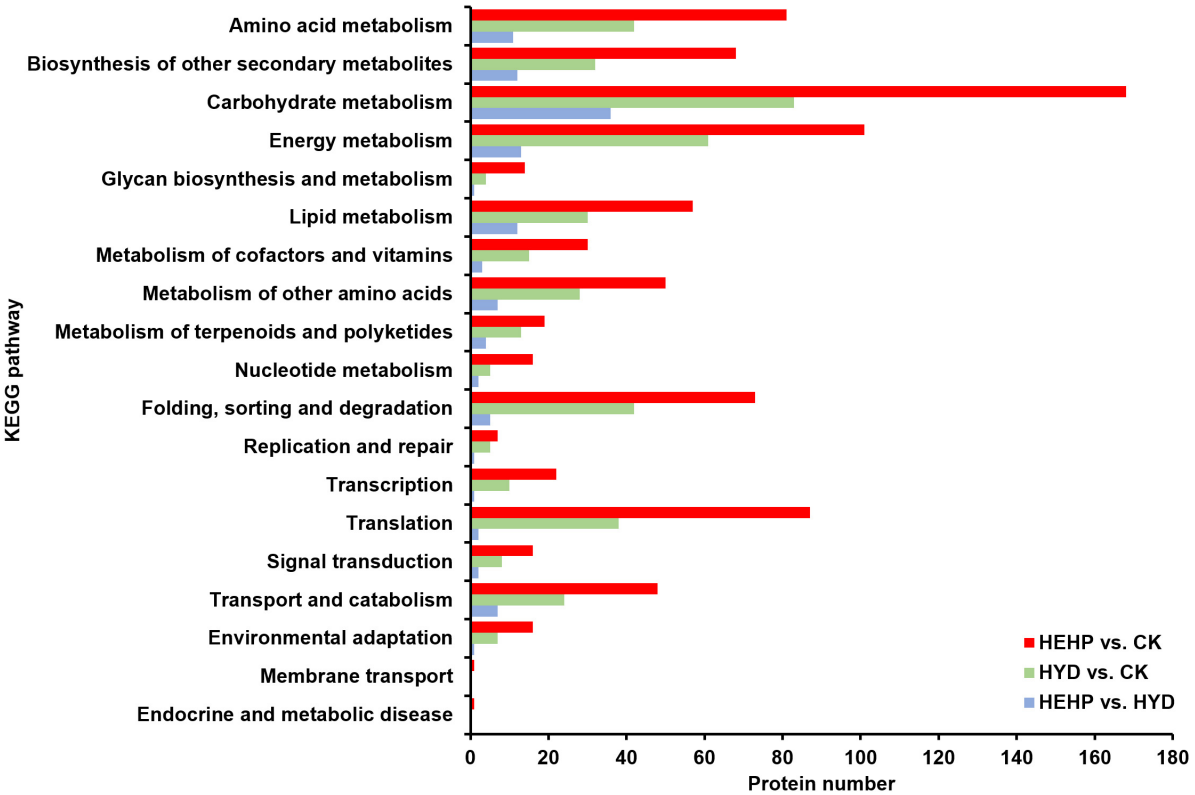


FIGURE 5 | Kyoto Encyclopedia of Genes and Genomes (KEGG) annotation of DAPs in HEHP vs. CK, HYD vs. CK and HEHP vs. HYD.

all the groups (HEHP vs. CK, HYD vs. CK and HEHP vs. HYD), while 661, 38 and 11 DAPs specifically appeared in the HEHP vs. CK, HYD vs. CK and HEHP vs. HYD groups, respectively (**Figure 3B**).

Gene Ontology and Kyoto Encyclopedia of Genes and Genomes Analysis of Abundant Proteins

Annotation of DAPs was performed using GO to clarify the functional distribution (**Figure 4**). GO analysis showed that among all three comparisons, the GO term with the maximum DAPs was "catalytic activity" (892 DAPs in HEHP vs. CK, 467 DAPs in HYD vs. CK, 137 DAPs in HEHP vs. CK) in the molecular function category, "cellular process" (642 DAPs in HEHP vs. CK, 343 DAPs in HYD vs. CK, 92 DAPs in HEHP vs. CK) and "metabolic process" (626 DAPs in HEHP vs. CK, 320 DAPs in HYD vs. CK, 101 DAPs in HEHP vs. CK) in the biological process category, and "cellular anatomical entity" (883 DAPs in HEHP vs. CK, 457 DAPs in HYD vs. CK, 118 DAPs in HEHP vs. CK) in the cellular component category.

KEGG pathway annotation showed that among all the three comparisons, pathways with the most DAPs were carbohydrate metabolism (168 DAPs in HEHP vs. CK, 83 DAPs in HYD vs. CK, 36 DAPs in HEHP vs. CK) and energy metabolism (101 DAPs in HEHP vs. CK, 61 DAPs in HYD vs. CK, 13 DAPs in HEHP vs. CK) (**Figure 5**).

To reveal the main metabolic pathways in which the DAPs participate, KEGG pathway enrichment analysis was also performed for upregulated and downregulated DAPs. The enriched KEGG pathways of each group are shown in **Supplementary Table 3**. For the upregulated DAPs, "ribosome" and "ribosome biogenesis in eukaryotes" were enriched in the HEHP vs. CK comparison (**Supplementary Table 3A**), while the "proteasome" was enriched in HYD vs. CK (**Supplementary Table 3B**). For the downregulated DAPs, "photosynthesis" was enriched in all three comparisons (**Supplementary Tables 3D,E,F**). "Photosynthesis - antenna proteins" was enriched in both the HEHP vs. CK and HEHP vs. HYD comparisons (**Supplementary Tables 3D,F**). In addition, "phenylpropanoid biosynthesis" was enriched in HEHP vs. CK (**Supplementary Table 3D**), and "carbon fixation in photosynthetic organisms" was enriched in HYD vs. CK (**Supplementary Table 3E**). In the HEHP vs. HYD comparisons, downregulated DAPs were also enriched in "starch and sucrose metabolism" (**Supplementary Table 3F**).

Differentially Abundant Proteins Involved in the Pathways Related to Protein Synthesis and Degradation

"Ribosome" was the most enriched KEGG pathway for the upregulated DAPs in the HEHP vs. CK comparison, which had 40 upregulated DAPs in total. The proteins involved

in the "ribosome" between HEHP and CK are shown in **Table 1**. DAPs increased in abundance that were involved in "ribosome" are mainly ribosomal proteins, including nine 40S ribosomal proteins and five 60S ribosomal proteins. The DAP with the largest FC was a ribosomal protein (A0A175YFY1). Additionally, among the four identified S-adenosylmethionine synthases (SAMs), three SAMs (A0A169WU65, A0A175YR74, A0A164YWI9) were found to have increased abundance in both the HYD and HEHP treatment, and the abundance of A0A175YR74 in the HEHP treatment was higher than that in the HYD treatment (**Supplementary Table 2**).

The most enriched KEGG pathway of the upregulated DAPs in HYD vs. CK was the "proteasome" with nine proteins in total, which indicates that HYD may promote the advancement of protein selective degradation. As shown in **Table 2**, more than half of the DAPs were identified as PCI domain-containing proteins, among which the DAP with the largest FC was a PCI domain-containing protein (A0A166FNJ4). Moreover, one 26S proteasome subunit protein (A0A164WP40) was identified. The ubiquitin-proteasome pathway is the main approach of intracellular protein degradation. We also found that one ubiquitin-activating enzyme (A0A166DB90), one ubiquitin ligase (A0A164Z661), and three ubiquitin hydrolases (A0A166FN67, A0A166EMX6, A0A175YJK6) were also upregulated by priming treatments (**Supplementary Table 2**).

TABLE 1 | Upregulated DAPs involved in ribosome (pathway ID: map03010) in HEHP vs. CK.

Accession	Description	FC (HEHP/CK)
A0A175YFY1	Ribosomal protein	2.873
A0A164XE28	40S ribosomal protein S12	2.422
A0A164ZZJ0	60S ribosomal protein L36	2.005
A0A165XFF0	S10_plectin domain-containing protein	1.904
A0A162AC45	S5 DRBM domain-containing protein	1.810
A0A166E4R6	60S ribosomal protein L36	1.670
A0A166AHI0	Ribosomal_L2_C domain-containing protein	1.661
A0A162BA60	40S ribosomal protein S26	1.635
A0A166D341	40S ribosomal protein S27	1.629
A0A161XAX4	40S ribosomal protein S3a	1.622
A0A165YZI7	40S ribosomal protein S8	1.611
A0A164USI6	Ribosomal_L28e domain-containing protein	1.610
A0A162AFC4	S5 DRBM domain-containing protein	1.581
A0A164ZFD9	40S ribosomal protein S3a	1.579
A0A175YRW0	40S ribosomal protein S8	1.568
A0A165YHA7	60S ribosomal protein L6	1.562
A0A166EPJ3	40S ribosomal protein SA	1.543
A0A161ZI33	Ribosomal_L18e/L15P domain-containing protein	1.537
A0A175YAJ3	60S acidic ribosomal protein P0	1.529
A0A175YPB6	60S ribosomal protein L29	1.528
A0A166ID45	40S ribosomal protein S25	1.526
A0A166FFE4	Ribosomal_S13_N domain-containing protein	1.522
A0A164WX80	Ubiquitin-like domain-containing protein	1.511

Uncharacterized proteins are not included in the table, the same below.

TABLE 2 | Upregulated DAPs involved in proteasomes (pathway ID: map03050) in HYD vs. CK.

Accession	Description	FC (HYD/CK)
A0A166FNJ4	PCI domain-containing protein	1.769
A0A166DU08	PCI domain-containing protein	1.723
A0A175YJ52	MPN domain-containing protein	1.672
A0A164WP40	26S proteasome regulatory subunit RPN11	1.656
A0A175YN30	PCI domain-containing protein	1.620
A0A164Z492	AAA domain-containing protein	1.620
A0A166GAD0	PCI domain-containing protein	1.592
A0A162BAD4	PCI domain-containing protein	1.563
A0A161XSF4	AAA domain-containing protein	1.521

Differentially Abundant Proteins Involved in Photosynthesis

Differentially abundant proteins that decreased in abundance were significantly enriched in "photosynthesis" among all three comparisons (36 DAPs in HEHP vs. CK, 23 DAPs in HYD vs. CK, 7 DAPs in HEHP vs. HYD), indicating that the expression of photosynthesis-related proteins was temporarily inhibited in the priming treatment. In the HYD vs. CK comparison, the identified DAPs involved in photosynthesis mainly contained photosystem II, photosystem I, ATP synthase subunit and ferredoxin-NADP reductase proteins (**Table 3**), while in the HEHP vs. CK comparison, in addition to the proteins mentioned

TABLE 3 | Downregulated DAPs involved in photosynthesis (pathway ID: map00195) in HYD vs. CK.

Accession	Description	FC (HYD/CK)
A0A166ACY1	23 kDa subunit of oxygen evolving system of photosystem II	0.371
A0A164UZK5	23 kDa subunit of oxygen evolving system of photosystem II	0.391
A0A175YCN0	ATP synthase subunit alpha	0.470
A0A175YC95	ATP synthase subunit beta	0.536
A0A165Y0P6	Plastocyanin	0.552
A0A165Y8N7	Photosystem II 10 kDa polypeptide, chloroplastic	0.560
A0A162AF45	Plastoquinol–plastocyanin reductase	0.561
A0A175YED7	Photosystem II protein D1	0.565
A0A175YDB1	Photosystem I P700 chlorophyll a apoprotein A2	0.576
A0A175YDE5	Photosystem II CP43 reaction center protein	0.578
A0A164XM95	Ferredoxin–NADP reductase, chloroplastic	0.590
A0A175YCC1	Photosystem II CP47 reaction center protein	0.622
A0A175YEF6	Photosystem I P700 chlorophyll a apoprotein A1	0.635
A0A175YCP3	Photosystem II D2 protein	0.640
A0A175YCE7	Photosystem I iron-sulfur center	0.643

above, the identified DAPs also contained cytochrome proteins (**Table 4**). In the HEHP vs. HYD comparison, the identified DAPs enriched in photosynthesis contained photosystem II, ATP synthase subunit and plastocyanin reductase proteins (**Table 5**).

TABLE 4 | Downregulated DAPs involved in photosynthesis (pathway ID: map00195) in HEHP vs. CK.

Accession	Description	FC (HEHP/CK)
A0A164UZK5	23 kDa subunit of oxygen evolving system of photosystem II	0.270
A0A166ACY1	23 kDa subunit of oxygen evolving system of photosystem II	0.294
A0A162AF45	Plastoquinol–plastocyanin reductase	0.315
A0A165Y8N7	Photosystem II 10 kDa polypeptide, chloroplastic	0.368
A0A175YCN0	ATP synthase subunit alpha	0.385
A0A164XM95	Ferredoxin–NADP reductase, chloroplastic	0.416
A0A175YED7	Photosystem II protein D1	0.434
A0A175YDB1	Photosystem I P700 chlorophyll a apoprotein A2	0.440
A0A175YDE5	Photosystem II CP43 reaction center protein	0.446
A0A165Y0P6	Plastocyanin	0.446
A0A175YDD0	ATP synthase CF0 B subunit	0.447
A0A175YC95	ATP synthase subunit beta	0.465
A0A166A8H7	Photosystem II 10 kDa polypeptide, chloroplastic	0.494
A0A175YCE7	Photosystem I iron-sulfur center	0.496
A0A175YCI3	Cytochrome f	0.515
A0A175YCP3	Photosystem II D2 protein	0.519
A0A175YEF6	Photosystem I P700 chlorophyll a apoprotein A1	0.521
A0A175YCS2	Cytochrome b6	0.536
A0A175YCC1	Photosystem II CP47 reaction center protein	0.560
A0A161Y8Y8	Ferredoxin–NADP reductase, chloroplastic	0.574
A0A161Y432	PSI-K	0.577
A0A161ZVZ6	Ferredoxin–NADP reductase, chloroplastic	0.583
A0A164VGA4	23 kDa subunit of oxygen evolving system of photosystem II	0.594
A0A166CIQ3	PSI subunit V	0.660

TABLE 5 | Downregulated DAPs involved in photosynthesis (pathway ID: map00195) in HEHP vs. HYD.

Accession	Description	FC (HEHP/HYD)
A0A162AF45	Plastoquinol–plastocyanin reductase	0.562
A0A175YCS2	Cytochrome b6	0.601
A0A166A8H7	Photosystem II 10 kDa polypeptide, chloroplastic	0.651
A0A165Y8N7	Photosystem II 10 kDa polypeptide, chloroplastic	0.657
A0A175YDD0	ATP synthase CF0 B subunit	0.667

Differentially Abundant Proteins Involved in Carbohydrate Metabolism

As shown in the analysis of KEGG pathway annotation, DAPs annotated as "carbohydrate metabolism" were the most among all three comparisons (**Figure 5**). The Venn analysis of these DAPs among the three comparisons suggests that almost all the DAPs in the HYD vs. CK and HEHP vs. HYD comparisons were included in the HEHP vs. CK comparison, except three DAPs in the HYD vs. CK comparison and two DAPs in the HYD vs. CK comparison (**Supplementary Figure 2**). This result indicates that the role of HEHP is realized by inducing more DAPs on the basis of HYD. From the DAPs annotated as "carbohydrate metabolism" (**Supplementary Table 4**), it is noteworthy that among the 14 DAPs encoding key enzymes of the glyoxylate cycle and respiratory metabolism, 12 DAPs increased in abundance in the priming treatments (**Figure 6**). One PFK (A0A166FRS9) was identified as upregulated DAP among all three comparisons. One isocitrate lyase (ICL: A0A165Z711) and one malic enzyme (ME: A0A166FRM4) were identified as upregulated DAPs in both the HEHP and HYD treatments compared with the control. Moreover, two succinate dehydrogenase subunits (SDH: A0A164VUP8 and A0A161XEK6), two SS (A0A175YLM7 and A0A164TL65), three fructose-bisphosphate aldolases (FPA: A0A161YFV4, A0A175YGJ3, A0A165YUW0), one GPDH (A0A164THB2), and one MS (A0A164XH24) were only significantly upregulated by HEHP treatment but were not upregulated by HYD treatment.

Activities of Key Enzymes Involved in Carbohydrate Metabolism

Based on the proteomic results, the activities of several key enzymes involved in carbohydrate metabolism were selected and further investigated, as shown in **Figure 7**. Except for SS, the activities of MS, GPDH and PFK were significantly increased by the HEHP and HYD treatments at S0, and for GPDH and PFK, the promoting effect of HEHP was significantly superior to HYD. In addition, the activities of all four enzymes were significantly increased by the HEHP treatment at S20, while for the activities of SS and PFK, there was no significant change between HYD and CK. The activities of all these enzymes in the HEHP treatment were also higher than those in the HYD treatment at S20.

Expression Pattern of Selected Differentially Abundant Proteins at the Transcriptional Level

To further explore the expression pattern of key genes encoding DAPs, ten DAPs were selected for qRT-PCR analysis at the transcriptional level from the seed samples at S0 and S20. As shown in **Figure 8A**, all of the selected genes were upregulated by HEHP treatment at S0, and the expression of some DAPs (A0A175YFY1, A0A166AHI0, A0A175YLM7, A0A166FRS9, A0A166IDX8, and A0A164THB2) at the transcriptional level in the HEHP treatment was significantly higher than that

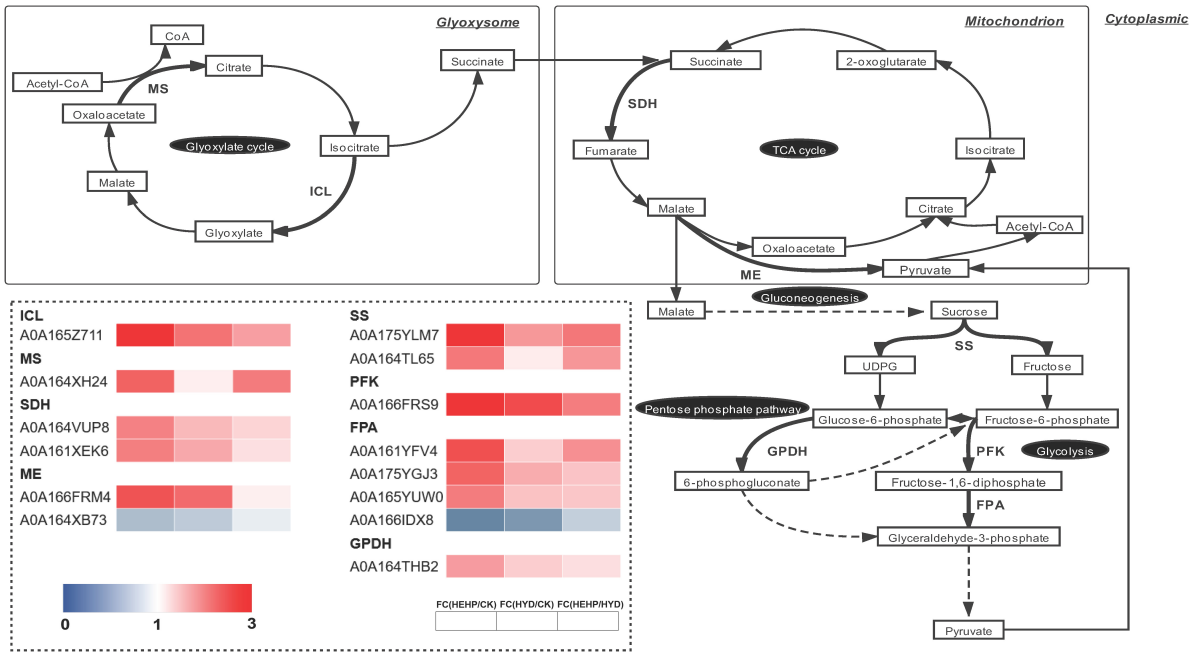


FIGURE 6 | Differentially abundant proteins involved in the glyoxylate cycle and respiratory metabolism. ICL, isocitrate lyase; MS, malate synthase; SDH, succinate dehydrogenase; ME, malic enzyme; SS, sucrose synthase; PFK, 6-phosphofructokinase; FPA, fructose-bisphosphate aldolase; GPDH, glucose 6-phosphate dehydrogenase.

in the HYD treatment. The expression of A0A164THB2 at the transcriptional level was not upregulated by HYD treatment. Moreover, it was found that the expression of five genes encoding A0A164WP40, A0A164XM95, A0A166FRM4, A0A175YLM7, and A0A166FRS9 was not upregulated by the priming treatments after imbibition at S20. From the correlation analysis between the results of the expression pattern of genes and the proteomic data, five DAPs (A0A175YFY1, A0A164WP40, A0A166AHI0, A0A175YLM7, and A0A166FRS9) showed a significant positive correlation with the gene expression data while only A0A166IDX8 showed a significant negative correlation with the gene expression data (**Figure 8B**).

DISCUSSION

Germination performance is an important criterion for seed lots as it takes into consideration seed vigor through the rate, timing, and uniformity of germination (Fabrissin et al., 2021). In the present work, the germination performance of carrot seeds was greatly promoted by HEHP treatment. Numerous seed priming techniques reported recently have applied exogenous growth regulators to obtain more significant effects, such as melatonin (Yan et al., 2020), ascorbic acid (Salemi et al., 2019) and some kinds of phytohormones (Hussain et al., 2016; Huang et al., 2020; Zhao et al., 2020; Chen et al., 2021). Nevertheless, the HEHP treatment used in our study is a physical method without using any exogenous growth regulator or chemical reagent, which means that this method entirely relies on activating the vigor of the seed itself to accelerate the germination process and will not cause any chemical substance penetration or residue on the seeds, thus, it is environmentally safe. Moreover, through applying EF treatment on the basis of HYD, the effect of seed priming was more remarkable compared with using HYD alone.

Priming Treatments Altered the Pattern of Protein Composition in Carrot Seeds

To the best of our knowledge, the present study is the first report on the response of the carrot seed proteome to seed priming by applying a TMT based proteomic technique. Both HYD and HEHP induced a large number of DAPs compared to the control seeds, while the HEHP treatment induced more DAPs than HYD, and DAPs between HEHP and HYD are mainly upregulated, which suggests that the HEHP treatment can further expand the effect of priming on the internal metabolic activities of seeds on the basis of HYD. Furthermore, compared with CK, DAPs induced by HEHP and HYD were mainly downregulated, indicating that priming treatments regulate seed germination largely by reducing the abundance of specific proteins.

Priming Treatments Induced Protein Synthesis and Degradation

Ribosomes are composed of ribosomal proteins and rRNA, which bind to mRNAs to participate in protein synthesis when seed germination begins (Zhang et al., 2016; Bai et al., 2020). Previous studies have suggested that biologically defective mutations in ribosomes tend to lead to a low germination rate (Abbasi et al., 2010; Maekawa et al., 2018). In our study, the upregulated DAPs between HEHP and CK were most

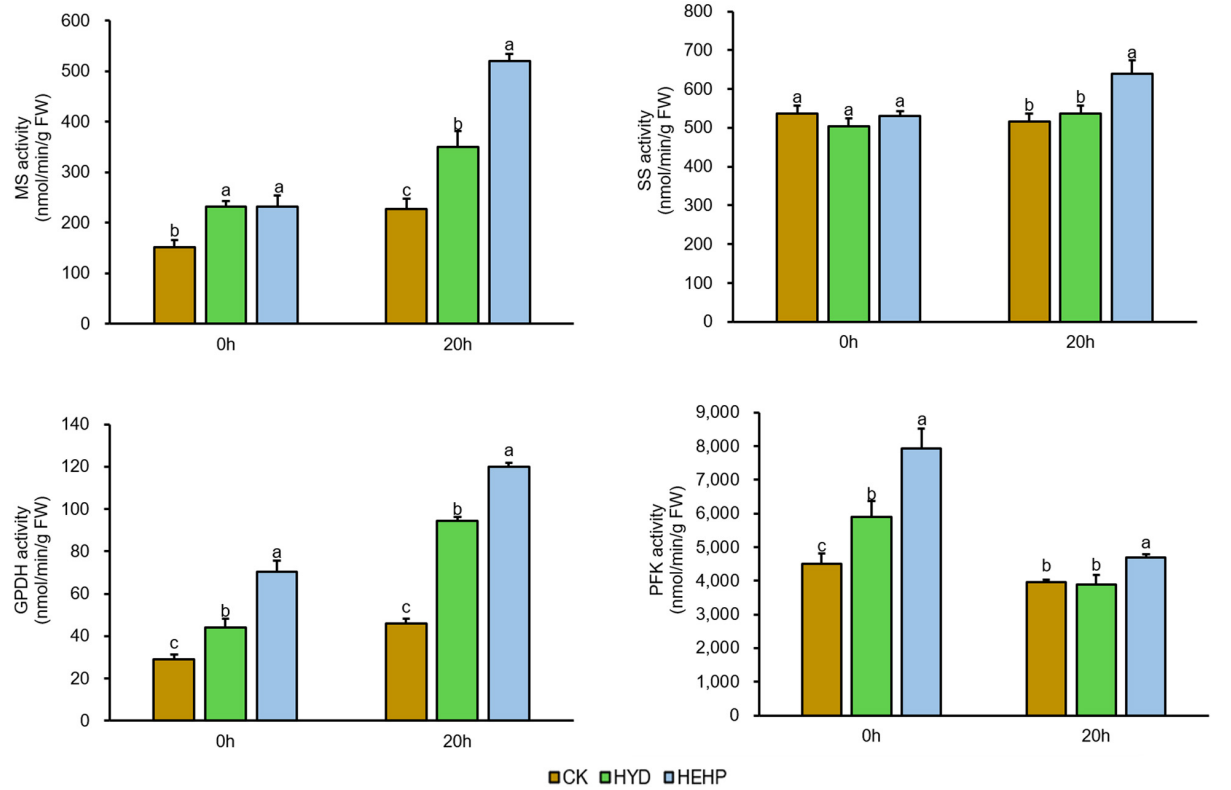


FIGURE 7 | Effects of different treatments on key enzyme activities. MS, malate synthase; SS, sucrose synthase; GPDH, glucose-6-phosphatedehydrogenase; PFK, 6-phosphofructokinase. Values represent the means ± SE from six biological replications. The different lowercase letters above the bars indicate significant differences in 95% probability level ($p < 0.05$, Duncan test was performed after ANOVA analysis).

enriched in "ribosome," and both ribosomal large subunit proteins and ribosomal small subunit proteins were found to significantly increase in abundance in HEHP, indicating that HEHP treatment plays a key role in improving the efficiency of protein synthesis. This result is consistent with previous research on cucumber (Zhang et al., 2017), rapeseed (Kubala et al., 2015) and *H. moellendorffii* (Li et al., 2019) seeds. In the present work, three SAMs were found to have increased abundance in priming treatments. Previous proteomic studies have also found that proteins related to methionine metabolism, including SAMs, increased accumulation during seed germination (Gallardo et al., 2002; Wang et al., 2012) and after priming (Fercha et al., 2013). SAMs use methionine to catalyze S-adenosylmethionine, which is involved in protein synthesis as a methyl donor (Takahashi et al., 2011). Previous study also pointed out that the recovery of seed vigor by seed priming is related to the maintenance of methionine metabolism (Yacoubi et al., 2013).

In addition to protein synthesis, selective protein degradation is also essential in the initiation process of seed germination (Muntz et al., 2001). The proteasome is a protease complex that degrades misfolded or damaged proteins, which can be ubiquitin-dependent (proteins are first labeled by ubiquitin) and directly involved in regulating the level of some key proteins (Sadanandom et al., 2012). The present results suggested that the upregulated DAPs between HYD and CK were most enriched in "proteasome," and DAPs identified as proteasome proteins, ubiquitin-activating enzymes, ubiquitin ligases, and ubiquitin hydrolases were significantly increased in abundance in priming treatments, which implied that the ubiquitin-proteasome pathway was advanced in priming treatment. A previous study also showed that the abundance of the 26S proteasome component increased when seeds germinated under suitable conditions, and proteasome activity was high during phase I to the end of phase II of germination (Xia et al., 2018). Additionally, five PCI domain-containing proteins were identified as upregulated DAPs. The PCI domain is frequently found in the subunits of the proteasome, and is related to the regulation of protein ubiquitination and degradation (Zhu et al., 2018).

Priming Treatments Decreased the Accumulation of Photosynthesis-Related Proteins in Carrot Seeds

In previous studies on seed germination, photosynthesis has rarely received much attention, because seed germination is an autotrophic process that mobilizes seed reserves to maintain growth, and photosynthesis does not occur (Deruyffelaere et al., 2015; Bai et al., 2020). In our study, it is noteworthy that in the HYD and HEHP treatments, we found that many

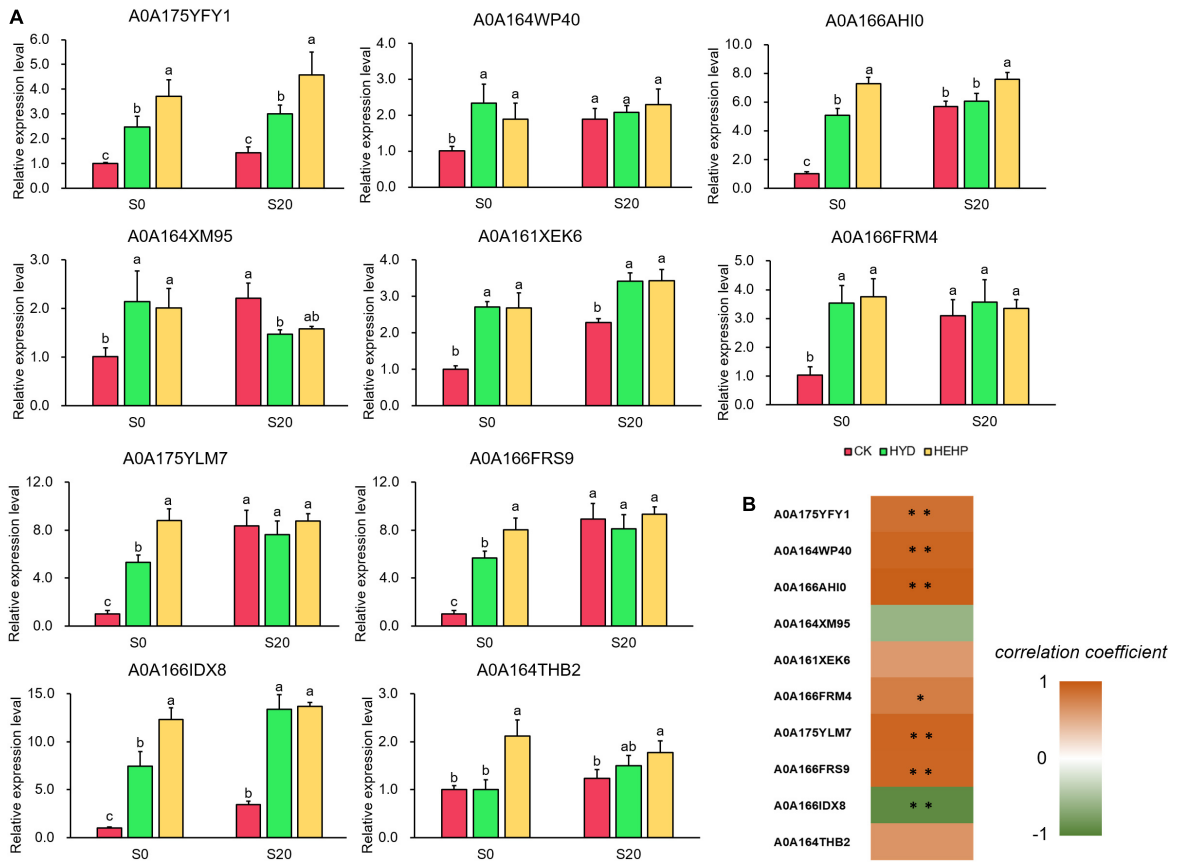


FIGURE 8 | Real-time quantitative PCR (qRT-PCR) analysis of 10 selected DAPs at S0 and S20. **(A)** Expression patterns at the transcriptional level of 10 selected DAPs from qRT-PCR; **(B)** correlation analysis between the results of qRT-PCR and the proteome. Spearman correlation coefficient was used. *Means significant at the level of $p < 0.05$. **Means significant at the level of $p < 0.01$. The different lowercase letters above the bars indicate significant differences in 95% probability level ($p < 0.05$, Duncan test was per-formed after ANOVA analysis).

DAPs involved in "photosynthesis" decreased in abundance, and the expression of some photosynthetic proteins in HEHP was even lower than that in HYD, such as A0A162AF45, A0A175YCS2, A0A166A8H7, A0A165Y8N7 and A0A175YDD0. Recent research on barley showed that photosynthesis-associated genes were downregulated during early germination (Zhu et al., 2020), which is similar to the present results. One possible explanation for this phenomenon is that due to the limited storage reserves of carrot seeds, the treatments of rapid germination (HEHP and HYD) need to concentrate all the material and energy supply in the early stage of germination to meet the requirements of germination. Therefore, the expression of photosynthesis-related proteins for heterotrophic growth is temporarily inhibited.

Priming Treatments, Especially Hydro-Electro Hybrid Priming, Initiated Storage Reserves Utilization and Respiratory Metabolism

Dry seeds contain all the components required for germination and seedling establishment until the seedlings reach an autotrophic state (Bai et al., 2020). Proteins related to energy metabolism may show increased abundance by seed priming (Araujo et al., 2021). Carrot seeds have an endosperm with a high lipid content (mainly in the form of triacylglycerol), which is degraded to provide carbon skeleton and energy for seedling growth during germination (Ngo-Duy et al., 2009; Pandey et al., 2019). A previous study in *Arabidopsis thaliana* showed that most sucrose used for respiratory metabolism is produced by storage lipid degradation rather than other soluble sugars in seeds (Pritchard et al., 2002). The utilization of triacylglycerol (TAG) in seeds depends on the glyoxylate cycle during germination, which uses β-acetyl-CoA produced by oxidation of fatty acids is converted to gluconeogenesis and then used for energy supply during germination (Eastmond and Graham, 2001). ICL and MS are the key enzymes in the glyoxylate cycle, and both were identified as upregulated DAPs in the HEHP treatment, which suggests that HEHP has prepared key substances for the utilization of seed storage. We further determined the MS activity and found that it increased significantly by the priming treatments. Existing studies have also found the expression and activities of ICL and MS were increased during germination (Bellieny-Rabelo et al., 2016; Faraoni et al., 2019). When stored

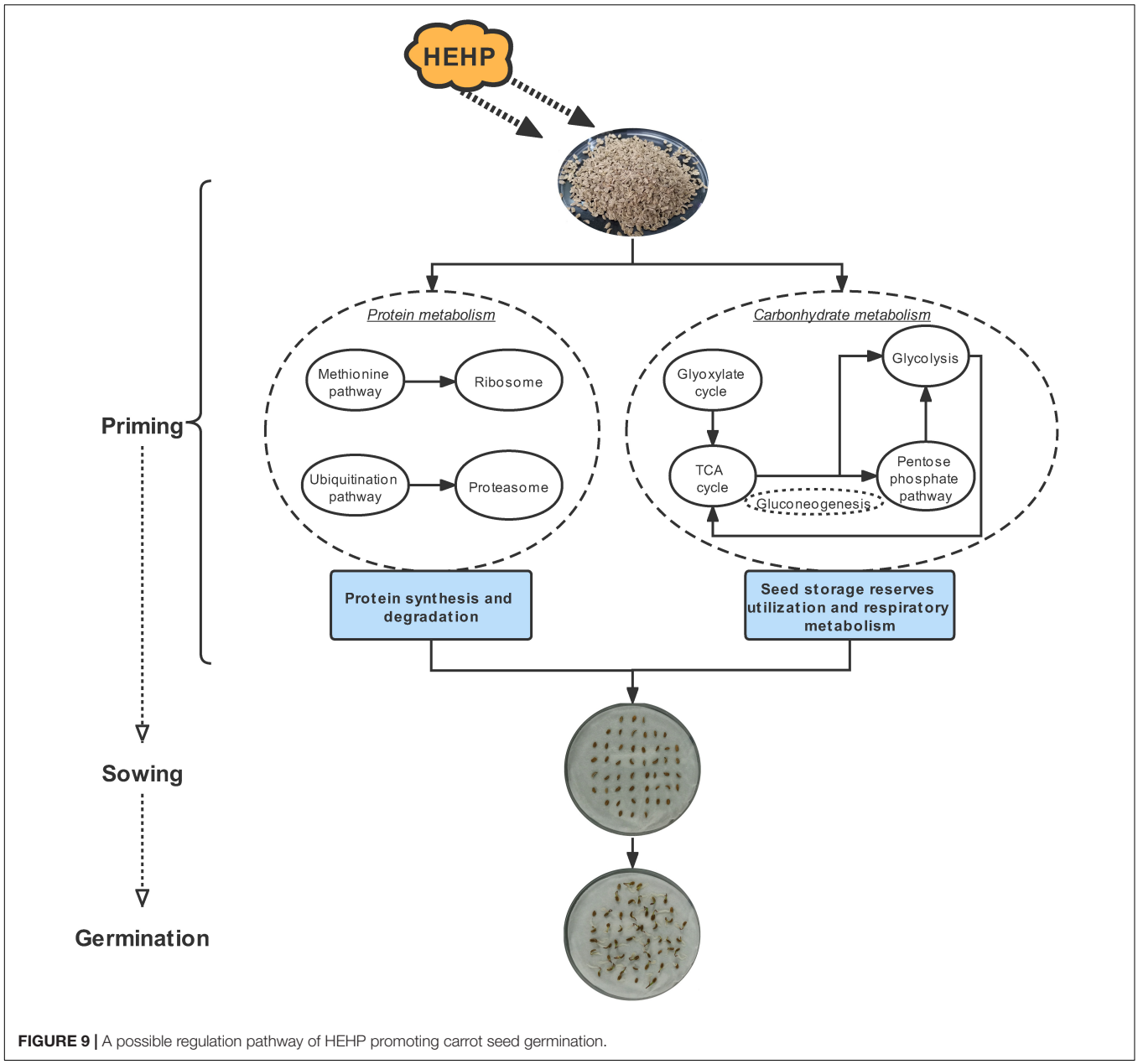


FIGURE 9 | A possible regulation pathway of HEHP promoting carrot seed germination.

lipids are converted to sucrose after gluconeogenesis, they can be decomposed into glucose and fructose by SS and then enter glycolysis or the pentose phosphate pathway (Yu et al., 2014; Huang et al., 2021). SS is also involved in the synthesis of the cell wall, which is an essential process underlying cellular expansion during germination (Wei et al., 2015; Bellieny-Rabelo et al., 2016). Our results showed that two SSs increased in abundance by HEHP and the activity of SS was also increased by HEHP treatment during imbibition. A recent study also indicated that SS participates in the release of seed dormancy (Liu et al., 2020). Early activation of respiratory pathways was also found, which is responsible for providing energy for cell activities and the carbon skeleton of macromolecular biosynthesis during seed germination. Both PFK and FPA were identified as upregulated DAPs in the HEHP treatment, which are involved in glycolysis and PFK is one of the rate-limiting enzymes of this pathway. PFK activity was also found to be increased by HEHP treatment at both S0 and S20. It has been proposed that glycolysis-related enzymes increase in expression during early germination and the energy needs of germination seem to be met mostly by glycolysis (Yang et al., 2007; Yu et al., 2014; Xu et al., 2016). As a key enzyme of the pentose phosphate pathway, GPDH was identified as an upregulated DAP in the HEHP treatment, the activity of which was also increased by the HEHP treatment. The pentose phosphate pathway is also considered to play a key role in energy supply in the early stage of seed germination (Huang et al., 2021). Both SDH and ME participate in the TCA cycle in mitochondria, which enables the intermediate products of the TCA cycle such

as malate to be completely oxidized and decomposed and can provide flexibility for the metabolism of phosphoenolpyruvate and malate (Lincoln and Eduardo, 2015). In brief, the increased abundance of these enzymes in priming treatments especially the HEHP treatment reflects that priming has induced material preparation for efficient energy supply during germination in advance. In addition, it is noteworthy that HYD did not increase the abundance of some proteins such as SDH subunits, SSs, FPAs and MA, while HEHP significantly increased the abundance of these proteins, and the results of enzyme activity determination were similar, which is another possible explanation for the more remarkable effect of HEHP than that of HYD.

The Effect of Priming Is Mainly Reflected Before Sowing

The results of the expression pattern of the genes encoding DAPs by qRT-PCR revealed that the expression level of some selected genes was not upregulated by the priming treatments after imbibition, indicating that the effect of priming is mainly reflected in regulating gene expression before sowing. When the carrot seeds are sown and begin to imbibe, the activation effect on gene expression will gradually dissipate, which means priming has completed its mission. The abundance of one DAP (A0A166IDX8) showed significant a negative correlation with gene expression data, which is also common in previous studies and is usually attributed to temporal differences in expression or posttranslational modification (Guo et al., 2012; Li et al., 2019; Yan and Mao, 2021).

CONCLUSION

The seed priming treatments used in the current study substantially shortened the germination process and improved the germination performance of carrot seeds. The application of HEHP contributed to promoting the germination of carrot seeds to a greater extent than hydropriming. Proteomic analysis revealed that the priming treatments significantly altered the protein composition of carrot seeds. The upregulated DAPs were mainly enriched in the pathways related to protein synthesis and degradation, while the downregulated DAPs were mainly enriched in photosynthesis-related pathways. Furthermore, given that the maximum DAPs were annotated in carbohydrate metabolism, we focused on the critical proteins involved in the glyoxylate cycle, sucrose metabolism and glycolysis induced by the priming treatments and found that some proteins encoding key enzymes of the above pathways showed enhanced abundance in priming treatments, especially the HEHP treatment, such as ICL, MS, PFK, FPA, SDH and ME. Meanwhile, the activities of the key enzymes were also enhanced by the priming treatments, and the HEHP treatment was able to improve enzyme activities to a greater extent. The effect of priming is mainly reflected before sowing and the activation effect on gene expression will gradually dissipate after imbibition. Overall, the optimal effect of HEHP is to regulate the synthesis and decomposition of proteins in seeds to meet the requirements of germination and initiate the utilization of seed storage reserves and respiratory metabolism (**Figure 9**). Future research is needed to focus on the changes in metabolites and other components of the above pathways in primed seeds during germination and to keep paying attention to the inhibitory effect of seed priming on photosynthesis-related pathways.

DATA AVAILABILITY STATEMENT

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found below: ProteomeXchange via PRIDE (Project accession: PXD030133).

AUTHOR CONTRIBUTIONS

SZ and DH designed the experiments. SZ and HZ carried out the experiments. SZ and XP took the samples. SZ and YJ analyzed the data and prepared the tables and figures. SZ drafted the manuscript. YJ and DH improved the manuscript in English. All authors contributed to the article and approved the submitted version.

FUNDING

This research was funded by Shanghai Municipal Agricultural Commission Program, China (No. 2019-02-08-00-08-F01124).

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpls.2022.824439/full#supplementary-material

REFERENCES

Abbasi, N., Kim, H. B., Park, N. I., Kim, H. S., Kim, Y. K., Park, Y. I., et al. (2010). Apum23, a nucleolar Puf domain protein, is involved in pre-ribosomal Rna processing and normal growth patterning in *Arabidopsis*. *Plant J.* 64, 960–976. doi: 10.1111/j.1365-313X.2010.04393.x

Araujo, G. D. S., Lopes, L. D. S., Paula-Marinho, S. D. O., Mesquita, R. O., Nagano, C. S., Vasconcelos, F. R., et al. (2021). H_2O_2 priming induces proteomic responses to defense against salt stress in maize. *Plant Mol. Biol.* 106, 33–48. doi: 10.1007/s11103-021-01127-x

Bai, B., van der Horst, S., Cordewener, J. H. G., America, T., Hanson, J., and Bentsink, L. (2020). Seed-stored mRNAs that are specifically associated to monosomes are translationally regulated during germination. *Plant Physiol.* 182, 378–392. doi: 10.1104/pp.19.00644

Bellieny-Rabelo, D., de Oliveira, E. A. G., Ribeiro, E. D., Costa, E. P., Oliveira, A. E. A., and Venancio, T. M. (2016). Transcriptome analysis uncovers key regulatory and metabolic aspects of soybean embryonic axes during germination. *Sci. Rep.* 6:36009. doi: 10.1038/srep36009

Bewley, J. D. (1997). Seed germination and dormancy. *Plant Cell* 9, 1055–1066. doi: 10.1105/tpc.9.7.1055

Carrera, E., Holman, T., Medhurst, A., Peer, W., Schmuths, H., Footitt, S., et al. (2007). Gene expression profiling reveals defined functions of the ATP-binding cassette transporter COMATOSE late in phase II of germination. *Plant Physiol.* 143, 1669–1679. doi: 10.1104/pp.107.096057

Chakma, S. P., Chileshe, S. M., Thomas, R., and Krishna, P. (2021). Cotton seed priming with brassinosteroid promotes germination and seedling growth. *Agron. Basel* 11:566. doi: 10.3390/agronomy11030566

Chen, H. N., Tao, L. Y., Shi, J. M., Han, X. R., and Cheng, X. G. (2021). Exogenous salicylic acid signal reveals an osmotic regulatory role in priming the seed germination of *Leymus chinensis* under salt-alkali stress. *Environ. Exper. Bot.* 188:104498. doi: 10.1016/j.envexpbot.2021.104498

Dekkers, B. J. W., Pearce, S., van Bolderen-Veldkamp, R. P., Marshall, A., Widera, P., Gilbert, J., et al. (2013). Transcriptional dynamics of two seed compartments with opposing roles in *Arabidopsis* seed germination. *Plant Physiol.* 163, 205–215. doi: 10.1104/pp.113.223511

Deruyffelaere, C., Bouchez, I., Morin, H., Guillot, A., Miquel, M., Froissard, M., et al. (2015). Ubiquitin-mediated proteasomal degradation of oleosins is involved in oil body mobilization during post-germinative seedling growth in *Arabidopsis*. *Plant Cell Physiol.* 56, 1374–1387. doi: 10.1093/pcp/pcv056

Eastmond, P. J., and Graham, I. A. (2001). Re-examining the role of the glyoxylate cycle in oilseeds. *Trends Plant Sci.* 6, 72–77. doi: 10.1016/s1360-1385(00)01835-5

Espanany, A., Fallah, S., and Tadayyon, A. (2016). Seed priming improves seed germination and reduces oxidative stress in black cumin (*Nigella sativa*) in presence of cadmium. *Industr. Crops Prod.* 79, 195–204. doi: 10.1016/j.indcrop.2015.11.016

Fabrissin, I., Sano, N., Seo, M., and North, H. M. (2021). Ageing beautifully: can the benefits of seed priming be separated from a reduced lifespan trade-off? *J. Exper. Bot.* 72, 2312–2333. doi: 10.1093/jxb/erab004

Faraoni, P., Sereni, E., Gnerucci, A., Cialdai, F., Monici, M., and Ranaldi, F. (2019). Glyoxylate cycle activity in *Pinus pinea* seeds during germination in altered gravity conditions. *Plant Physiol. Biochem.* 139, 389–394. doi: 10.1016/j.plaphy.2019.03.042

Fercha, A., Capriotti, A. L., Caruso, G., Cavaliere, C., Gherroucha, H., Samperi, R., et al. (2013). Gel-free proteomics reveal potential biomarkers of priming-induced salt tolerance in durum wheat. *J. Proteom.* 91, 486–499. doi: 10.1016/j.jprot.2013.08.010

Finch-Savage, W. E., and Bassel, G. W. (2016). Seed vigour and crop establishment: extending performance beyond adaptation. *J. Exper. Bot.* 67, 567–591. doi: 10.1093/jxb/erv490

Finch-Savage, W. E., and Leubner-Metzger, G. (2006). Seed dormancy and the control of germination. *New Phytol.* 171, 501–523. doi: 10.1111/j.1469-8137.2006.01787.x

Gallardo, K., Job, C., Groot, S., Puype, M., Demol, H., and Job, V. D. (2002). Proteomics of *Arabidopsis* seed germination. A comparative study of wild-type and gibberellin-deficient seeds. *Plant Physiol.* 129, 823–837. doi: 10.2307/4280508

Gallardo, K., Job, C., Groot, S. P. C., Puype, M., Demol, H., Vandekerckhove, J., et al. (2001). Proteomic analysis of *Arabidopsis* seed germination and priming. *Plant Physiol.* 126, 835–848. doi: 10.1104/pp.126.2.835

Gu, J. W., Chao, H. B., Gan, L., Guo, L. X., Zhang, K., Li, Y. H., et al. (2016). Proteomic dissection of seed germination and seedling establishment in *Brassica napus*. *Front. Plant Sci.* 7:1482. doi: 10.3389/fpls.2016.01482

Guo, G., Lv, D., Yan, X., Subburaj, S., Ge, P., Li, X., et al. (2012). Proteome characterization of developing grains in bread wheat cultivars (*Triticum aestivum* L.). *BMC Plant Biol.* 12:147. doi: 10.1186/1471-2229-12-147

Han, C., Zhen, S., Zhu, G., Bian, Y., and Yan, Y. (2017). Comparative metabolome analysis of wheat embryo and endosperm reveals the dynamic changes of metabolites during seed germination. *Plant Physiol. Biochem.* 115, 320–327. doi: 10.1016/j.plaphy.2017.04.013

Huang, L. P., Zhang, L., Zeng, R. E., Wang, X. Y., Zhang, H. J., Wang, L. D., et al. (2020). Brassinosteroid priming improves peanut drought tolerance via eliminating inhibition on genes in photosynthesis and hormone signaling. *Genes* 11:919. doi: 10.3390/genes11080919

Huang, X. L., Tian, T., Chen, J. Z., Wang, D., Tong, B. L., and Liu, J. M. (2021). Transcriptome analysis of Cinnamomum migao seed germination in medicinal plants of Southwest China. *BMC Plant Biol.* 21:270. doi: 10.1186/s12870-021-03020-7

Hussain, S., Khan, F., Hussain, H. A., and Nie, L. (2016). Physiological and biochemical mechanisms of seed priming-induced chilling tolerance in rice cultivars. *Front. Plant Sci.* 7:116. doi: 10.3389/fpls.2016.00116

Kubala, S., Garnczarska, M., Wojtyla, L., Clippe, A., Kosmala, A., Zmienko, A., et al. (2015). Deciphering priming-induced improvement of rapeseed (*Brassica napus* L.) germination through an integrated transcriptomic and proteomic approach. *Plant Sci.* 231, 94–113. doi: 10.1016/j.plantsci.2014.11.008

Li, F. H., Yu, P., Song, C. H., Wu, J. J., Tian, Y., Wu, X. F., et al. (2019). Differential protein analysis of *Heracleum moellendorffii* Hance seeds during stratification. *Plant Physiol. Biochem.* 145, 10–20. doi: 10.1016/j.plaphy.2019.10.002

Lincoln, T., and Eduardo, Z. (2015). *Plant Physiology*. Beijing: Science Press.

Liu, X., Huang, X., Kong, X. X., Zhang, J., Wang, J. Z., Yang, M. L., et al. (2020). Sucrose synthase is involved in the carbohydrate metabolism-based regulation of seed dormancy release in *Pyrus calleryana* Decne. *J. Hortic. Sci. Biotechnol.* 95, 590–599. doi: 10.1080/14620316.2020.1740612

Macovei, A., Pagano, A., Leonetti, P., Carbonera, D., Balestrazzi, A., and Araujo, S. S. (2017). Systems biology and genome-wide approaches to unveil the molecular players involved in the pre-germinative metabolism: implications on seed technology traits. *Plant Cell Rep.* 36, 669–688. doi: 10.1007/s00299-016-2060-5

Maekawa, S., Ishida, T., and Yanagisawa, S. (2018). Reduced expression of APUM24, encoding a novel rRNA processing factor, induces sugar-dependent nucleolar stress and altered sugar responses in *Arabidopsis thaliana*. *Plant Cell* 30, 209–227. doi: 10.1105/tpc.17.00778

Marcos-Filho, J. (2015). Seed vigor testing: an overview of the past, present and future perspective. *Sci. Agric.* 72, 363–374. doi: 10.1590/0103-9016-2015-0007

Marthandan, V., Geetha, R., Kumutha, K., Renganathan, V. G., Karthikeyan, A., and Ramalingam, J. (2020). Seed priming: a feasible strategy to enhance drought tolerance in crop plants. *Intern. J. Mol. Sci.* 21:8258. doi: 10.3390/ijms21218258

Muntz, K., Belozersky, M. A., Dunaevsky, Y. E., Schlereth, A., and Tiedemann, J. (2001). Stored proteinases and the initiation of storage protein mobilization in seeds during germination and seedling growth. *J. Exper. Bot.* 52, 1741–1752. doi: 10.1093/jexbot/52.362.1741

Nakabayashi, K., Okamoto, M., Koshiba, T., Kamiya, Y., and Nambara, E. (2005). Genome-wide profiling of stored mRNA in *Arabidopsis thaliana* seed germination: epigenetic and genetic regulation of transcription in seed. *Plant J.* 41, 697–709. doi: 10.1111/j.1365-313X.2005.02337.x

Nakao, Y., Sone, C., and Sakagami, J. I. (2020). Genetic diversity of hydro priming effects on rice seed emergence and subsequent growth under different moisture conditions. *Genes* 11:13. doi: 10.3390/genes11090994

Ngo-Duy, C.-C., Destaillats, F., Keskitalo, M., Arul, J., and Angers, P. (2009). Triacylglycerols of Apiaceae seed oils: composition and regiodistribution of fatty acids. *Eur. J. Lipid Sci. Technol.* 111, 164–169. doi: 10.1002/ejlt.200800178

Nonogaki, H., Bassel, G. W., and Bewley, J. D. (2010). Germination-still a mystery. *Plant Sci.* 179, 574–581. doi: 10.1016/j.plantsci.2010.02.010

Pandey, S., Kumari, A., Shree, M., Kumar, V., Singh, P., Bharadwaj, C., et al. (2019). Nitric oxide accelerates germination via the regulation of respiration in chickpea. *J. Exper. Bot.* 70, 4539–4555. doi: 10.1093/jxb/erz185

Pritchard, S. L., Charlton, W. L., Baker, A., and Graham, I. A. (2002). Germination and storage reserve mobilization are regulated independently in *Arabidopsis*. *Plant J.* 31, 639–647. doi: 10.1046/j.1365-313X.2002.01376.x

Rajjou, L., Belghazi, M., Huguet, R., Robin, C., Moreau, A., Job, C., et al. (2006). Proteomic investigation of the effect of salicylic acid on *Arabidopsis* seed germination and establishment of early defense mechanisms. *Plant Physiol.* 141, 910–923. doi: 10.1104/pp.106.082057

Rajjou, L., Gallardo, K., Debeaujon, I., Vandekerckhove, J., Job, C., and Job, D. (2004). The effect of alpha-amanitin on the *Arabidopsis* seed proteome highlights the distinct roles of stored and neosynthesized mRNAs during germination. *Plant Physiol.* 134, 1598–1613. doi: 10.1104/pp.103.036293

Sadanandom, A., Bailey, M., Ewan, R., Lee, J., and Nelis, S. (2012). The ubiquitin-proteasome system: central modifier of plant signalling. *New Phytol.* 196, 13–28. doi: 10.1111/j.1469-8137.2012.04266.x

Salemi, F., Esfahani, M. N., and Lam-Son Phan, T. (2019). Mechanistic insights into enhanced tolerance of early growth of alfalfa (*Medicago sativa* L.) under low water potential by seed-priming with ascorbic acid or polyethylene glycol solution. *Industr. Crops Prod.* 137, 436–445. doi: 10.1016/j.indcrop.2019.05.049

Sano, N., Permana, H., Kumada, R., Shinozaki, Y., Tanabata, T., Yamada, T., et al. (2012). Proteomic analysis of embryonic proteins synthesized from long-lived mRNAs during germination of rice seeds. *Plant Cell Physiol.* 53, 687–698. doi: 10.1093/pcp/pcs024

Singh, H., Jassal, R. K., Kang, J. S., Sandhu, S. S., and Grewal, K. (2015). Seed priming techniques in field crops -A review. *Agric. Rev.* 36, 251–264. doi: 10.18805/ag.v36i4.6662

Srivastava, A. K., Kumar, J. S., and Suprasanna, P. (2021). Seed 'primeomics': plants memorize their germination under stress. *Biol. Rev.* 96, 1723–1743. doi: 10.1111/brv.12722

Takahashi, H., Kopriva, S., Giordano, M., Saito, K., and Hell, R. (2011). Sulfur assimilation in photosynthetic organisms: molecular functions and regulations of transporters and assimilatory enzymes. *Annu. Rev. Plant Biol.* 62, 157–184. doi: 10.1146/annurev-arplant-042110-103921

Wang, W. Q., Mller, I. M., and Song, S. Q. (2012). Proteomic analysis of embryonic axis of *Pisum sativum* seeds during germination and identification of proteins associated with loss of desiccation tolerance. *J. Proteom.* 77, 68–86. doi: 10.1016/j.jprot.2012.07.005

Wei, Z., Qu, Z., Zhang, L., Zhao, S., Bi, Z., Ji, X., et al. (2015). Overexpression of poplar xylem sucrose synthase in tobacco leads to a thickened cell wall and increased height. *PLoS One* 10:e0120669. doi: 10.1371/journal.pone.0120669

Xia, Q., Maharajah, P., Cueff, G., Rajjou, L., Prodhomme, D., Gibon, Y., et al. (2018). Integrating proteomics and enzymatic profiling to decipher seed metabolism affected by temperature in seed dormancy and germination. *Plant Sci.* 269, 118–125. doi: 10.1016/j.plantsci.2018.01.014

Xu, H. H., Liu, S. J., Song, S. H., Wang, R. X., Wang, W. Q., and Song, S. Q. (2016). Proteomics analysis reveals distinct involvement of embryo and endosperm proteins during seed germination in dormant and non-dormant rice seeds. *Plant Physiol. Biochem.* 103, 219–242. doi: 10.1016/j.plaphy.2016.03.007

Yacoubi, R., Job, C., Belghazi, M., Chaibi, W., and Job, D. (2013). Proteomic analysis of the enhancement of seed vigour in osmoprimed alfalfa seeds germinated under salinity stress. *Seed Sci. Res.* 23, 99–110. doi: 10.1017/s0960258513000093

Yan, H., Jia, S., and Mao, P. (2020). Melatonin priming alleviates aging-induced germination inhibition by regulating beta-oxidation, protein translation, and antioxidant metabolism in oat (*Avena sativa* L.) seeds. *Intern. J. Mol. Sci.* 21:1898. doi: 10.3390/ijms21051898

Yan, H., and Mao, P. (2021). Comparative time-course physiological responses and proteomic analysis of melatonin priming on promoting germination in aged oat (*Avena sativa* L.) seeds. *Intern. J. Mol. Sci.* 22:811. doi: 10.3390/ijms22020811

Yan, M. (2016). Hydro-priming increases seed germination and early seedling growth in two cultivars of Napa cabbage (*Brassica rapa* subsp pekinensis) grown under salt stress. *J. Hortic. Sci. Biotechnol.* 91, 421–426. doi: 10.1080/14620316.2016.1162031

Yang, P. F., Li, X. J., Wang, X. Q., Chen, H., Chen, F., and Shen, S. H. (2007). Proteomic analysis of rice (*Oryza sativa*) seeds during germination. *Proteomics* 7, 3358–3368. doi: 10.1002/pmic.200700207

Yu, Y. L., Guo, G. F., Lv, D. W., Hu, Y. K., Li, J. R., Li, X. H., et al. (2014). Transcriptome analysis during seed germination of elite Chinese bread wheat cultivar Jimai 20. *BMC Plant Biol.* 14:20. doi: 10.1186/1471-2229-14-20

Zhang, G. L., Zhu, Y., Fu, W. D., Wang, P., Zhang, R. H., Zhang, Y. L., et al. (2016). iTRAQ protein profile differential analysis of dormant and germinated grassbur twin seeds reveals that ribosomal synthesis and carbohydrate metabolism promote germination possibly through the PI3K pathway. *Plant Cell Physiol.* 57, 1244–1256. doi: 10.1093/pcp/pcw074

Zhang, N., Zhang, H. J., Sun, Q. Q., Cao, Y. Y., Li, X. S., Zhao, B., et al. (2017). Proteomic analysis reveals a role of melatonin in promoting cucumber seed germination under high salinity by regulating energy production. *Sci. Rep.* 7:503. doi: 10.1038/s41598-017-00566-1

Zhao, T., Deng, X., Xiao, Q., Han, Y., Zhu, S., and Chen, J. (2020). IAA priming improves the germination and seedling growth in cotton (*Gossypium hirsutum* L.) via regulating the endogenous phytohormones and enhancing the sucrose metabolism. *Industr. Crops Prod.* 155:112788. doi: 10.1016/j.indcrop.2020.112788

Zhao, Y., Hu, M., Gao, Z., Chen, X., and Huang, D. (2018). Biological mechanisms of a novel hydro-electro hybrid priming recovers potential vigor of onion seeds. *Environ. Exper. Bot.* 150, 260–271. doi: 10.1016/j.envexpbot.2018.04.002

Zhu, Y., Lei, Q., Li, D., Zhang, Y., Jiang, X. G., Hu, Z. H., et al. (2018). Proteomic and biochemical analyses reveal a novel mechanism for promoting protein Ubiquitination and degradation by UFBP1, a key component of Ufmylation. *J. Proteome Res.* 17, 1509–1520. doi: 10.1021/acs.jproteome.7b00843

Zhu, Y. Q., Berkowitz, O., Selinski, J., Hartmann, A., Narsai, R., Wang, Y., et al. (2020). Conserved and opposite transcriptome patterns during germination in *Hordeum vulgare* and *Arabidopsis thaliana*. *Intern. J. Mol. Sci.* 21:7404. doi: 10.3390/ijms21197404

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

frontiers | Frontiers in Plant Science

REVIEW
published: 25 April 2022
doi: 10.3389/fpls.2022.865361

Histone Modification and Chromatin Remodeling During the Seed Life Cycle

*Xiali Ding[1], Xuhui Jia[1,2], Yong Xiang[1] and Wenhui Jiang[1]**

[1] Guangdong Laboratory for Lingnan Modern Agriculture, Genome Analysis Laboratory of the Ministry of Agriculture, Agricultural Genomics Institute at Shenzhen, Chinese Academy of Agricultural Sciences (CAAS), Shenzhen, China, [2] College of Life Science and Technology, Guangxi University, Nanning, China

Seeds are essential for the reproduction and dispersion of spermatophytes. The seed life cycle from seed development to seedling establishment proceeds through a series of defined stages regulated by distinctive physiological and biochemical mechanisms. The role of histone modification and chromatin remodeling in seed behavior has been intensively studied in recent years. In this review, we summarize progress in elucidating the regulatory network of these two kinds of epigenetic regulation during the seed life cycle, especially in two model plants, rice and Arabidopsis. Particular emphasis is placed on epigenetic effects on primary tissue formation (e.g., the organized development of embryo and endosperm), pivotal downstream gene expression (e.g., transcription of *DOG1* in seed dormancy and repression of seed maturation genes in seed-to-seedling transition), and environmental responses (e.g., seed germination in response to different environmental cues). Future prospects for understanding of intricate interplay of epigenetic pathways and the epigenetic mechanisms in other commercial species are also proposed.

Keywords: histone modification, chromatin remodeling, seed development, seed dormancy, seed germination, seedling establishment

OPEN ACCESS

Edited by:
Bing Bai,
University of Copenhagen, Denmark

Reviewed by:
Henk Hilhorst,
Wageningen University and Research, Netherlands
Wilco Ligterink,
KeyGene, Netherlands

***Correspondence:**
Wenhui Jiang
jiangwenhui@caas.cn

Specialty section:
This article was submitted to Plant Proteomics and Protein Structural Biology, a section of the journal Frontiers in Plant Science

Received: *29 January 2022*
Accepted: *21 March 2022*
Published: *25 April 2022*

Citation:
Ding X, Jia X, Xiang Y and Jiang W (2022) Histone Modification and Chromatin Remodeling During the Seed Life Cycle. Front. Plant Sci. 13:865361. doi: 10.3389/fpls.2022.865361

INTRODUCTION

Well-developed seeds assure species dispersion of parent plants and serve as important sources of human food. Progression from seed development to seedling establishment is the crucial phase of ontogenesis in spermatophytes. It involves a series of sequential physio-morphological state changes and includes several biological stages. The process starts with seed development and followed by maturation, when seed desiccation and seed dormancy are achieved in some species. After that, seeds germinate in a suitable environment, which marks the initiation of seedling establishment.

Proper seed development is inseparable from the organized establishment of all tissues (such as embryo, endosperm and seed coat), which is coordinated by changes in hormone levels and gene expression (Zhou et al., 2013; Figueiredo and Köhler, 2018). Seed maturation proteins in the LAFL regulatory network—LEAFY COTYLEDON 1 (LEC1), ABSCISIC ACID INSENSITIVE3 (ABI3), FUSCA3 (FUS3), and LEAFY COTYLEDON 2 (LEC2)—play predominant roles in triggering and maintaining embryonic cell fate by fine-tuning the expression of genes involved in the

accumulation of storage protein and lipid reserves in the embryo (Giraudat et al., 1992; Keith et al., 1994; Lotan et al., 1998; Stone et al., 2001; Yamamoto et al., 2009; Lepiniec et al., 2018).

In some species, during the maturation stage of seed development, dormancy gradually increases, peaking in freshly matured seeds. Dormancy enables seeds to adapt to the environment and plants to maintain reproduction (Née et al., 2017). *DELAY OF GERMINATION1* (*DOG1*) is the master regulator of primary dormancy in Arabidopsis (*Arabidopsis thaliana*); the encoded protein is a temperature detector that directs dormancy cycling in response to seasonal changes (Alonso-Blanco et al., 2003; Bentsink et al., 2006; Nakabayashi et al., 2012; Graeber et al., 2013; Footitt et al., 2015).

In appropriate condition, seeds can germinate once exposed to water. Germination behaviors, including germination rate and efficiency, differ between species and varieties in response to environment cues or abiotic stress. These differences are mediated mainly through the antagonistic roles of the plant hormones gibberellic acid (GA) and abscisic acid (ABA) (Jacobsen et al., 2002; Oh et al., 2007; Holdsworth et al., 2008; Seo et al., 2009; Vaistij et al., 2013; Shu et al., 2016).

Seed germination marks the initiation of the seed-to-seedling developmental transition. In this process, the sources of seedling nutrition and energy acquisition gradually transition from consumption of seed storage substances to photoautotrophy, in conjunction with significant alteration of biosynthetic and signaling pathways (Zanten et al., 2013; Jia et al., 2014). Correspondingly, suppression of seed maturation genes, the LAFL, and activation of those involved in vegetative growth is indispensable to avoid ectopic proliferation of embryonic tissues and thus maintain the normal vegetative morphology of seedlings (Parcy and Giraudat, 1997; Lotan et al., 1998; Stone et al., 2001; Gazzarrini et al., 2004; Braybrook et al., 2006; Yang et al., 2013).

The different stages of the seed life cycle are not isolated and are each under precise control. Accurate DNA processing and subsequent gene transcript levels are tightly linked to chromatin status, which is regulated by epigenetic modification. Epigenetic changes, including DNA methylation, histone modifications, chromatin remodeling, and the activities of small RNAs, affect plants in many ways (Goldberg et al., 2007). Research on the effects of epigenetic regulation in seed biology has recently increased. This review focuses mainly on the effects of two types of key regulatory factors, histone modifiers and chromatin remodelers, on the pivotal phase of the seed life cycle (**Table 1**).

Histone modifications, which are usually added to the N-terminus of the histone protein tail, can either regulate chromatin state directly or act as hotspots for the recruitment of other effectors to chromatin. Different histone modifications, including acetylation, methylation, ubiquitylation, and phosphorylation, comprise a "histone code" that provides a flexible method of governing gene transcription in response to developmental or environmental cues (Strahl and Allis, 2000; Turner, 2000; Dutnall, 2003). Histone modifications are reversible, being added and removed by "writer" and "eraser" enzyme complexes, respectively, that execute distinct functions on chromatin, promoting either active transcription or gene silencing (Kumar et al., 2021). For example, trimethylation of histone H3 lysine 4 (H3K4me3) and 36 (H3K36me3) are usually associated with gene activation, whereas H3K27me3, which is directly regulated by the classical polycomb repressive complex 2 (PRC2), correlates with heterochromatinization and transcriptional silencing (Berger, 2007; Zhang et al., 2007a,b, 2009; Mozgova and Hennig, 2015; Liu et al., 2019).

ATP-dependent chromatin remodeling complexes, with members of DNA-dependent ATPases as core subunits, utilize energy from ATP hydrolysis to disrupt the contacts between histones and DNA, thereby regulating dynamic access to packaged DNA. They mediate DNA replication, damage repair, and gene expression by changing the positions and occupancy of nucleosomes, introducing histone variants, and cooperating with histone-modifying factors (Goldberg et al., 2007; Hargreaves and Crabtree, 2011; Li et al., 2016). In eukaryotes, different epigenetic regulations often work coordinately to achieve cooperative or antagonistic modes of regulation (Li et al., 2015).

In this review, we summarize the histone modification and chromatin remodeling that occur during the seed life cycle, from seed development to seedling establishment. We hope to thereby pave the way toward a fundamental understanding and integration of the complex networks of epigenetic regulation acting in seed biology.

EPIGENETIC REGULATION OF SEED DEVELOPMENT

The production of viable seed is important for plant dispersal, and is a major focus of crop breeders because of its direct association with grain yield. Seed development involves the sequential and orderly formation of various structures, including the embryo, endosperm, and seed coat. Among angiosperms, monocotyledonous and dicotyledonous plants show both similarities and differences in seed development (Zhou et al., 2013). In this section, we primarily discuss epigenetic effects on seed formation in Arabidopsis and rice (*Oryza sativa*), the model dicot and monocot species, respectively (**Figure 1A**).

Histone deacetylases (HDACs) control seed-setting rate through affecting acetylation levels of target genes. AtHD2A, a member of the HD2 subfamily of HDAC proteins, is highly expressed in flowers and siliques. Silencing of *AtHD2A* expression aborts seed development (Wu et al., 2000). Similarly, the *Athda7-2* mutant causes both degeneration of micropilar nuclei at the four-nucleate embryo sac stage and an overall delay of embryo development, ultimately decreasing seed fertility (Cigliano et al., 2013). Meanwhile, OsSRT1, a NAD + -dependent HDAC in rice, represses the expression of *RICE STARCH REGULATOR1* (*RSR1*) and amylase genes, thus maintaining starch accumulation in developing seeds (Zhang et al., 2016). In maize, increases in acetylated histones H3 and H4 accompanied by decreases in H3K9me2 are observed in *hda108* mutants, resulting in a wide range of plant damage, including impaired fertility of cobs (Forestan et al., 2018).

As a nutritional supply tissue, the endosperm is indispensable for seed development. Disordered timing of endosperm development leads to seed failure in interploid and interspecific

TABLE 1 | List of histone modifiers, chromatin remodelers and associated regulators involved in seed life cycle.

Gene Name	Species	Locus	Seed development	Seed dormancy	Seed germination	Seedling establishment	References
HD2A	*A. thaliana*	AT3G44750	√		√		Wu et al., 2000; Colville et al., 2011
HD2C	*A. thaliana*	AT5G03740			√		Colville et al., 2011; Luo et al., 2012
HDA7	*A. thaliana*	AT5G35600	√				Cigliano et al., 2013
HDA9	*A. thaliana*	AT3G44680			√		Baek et al., 2020
HDA6	*A. thaliana*	AT5G63110			√	√	Tanaka et al., 2008; Chen and Wu, 2010; Chen L. T. et al., 2010; Luo et al., 2012
HDA19	*A. thaliana*	AT4G38130		√	√		Tanaka et al., 2008; Chen and Wu, 2010; Zhou et al., 2020
HDA15	*A. thaliana*	AT3G18520			√		Gu et al., 2017
ZmHDA108	Z. mays	GRMZM2G136067	√				Forestan et al., 2018
OsHDA705	*O. sativa L.*	Os08g25570			√		Zhao et al., 2016
OsHDT701	*O. sativa L.*	Os05g51830			√		Zhao et al., 2014
OsSRT1	*O. sativa L.*	LOC_Os04g20270	√				Zhang et al., 2016
OsGW6a	*O. sativa L.*	LOC_Os06g44100	√				Song et al., 2015
EFS	*A. thaliana*	AT1G77300	√		√	√	Tang et al., 2012; Lee et al., 2014; Cheng et al., 2018
SUVH5	*A. thaliana*	AT2G35160		√	√		Gu et al., 2019; Zhou et al., 2020
SUVH4	*A. thaliana*	AT5G13960		√	√		Zheng et al., 2012
ULT1	*A. thaliana*	AT4G28190				√	Xu et al., 2018
ATX1	*A. thaliana*	AT1G66240				√	Xu et al., 2018
LDL1, LDL2	*A. thaliana*	AT1G62830, AT3G13682		√			Zhao et al., 2015
REF6	*A. thaliana*	AT3G48430		√			Chen H. et al., 2020
JMJ20, JMJ22	*A. thaliana*	AT5G63080, AT5G06550			√		Cho et al., 2012
JMJ17	*A. thaliana*	AT1G63490			√		Wang et al., 2021
HUB1, HUB2	*A. thaliana*	AT2G44950, AT1G55250		√			Liu et al., 2007; Liu et al., 2011
MEA	*A. thaliana*	AT1G02580	√				Yadegari et al., 2000; Köhler et al., 2003b; Makarevich et al., 2006
FIE	*A. thaliana*	AT3G20740	√			√	Ohad et al., 1999; Yadegari et al., 2000; Wang et al., 2006; Bouyer et al., 2011
FIS2	*A. thaliana*	AT2G35670	√				Chaudhury et al., 1997; Hehenberger et al., 2012
MSI1	*A. thaliana*	AT5G58230	√				Köhler et al., 2003a
SWN	*A. thaliana*	AT4G02020	√			√	Chanvivattana et al., 2004; Schubert et al., 2005; Wang et al., 2006; Makarevich et al., 2006; Yang et al., 2013
CLF	*A. thaliana*	AT2G23380	√	√		√	Schubert et al., 2005; Makarevich et al., 2006; Yang et al., 2013; Liu J. et al., 2016; Chen N. et al., 2020
VRN2	*A. thaliana*	AT4G16845				√	Schubert et al., 2005
EMF2	*A. thaliana*	AT5G51230				√	Moon et al., 2003; Schubert et al., 2005; Tang et al., 2012;
OsFIE2	*O. sativa L.*	LOC_Os08g04270	√	√			Luo et al., 2009; Nallamilli et al., 2013; Li et al., 2014; Liu X. et al., 2016; Cheng et al., 2020;
OsFIE1	*O. sativa L.*	LOC_Os08g04290	√	√			Luo et al., 2009; Folsom et al., 2014; Huang et al., 2016; Cheng et al., 2020
OsSDG711	*O. sativa L.*	LOC_Os06g16390	√				Liu et al., 2021
OsEMF2a	*O. sativa L.*	LOC_Os04g08034	√				Luo et al., 2009; Tonosaki et al., 2021; Cheng et al., 2021
AtBMI1a, AtBMI1b, AtBMI1c	*A. thaliana*	AT2G30580, AT1G06770, AT3G23060				√	Bratzel et al., 2010; Chen D. et al., 2010; Yang et al., 2013
AtRING1a, AtRING1b	*A. thaliana*	AT5G44280, AT1G03770				√	Bratzel et al., 2010; Chen D. et al., 2010
EMF1	*A. thaliana*	AT5G11530				√	Moon et al., 2003; Kim et al., 2012; Xu et al., 2018
PKL	*A. thaliana*	AT2G25170		√		√	Ogas et al., 1997; Ogas et al., 1999; Dean Rider et al., 2003; Henderson et al., 2004; Li et al., 2005; Carter et al., 2018; Zha et al., 2020
BRM	*A. thaliana*	AT2G46020			√		Han et al., 2012
SWI3B	*A. thaliana*	AT2G33610			√		Saez et al., 2008
CHR12, CHR23	*A. thaliana*	AT3G06010, AT5G19310			√		Leeggangers et al., 2015
EBS	*A. thaliana*	AT4G22140		√	√		Narro-Diego et al., 2017; Li et al., 2020
S2Lb	*A. thaliana*	AT5G66240	√	√			Fiorucci et al., 2019
HDC1	*A. thaliana*	AT5G08450			√		Perrella et al., 2013
PWR	*A. thaliana*	AT3G52250			√		Yang et al., 2019
VAL1, VAL2	*A. thaliana*	AT2G30470, AT4G32010		√		√	Suzuki et al., 2007; Yang et al., 2013; Chen N. et al., 2020

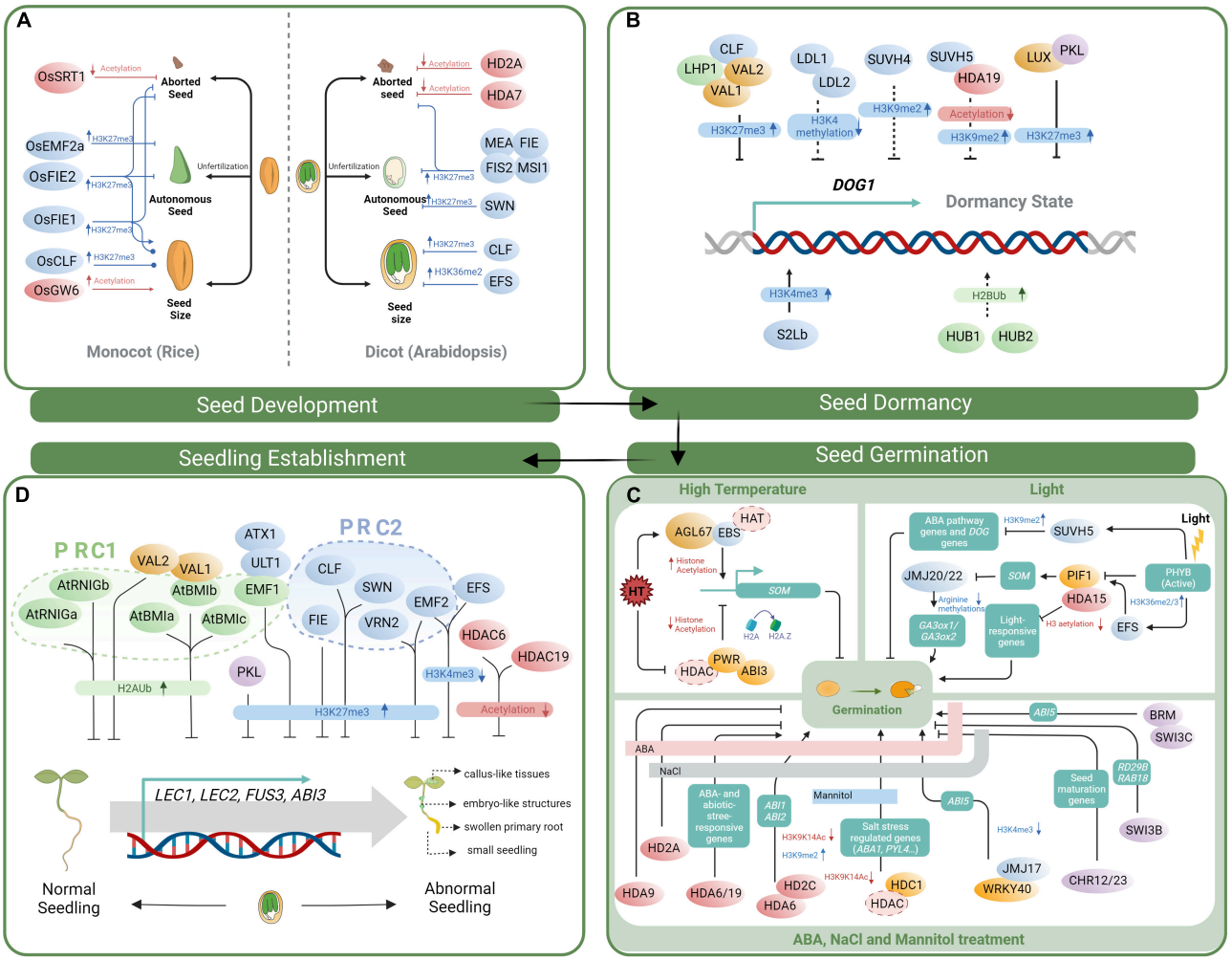


FIGURE 1 | Model of histone modification and chromatin remodeling during various stages of the seed life cycle. **(A)** Epigenetic regulators act of pivotal seed development in Arabidopsis and rice. Line with a dot in tail end represent an unclear regulatory pathway. **(B)** Chromatin regulation of seed dormancy through mediating the expression of *DOG1*. Solid lines represent regulator that can modulate *DOG1* expression directly. Dash lines represent those regulatory mechanisms that are still unclear. **(C)** The roles of histone modifiers and chromatin remodelers in regulating seed germination in response to light, temperature and abiotic stress. In the box below, the bold line with pink, gray and blue colors represent ABA, NaCl and mannitol treatment, respectively. Ellipse with dotted outline represents that the specific regulator is not clear. **(D)** Epigenetic factors act in seed-to-seedling transition. Illustrated histone modifiers and chromatin remodelers act in repression of ectopic expression of LAFL genes, thereby maintaining normal vegetative morphology of seedlings. The histone modifiers for histone acetylation, methylation, and ubiquitination are marked in red, blue and green, respectively. Chromatin remodelers and epigenetic associated regulators are marked in purple and orange, respectively. Downstream genes are shown in boxes with aquamarine blue.

hybrids, directly impeding crop breeding (Walia et al., 2009; Ishikawa et al., 2011; Kradolfer et al., 2013; Sekine et al., 2013). During cell proliferation, the PRC2 complex adds the repressive mark H3K27me3 on endosperm-related transcripts. In Arabidopsis, mutations of genes encoding members of the FERTILIZATION-INDEPENDENT SEED (FIS)–PRC2 complex [*FERTILIZATION-INDEPENDENT ENDOSPERM* (*FIE*), *MEDEA* (*MEA*), *FIS2*, *MULTICOPY SUPRESSOR OF IRA1 (MSI1)*] decrease H3K27me3 accumulation and impair cellularization of endosperm, and the mutants are characterized by a gametophytic maternal effect. After fertilization, embryonic cell proliferation and morphogenesis are inhibited in these mutants, decreasing their seed-setting rate. Notably, endosperm proliferation can also initiate in mutants in the absence of fertilization, but with arrested embryo development, producing non-functional autonomous seeds (Chaudhury et al., 1997; Grossniklaus et al., 1998; Kiyosue et al., 1999; Ohad et al., 1999; Yadegari et al., 2000; Sørensen et al., 2001; Köhler et al., 2003a,b). The downstream genes regulated by the FIS–PRC2 complex include type I MADS-box transcription factor genes, which encode key regulators in endosperm formation (Kang et al., 2008; Hehenberger et al., 2012; Figueiredo et al., 2015; Zhang et al., 2018). Moreover, the seed-abortion phenotype of a *mea* mutant can be alleviated by reducing the expression of the type I MADS-box gene *PHERES1* (*PHE1*) (Köhler et al., 2003b). Similarly, maternal loss of *AGAMOUS-LIKE62* (*AGL62*)

can also rescue delayed cellularization of endosperm cells, normalizing seed development in a *fis2* mutant (Hehenberger et al., 2012). Aside from the FIS–PRC2 complex, SWINGER (SWN), a subunit of the EMBRYONIC FLOWER 1 (EMF)–PRC2 complex also participates in the initiation of endosperm development. Although a *swn* mutant shows no identifiable developmental defect, a *swn mea* double mutant has an enhanced autonomous seed formation phenotype compared to the *mea* single mutant, indicating that SWN and MEA work redundantly (Wang et al., 2006).

In cereal, the PRC2 complex is evolutionarily conserved, but shows differences in combination of subunits and in specific function compared with that in Arabidopsis, as exemplified by the absence of some homologs of Arabidopsis *FIS* genes (*MEA* and *FIS2*) in cereal genomes (Rossi et al., 2001; Springer et al., 2002; Danilevskaya et al., 2003; Spillane et al., 2007; Luo et al., 2009; Rodrigues et al., 2010). Rice has two *FIE* homologs, *OsFIE1* and *OsFIE2*, with different expression patterns: *OsFIE1* is specifically expressed in endosperm, whereas *OsFIE2* is expressed in all tissues tested. In earlier research, no autonomous endosperm development was observed in *OsFIE1* and *OsFIE2* loss-of-function plants with emasculated florets (Luo et al., 2009; Nallamilli et al., 2013). However, other studies revealed that the autonomous endosperm phenotype could be occasionally detected in unfertilized lines with *OsFIE2* defect, along with impaired cellularization, suggesting that *OsFIE2* may retain functional similarity to its Arabidopsis homolog (Li et al., 2014; Cheng et al., 2020).

Moreover, rice also has two homologs of the Arabidopsis *PcG* gene *EMBRYONIC FLOWER2* (*EMF2*): *EMF2a* and *EMF2b*. Arabidopsis EMF2 is required to maintain vegetative development in Arabidopsis (Moon et al., 2003). Rice EMF2a, a maternally expressed gene in the endosperm, is indispensable for early seed development; delayed cellularization of endosperm and subsequent autonomous endosperm is observed in emasculated spikelets of an *osemf2a* mutant. Not surprisingly, the loss of OsEMF2a function reduces H3K27me3 modifications at various type I MADS-box genes, several of which (e.g., *OsMADS77* and *OsMADS89*) may control the timing of cellularization in endosperm in a similar manner to *AGL62* and *PHE1* in Arabidopsis (Luo et al., 2009; Cheng et al., 2021; Tonosaki et al., 2021).

Seed size and weight are crucial agronomical traits, tightly linked with grain yield in crop breeding (Sweeney and McCouch, 2007). In Arabidopsis, the inner space of mature seed is mostly occupied by the embryo, and nutrients are mainly stored in the cotyledons (Zhou et al., 2013). Mutation of *EARLY FLOWERING IN SHORT DAYS* (*EFS*), also called *SDG8*, encoding the major contributor to H3K36 methylation, leads to the formation of larger embryos, resulting in enlarged seeds in Arabidopsis (Cheng et al., 2018). Similarly, larger and heavier seeds are also observed in *clf-28* lines (mutants of *CURLY LEAF* [*CLF*], which encodes the core unit of the PRC2 complex), along with a large-scale, dynamic change in H3K27me3 level during embryonic development (Liu J. et al., 2016). Unlike in dicotyledons, endosperm in monocotyledon crops is not gradually consumed during seed maturation, but filled with large amounts of starch and nutrients for storage (Olsen, 2004; Agarwal et al., 2011; Zhou et al., 2013). In rice, unlike in the Arabidopsis *clf-28* mutant, which has enlarged seeds, *OsSDG711* (*OsCLF*) downregulation lines have smaller seeds, accompanied by altered expression of starch-related genes (Liu et al., 2021). A similar phenotype of smaller seed size and reduced contents of multiple storage proteins is also observed in reduction lines of *OsFIE2* or *OsFIE1* (Nallamilli et al., 2013; Huang et al., 2016; Liu X. et al., 2016). However, overexpression of *SDG711* or *OsFIE1* also decrease seed size, but the regulatory mechanisms are not fully elucidated (Folsom et al., 2014; Liu et al., 2021). Another tissue with important effects on grain size and weight is the spikelet hull. In rice, the grain is physically restricted by the size of the hull. The quantitative trait locus (QTL) *GRAIN WEIGHT ON CHROMOSOME 6* (*GW6a*), also called *OsglHAT1*, encodes a histone acetyltransferase that regulates grain weight, hull size, yield, and plant biomass. Elevated *GW6a* expression enhances grain yield by enlarging spikelet hulls *via* increasing cell number and accelerating grain filling (Song et al., 2015).

In general, the epigenetic regulation of major seed traits in two model plants, Arabidopsis and rice, partially overlaps but also shows some divergence. The comparison of epigenetic machinery in seed development is also extended to other crop plants, like soybean and maize (Lu et al., 2015; Lin et al., 2017). Therefore, the similarity and divergence highlight the need for more knowledge of the complex network of epigenetic regulation influencing seed development programs in different species.

EPIGENETIC REGULATION OF A CRITICAL SEED DORMANCY PATHWAY

Seed dormancy is an innate state in which the seed is unable to germinate, even under favorable conditions. Entry into dormancy is determined primarily by genetic factors and is also influenced by the environment surrounding the mother plant (Finch-Savage and Leubner-Metzger, 2006). Seeds in the soil seed bank (SSB) can sense seasonal signals and continually adjust their dormancy levels in order to complete germination at a suitable time of the year (Baskin and Baskin, 2006; Walck et al., 2011; Finch-Savage and Footitt, 2017).

Several histone-modification genes regulating histone ubiquitination, methylation, and acetylation exhibit dynamic expression patterns in response to seasonal change that are correlated with dormancy cycling in the SSB. Moreover, the accumulation of two antagonistic histone marks, H3K4me3 and H3K27me3, on key dormancy genes changes dynamically accompanied by changes in dormancy level, suggesting that chromatin regulators play pivotal roles in this process (Footitt et al., 2015). DOG1 is a master regulator of primary dormancy that acts in concert with ABA and HEME to delay germination (Née et al., 2017; Carrillo-Barral et al., 2020). During release from dormancy, the active H3K4me3 mark on *DOG1* chromatin is removed when the seeds are exposed to light; meanwhile, the repressive H3K27me3 mark accumulates on *DOG1* in seeds of the SSB, and light exposure amplifies this accumulation (Müller et al., 2012; Footitt et al., 2015).

The components of polycomb-group proteins, including CLF and LIKE HETEROCHROMATIN PROTEIN 1 (LHP1), are recruited by B3-domain-containing transcriptional repressors, HSI2/VAL1 (HIGH-LEVEL EXPRESSION OF SUGAR INDUCIBLE2/VIVIPAROUS-1/ABA3-LIKE1) and HSL1 (HSI2-LIKE1)/VAL2, to RY elements in the *DOG1* promoter. Hence, they accelerate the deposition of H3K27me3 marks and subsequent repression of *DOG1* expression (Chen N. et al., 2020). PICKLE (PKL) is an ATP-dependent chromatin-remodeling factor that promotes the deposition of H3K27me3 (Zhang et al., 2008, 2012). LUX ARRHYTHMO (LUX), a member of the evening complex (EC) of the circadian clock, physically interacts with PKL and recruits it to the chromatin region of *DOG1*. Correspondingly, levels of the repressive mark H3K27me3 at specific *DOG1* chromatin loci are greatly reduced in the *lux* and *pkl* mutants, increasing dormancy compared with the wild type. Moreover, these phenotypes are abolished when the mother plants are grown under continuous light. Thus, there may exist a regulatory mechanism in which EC proteins coordinate with PKL to transmit circadian signals, thereby directly regulating *DOG1* expression and seed dormancy during seed maturation (Zha et al., 2020). On the other hand, the binding site of LUX is close to the transcriptional start site of the non-coding antisense transcript *asDOG1*, which suppresses the expression of the *DOG1* sense transcript, and *asDOG1* transcription is decreased in the *pkl-1* mutant (Fedak et al., 2016; Zha et al., 2020). It suggests that the effect of PKL-LUX repression of *DOG1* transcription maybe much more elaborate than might be expected. In contrast to the H3K27me3 deposition functions of the LUX-PKL regulatory complex, RELATIVE OF EARLY FLOWERING6 (REF6), a key H3K27me3 demethylase that binds directly to the ABA catabolism genes *CYP707A1* and *CYP707A3*, is responsible for reducing their H3K27me3 levels. Correspondingly, the seeds of *ref6* mutants display enhanced dormancy, associated with increased endogenous ABA content (Chen H. et al., 2020).

Beyond H3K27me3, H3K9 and H3K4 methylation are also involved in *DOG1* and seed dormancy control. The global accumulation of H3K9 dimethylation is catalyzed by KYP/SUVH4, a Su(var)-type methyltransferase. Mutations in the *KYP* increase *DOG1* and *ABI3* expression, promoting seed dormancy. On the other hand, the sensitively of seed germination to ABA and paclobutrazol (PAC) is also increased in *kyp-2* mutant (Zheng et al., 2012). *SUVH5*, a homolog of *SUVH4*, interacts with the histone deacetylase HDA19 *in vivo* and *in vitro*. Mutants of both *SUVH5* and *HDA19* increase histone H3 acetylation (H3ac) but decrease H3K9me2, therefore enhancing *DOG1* expression and seed dormancy (Zhou et al., 2020). LYSINE-SPECIFIC DEMETHYLASE 1-LIKE1 (LDL1) and LDL2, two Arabidopsis histone demethylases, reduce the level of the histone H3-Lys 4 methylation in chromatin. They act redundantly to repress genes related to seed dormancy, including *DOG1*, *ABA2*, and *ABI3*, and *LDL1* or *LDL2* overexpression lines cause reduced seed dormancy (Zhao et al., 2015).

In yeast, H3K4me3 deposition is regulated by the SET1 histone methyltransferase (HMT) embedded in a so-called COMPlex of Proteins Associated with Set1 (COMPASS), in a process dependent on H2B monoubiquitination (H2Bub) (Miller et al., 2001; Sun and Allis, 2002). Similarly, Arabidopsis possesses homologs of all known COMPASS subunits, which potentially form several COMPASS-like complexes (Jiang et al., 2009, 2011; Fletcher, 2017). Genetic knockout of *SWD2-LIKE b* (*S2Lb*), the Arabidopsis homolog of Swd2 axillary subunit of yeast COMPASS, triggers pleiotropic developmental phenotypes, including reduced fertility and seed dormancy, accompanied by decreased H3K4me3 deposition and barely detectable *DOG1* expression (Fiorucci et al., 2019). However, even though H2B monoubiquitination regulated by HISTONE MONOUBIQUITINATION1 (HUB1) and HUB2 also increases *DOG1* expression and seed dormancy, the classical H2Bub–H3K4me3 *trans*-histone crosstalk seems to be lacking in Arabidopsis, since global H3K4me3 enrichment and the occupancy of an S2Lb-GFP (green fluorescent protein) fusion on target genes do not show obvious differences in the *hub1-3* background as compared with that in wild-type (Liu et al., 2007; Fiorucci et al., 2019).

Unlike direct regulators of histone modification, histone readers are involved in recognizing these marks and transferring the information to subsequent regulator units. The reader EARLY BOLTING IN SHORT DAYS (EBS) specifically recognizes the H3K4me2/3 mark and interacts with HDAC proteins such as HDA6 to modulate gene expression. Loss of function mutation of *EBS* reduces seed dormancy, and mutation of *EBS* homolog *SHORT LIFE* (*SHL*) deepens the seed dormancy alteration. However, EBS acts independently of two other types of dormancy regulators, HUB proteins and ARABIDOPSIS TRITHORAX-RELATED7 (ATXR7), and it does not affect the expression of *DOG1* (Liu et al., 2007, 2011; Narro-Diego et al., 2017).

Overall, studies of the epigenetic regulation of seed dormancy have mainly focused on *DOG1* expression (**Figure 1B**), and other regulatory pathways need further investigation to further complete the picture.

SEED GERMINATION IN RESPONSE TO DIVERGENT ENVIRONMENTAL CUES

Seed germination is an important physiological event that marks a transition from the quiet status of seeds to the active statues of seedlings, during which many processes are reprogrammed. The condensed chromatin state may diminish during seed germination (van Zanten et al., 2011; Zanten et al., 2013), providing a suitable environment for activating gene expression and physiological metabolisms that facilitate the process. Seeds can adjust their germination strategies in response to external environmental cues. In this section, we highlight current knowledge of the roles of histone modifiers and chromatin remodelers in regulating seed germination in response to light, temperature, and abiotic stresses (**Figure 1C**).

Light and temperature are two main exogenous factors determining plant growth, development, and productivity, including seed germination. In this phase, one of the classical light signal transport chains is the PHYTOCHROME B (PHYB)–PHYTOCHROME INTERACTING FACTOR1 (PIF1)/PIL5 pathway. In Arabidopsis, PHYB destabilizes PIF1 to regulate

light-responses seed germination through affecting genes expression in ABA and GA pathways. *SOMNUS* (*SOM*) is an important PIF1 downstream target that negatively regulates seed germination (Oh et al., 2006; Kim et al., 2008; Luo et al., 2014; Vaistij et al., 2018). Chromatin remodeling and histone modification that participate in light-regulated seed germination mainly function by affecting this PHYB-dependent pathway.

SUVH5, an H3K9 methyltransferase, acts as a positive regulator of PHYB-dependent seed germination in Arabidopsis. It functions by repressing the transmission of ABA signal and ABA biosynthesis, as well as suppressing the expression a family of *DOG1* genes *via* H3K9me2 in imbibed seeds (Gu et al., 2019). Mutation of *EFS/SDG8*, another HMT gene, decreases H3K36me2 and H3K36me3 levels at the *PIF1* locus, resulting in reduced *PIF1* expression in imbibed seeds (Lee et al., 2014). Meanwhile, HDA15 can be recruited by PIF1 and to form a repression module that regulates light-dependent seed germination by decreasing histone H3 acetylation levels and the corresponding transcription of light-responsive genes (Gu et al., 2017). JMJ20 and JMJ22, two histone demethylation enzymes, act redundantly as positive regulators of seed germination. When PHYB is inactive, JMJ20 and JMJ22 are directly suppressed by the zinc-finger protein SOM, and the repression will be released upon PHYB activation by light. Derepressed JMJ20/JMJ22 increase seed germination rate through the removal of repressive histone arginine methylations at *GIBBERELLIN 3-OXIDASE1* (*GA3ox1*) and *GA3ox2* (Cho et al., 2012). Therefore, light treatment promotes seed germination in a process that may be partially regulated by the PHYB–PIF1–SOM–JMJ20/JMJ22 pathway.

Moreover, another JmjC-domain demethylase JMJ17 participates in ABA response in seed germination through co-regulation with WRKY DNA-BINDING PROTEIN 40 (WRKY40), HYPOCOTYL5 (HY5), and ABI5. An elevated level of H3K4me3 at *ABI5* has been detected in *jmj17* and *wrky40* mutants. In the presence of ABA, WRKY40 and JMJ17 are released from ABI5 chromatin, which allows HY5 to induce ABI5 expression. Because HY5 is another crucial factor that helps promote photomorphogenesis, the transcriptional switch composed of JMJ17–WRKY40 and HY5–ABI5 modules may play an essential role in the integration of light and ABA signaling (Wang et al., 2021).

Temperature is another critical environmental cue affecting seed germination (Toh et al., 2008). SOM participates in thermoinhibition of seed germination by altering ABA and GA metabolism (Park et al., 2011; Lim et al., 2013; Chang et al., 2018). EBS, the histone mark reader, can be recruited by AGL67 to the *SOM* locus, thus recognizing H3K4me3 at the *SOM* promoter. Under high temperature (HT), the AGL67–EBS complex is highly enriched around the *SOM* promoter, leading to deposition of the activation mark H4K5 acetylation on *SOM* and ultimately inhibiting seed germination (Li et al., 2020). POWERDRESS (PWR), a protein with a SANT-domain, interacts with ABI3 and HDAC proteins to modify histone acetylation status and the level of nucleosome histone H2A.Z incorporation in the target loci. The complex inhibits *SOM* expression by reducing H4 acetylation deposition and increasing nucleosome H2A.Z content at the *SOM* locus, thus promoting the thermotolerance of seed germination. Under HT, the *PWR* transcript decreased, resulting in releasing of *SOM* from repression state (Yang et al., 2019).

ABA, as a barrier to germination, plays a pivotal role in plant response to abiotic stresses, such as drought and salt (Zhu, 2016). Members in the SWITCH2(SWI2)/SNF2 chromatin-remodeling complexes affect seed germination under ABA treatment. BRAHMA (BRM), the core SWI2/SNF2 ATPases within the complex, directly repress the expression of *ABI5*, the *brm-3* mutant presents ABA hypersensitivity in seed germination (Han et al., 2012). SWITCH SUBUNIT3 (SWI3) proteins (called SWI3A–D) in Arabidopsis are also important subunits of SWI2/SNF2-dependent chromatin-remodeling complexes (Sarnowski et al., 2005). BRM and SWI3C show strong direct physical interaction and null *swi3c-2* mutants show an ABA-hypersensitive phenotype similar to *brm*. These observations suggest that SWI2C may be a dedicated BRM complex component (Hurtado et al., 2006; Han et al., 2012). In contrast, mutants of *SWI3B* show reduced sensitivity to ABA-mediated inhibition of seed germination, with reduced expression of the ABA-responsive genes *RAB18* and *RD29B* (Saez et al., 2008). Furthermore, in overexpression lines of *AtCHR12* or *AtCHR23*, another two SWI2/SNF2 ATPase genes, the phenotype of reduced germination is pronounced under ABA and NaCl treatment, coinciding with increased transcription of seed maturation genes (Leeggangers et al., 2015).

In addition to chromatin remodelers, HDAC proteins are also involved in abiotic-stress-responsive seed germination. In Arabidopsis, germination of the *hd2c* mutant is restrained under ABA and salinity stress, while the *hd2a* mutant is insensitive to ABA (Colville et al., 2011). HD2C interacts with HDA6 and binds to histone H3. The expressions of *ABI1* and *ABI2* are decreased along with increased H3K9K14Ac and decreased H3K9me2 modification in *hda6*, *hd2c*, and *hda6 hd2c-1* (Luo et al., 2012). Moreover, HDA6 and HDA19 may play redundant roles in modulating seed germination response to abiotic stress by increasing the expression of ABA- and abiotic-stress-responsive genes (Chen and Wu, 2010; Chen L. T. et al., 2010). Arabidopsis HISTONE DEACETYLATION COMPLEX1 (HDC1) is the rate-limiting component of the histone deacetylation complex that physically interacts with HDAC proteins to desensitize plant germination to salt, mannitol, ABA, and PAC treatments (Perrella et al., 2013). By contrast, *hda9-1* and *hda9-2* mutants show increased germination in response to ABA treatment and HDA9 forms a complex with ABI4 to regulate the expression of the ABA catabolic genes *CYP707A1* and *CYP707A2* (Baek et al., 2020). In rice, plants overexpressing *HDA705* or *HDT701* show not only delayed germination under ABA, NaCl, or polyethylene glycol (PEG) treatment, but also stronger resistance to drought stress as seedlings (Zhao et al., 2014, 2016).

EPIGENETIC FACTORS ACTING IN SEED-TO-SEEDLING TRANSITION

After seeds germination, plants undergo an irreversible transition from embryo to seedling development, accompanied

by repression of embryonic traits and emergence of vegetative tissue. Expression change of the seed-maturation genes collectively known as LAFL is important for the switch of the developmental program. In loss-of-function mutants of these genes, embryos skip late-embryonic development and enter the vegetative program prematurely (Keith et al., 1994; West et al., 1994; Nambara et al., 2000). However, when some of these genes are misexpressed in vegetative tissues, abnormally developed seedlings emerge that show induced ectopic deposition of seed storage proteins and even somatic embryo or callus formation (Parcy and Giraudat, 1997; Lotan et al., 1998; Stone et al., 2001; Gazzarrini et al., 2004; Braybrook et al., 2006; Yang et al., 2013). The factors involved in chromatin remodeling and histone modification protect normal seedling morphology mainly by repressing the transcription of LAFL genes (**Figure 1D**).

In the PRC1 complex, two types of ring-finger proteins, AtRING1s and AtBMI1s, are major subunits that directly catalyze H2A monoubiquitination (H2Aub). The mutants *Atring1a Atring1b* and *Atbmi1a Atbmi1b* show ectopic expression of seed-maturation genes and indeterminate embryonic traits at the seedling stage (Bratzel et al., 2010; Chen D. et al., 2010; Yang et al., 2013). Moreover, mutations of *EMF1* or *EMF2*, two PcG proteins, can strengthen the phenotype of *Atbmi1a Atbmi1b* and expression of the seed maturation genes increased obviously in the *emf1* mutant, suggesting these regulators may collaborate in repression of the maturation program after germination (Moon et al., 2003; Kim et al., 2012; Yang et al., 2013). VAL proteins are B3-type transcription factors that interact with AtBMI1 proteins. *val1/2* seedlings display phenotypic defects similar to those of *Atbmi1a/b/c* mutants, accompanied by strongly reduced H2Aub levels at seed-maturation genes and concomitant derepressed gene transcription (Suzuki et al., 2007; Yang et al., 2013). On the other hand, the levels of H3K27me3 at *LEC1*, *FUS3*, and *ABI3* are also strongly decreased in *val1/2* and *Atbmi1a/b/c* mutants (Yang et al., 2013). VAL proteins can recruit the PRC2 subunit CLF and promote the placement of H3K27me3 on target loci (Chen N. et al., 2020; Yuan et al., 2021), indicative of genetic and physical interaction between the PRC1 and PRC2 complexes.

The Arabidopsis PRC2 complex, which catalyzes H3K27me3 addition, represses seed maturation genes, as evidenced by somatic embryo emergence in vegetative tissues of double mutants deficient in redundant PRC2 subunits, i.e., CLF and SWN or EMF2 and VERNALIZATION2 (VRN2) (Chanvivattana et al., 2004; Schubert et al., 2005; Makarevich et al., 2006; Yang et al., 2013). A single mutant of Arabidopsis FIE also gives rise to the degeneration of vegetative cells into neoplastic, callus-like structures in seedlings with abolished H3K27me3 deposition (Kinoshita et al., 2001; Bouyer et al., 2011). It should be noted that EMF2 or FIE are mainly worked in repressing flower formation upon germination, though discussion of this function is outside the scope of this manuscript (Kinoshita et al., 2001; Moon et al., 2003). Furthermore, the chromatin remodeler PKL, which has a potential role in the retention of H3K27me3, acts throughout the seedling, repressing embryonic traits. Loss of PKL function reduces levels of H3K27me3 and ectopic expression of the embryo-specific genes *LEC1*, *LEC2*, and *FUS3*, resulting in seedlings with swollen primary roots, referred to as pickle roots (Ogas et al., 1997, 1999; Dean Rider et al., 2003; Henderson et al., 2004; Li et al., 2005; Carter et al., 2018).

Therefore, the PRC1 and PRC2 repressive system are tightly integrated in the transition from seed to seedling. Moreover, many other chromatin regulators also participate in this process, and some show crosstalk with the PcG working program.

In Arabidopsis, trithorax group (trxG) proteins catalyze H3K4 methylation, which play roles opposite to that of H3K27me3. Correspondingly, the trxG members (ATX1) and ULTRAPETALA1 (ULT1) counteract the effect of CLF in floral repression (Carles and Fletcher, 2009; Alvarez-Venegas, 2010). However, removal of either or both *ATX1* and *ULT1* fails to rescue the defects exhibited by an *emf1* mutant, but promote H3K27me3, causing a swollen, pickle-like root phenotype in seedlings of *emf1 atx1 ult1* triple mutants. Yeast two-hybrid assays reveal that ULT1 physically interacts with ATX1 and EMF1, and both ATX1 and ULT1 are able to bind the chromatin of seed genes, including *LEC2* and *ABI3* (Xu et al., 2018). This suggests a new, more complex framework whereby trxG acts in concert with PcG to maintain chromatin integrity and prevent seed maturation gene expression after germination. A mutation of *SDG8/EFS*, encoding an HMT that mainly mediates H3K36 methylation, acts synergistically with *emf2* to induce the deposition of the active mark H3K4me3 on seed-maturation loci, leading to the emergence of embryonic traits (Tang et al., 2012). However, the mechanism and the specific pathway whereby the activating H3K4me3 marks are deposited in the *sdg8 emf2* double mutant is still unclear.

Arrested growth and the formation of embryo-like structures on vegetative tissues can also be observed in a *hda6 hda19* double mutant. Moreover, the disturbed cell fate seen in the *hda6* mutant upon treatment with the HDAC inhibitor trichostatin A (TSA) is rescued by mutations of *lec1*, *fus3*, and *abi3*, indicating that acetylation also has effects at the seedling development stage (Tanaka et al., 2008). In animals, CHD3 chromatin remodelers are components of RPD3-containing HDAC complexes (Tong et al., 1998; Zhang et al., 1998; Fukaki et al., 2006). Therefore, whether there are connections between the HDACs and the CHD3 protein PKL that influence the repression of embryonic properties is a question in need of further study.

CONCLUSION AND FUTURE PERSPECTIVES

As plants progress from seed development through seedling establishment, gene expression is dynamically affected by histone modifications and chromatin states. These epigenetic regulations include a combination of synergistic and antagonistic crosstalk between histone-modifying enzymes through specific connecting factors. By screening and analyzing regulators in the nodes of the epigenetic network, it is possible to uncover the comprehensive changes in the epigenetic modifications of pivotal genes during development.

Additionally, loss of function of epigenetic regulatory genes often gives rise to pleiotropic effects, with changes in chromatin state at the whole-genome level. To some extent, these extensive

effects are mediated by various specific co-regulators that participate in diverse biological processes. Further studies to identify the working partners of epigenetic regulatory proteins will thus provide further information about the processes governing specific pathways at the epigenetic level.

Furthermore, gene transcription regulation mediated by histone modifiers or chromatin remodelers might be the later step in the cascades of environmental signal transition. Specific knowledge about how the epigenetic regulators receive these signals needs to be uncovered.

Finally, investigations to date of the effects of histone modification and chromatin remodeling on seed developmental programs have mainly focused on the model plant Arabidopsis, with few studies performed in commercial species, such as grain, fruit, and vegetable crop species. Therefore, research on classical epigenetic regulatory pathways and associated components in model plants need to be extended to other, economically important species to accelerate its application to molecular breeding for agricultural production. Moreover, better understanding of the conserved and diverse regulatory mechanisms acting in different plant species will enhance knowledge of the complex epigenetic regulatory mechanisms controlling this process.

AUTHOR CONTRIBUTIONS

WJ: writing-original draft. YX, XD, and XJ: review and editing. YX: supervision. All authors contributed to the article and approved the submitted version.

FUNDING

This research was supported by the Key-Area Research and Development Program of Guangdong Province (Grant No. 2021B0707010006), China Postdoctoral Science Foundation (Grant No. 2019M650914), The Science, Technology and Innovation Commission of Shenzhen Municipality (Grant Nos. KCXFZ20201221173203009 and JCYJ20200109150713553), and Dapeng District Industry Development Special Funds (Grant Nos. KJYF202101-09 and RCTD20180102).

ACKNOWLEDGMENTS

We apologize to those whose work was not cited because of space constrains.

REFERENCES

Agarwal, P., Kapoor, S., and Tyagi, A. K. (2011). Transcription factors regulating the progression of monocot and dicot seed development. *Bioessays* 33, 189–202. doi: 10.1002/bies.201000107

Alonso-Blanco, C., Bentsink, L., Hanhart, C. J., Blankestijn-de Vries, H., and Koornneef, M. (2003). Analysis of natural allelic variation at seed dormancy loci of *Arabidopsis thaliana*. *Genetics* 164, 711–729. doi: 10.1093/genetics/164.2.711

Alvarez-Venegas, R. (2010). Regulation by polycomb and trithorax group proteins in *Arabidopsis*. *Arabidopsis Book* 8:e0128. doi: 10.1199/tab.0128

Baek, D., Shin, G., Kim, M. C., Shen, M., Lee, S. Y., and Yun, D. J. (2020). Histone deacetylase HDA9 with ABI4 contributes to abscisic acid homeostasis in drought stress response. *Front. Plant. Sci.* 11:143. doi: 10.3389/fpls.2020.00143

Baskin, C. C., and Baskin, J. M. (2006). The natural history of soil seed banks of arable land. *Weed Sci.* 54, 549–557. doi: 10.1614/WS-05-034R.1

Bentsink, L., Jowett, J., Hanhart, C. J., and Koornneef, M. (2006). Cloning of DOG1, a quantitative trait locus controlling seed dormancy in *Arabidopsis*. *Proc. Natl. Acad. Sci. U.S.A.* 103, 17042–17047. doi: 10.1073/pnas.0607877103

Berger, S. L. (2007). The complex language of chromatin regulation during transcription. *Nature* 447, 407–412. doi: 10.1038/nature05915

Bouyer, D., Roudier, F., Heese, M., Andersen, E. D., Gey, D., Nowack, M. K., et al. (2011). Polycomb repressive complex 2 controls the embryo-to-seedling phase transition. *PLoS Genet.* 7:e1002014. doi: 10.1371/journal.pgen.1002014

Bratzel, F., López-Torrejón, G., Koch, M., Del Pozo, J. C., and Calonje, M. (2010). Keeping cell identity in *Arabidopsis* requires PRC1 RING-finger homologs that catalyze H2A monoubiquitination. *Curr. Biol.* 20, 1853–1859. doi: 10.1016/j.cub.2010.09.046

Braybrook, S. A., Stone, S. L., Park, S., Bui, A. Q., Le, B. H., Fischer, R. L., et al. (2006). Genes directly regulated by LEAFY COTYLEDON2 provide insight into the control of embryo maturation and somatic embryogenesis. *Proc. Natl. Acad. Sci. U.S.A.* 103, 3468–3473. doi: 10.1073/pnas.0511331103

Carles, C. C., and Fletcher, J. C. (2009). The SAND domain protein ULTRAPETALA1 acts as a trithorax group factor to regulate cell fate in plants. *Genes Dev.* 23, 2723–2728. doi: 10.1101/gad.1812609

Carrillo-Barral, N., Rodríguez-Gacio, M. D. C., and Matilla, A. J. (2020). Delay of germination-1 (DOG1): a key to understanding seed dormancy. *Plants (Basel)* 9:480. doi: 10.3390/plants9040480

Carter, B., Bishop, B., Ho, K. K., Huang, R., Jia, W., Zhang, H., et al. (2018). The chromatin remodelers PKL and PIE1 act in an epigenetic pathway that determines H3K27me3 homeostasis in *Arabidopsis*. *Plant Cell* 30, 1337–1352. doi: 10.1105/tpc.17.00867

Chang, G., Wang, C., Kong, X., Chen, Q., Yang, Y., and Hu, X. (2018). AFP2 as the novel regulator breaks high-temperature-induced seeds secondary dormancy through ABI5 and SOM in *Arabidopsis thaliana*. *Biochem. Biophys. Res. Commun.* 501, 232–238. doi: 10.1016/j.bbrc.2018.04.222

Chanvivattana, Y., Bishopp, A., Schubert, D., Stock, C., Moon, Y. H., Sung, Z. R., et al. (2004). Interaction of polycomb-group proteins controlling flowering in *Arabidopsis*. *Development* 131, 5263–5276. doi: 10.1242/dev.01400

Chaudhury, A. M., Ming, L., Miller, C., Craig, S., Dennis, E. S., and Peacock, W. J. (1997). Fertilization-independent seed development in *Arabidopsis thaliana*. *Proc. Natl. Acad. Sci. U.S.A.* 94, 4223–4228. doi: 10.1073/pnas.94.8.4223

Chen, D., Molitor, A., Liu, C., and Shen, W. H. (2010). The *Arabidopsis* PRC1-like ring-finger proteins are necessary for repression of embryonic traits during vegetative growth. *Cell Res.* 20, 1332–1344. doi: 10.1038/cr.2010.151

Chen, H., Tong, J., Fu, W., Liang, Z., Ruan, J., Yu, Y., et al. (2020). The H3K27me3 demethylase RELATIVE OF EARLY FLOWERING6 suppresses seed dormancy by inducing abscisic acid catabolism. *Plant Physiol.* 184, 1969–1978. doi: 10.1104/pp.20.01255

Chen, L. T., Luo, M., Wang, Y. Y., and Wu, K. (2010). Involvement of *Arabidopsis* histone deacetylase HDA6 in ABA and salt stress response. *J. Exp. Bot.* 61, 3345–3353. doi: 10.1093/jxb/erq154

Chen, L. T., and Wu, K. (2010). Role of histone deacetylases HDA6 and HDA19 in ABA and abiotic stress response. *Plant Signal. Behav.* 5, 1318–1320. doi: 10.4161/psb.5.10.13168

Chen, N., Wang, H., Abdelmageed, H., Veerappan, V., Tadege, M., and Allen, R. D. (2020). HSI2/VAL1 and HSL1/VAL2 function redundantly to repress DOG1 expression in *Arabidopsis* seeds and seedlings. *New Phytol.* 227, 840–856. doi: 10.1111/nph.16559

Cheng, L., Shafiq, S., Xu, W., and Sun, Q. (2018). EARLY FLOWERING IN SHORT DAYS (EFS) regulates the seed size in *Arabidopsis*. *Sci. China Life Sci.* 61, 214–224. doi: 10.1007/s11427-017-9236-x

Cheng, X., Pan, M., E, Z., Zhou, Y., Niu, B., and Chen, C. (2020). Functional divergence of two duplicated fertilization independent endosperm genes in rice with respect to seed development. *Plant J.* 104, 124–137. doi: 10.1111/tpj.14911

Cheng, X., Pan, M., E, Z., Zhou, Y., Niu, B., and Chen, C. (2021). The maternally expressed polycomb group gene OsEMF2a is essential for endosperm cellularization and imprinting in rice. *Plant Commun.* 2:100092. doi: 10.1016/j.xplc.2020.100092

Cho, J. N., Ryu, J. Y., Jeong, Y. M., Park, J., Song, J. J., Amasino, R. M., et al. (2012). Control of seed germination by light-induced histone arginine demethylation activity. *Dev. Cell* 22, 736–748. doi: 10.1016/j.devcel.2012.01.024

Cigliano, R. A., Cremona, G., Paparo, R., Termolino, P., Perrella, G., Gutzat, R., et al. (2013). Histone deacetylase AtHDA7 is required for female gametophyte and embryo development in *Arabidopsis*. *Plant Physiol.* 163, 431–440. doi: 10.1104/pp.113.221713

Colville, A., Alhattab, R., Hu, M., Labbé, H., Xing, T., and Miki, B. (2011). Role of HD2 genes in seed germination and early seedling growth in *Arabidopsis*. *Plant Cell Rep.* 30, 1969–1979. doi: 10.1007/s00299-011-1105-z

Danilevskaya, O. N., Hermon, P., Hantke, S., Muszynski, M. G., Kollipara, K., and Ananiev, E. V. (2003). Duplicated fie genes in maize: expression pattern and imprinting suggest distinct functions. *Plant Cell* 15, 425–438. doi: 10.1105/tpc.006759

Dean Rider, S. Jr., Henderson, J. T., Jerome, R. E., Edenberg, H. J., Romero-Severson, J., and Ogas, J. (2003). Coordinate repression of regulators of embryonic identity by PICKLE during germination in *Arabidopsis*. *Plant J.* 35, 33–43. doi: 10.1046/j.1365-313x.2003.01783.x

Dutnall, R. N. (2003). Cracking the histone code: one, two, three methyls, you're out! *Mol. Cell* 12, 3–4. doi: 10.1016/s1097-2765(03)00282-x

Fedak, H., Palusinska, M., Krzyczmonik, K., Brzezniak, L., Yatusevich, R., Pietras, Z., et al. (2016). Control of seed dormancy in *Arabidopsis* by a cis-acting noncoding antisense transcript. *Proc. Natl. Acad. Sci U.S.A.* 113, 7846–7855. doi: 10.1073/pnas.1608827113

Figueiredo, D. D., Batista, R. A., Roszak, P. J., and Köhler, C. (2015). Auxin production couples endosperm development to fertilization. *Nat. Plants* 1:15184. doi: 10.1038/nplants.2015.184

Figueiredo, D. D., and Köhler, C. (2018). Auxin: a molecular trigger of seed development. *Genes. Dev.* 32, 479–490. doi: 10.1101/gad.312546.118

Finch-Savage, W. E., and Footitt, S. (2017). Seed dormancy cycling and the regulation of dormancy mechanisms to time germination in variable field environments. *J. Exp. Bot.* 68, 843–856. doi: 10.1093/jxb/erw477

Finch-Savage, W. E., and Leubner-Metzger, G. (2006). Seed dormancy and the control of germination. *New Phytol.* 171, 501–523. doi: 10.1111/j.1469-8137.2006.01787.x

Fiorucci, A. S., Bourbousse, C., Concia, L., Rougée, M., Deton-Cabanillas, A. F., Zabulon, G., et al. (2019). *Arabidopsis* S2Lb links AtCOMPASS-like and SDG2 activity in H3K4me3 independently from histone H2B monoubiquitination. *Genome Biol.* 20:100. doi: 10.1186/s13059-019-1705-4

Fletcher, J. C. (2017). State of the art: trxG factor regulation of post-embryonic plant development. *Front. Plant. Sci.* 8:1925. doi: 10.3389/fpls.2017.01925

Folsom, J. J., Begcy, K., Hao, X., Wang, D., and Walia, H. (2014). Rice fertilization-independent endosperm1 regulates seed size under heat stress by controlling early endosperm development. *Plant Physiol.* 165, 238–248. doi: 10.1104/pp.113.232413

Footitt, S., Muller, K., Kermode, A. R., and Finch-Savage, W. E. (2015). Seed dormancy cycling in *Arabidopsis*: chromatin remodelling and regulation of DOG1 in response to seasonal environmental signals. *Plant J.* 81, 413–425. doi: 10.1111/tpj.12735

Forestan, C., Farinati, S., Rouster, J., Lassagne, H., Lauria, M., Dal Ferro, N., et al. (2018). Control of maize vegetative and reproductive development, fertility, and rRNAs silencing by HISTONE DEACETYLASE 108. *Genetics* 208, 1443–1466. doi: 10.1534/genetics.117.300625

Fukaki, H., Taniguchi, N., and Tasaka, M. (2006). PICKLE is required for SOLITARY-ROOT/IAA14-mediated repression of ARF7 and ARF19 activity during *Arabidopsis* lateral root initiation. *Plant J.* 48, 380–389. doi: 10.1111/j.1365-313X.2006.02882.x

Gazzarrini, S., Tsuchiya, Y., Lumba, S., Okamoto, M., and McCourt, P. (2004). The transcription factor FUSCA3 controls developmental timing in *Arabidopsis* through the hormones gibberellin and abscisic acid. *Dev. Cell* 7, 373–385. doi: 10.1016/j.devcel.2004.06.017

Giraudat, J., Hauge, B. M., Valon, C., Smalle, J., Parcy, F., and Goodman, H. M. (1992). Isolation of the *Arabidopsis* ABI3 gene by positional cloning. *Plant Cell* 4, 1251–1261. doi: 10.1105/tpc.4.10.1251

Goldberg, A. D., Allis, C. D., and Bernstein, E. (2007). Epigenetics: a landscape takes shape. *Cell* 128, 635–638. doi: 10.1016/j.cell.2007.02.006

Graeber, K., Voegele, A., Büttner-Mainik, A., Sperber, K., Mummenhoff, K., and Leubner-Metzger, G. (2013). Spatiotemporal seed development analysis provides insight into primary dormancy induction and evolution of the *Lepidium* delay of germination1 genes. *Plant Physiol.* 161, 1903–1917. doi: 10.1104/pp.112.213298

Grossniklaus, U., Vielle-Calzada, J. P., Hoeppner, M. A., and Gagliano, W. B. (1998). Maternal control of embryogenesis by MEDEA, a polycomb group gene in *Arabidopsis*. *Science* 280, 446–450. doi: 10.1126/science.280.5362.446

Gu, D., Chen, C. Y., Zhao, M., Zhao, L., Duan, X., Duan, J., et al. (2017). Identification of HDA15-PIF1 as a key repression module directing the transcriptional network of seed germination in the dark. *Nucleic. Acids Res.* 45, 7137–7150. doi: 10.1093/nar/gkx283

Gu, D., Ji, R., He, C., Peng, T., Zhang, M., Duan, J., et al. (2019). *Arabidopsis* histone methyltransferase SUVH5 is a positive regulator of light-mediated seed germination. *Front. Plant. Sci.* 10:841. doi: 10.3389/fpls.2019.00841

Han, S. K., Sang, Y., Rodrigues, A., Wu, M. F., Rodriguez, P. L., and Wagner, D. (2012). The SWI2/SNF2 chromatin remodeling ATPase BRAHMA represses abscisic acid responses in the absence of the stress stimulus in *Arabidopsis*. *Plant Cell* 24, 4892–4906. doi: 10.1105/tpc.112.105114

Hargreaves, D. C., and Crabtree, G. R. (2011). ATP-dependent chromatin remodeling: genetics, genomics and mechanisms. *Cell Res.* 21, 396–420. doi: 10.1038/cr.2011.32

Hehenberger, E., Kradolfer, D., and Köhler, C. (2012). Endosperm cellularization defines an important developmental transition for embryo development. *Development* 139, 2031–2039. doi: 10.1242/dev.077057

Henderson, J. T., Li, H. C., Rider, S. D., Mordhorst, A. P., Romero-Severson, J., Cheng, J. C., et al. (2004). PICKLE acts throughout the plant to repress expression of embryonic traits and may play a role in gibberellin-dependent responses. *Plant Physiol.* 134, 995–1005. doi: 10.1104/pp.103.030148

Leeggangers, H. A., Folta, A., Muras, A., Nap, J. P., and Mlynarova, L. (2015). Reduced seed germination in *Arabidopsis* over-expressing SWI/SNF2 ATPase genes. *Physiol. Plant* 153, 318–326. doi: 10.1111/ppl.12231

Holdsworth, M. J., Bentsink, L., and Soppe, W. J. J. (2008). Molecular networks regulating *Arabidopsis* seed maturation, after-ripening, dormancy and germination. *New Phytol.* 179, 33–54. doi: 10.1111/j.1469-8137.2008.02437.x

Huang, X., Lu, Z., Wang, X., Ouyang, Y., Chen, W., Xie, K., et al. (2016). Imprinted gene OsFIE1 modulates rice seed development by influencing nutrient metabolism and modifying genome H3K27me3. *Plant J.* 87, 305–317. doi: 10.1111/tpj.13202

Hurtado, L., Farrona, S., and Reyes, J. C. (2006). The putative SWI/SNF complex subunit BRAHMA activates flower homeotic genes in *Arabidopsis thaliana*. *Plant Mol. Biol.* 62, 291–304. doi: 10.1007/s11103-006-9021-2

Ishikawa, R., Ohnishi, T., Kinoshita, Y., Eiguchi, M., Kurata, N., and Kinoshita, T. (2011). Rice interspecies hybrids show precocious or delayed developmental transitions in the endosperm without change to the rate of syncytial nuclear division. *Plant J.* 65, 798–806. doi: 10.1111/j.1365-313X.2010.04466.x

Jacobsen, J. V., Pearce, D. W., Poole, A. T., Pharis, R. P., and Mander, L. N. (2002). Abscisic acid, phaseic acid and gibberellin contents associated with dormancy and germination in barley. *Physiol. Plant* 115, 428–441. doi: 10.1034/j.1399-3054.2002.1150313.x

Jia, H., Suzuki, M., and McCarty, D. R. (2014). Regulation of the seed to seedling developmental phase transition by the LAFL and VAL transcription factor networks. *Wiley Interdiscip. Rev. Dev. Biol.* 3, 135–145. doi: 10.1002/wdev.126

Jiang, D., Gu, X., and He, Y. (2009). Establishment of the winter-annual growth habit via FRIGIDA-mediated histone methylation at FLOWERING LOCUS C in *Arabidopsis*. *Plant Cell* 21, 1733–1746. doi: 10.1105/tpc.109.067967

Jiang, D., Kong, N. C., Gu, X., Li, Z., and He, Y. (2011). *Arabidopsis* COMPASS-like complexes mediate histone H3 lysine-4 trimethylation to control floral transition and plant development. *PLoS Genet.* 7:e1001330. doi: 10.1371/journal.pgen.1001330

Kang, I. H., Steffen, J. G., Portereiko, M. F., Lloyd, A., and Drews, G. N. (2008). The AGL62 MADS domain protein regulates cellularization during endosperm development in *Arabidopsis*. *Plant Cell* 20, 635–647. doi: 10.1105/tpc.107.055137

Keith, K., Kraml, M., Dengler, N. G., and McCourt, P. (1994). fusca3: a heterochronic mutation affecting late embryo development in *Arabidopsis*. *Plant Cell* 6, 589–600. doi: 10.1105/tpc.6.5.589

Kim, D. H., Yamaguchi, S., Lim, S., Oh, E., Park, J., Hanada, A., et al. (2008). SOMNUS, a CCCH-type zinc finger protein in *Arabidopsis*, negatively regulates light-dependent seed germination downstream of PIL5. *Plant Cell* 20, 1260–1277. doi: 10.1105/tpc.108.058859

Kim, S. Y., Lee, J., Eshed-Williams, L., Zilberman, D., and Sung, Z. R. (2012). EMF1 and PRC2 cooperate to repress key regulators of *Arabidopsis* development. *PLoS Genet.* 8:e1002512. doi: 10.1371/journal.pgen.1002512

Kinoshita, T., Harada, J. J., Goldberg, R. B., and Fischer, R. L. (2001). Polycomb repression of flowering during early plant development. *Proc. Natl. Acad. Sci. U.S.A.* 98, 14156–14161. doi: 10.1073/pnas.241507798

Kiyosue, T., Ohad, N., Yadegari, R., Hannon, M., Dinneny, J., Wells, D., et al. (1999). Control of fertilization-independent endosperm development by the MEDEA polycomb gene in *Arabidopsis*. *Proc. Natl. Acad. Sci U.S.A.* 96, 4186–4191. doi: 10.1073/pnas.96.7.4186

Köhler, C., Hennig, L., Spillane, C., Pien, S., Gruissem, W., and Grossniklaus, U. (2003b). The polycomb-group protein MEDEA regulates seed development by controlling expression of the MADS-box gene PHERES1. *Genes Dev.* 17, 1540–1553. doi: 10.1101/gad.257403

Köhler, C., Hennig, L., Bouveret, R., Gheyselinck, J., Grossniklaus, U., and Gruissem, W. (2003a). *Arabidopsis* MSI1 is a component of the MEA/FIE Polycomb group complex and required for seed development. *EMBO. J.* 22, 4804–4814. doi: 10.1093/emboj/cdg444

Kradolfer, D., Wolff, P., Jiang, H., Siretskiy, A., and Köhler, C. (2013). An imprinted gene underlies postzygotic reproductive isolation in *Arabidopsis thaliana*. *Dev. Cell* 26, 525–535. doi: 10.1016/j.devcel.2013.08.006

Kumar, V., Thakur, J. K., and Prasad, M. (2021). Histone acetylation dynamics regulating plant development and stress responses. *Cell Mol. Life Sci.* 78, 4467–4486. doi: 10.1007/s00018-021-03794-x

Lee, N., Kang, H., Lee, D., and Choi, G. (2014). A histone methyltransferase inhibits seed germination by increasing PIF1 mRNA expression in imbibed seeds. *Plant J.* 78, 282–293. doi: 10.1111/tpj.12467

Lepiniec, L., Devic, M., Roscoe, T. J., Bouyer, D., Zhou, D. X., Boulard, C., et al. (2018). Molecular and epigenetic regulations and functions of the LAFL transcriptional regulators that control seed development. *Plant Reprod.* 31, 291–307. doi: 10.1007/s00497-018-0337-2

Li, C., Gu, L., Gao, L., Chen, C., Wei, C. Q., Qiu, Q., et al. (2016). Concerted genomic targeting of H3K27 demethylase REF6 and chromatin-remodeling ATPase BRM in *Arabidopsis*. *Nat. Genet.* 48, 687–693. doi: 10.1038/ng.3555

Li, H. C., Chuang, K., Henderson, J. T., Rider, S. D. Jr., Bai, Y., Zhang, H., et al. (2005). PICKLE acts during germination to repress expression of embryonic traits. *Plant J.* 44, 1010–1022. doi: 10.1111/j.1365-313X.2005.02602.x

Li, P., Zhang, Q., He, D., Zhou, Y., Ni, H., Tian, D., et al. (2020). AGAMOUS-LIKE67 cooperates with the histone mark reader EBS to modulate seed germination under high temperature. *Plant Physiol.* 184, 529–545. doi: 10.1104/pp.20.00056

Li, S., Zhou, B., Peng, X., Kuang, Q., Huang, X., Yao, J., et al. (2014). OsFIE2 plays an essential role in the regulation of rice vegetative and reproductive development. *New Phytol.* 201, 66–79. doi: 10.1111/nph.12472

Li, X., Jiang, Y., Ji, Z., Liu, Y., and Zhang, Q. (2015). BRHIS1 suppresses rice innate immunity through binding to monoubiquitinated H2A and H2B variants. *EMBO Rep.* 16, 1192–1202. doi: 10.15252/embr.201440000

Lim, S., Park, J., Lee, N., Jeong, J., Toh, S., Watanabe, A., et al. (2013). ABA-insensitive3, ABA-insensitive5, and DELLAs Interact to activate the expression of SOMNUS and other high-temperature-inducible genes in imbibed seeds in *Arabidopsis*. *Plant Cell* 25, 4863–4878. doi: 10.1105/tpc.113.118604

Lin, J. Y., Le, B. H., Chen, M., Henry, K. F., Hur, J., Hsieh, T. F., et al. (2017). Similarity between soybean and *Arabidopsis* seed methylomes and loss of non-CG methylation does not affect seed development. *Proc. Natl. Acad. Sci U.S.A.* 114, E9730–E9739. doi: 10.1073/pnas.1716758114

Liu, B., Liu, Y., Wang, B., Luo, Q., Shi, J., Gan, J., et al. (2019). The transcription factor OsSUF4 interacts with SDG725 in promoting H3K36me3 establishment. *Nat. Commun.* 10:2999. doi: 10.1038/s41467-019-10850-5

Liu, J., Deng, S., Wang, H., Ye, J., Wu, H. W., Sun, H. X., et al. (2016). CURLY LEAF regulates gene sets coordinating seed size and lipid biosynthesis. *Plant Physiol.* 171, 424–436. doi: 10.1104/pp.15.01335

Liu, X., Luo, J., Li, T., Yang, H., Wang, P., Su, L., et al. (2021). SDG711 is involved in rice seed development through regulation of starch metabolism gene expression in coordination with other histone modifications. *Rice (N Y)* 14:25. doi: 10.1186/s12284-021-00467-y

Liu, X., Wei, X., Sheng, Z., Jiao, G., Tang, S., Luo, J., et al. (2016). Polycomb protein OsFIE2 affects plant height and grain yield in rice. *PLoS One* 11:e0164748. doi: 10.1371/journal.pone.0164748

Liu, Y., Geyer, R., van Zanten, M., Carles, A., Li, Y., Hörold, A., et al. (2011). Identification of the *Arabidopsis* REDUCED DORMANCY 2 gene uncovers a role for the polymerase associated factor 1 complex in seed dormancy. *PLoS One* 6:e22241. doi: 10.1371/journal.pone.0022241

Liu, Y., Koornneef, M., and Soppe, W. J. (2007). The absence of histone H2B monoubiquitination in the *Arabidopsis* hub1 (rdo4) mutant reveals a role for chromatin remodeling in seed dormancy. *Plant Cell* 19, 433–444. doi: 10.1105/tpc.106.049221

Lotan, T., Ohto, M., Yee, K. M., West, M. A., Lo, R., Kwong, R. W., et al. (1998). *Arabidopsis* LEAFY COTYLEDON1 is sufficient to induce embryo development in vegetative cells. *Cell* 93, 1195–1205. doi: 10.1016/s0092-8674(00)81463-4

Lu, X., Wang, W., Ren, W., Chai, Z., Guo, W., Chen, R., et al. (2015). Genome-wide epigenetic regulation of gene transcription in maize seeds. *PLoS One* 10:e0139582. doi: 10.1371/journal.pone.0139582

Luo, M., Platten, D., Chaudhury, A., Peacock, W. J., and Dennis, E. S. (2009). Expression, imprinting, and evolution of rice homologs of the polycomb group genes. *Mol. Plant* 2, 711–723. doi: 10.1093/mp/ssp036

Luo, M., Wang, Y. Y., Liu, X., Yang, S., Lu, Q., Cui, Y., et al. (2012). HD2C interacts with HDA6 and is involved in ABA and salt stress response in *Arabidopsis*. *J. Exp. Bot.* 63, 3297–3306. doi: 10.1093/jxb/ers059

Luo, Q., Lian, H. L., He, S. B., Li, L., Jia, K. P., and Yang, H. Q. (2014). COP1 and phyB physically interact with PIL1 to regulate its stability and photomorphogenic development in *Arabidopsis*. *Plant Cell* 26, 2441–2456. doi: 10.1105/tpc.113.121657

Makarevich, G., Leroy, O., Akinci, U., Schubert, D., Clarenz, O., Goodrich, J., et al. (2006). Different polycomb group complexes regulate common target genes in *Arabidopsis*. *EMBO Rep.* 7, 947–952. doi: 10.1038/sj.embor.7400760

Miller, T., Krogan, N. J., Dover, J., Erdjument-Bromage, H., Tempst, P., Johnston, M., et al. (2001). COMPASS: a complex of proteins associated with a trithorax-related SET domain protein. *Proc. Natl. Acad. Sci U.S.A.* 98, 12902–12907. doi: 10.1073/pnas.231473398

Moon, Y. H., Chen, L., Pan, R. L., Chang, H. S., Zhu, T., Maffeo, D. M., et al. (2003). EMF genes maintain vegetative development by repressing the flower program in *Arabidopsis*. *Plant Cell* 15, 681–693. doi: 10.1105/tpc.007831

Mozgova, I., and Hennig, L. (2015). The polycomb group protein regulatory network. *Annu. Rev. Plant Biol.* 66, 269–296. doi: 10.1146/annurev-arplant-043014-115627

Müller, K., Bouyer, D., Schnittger, A., and Kermode, A. R. (2012). Evolutionarily conserved histone methylation dynamics during seed life-cycle transitions. *PLoS One* 7:e51532. doi: 10.1371/journal.pone.0051532

Nakabayashi, K., Bartsch, M., Xiang, Y., Miatton, E., Pellengahr, S., Yano, R., et al. (2012). The time required for dormancy release in *Arabidopsis* is determined by DELAY OF GERMINATION1 protein levels in freshly harvested seeds. *Plant Cell* 24, 2826–2838. doi: 10.1105/tpc.112.100214

Nallamilli, B. R., Zhang, J., Mujahid, H., Malone, B. M., Bridges, S. M., and Peng, Z. (2013). Polycomb group gene OsFIE2 regulates rice (*Oryza sativa*) seed development and grain filling via a mechanism distinct from *Arabidopsis*. *PLoS Genet.* 9:e1003322. doi: 10.1371/journal.pgen.1003322

Nambara, E., Hayama, R., Tsuchiya, Y., Nishimura, M., Kawaide, H., Kamiya, Y., et al. (2000). The role of ABI3 and FUS3 loci in *Arabidopsis thaliana* on phase transition from late embryo development to germination. *Dev. Biol.* 220, 412–423. doi: 10.1006/dbio.2000.9632

Narro-Diego, L., López-González, L., Jarillo, J. A., and Piñeiro, M. (2017). The PHD-containing protein EARLY BOLTING IN SHORT DAYS regulates seed dormancy in *Arabidopsis*. *Plant Cell Environ.* 40, 2393–2405. doi: 10.1111/pce.13046

Née, G., Xiang, Y., and Soppe, W. J. (2017). The release of dormancy, a wake-up call for seeds to germinate. *Curr. Opin. Plant Biol.* 35, 8–14. doi: 10.1016/j.pbi.2016.09.002

Ogas, J., Cheng, J. C., Sung, Z. R., and Somerville, C. (1997). Cellular differentiation regulated by gibberellin in the *Arabidopsis thaliana* pickle mutant. *Science* 277, 91–94. doi: 10.1126/science.277.5322.91

Ogas, J., Kaufmann, S., Henderson, J., and Somerville, C. (1999). PICKLE is a CHD3 chromatin-remodeling factor that regulates the transition from embryonic to vegetative development in *Arabidopsis*. *Proc. Natl. Acad. Sci U.S.A.* 96, 13839–13844. doi: 10.1073/pnas.96.24.13839

Oh, E., Yamaguchi, S., Hu, J., Yusuke, J., Jung, B., Paik, I., et al. (2007). PIL5, a phytochrome-interacting bHLH protein, regulates gibberellin responsiveness by binding directly to the GAI and RGA promoters in *Arabidopsis* seeds. *Plant Cell* 19, 1192–1208. doi: 10.1105/tpc.107.050153

Oh, E., Yamaguchi, S., Kamiya, Y., Bae, G., Chung, W. I., and Choi, G. (2006). Light activates the degradation of PIL5 protein to promote seed germination through gibberellin in *Arabidopsis*. *Plant J.* 47, 124–139. doi: 10.1111/j.1365-313X.2006.02773.x

Ohad, N., Yadegari, R., Margossian, L., Hannon, M., Michaeli, D., Harada, J. J., et al. (1999). Mutations in FIE, a WD polycomb group gene, allow endosperm development without fertilization. *Plant Cell* 11, 407–416. doi: 10.1105/tpc.11.3.407

Olsen, O. A. (2004). Nuclear endosperm development in cereals and *Arabidopsis thaliana*. *Plant Cell* 16(Suppl), S214–S227. doi: 10.1105/tpc.017111

Parcy, F., and Giraudat, J. (1997). Interactions between the ABI1 and the ectopically expressed ABI3 genes in controlling abscisic acid responses in *Arabidopsis* vegetative tissues. *Plant J.* 11, 693–702. doi: 10.1046/j.1365-313x.1997.11040693.x

Park, J., Lee, N., Kim, W., Lim, S., and Choi, G. (2011). ABI3 and PIL5 collaboratively activate the expression of SOMNUS by directly binding to its promoter in imbibed *Arabidopsis* seeds. *Plant Cell* 23, 1404–1415. doi: 10.1105/tpc.110.080721

Perrella, G., Lopez-Vernaza, M. A., Carr, C., Sani, E., Gosselé, V., Verduyn, C., et al. (2013). Histone deacetylase complex1 expression level titrates plant growth and abscisic acid sensitivity in *Arabidopsis*. *Plant Cell* 25, 3491–3505. doi: 10.1105/tpc.113.114835

Rodrigues, J. C., Luo, M., Berger, F., and Koltunow, A. M. (2010). Polycomb group gene function in sexual and asexual seed development in angiosperms. *Sex Plant Reprod.* 23, 123–133. doi: 10.1007/s00497-009-0131-2

Rossi, V., Varotto, S., Locatelli, S., Lanzanova, C., Lauria, M., Zanotti, E., et al. (2001). The maize WD-repeat gene ZmRbAp1 encodes a member of the MSI/RbAp sub-family and is differentially expressed during endosperm development. *Mol. Genet. Genomics* 265, 576–584. doi: 10.1007/s004380100461

Saez, A., Rodrigues, A., Santiago, J., Rubio, S., and Rodriguez, P. L. (2008). HAB1-SWI3B interaction reveals a link between abscisic acid signaling and putative SWI/SNF chromatin-remodeling complexes in *Arabidopsis*. *Plant Cell* 20, 2972–2988. doi: 10.1105/tpc.107.056705

Sarnowski, T. J., Ríos, G., Jásik, J., Swiezewski, S., Kaczanowski, S., Li, Y., et al. (2005). SWI3 subunits of putative SWI/SNF chromatin-remodeling complexes play distinct roles during *Arabidopsis* development. *Plant Cell* 17, 2454–2472. doi: 10.1105/tpc.105.031203

Schubert, D., Clarenz, O., and Goodrich, J. (2005). Epigenetic control of plant development by polycomb-group proteins. *Curr. Opin. Plant Biol.* 8, 553–561. doi: 10.1016/j.pbi.2005.07.005

Sekine, D., Ohnishi, T., Furuumi, H., Ono, A., Yamada, T., Kurata, N., et al. (2013). Dissection of two major components of the post-zygotic hybridization barrier in rice endosperm. *Plant J.* 76, 792–799. doi: 10.1111/tpj.12333

Seo, M., Nambara, E., Choi, G., and Yamaguchi, S. (2009). Interaction of light and hormone signals in germinating seeds. *Plant Mol. Biol.* 69, 463–472. doi: 10.1007/s11103-008-9429-y

Shu, K., Liu, X. D., Xie, Q., and He, Z. H. (2016). Two faces of one seed: hormonal regulation of dormancy and germination. *Mol. Plant* 9, 34–45. doi: 10.1016/j.molp.2015.08.010

Song, X. J., Kuroha, T., Ayano, M., Furuta, T., Nagai, K., Komeda, N., et al. (2015). Rare allele of a previously unidentified histone H4 acetyltransferase enhances grain weight, yield, and plant biomass in rice. *Proc. Natl. Acad. Sci. U.S.A.* 112, 76–81. doi: 10.1073/pnas.1421127112

Sørensen, M. B., Chaudhury, A. M., Robert, H., Bancharel, E., and Berger, F. (2001). Polycomb group genes control pattern formation in plant seed. *Curr. Biol.* 11, 277–281. doi: 10.1016/s0960-9822(01)00072-0

Spillane, C., Schmid, K. J., Laoueillé-Duprat, S., Pien, S., Escobar-Restrepo, J. M., Baroux, C., et al. (2007). Positive darwinian selection at the imprinted MEDEA locus in plants. *Nature* 448, 349–352. doi: 10.1038/nature05984

Springer, N. M., Danilevskaya, O. N., Hermon, P., Helentjaris, T. G., Phillips, R. L., Kaeppler, H. F., et al. (2002). Sequence relationships, conserved domains, and expression patterns for maize homologs of the polycomb group genes E(z), esc, and E(Pc). *Plant Physiol.* 128, 1332–1345. doi: 10.1104/pp.010742

Stone, S. L., Kwong, L. W., Yee, K. M., Pelletier, J., Lepiniec, L., Fischer, R. L., et al. (2001). LEAFY COTYLEDON2 encodes a B3 domain transcription factor that induces embryo development. *Proc. Natl. Acad. Sci U.S.A.* 98, 11806–11811. doi: 10.1073/pnas.201413498

Strahl, B. D., and Allis, C. D. (2000). The language of covalent histone modifications. *Nature* 403, 41–45. doi: 10.1038/47412

Sun, Z. W., and Allis, C. D. (2002). Ubiquitination of histone H2B regulates H3 methylation and gene silencing in yeast. *Nature* 418, 104–108. doi: 10.1038/nature00883

Suzuki, M., Wang, H. H., and McCarty, D. R. (2007). Repression of the LEAFY COTYLEDON 1/B3 regulatory network in plant embryo development by VP1/ABSCISIC ACID INSENSITIVE 3-LIKE B3 genes. *Plant Physiol.* 143, 902–911. doi: 10.1104/pp.106.092320

Sweeney, M., and McCouch, S. (2007). The complex history of the domestication of rice. *Ann. Bot.* 100, 951–957. doi: 10.1093/aob/mcm128

Tanaka, M., Kikuchi, A., and Kamada, H. (2008). The *Arabidopsis* histone deacetylases HDA6 and HDA19 contribute to the repression of embryonic properties after germination. *Plant Physiol.* 146, 149–161. doi: 10.1104/pp.107.111674

Tang, X., Lim, M. H., Pelletier, J., Tang, M., Nguyen, V., Keller, W. A., et al. (2012). Synergistic repression of the embryonic programme by SET DOMAIN GROUP 8 and EMBRYONIC FLOWER 2 in *Arabidopsis* seedlings. *J. Exp. Bot.* 63, 1391–1404. doi: 10.1093/jxb/err383

Toh, S., Imamura, A., Watanabe, A., Nakabayashi, K., Okamoto, M., Jikumaru, Y., et al. (2008). High temperature-induced abscisic acid biosynthesis and its role in the inhibition of gibberellin action in *Arabidopsis* seeds. *Plant Physiol.* 146, 1368–1385. doi: 10.1104/pp.107.113738

Tong, J. K., Hassig, C. A., Schnitzler, G. R., Kingston, R. E., and Schreiber, S. L. (1998). Chromatin deacetylation by an ATP-dependent nucleosome remodelling complex. *Nature* 395, 917–921. doi: 10.1038/27699

Tonosaki, K., Ono, A., Kunisada, M., Nishino, M., Nagata, H., Sakamoto, S., et al. (2021). Mutation of the imprinted gene OsEMF2a induces autonomous endosperm development and delayed cellularization in rice. *Plant Cell* 33, 85–103. doi: 10.1093/plcell/koaa006

Turner, B. M. (2000). Histone acetylation and an epigenetic code. *Bioessays* 22, 836–845. doi: 10.1002/1521-1878(200009)22:9<836::Aid-bies9<3.0.Co;2-x

Vaistij, F. E., Barros-Galvão, T., Cole, A. F., Gilday, A. D., He, Z., Li, Y., et al. (2018). MOTHER-OF-FT-AND-TFL1 represses seed germination under far-red light by modulating phytohormone responses in *Arabidopsis thaliana*. *Proc. Natl. Acad. Sci. U.S.A.* 115, 8442–8447. doi: 10.1073/pnas.1806460115

Vaistij, F. E., Gan, Y., Penfield, S., Gilday, A. D., Dave, A., He, Z., et al. (2013). Differential control of seed primary dormancy in *Arabidopsis* ecotypes by the transcription factor SPATULA. *Proc. Natl. Acad. Sci U.S.A.* 110, 10866–10871. doi: 10.1073/pnas.1301647110

van Zanten, M., Koini, M. A., Geyer, R., Liu, Y., Brambilla, V., Bartels, D., et al. (2011). Seed maturation in *Arabidopsis thaliana* is characterized by nuclear size reduction and increased chromatin condensation. *Proc. Natl. Acad. Sci U.S.A.* 108, 20219–20224. doi: 10.1073/pnas.1117726108

Walck, J. L., Hidayati, S. N., Dixon, K. W., Thompson, K. A., and Poschlod, P. (2011). Climate change and plant regeneration from seed. *Glob. Change Biol.* 17, 2145–2161. doi: 10.1111/j.1365-2486.2010.02368.x

Walia, H., Josefsson, C., Dilkes, B., Kirkbride, R., Harada, J., and Comai, L. (2009). Dosage-dependent deregulation of an AGAMOUS-LIKE gene cluster contributes to interspecific incompatibility. *Curr. Biol.* 19, 1128–1132. doi: 10.1016/j.cub.2009.05.068

Wang, D., Tyson, M. D., Jackson, S. S., and Yadegari, R. (2006). Partially redundant functions of two SET-domain polycomb-group proteins in controlling initiation of seed development in *Arabidopsis*. *Proc. Natl. Acad. Sci U.S.A.* 103, 13244–13249. doi: 10.1073/pnas.0605551103

Wang, T. J., Huang, S., Zhang, A., Guo, P., Liu, Y., Xu, C., et al. (2021). JMJ17-WRKY40 and HY5-ABI5 modules regulate the expression of ABA-responsive genes in *Arabidopsis*. *New Phytol.* 230, 567–584. doi: 10.1111/nph.17177

West, M., Yee, K. M., Danao, J., Zimmerman, J. L., Fischer, R. L., Goldberg, R. B., et al. (1994). LEAFY COTYLEDON1 is an essential regulator of late embryogenesis and cotyledon identity in *Arabidopsis*. *Plant Cell* 6, 1731–1745. doi: 10.1105/tpc.6.12.1731

Wu, K., Tian, L., Malik, K., Brown, D., and Miki, B. (2000). Functional analysis of HD2 histone deacetylase homologues in *Arabidopsis thaliana*. *Plant J.* 22, 19–27. doi: 10.1046/j.1365-313x.2000.00711.x

Xu, F., Kuo, T., Rosli, Y., Liu, M. S., Wu, L., Chen, L. O., et al. (2018). Trithorax group proteins act together with a polycomb group protein to maintain chromatin integrity for epigenetic silencing during seed germination in *Arabidopsis*. *Mol. Plant* 11, 659–677. doi: 10.1016/j.molp.2018.01.010

Yadegari, R., Kinoshita, T., Lotan, O., Cohen, G., Katz, A., Choi, Y., et al. (2000). Mutations in the FIE and MEA genes that encode interacting polycomb proteins cause parent-of-origin effects on seed development by distinct mechanisms. *Plant Cell* 12, 2367–2382. doi: 10.1105/tpc.12.12.2367

Yamamoto, A., Kagaya, Y., Toyoshima, R., Kagaya, M., Takeda, S., and Hattori, T. (2009). *Arabidopsis* NF-YB subunits LEC1 and LEC1-LIKE activate transcription by interacting with seed-specific ABRE-binding factors. *Plant J.* 58, 843–856. doi: 10.1111/j.1365-313X.2009.03817.x

Yang, C., Bratzel, F., Hohmann, N., Koch, M., Turck, F., and Calonje, M. (2013). VAL- and AtBMI1-mediated H2Aub initiate the switch from embryonic to postgerminative growth in *Arabidopsis*. *Curr. Biol.* 23, 1324–1329. doi: 10.1016/j.cub.2013.05.050

Yang, W., Chen, Z., Huang, Y., Chang, G., Li, P., Wei, J., et al. (2019). Powerdress as the novel regulator enhances *Arabidopsis* seeds germination tolerance to high temperature stress by histone modification of SOM locus. *Plant Sci.* 284, 91–98. doi: 10.1016/j.plantsci.2019.04.001

Yuan, L., Song, X., Zhang, L., Yu, Y., Liang, Z., Lei, Y., et al. (2021). The transcriptional repressors VAL1 and VAL2 recruit PRC2 for genome-wide Polycomb silencing in *Arabidopsis*. *Nucleic Acids Res.* 49, 98–113. doi: 10.1093/nar/gkaa1129

Zanten, M.v, Liu, Y.-x, and Soppe, W. J. J. (2013). "Epigenetic signalling during the life of seeds," in *Epigenetic Memory and Control in Plants*, eds G. Grafi and N. Ohad (Berlin: Springer), 127–153. doi: 10.1007/978-3-642-35227-0_7

Zha, P., Liu, S., Li, Y., Ma, T., Yang, L., Jing, Y., et al. (2020). The evening complex and the chromatin-remodeling factor PICKLE coordinately control seed dormancy by directly repressing DOG1 in *Arabidopsis*. *Plant Commun.* 1:100011. doi: 10.1016/j.xplc.2019.100011

Zhang, H., Bishop, B., Ringenberg, W., Muir, W. M., and Ogas, J. (2012). The CHD3 remodeler PICKLE associates with genes enriched for trimethylation of histone H3 lysine 27. *Plant Physiol.* 159, 418–432. doi: 10.1104/pp.112.194878

Zhang, H., Lu, Y., Zhao, Y., and Zhou, D. X. (2016). OsSRT1 is involved in rice seed development through regulation of starch metabolism gene expression. *Plant Sci.* 248, 28–36. doi: 10.1016/j.plantsci.2016.04.004

Zhang, H., Rider, S. D. Jr., Henderson, J. T., Fountain, M., Chuang, K., Kandachar, V., et al. (2008). The CHD3 remodeler PICKLE promotes trimethylation of histone H3 lysine 27. *J. Biol. Chem.* 28, 22637–22648. doi: 10.1074/jbc.M802129200

Zhang, S., Wang, D., Zhang, H., Skaggs, M. I., Lloyd, A., Ran, D., et al. (2018). FERTILIZATION-INDEPENDENT SEED-Polycomb repressive complex 2 plays a dual role in regulating type I MADS-Box genes in early endosperm development. *Plant Physiol.* 177, 285–299. doi: 10.1104/pp.17.00534

Zhang, X., Bernatavichute, Y. V., Cokus, S., Pellegrini, M., and Jacobsen, S. E. (2009). Genome-wide analysis of mono-, di- and trimethylation of histone H3 lysine 4 in *Arabidopsis thaliana*. *Genome Biol.* 10:R62. doi: 10.1186/gb-2009-10-6-r62

Zhang, X., Clarenz, O., Cokus, S., Bernatavichute, Y. V., Pellegrini, M., Goodrich, J., et al. (2007a). Whole-genome analysis of histone H3 lysine 27 trimethylation in *Arabidopsis*. *PLoS Biol.* 5:e129. doi: 10.1371/journal.pbio.0050129

Zhang, X., Germann, S., Blus, B. J., Khorasanizadeh, S., Gaudin, V., and Jacobsen, S. E. (2007b). The *Arabidopsis* LHP1 protein colocalizes with histone H3 Lys27 trimethylation. *Nat. Struct. Mol. Biol.* 14, 869–871. doi: 10.1038/nsmb1283

Zhang, Y., LeRoy, G., Seelig, H. P., Lane, W. S., and Reinberg, D. (1998). The dermatomyositis-specific autoantigen Mi2 is a component of a complex containing histone deacetylase and nucleosome remodeling activities. *Cell* 95, 279–289. doi: 10.1016/s0092-8674(00)81758-4

Zhao, J., Li, M., Gu, D., Liu, X., Zhang, J., Wu, K., et al. (2016). Involvement of rice histone deacetylase HDA705 in seed germination and in response to ABA and abiotic stresses. *Biochem. Biophys. Res. Commun.* 470, 439–444. doi: 10.1016/j.bbrc.2016.01.016

Zhao, J., Zhang, J., Zhang, W., Wu, K., Zheng, F., Tian, L., et al. (2014). Expression and functional analysis of the plant-specific histone deacetylase HDT701 in rice. *Front. Plant. Sci.* 5:764. doi: 10.3389/fpls.2014.00764

Zhao, M., Yang, S., Liu, X., and Wu, K. (2015). *Arabidopsis* histone demethylases LDL1 and LDL2 control primary seed dormancy by regulating DELAY OF GERMINATION 1 and ABA signaling-related genes. *Front. Plant. Sci.* 6:159. doi: 10.3389/fpls.2015.00159

Zheng, J., Chen, F., Wang, Z., Cao, H., Li, X., Deng, X., et al. (2012). A novel role for histone methyltransferase KYP/SUVH4 in the control of *Arabidopsis* primary seed dormancy. *New Phytol.* 193, 605–616. doi: 10.1111/j.1469-8137.2011.03969.x

Zhou, S. R., Yin, L. L., and Xue, H. W. (2013). Functional genomics based understanding of rice endosperm development. *Curr. Opin. Plant Biol.* 16, 236–246. doi: 10.1016/j.pbi.2013.03.001

Zhou, Y., Yang, P., Zhang, F., Luo, X., and Xie, J. (2020). Histone deacetylase HDA19 interacts with histone methyltransferase SUVH5 to regulate seed dormancy in *Arabidopsis*. *Plant Biol. (Stuttg.)* 22, 1062–1071. doi: 10.1111/plb.13158

Zhu, J. K. (2016). Abiotic stress signaling and responses in plants. *Cell* 167, 313–324. doi: 10.1016/j.cell.2016.08.029

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

frontiers | Frontiers in Plant Science

ORIGINAL RESEARCH
published: 02 June 2022
doi: 10.3389/fpls.2022.907244

Tandem Mass Tag-Based Quantitative Proteomics Reveals Implication of a Late Embryogenesis Abundant Protein (BnLEA57) in Seed Oil Accumulation in *Brassica napus* L.

Zhongjing Zhou[1], Baogang Lin[2], Jinjuan Tan[1], Pengfei Hao[2], Shuijin Hua[2] and Zhiping Deng[1]**

[1] State Key Laboratory for Managing Biotic and Chemical Threats to the Quality and Safety of Agro-Products, Institute of Virology and Biotechnology, Zhejiang Academy of Agricultural Sciences, Hangzhou, China, [2] Zhejiang Key Laboratory of Digital Dry Land Crops, Institute of Crops and Nuclear Technology Utilization, Zhejiang Academy of Agricultural Sciences, Hangzhou, China

OPEN ACCESS

Edited by:
Pingfang Yang,
Hubei University, China

Reviewed by:
Liang Guo,
Huazhong Agricultural University, China
Tiago Santana Balbuena,
São Paulo State University, Brazil

***Correspondence:**
Shuijin Hua
sjhua1@163.com
Zhiping Deng
zhipingdeng@zaas.ac.cn

Specialty section:
This article was submitted to Plant Proteomics and Protein Structural Biology, a section of the journal Frontiers in Plant Science

Received: 29 March 2022
Accepted: 03 May 2022
Published: 02 June 2022

Citation:
Zhou Z, Lin B, Tan J, Hao P, Hua S and Deng Z (2022) Tandem Mass Tag-Based Quantitative Proteomics Reveals Implication of a Late Embryogenesis Abundant Protein (BnLEA57) in Seed Oil Accumulation in Brassica napus L. Front. Plant Sci. 13:907244. doi: 10.3389/fpls.2022.907244

Enhancing oil content is one of the major goals in *Brassica napus* breeding; however, genetic regulation of seed oil content in plants is complex and not fully elucidated. In this study, we report proteins that were differentially accumulated in immature seeds of 35 days after anthesis between two recombinant inbred lines with contrasting seed oil content, high oil content line (HOCL) and low oil content line (LOCL) using a multiplex isobaric tandem mass tags (TMT)-based quantitative proteomic approach. Over 4,600 proteins were quantified in seeds of the two lines, and 342 proteins showed differential accumulation between seeds of HOCL and LOCL. Gene Ontology enrichment analysis revealed that the differentially accumulated proteins were enriched in proteins involved in lipid biosynthesis and metabolism, photosynthesis, and nutrient reservoir activity. Western blot confirmed the increased abundance of a late embryogenesis abundant protein (BnLEA57) in HOCL seeds compared with LOCL seeds, and overexpression of either *BnLEA57* gene or its homology *BnLEA55* in transgenic *Arabidopsis thaliana* enhanced oil content in *Arabidopsis* seeds. Our work provides new insights into the molecular regulatory mechanism of seed oil content in *B. napus*.

Keywords: seeds, quantitative proteomics, oil content, rapeseed, *Brassica napus*, LEA protein, late embryogenesis abundant protein, TMT

INTRODUCTION

Rapeseed (*Brassica napus* L.) is an important oil-producing crop in the world because of large demand for both edible and industrial purposes (Dupont et al., 1989). Furthermore, *B. napus* oil has been pursued as an attractive alternative for renewable biofuels (Sharma et al., 2015). Improving seed oil content in *B. napus* is still one of the most important breeding objectives (Hua et al., 2016).

Brassica napus seeds accumulate oils during seed development, and the storage of seed oil is mainly in the embryo in the form of triacylglycerol (TAG) (Bai et al., 2020). The biosynthetic

pathway of the fatty acid and TAG in higher plants has been well elucidated, including in *Arabidopsis thaliana*, *Glycine max*, and *Brassica napus* (Maeo et al., 2009; Woodfield et al., 2019; Chen et al., 2020); however, the regulatory mechanism of seed oil accumulation is only partly understood. Modification of key genes involved in fatty acid biosynthesis, lipid biosynthesis, or glycolysis is known to affect seed oil accumulation (Hua et al., 2016; Herman, 2017). Several transcriptional factors that include *Leafy Cotyledon1* (*LEC1*) (Mu et al., 2008), *Leafy Cotyledon2* (*LEC2*) (Stone et al., 2008), *ABA-Insensitive3* (*ABI3*) (Giraudat et al., 1992), *FUSCA3* (*FUS3*) (Wang and Perry, 2013), *Transparent Testa2* (*TT2*) (Wang et al., 2014), and *Wrinkled1* (*WRI1*) (Liu et al., 2010) were shown to regulate lipid biosynthesis and seed oil accumulation. In addition, a few studies have demonstrated the maternal effects on seed oil content, such as *Apetala2* (*AP2*) (Jofuku et al., 2005) and *Transparent Testa8* (*TT8*) (Chen et al., 2014) in *Arabidopsis*.

Molecular genetic studies have revealed a large number of loci involved in controlling *B. napus* oil content, but the identification of candidate genes from QTLs controlling seed oil content is rare because of the presence of large intervals and small effect of each QTL on oil accumulation (Rahman et al., 2013; Hua et al., 2016). The omics strategy such as transcriptome studies has identified a large number of genes associated with seed oil accumulation (Weselake et al., 2009; Tan et al., 2011; Huang et al., 2017; Tang et al., 2021). By combing comprehensive genome- and transcriptome-wide association studies using 505 inbred lines, Tang et al. (2021) identified two seed-oil-content QTLs as two homologous putative methyltransferase *BnPMT6*s and validated their functions using genetic studies on CRISPR/Cas9 mutants of each gene.

Seed oil accumulation in *B. napus* is accomplished by a complex array of physiological and biological processes, and members of late embryogenesis abundant (LEA) proteins are associated in this process. There are three main stages during seed oil accumulation in *B. napus*. The first stage is endosperm development. At this stage, considerable soluble sugars accumulate in the developing seeds mainly for osmotic regulation for seed expansion, substrate preparation for oil and protein biosynthesis, and signaling molecules for oil and other secondary metabolism (Lorenz et al., 2014). The second stage is embryo development. Approximately 20 days after anthesis, the embryo initiates, and then, the endosperm is quickly and thoroughly absorbed by the embryo. Then, the seed is full of the embryo and a layer of seed coat surrounding the embryo (Schwender and Ohlrogge, 2002). Most of the seed storages such as oil and proteins are mainly synthesized at this stage (Schwender and Hay, 2012). The third stage is seed dehydration and maturation. During seed maturation, considerable water in the seed will be dehydrated because excessive water will affect seed vitality. In *B. napus* seed, the oil does not exist as the formation of free fatty acids but in the form of oil body. The synthesized fatty acid normally is transported into endoplasmic reticulum membrane for further folding into oil body (He and Wu, 2009). The process is always accompanied by dehydration although the exact functions of dehydration are not very clear (Quettier and Eastmond, 2009). In previous investigations, a large family of small and highly hydrophilic proteins named late embryogenesis abundant (LEA) proteins was identified in plant tissues such as in maturating seeds (Liang et al., 2016, 2019). LEA proteins are mostly associated with the drought and freezing tolerance in plants (Liang et al., 2019) and are suggested to stabilize membranes and macromolecules by acting as water-binding molecules (Wang et al., 2003; Chakrabortee et al., 2012). But the exact functions of LEA proteins are largely unknown, especially during seed development and maturation in *B. napus*.

With the rapid improvement in the sensitivity, reproducibility and high throughput of liquid chromatography with tandem mass spectrometry (LC-MS/MS) -based quantitative proteomics, quantitative proteomics has become an important tool in identifying candidate proteins in improving plant productivity and stress responses (Li et al., 2011; Hu et al., 2015; Zhao et al., 2017; Wu et al., 2021; Zhu et al., 2022). In this study, TMT-based quantitative proteomics was used to uncover differentially accumulated proteins (DAPs) between two recombinant breeding lines differing in oil content, and an LEA protein BnLEA57 was found to be about two-fold in abundance in the higher oil content line (HOCL) compared to the low oil content line (LOCL). The function of the *BnLEA57* gene and its close homolog *BnLEA55* gene in seed oil accumulation was further analyzed through the ectopic expression in *Arabidopsis thaliana*. The result shed new insights on elucidating the molecular function of the LEA proteins on the oil biosynthesis in *B. napus*.

RESULTS

Identification of Differentially Accumulated Proteins in Developing Seeds Between High Oil Content Line and Low Oil Content Line

To identify DAPs in the developing seeds of HOCL and LOCL in *B. napus*, seeds at 35 days after anthesis (DAA), which were at the critical stage of seed oil accumulation (Hua et al., 2014), were sampled for proteomic analysis. Tandem mass tag (TMT)-based quantitation was used to compare the seed proteome between HOCL and LOCL, following the general workflow in **Figure 1**. A total of 5,106 *B. napus* proteins were identified with FDR at 1% protein level, and 4,638 proteins were quantified across the three biological replicates for both HOCL and LOCL (no imputation was performed) (**Supplementary Table 1**). The principal component analysis showed that the three replicates of HOCL and LOCL were clustered together, respectively, but both clusters were well separated in component I, indicating the reproducible differences present in the proteome between both lines (**Figure 2A**). The data of DAPs were further analyzed according to the fold change and adjusted *p*-value. The result of volcano plot (**Figure 2B**) showed that 342 proteins were significantly different between HOCL and LOCL with a fold change >1.2 or <0.83 and adjusted *p*-value less than 0.05. Among those proteins, there were 199 proteins up-accumulated and 143

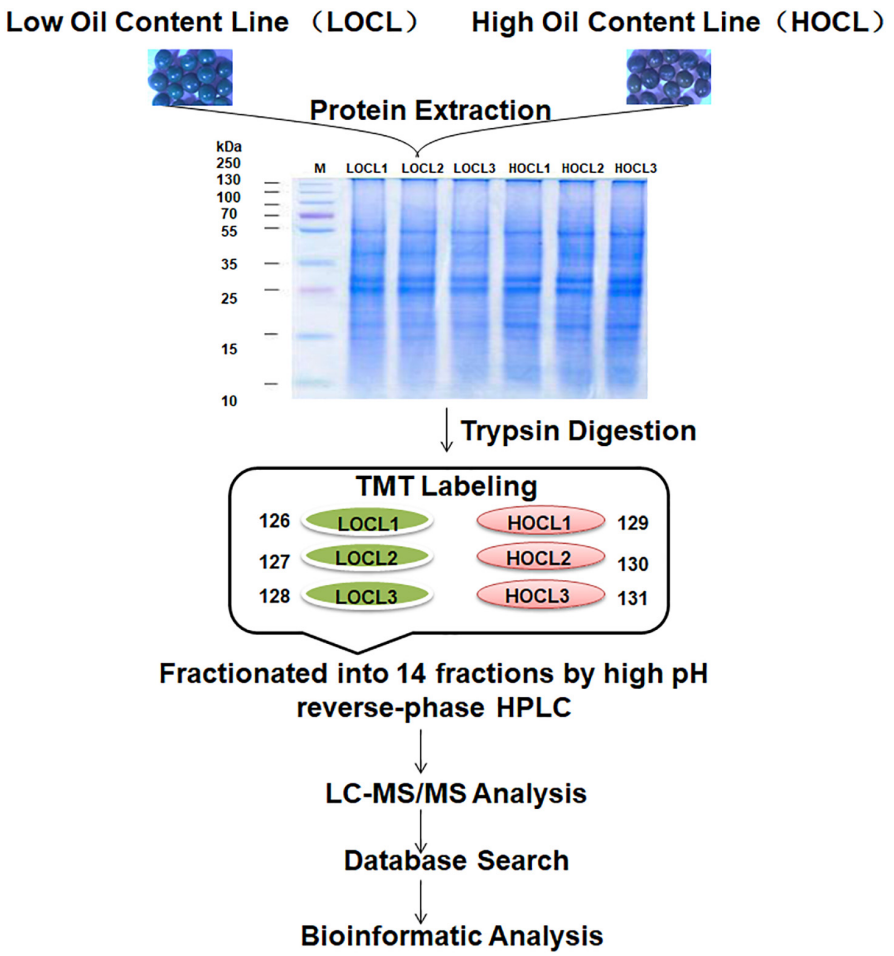


FIGURE 1 | Workflow of TMT labeling-based proteome profiling in *B. napus* seeds. Total proteins were extracted from 35-DAA seeds of two recombinant inbred lines of the *B. napus* with high oil content (HOCL) and low oil content (LOCL) (each line with three biological repeats), digested with trypsin after protein quality was checked on SDS-PAGE. The peptides were individually labeled with 6-plex TMT reagent (LOCL: 126, 127, and 128 labels; HOCL: 129, 130, and 131 labels) before combined, fractionated, and analyzed by LC-MS/MS.

proteins down-accumulated in the HOCL compared to the LOCL (**Supplementary Table 1**).

Gene Ontology Enrichment Analysis of the Differentially Accumulated Proteins

The functions of the DAPs were further analyzed by Gene Ontology (GO) enrichment. The DAPs were classified by GO terms according to three categories, namely, biological process, molecular function, and cellular component. In biological process, the largest functional enriched groups were lipid biosynthesis and metabolism (**Figure 3A**). Among lipid metabolic proteins, two biotin carboxyl carrier proteins of acetyl-CoA carboxylases (BCCPs) were significantly up-accumulated in the seeds of HOCL (**Table 1**). There were two acyl carrier proteins downregulated whereas one upregulated in the seed cells of HOCL. A protein encoding beta-ketoacyl-(acyl-carrier-protein) synthase (KAS) was also markedly upregulated (**Table 1**). Another important type of the DAPs was related to photosynthesis. The proteins were involved in the whole process of photosynthesis in the seed cells including antenna proteins (LHCA1 and LHCB5), photosystem II proteins (PsbE, PsbQ, and PsbR), photosystem I proteins (PsaL and PsaN), ATPase (AtpH), and ribulose bisphosphate carboxylase (RbcL). Importantly, all the proteins except two ribulose bisphosphate carboxylase small subunits were significantly higher in the seed cells of HOCL (**Table 1**). For the DAPs of essential components of the photosynthetic apparatus, all of the proteins involving in isoprenoid biosynthesis were upregulated except one zeta-carotene desaturase (**Supplementary Table 2**). Amazingly, all the DAPs associated with cell wall modification, the polysaccharide metabolism, pectin catabolic–metabolic process, and galacturonan metabolic process were the members of pectinesterase family, suggesting a role of cell wall modification on oil accumulation (**Supplementary Table 2**).

For cellular component, GO-enrichment-identified terms include extracellular region and chlorophyll thylakoid membrane. DAPs associated with chlorophyll thylakoid

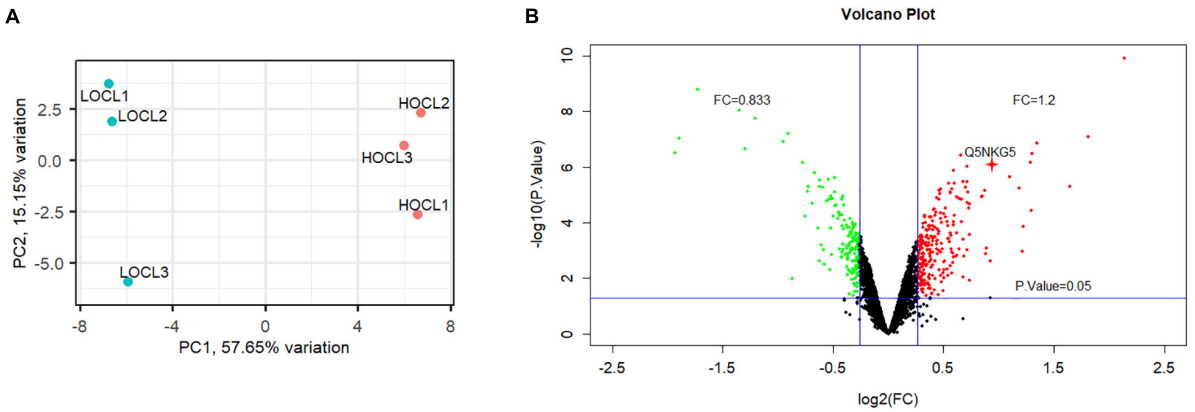


FIGURE 2 | Principal component analysis and volcano plot of the proteome in seeds of HOCL and LOC. **(A)** Principal component analysis of the proteome in seeds of HOCL and LOCL. Percentages (%) indicate the percentage of overall variance captured by each principal component. PC1: principal component 1; PC2: principal component 2. **(B)** Volcano plot of DAPs between the seeds of HOCL and LOCL. Threshold cutoff (>1.2 or <0.83) determined for log2 fold change ratios are represented by two vertical blue lines. Adjusted *p*-value cutoff of 0.05 is represented by a horizontal blue line. Red dots represent up-accumulated proteins, and green dots indicate down-accumulated proteins, whereas back dots represent proteins that do not meet the threshold or adjusted *p*-value, in HOCL compared to LOCL. The red cross indicates BnLEA57 protein.

membrane term were mostly involved in photosynthesis, and all the 13 DAPs were upregulated in HOCL (**Supplementary Table 2**). A total of 18 DAPs were extracellular region proteins, 11 of which were up-accumulated in HOCL. Functions of these extracellular region proteins in seed oil accumulation are mostly unexplored, which include trypsin inhibitor, pectin acetylesterase, and peroxidase.

For the molecular function terms, the largest enriched group in DAPs was the nutrient reservoir activity, which included 12 DAPs (**Table 1**), nine of which were up-accumulated in the HOCL. These proteins included seed storage proteins on the endoplasmic reticulum and are the members of the cruciferin (11S globulin) or napin (1.7–2S albumin) family. The second largest group was enzyme inhibitor activity which included five pectinesterase, three pectinesterase inhibitors, one Kunitz-type soybean trypsin inhibitor (STI), and one cysteine proteinase inhibitor, four of which showed up-accumulation in HOCL (**Supplementary Table 2**).

Immunoblot Analysis of BnLEA57 in Developing Seeds of *B. napus*

Among the DAPs, one late embryogenesis abundant protein, BnLEA57 (BnaC05g37670D), was chosen for further analysis because of its much higher abundance in the HOCL (the expression ratio between HOCL and LOCL was 1.91) and the implication of LEA proteins in dehydration stress response (Liang et al., 2016). Its accumulation in the developing seeds in the two lines was further examined by immunoblot analysis. The result showed that BnLEA57 protein abundance rose from 35 to 45 DAA and sharply increased from 45 to 55 DAA in the developing seeds of both lines. In addition, BnLEA57 protein accumulation in 35- and 45-, DAA seeds was consistently higher in HOCL than in LOCL (**Figure 4**), supporting the proteomic analysis results.

Overexpression of the *BnLEA57* and *BnLEA55* Increased Seed Oil Content in *Arabidopsis*

To further characterize the *BnLEA57* function on seed oil deposition, we overexpressed *BnLEA57* (*BnaC05g37670D*) and its closest *B. napus* homolog *BnLEA55* (*BnaA05g23860D*) in *Arabidopsis* plants, respectively. Less sequence similarity was observed among the 108 known LEA proteins in *B. napus* (Liang et al., 2016), but BnLEA57 and BnLEA55 that shared 96.1% of sequence identity (**Figure 5**) were clustered together in a subgroup in the LEA_4 family. BnLEA57 and BnLEA55 shared amino acid sequence identity of 84.3 and 86% with their ortholog LEA76 (AT3G15670) in *Arabidopsis*, respectively (**Figure 5**). Overexpression of either *BnLEA57* or *BnLEA57* in *Arabidopsis* increased seed oil content in *Arabidopsis*, which is consistent with the increased *LEA57*/*LEA57* transcript levels compared with wild type (**Figures 6A,B**). The mean oil content was 38.9% in wild type seeds, 41.8% in *35S::BnLEA57* seeds, and 41.4% in *35S::BnLEA55* seeds. The results revealed the positive regulation of seed oil content by both *BnLEA57* and *BnLEA55* genes.

DISCUSSION

Fatty acid biosynthesis and oil body formation in *B. napus* seeds are the intricate and complicated networks of physiological and biochemical processes, and the underlying molecular mechanism remains to be further unveiled and characterized. In this study, TMT-based quantitative proteomics was applied to identify the proteins mediating seed oil accumulation *via* a pair of *B. napus* lines with contrasting seed oil content. GO enrichment analysis of DAPs suggests that proteins involved in lipid metabolism, photosynthesis, and nutrient reservoir activity are associated with contrasting seed oil accumulation between HOCL and LOCL (**Figure 3**). In addition, a role of LEA protein BnLEA57 and its

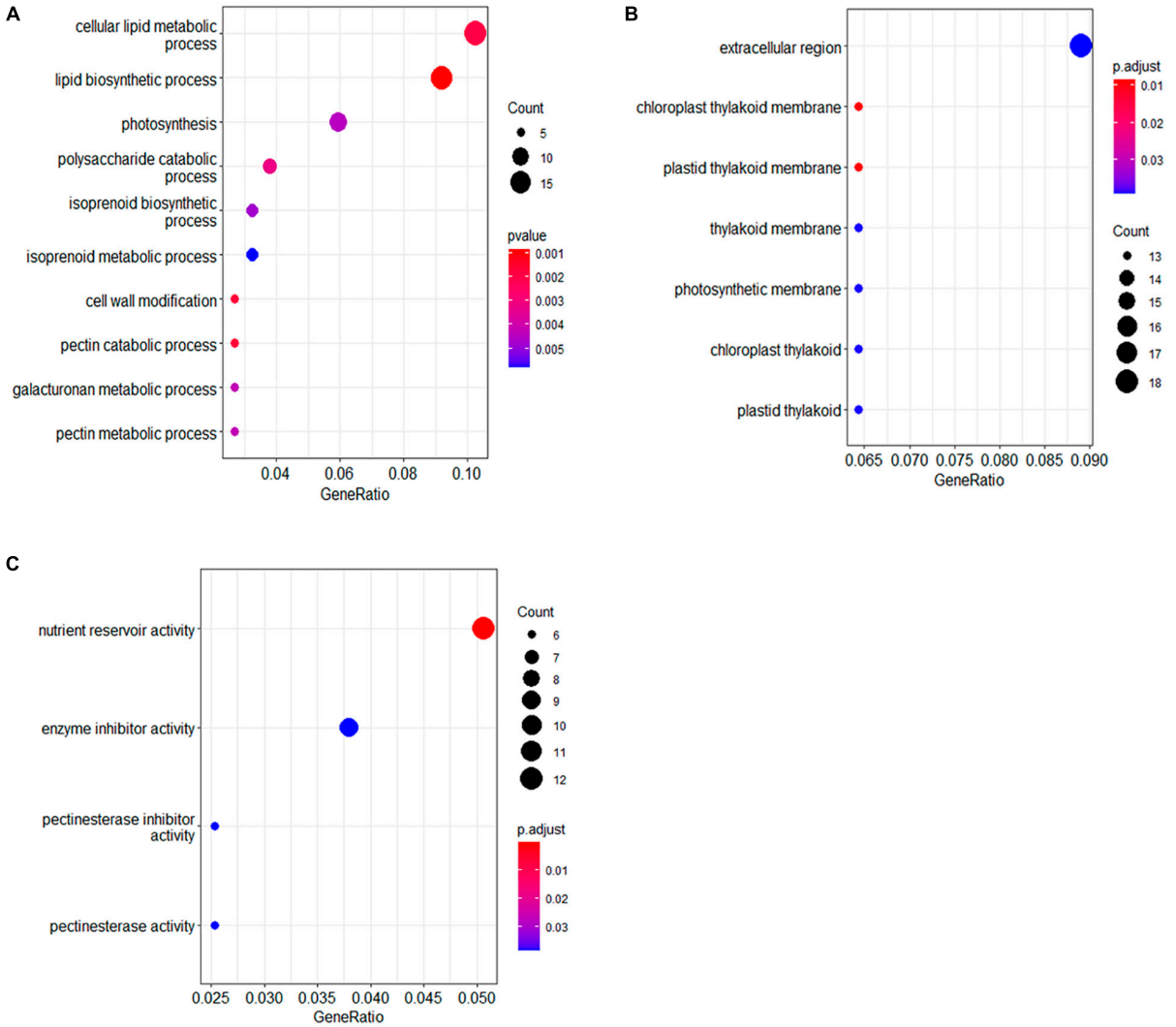


FIGURE 3 | Gene Ontology enrichment analysis for the differentially accumulated proteins between HOCL and LOCL. Gene Ontology enrichment analysis was categorized according to biology process [BP, **(A)**], cellular component [CC, **(B)**], and molecular function [MF, **(C)**].

homolog BnLEA55 in seed accumulation was further supported by their ectopic expression in *Arabidopsis*.

Our proteomic study suggests that higher photosynthesis capacity in the seeds contributes to higher lipid biosynthesis and oil accumulation in HOCL. Previous physiological studies comparing HOCL with LOCL, the same inbred lines used in this study, indicate higher chlorophyll content in the siliques of HOCL, supporting that higher photosynthetic capacity in siliques contribute to increased oil content in HOCL seeds. Results from this proteomic study support that photosynthesis from seeds could also contribute to oil accumulation in seeds. Among the DAPs, upregulation of protein abundance of two photosynthesis marker genes (chlorophyll A/B-binding proteins) (Martin et al., 2002) and two fatty acid biosynthesis marker genes (BCCP) (Ruuska et al., 2002) were observed (**Table 1**). In addition, among DAPs, GO-enriched terms in biological process include photosynthesis (9 of 11 proteins were up-accumulated in HOCL), and the enriched terms in the cellular process include thylakoid membranes (of which 13 proteins were all up-accumulated in HOCL, also involved in photosynthesis), supporting that higher photosynthesis activity contributes to higher seed oil content in HOCL. During fatty acid biosynthesis, considerable carbohydrates as substrate are required. Although it was estimated that about 70% of the photoassimilates were from silique wall (King et al., 1998), the remains were from transportation from other tissues such as leaf and stem and from developing seeds itself. A few lines of evidence showed that photosynthetic organelles were observed and the photosynthetic capability was found in developing seed cells (Eastmond et al., 1996; King et al., 1998; Ruuska et al., 2004; Houston et al., 2009). In addition, photosynthesis in seeds could enhance seed oil biosynthesis by increasing the carbon utilization efficiency (Schwender et al., 2004; Hay and Schwender, 2011) and by providing adenosine triphosphate (ATP), reductants NADPH,

TABLE 1 | Selected DAPs in immature seeds of 35DAA between HOCL and LOCL.

Accession	Description	Gene locus	GO description	Ratio	Adjusted *p*-value
A0A078E790	Acyl carrier protein	BnaAnng23710D	Lipid biosynthetic process	0.77	1.66E-02
A0A078FHD7	3-ketoacyl-CoA synthase	BnaA09g40640D	Lipid biosynthetic process	0.63	3.98E-04
A0A078FXF0	Transferase activity, transferring alkyl or aryl	BnaA06g35250D	Lipid biosynthetic process	1.22	1.44E-02
A0A078G4T7	Acyl carrier protein	BnaA09g03610D	Lipid biosynthetic process	1.24	1.29E-02
A0A078GDQ4	Hypothetical protein	BnaA05g30440D	Lipid biosynthetic process	1.20	1.22E-02
A0A078HJT1	Biotin carboxyl carrier protein of acetyl-CoA carboxylase	BnaC02g06560D	Lipid biosynthetic process	1.45	5.78E-03
A0A078HL12	Alkyl transferase	BnaC03g26600D	Lipid biosynthetic process	1.22	3.36E-02
A0A078HTG3	Phosphoethanolamine N-methyltransferase activity	BnaC06g02100D	Lipid biosynthetic process	0.80	7.71E-03
A0A078HTK0	Zeta-carotene desaturase	BnaC05g46990D	Lipid biosynthetic process	0.83	1.19E-02
A0A078HX24	Glycerol-3-phosphate acyltransferase, chloroplastic	BnaC05g28890D	Lipid biosynthetic process	0.83	3.18E-02
A0A078I1J8	Beta-ketoacyl-[acyl-carrier-protein] synthase I	BnaA02g24400D	Lipid biosynthetic process	1.65	9.10E-04
A0A078IU40	Xanthoxin dehydrogenase	BnaA06g01590D	Lipid biosynthetic process	1.57	1.33E-04
A0A078JFC7	Diphosphomevalonate decarboxylase	BnaCnng46900D	Lipid biosynthetic process	1.29	1.78E-02
A0A0K0K5D2	Zeta-carotene desaturase	BnaAnng10690D	Lipid biosynthetic process	2.55	7.34E-05
G4WXD9	Biotin carboxyl carrier protein of acetyl-CoA carboxylase	BnaC03g07000D	Lipid biosynthetic process	1.24	2.90E-02
P10352	Acyl carrier protein, chloroplastic, ACL1.A1	BnaA09g03610D	Lipid biosynthetic process	0.65	6.12E-04
Q8GV01	Lipoxygenase	BnaC06g26190D	Lipid biosynthetic process	0.80	5.67E-03
A0A078F5L4	Protein RALF-like 33	BnaC07g33210D	Extracellular region	0.79	1.76E-02
A0A078FGJ4	Mannan endo-1,4-beta-mannosidase	BnaA03g00090D	Extracellular region	0.80	1.06E-02
A0A078FNY0	FAD-binding PCMH-type	BnaA09g28890D	Extracellular region	1.32	5.78E-03
A0A078FSV0	FAD-binding PCMH-type	BnaA09g28900D	Extracellular region	1.36	1.20E-03
A0A078G1F3	Peroxidase OS = *Brassica napus*	BnaC01g31810D	Extracellular region	0.75	1.20E-03
A0A078GK30	Peroxidase OS = *Brassica napus*	BnaA04g12860D	Extracellular region	1.27	1.02E-02
A0A078GRW5	Carboxypeptidase	BnaA01g06330D	Extracellular region	0.80	4.57E-02
A0A078GYJ4	S-protein homolog	BnaA04g15920D	Extracellular region	1.29	1.08E-02
A0A078H673	(Rape) hypothetical protein	BnaC03g27700D	Extracellular region	0.62	1.60E-03
A0A078HGM0	Dirigent protein	BnaC01g37520D	Extracellular region	0.75	1.70E-03
A0A078HT58	(Rape) hypothetical protein	BnaC05g14950D	Extracellular region	1.37	1.83E-02
A0A078I5K0	Ribonuclease T(2)	BnaA06g33900D	Extracellular region	1.29	2.12E-02
A0A078IG99	Cysteine-type peptidase	BnaA06g16650D	Extracellular region	1.47	9.10E-04
A0A078IRK4	S-protein homolog	BnaA07g37790D	Extracellular region	1.30	2.47E-02
A0A078ITV4	Fn3_like domain-containing protein	BnaC06g27770D	Extracellular region	1.39	2.01E-02
A0A078J4F0	Pectin acetylesterase	BnaA04g27440D	Extracellular region	1.21	1.86E-02
A0A078JZ92	Gibberellin-regulated protein 8	BnaCnng75550D	Extracellular region	1.42	3.05E-02
Q2MD61	Rapeseed trypsin inhibitor-3, RTI-3	BnaC04g49020D	Extracellular region	0.26	1.24E-04
A0A078FMD5	2S SEED STORAGE PROTEIN 1-RELATED	BnaA08g14120D	Nutrient reservoir activity	1.50	3.22E-02
A0A078FMU9	11S globulin	BnaA08g13680D	Nutrient reservoir activity	1.31	4.17E-02
A0A078GA98	11S globulin	BnaA06g36310D	Nutrient reservoir activity	1.47	1.25E-02
A0A078GXC5	2S SEED STORAGE PROTEIN 1-RELATED	BnaC01g19300D	Nutrient reservoir activity	1.27	1.47E-02
A0A078IBF3	Napin/2S seed storage protein/Conglutin	BnaA03g48490D	Nutrient reservoir activity	1.64	6.37E-04
A0A078K0Y7	11S globulin	BnaAnng37930D	Nutrient reservoir activity	0.53	5.97E-05
P01090	Napin-2	BnaA01g16200D	Nutrient reservoir activity	1.21	3.48E-02
P17333	Napin, NAP1	BnaC01g19300D	Nutrient reservoir activity	1.33	1.03E-02
P33524	Cruciferin BnC2	BnaA02g22500D	Nutrient reservoir activity	0.66	1.45E-02
Q39344	*Brassica napus* napB napin	BnaA01g17200D	Nutrient reservoir activity	0.72	9.10E-04
Q7XB53	11S globulin (Fragment)	BnaC07g48660D	Nutrient reservoir activity	1.56	1.73E-02
Q9S9E6	Napin large chain L2B	BnaC01g19330D	Nutrient reservoir activity	1.23	3.77E-02
A0A075M3Q5	Ribulose bisphosphate carboxylase small subunit	BnaA02g12840D	Photosynthesis	0.82	3.96E-02
A0A078F3B2	Chlorophyll a-b binding protein, LHCA1	BnaC06g15470D	Photosynthesis	1.25	4.70E-02
A0A078HMF9	PSI subunit V, PsaL	BnaA04g07250D	Photosynthesis	1.25	3.36E-02
A0A078HR87	Photosystem I subunit PsaN	BnaC02g42890D	Photosynthesis	1.23	1.45E-02
A0A078IAV5	Photosystem II oxygen-evolving enhancer protein, PsbQ	BnaA09g19870D	Photosynthesis	1.27	1.03E-02
A0A078IFG6	F-type H+-transporting ATPase, AtpH	BnaA09g23090D	Photosynthesis	1.38	7.71E-03
A0A078J8I6	Chlorophyll a-b binding protein, LHCB5	BnaC02g46810D	Photosynthesis	1.60	1.14E-02
A0A078JXE7	Photosystem II 10 kDa polypeptide, PsbR	BnaAnng40550D	Photosynthesis	1.66	2.13E-03
A0A078JZM9	Ribulose bisphosphate carboxylase small subunit	BnaAnng36210D	Photosynthesis	0.83	2.77E-02
D1L8Q5	Cytochrome b559 subunit alpha, PsbE	BnaC04g29450D	Photosynthesis	1.62	7.83E-04
Q71SX0	Ribulose bisphosphate carboxylase large chain, RbcL	BnaA01g34300D	Photosynthesis	1.25	1.22E-02

DAPs involved in lipid biosynthetic process, extracellular region, nutrient reservoir activity, and photosynthesis were included. Ratio indicates the average protein abundance ratio of HOCL/LOCL.

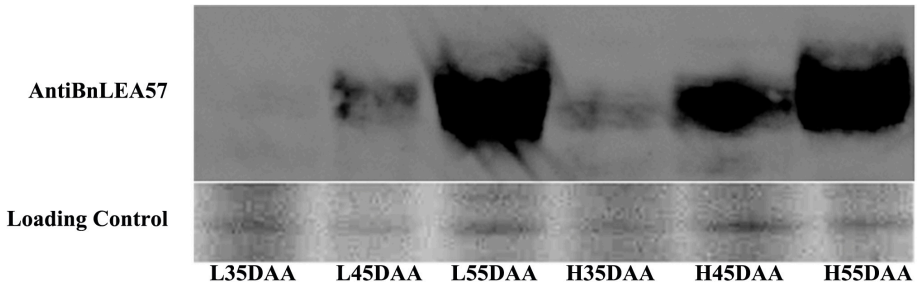


FIGURE 4 | Protein abundance of BnLEA57 in different seed development stages and in different lines (HOCL and LOCL). Seed protein samples were probed with an anti-BnLEA57 antibody **(top lane)**. Coomassie blue-stained RuBisCO large subunit (RbcL) was used as a loading control **(bottom lane)**.

```
BnaC05g37670D  MASNQQSYKAGETRGKTQEKTGQAMGAMRDKAEEGKDKTSQTAQIAQQKAQETAQAAKEK 60
BnaA05g23860D  MASNQQSYKAGETRGKTQEKTGQAMGAMRDKAEEGRDKTSQTAQTAQQKAQETAQAAKEK 60
AT3G15670      MASNQQSYKAGETRGKAQEKTGQAMGTMRDKAEEGRDKTSQTAQTAQQKAHETAQSAKDK 60
               **********************:*********:********:******** *****:****:**:*

BnaC05g37670D  TSQASQTTQQKAQETAQATKDKTSQAAQTTQQKAHETTQAAKDKTSQAAQTAQEKARETK 120
BnaA05g23860D  TSQAAQTTQQKAHETTQATKDKTSQAAQTTQQKAHETTQAAKDKTSQAAKTAQEKAHETK 120
AT3G15670      TSQTAQAAQQKAHETAQSAKEKTSQTAQTAQQKAHETTQAAKEKTSQAG----DKAREAK 116
               ***:::*::****:**:*::*:****:***:************:*****.     :**:*:*

BnaC05g37670D  DKTGSYMSETGEAIKQKAQNAAQYTKETAQEAAQYTKETAEAGRDKTGGFLSQTGEQVKQ 180
BnaA05g23860D  DKTGSYMSETGEAIKQKAQNAAQYTKETAQEAAQYTKETAEAGRDKTGGFLSQTGEQVKQ 180
AT3G15670      DKAGSYLSETGEAIKNKAQDAAQYTKETAQGAAQYTKETAEAGRDKTGGFLSQTGEHVKQ 176
               **:***:*********:***:********** *************************:***

BnaC05g37670D  MAMGAADAVKHTVGMATEEEDREHYPGTTTTTTGTTRTTDPTHHTYQRK 229
BnaA05g23860D  MAMGAADAVKHTFGMATEEEDKEHYPGTTTTTTGTTRTTDPTHHTYQRK 229
AT3G15670      MAMGAADAVKHTFGMATEEEDKEHYPGSTTTTTATTRTTDPTHQTYQRK 225
               ************. ********:*****:*****. *********:*****
```

FIGURE 5 | Alignment of the amino acid sequences of BnLEA57 (BnaC05g37670D), BnLEA55 (BnaA05g23860D) and their *Arabidopsis* ortholog AtLEA76 (AT3G15670). Residues identical in all the three LEAs are denoted with an asterisk (*), residues shared by two LEAs are indicated with a dot (.) or a colon (:).

and NADH, which are the co-factors required for biosynthesis of fatty acids in chloroplasts (Goffman et al., 2005).

Among those DAPs, we found a late embryogenesis abundant protein BnLEA57 that was up-accumulated in the HOCL and showed that BnLEA57 and its homolog BnLEA55 positively regulate seed oil accumulation. It is known that LEAs are involved in the abiotic stress such as water deficiency and salt resistance (Dalal et al., 2009; Cuevas-Velazquez et al., 2014), but the exact molecular mechanism is poorly understood. In *B. napus,* Liang et al. (2019) found that overexpression of members of LEA_3 family (*BnLEA3.1*, *BnLEA3.2*, *BnLEA3.3*, and *BnLEA3.4*) increased oil content in *Arabidopsis* and *B. napus* seeds and suggested that LEA proteins increase seed oil accumulation by enhancing photosynthetic efficiency and improving membrane stability. The enhancement of photosynthetic capacity may result in the increment of substrate for fatty acid biosynthesis, namely, carbohydrate. The result was in accordance with our proteomic analysis. The LEA_4 family proteins BnLEA57 and BnLEA55 in our experiment were also shown to improve seed oil content when overexpressed in *Arabidopsis*. There were 108 LEA genes identified in *B. napus* genome, and many of them showed increased expression at late-stage seeds (Liang et al., 2016). Further investigations are required to reveal the accurate molecular mechanism of BnLEA57 and BnLEA55 on seed oil accumulation and to characterize whether different members of LEA proteins work individually or together on seed maturity and oil accumulation.

MATERIALS AND METHODS

Plant Materials and Sampling

A total of two recombinant inbred lines of *B. napus* with high oil content (50.4%, HOCL) and low oil content (41.4%, LOCL) lines, two recombinant inbred lines at the F8 generations derived from parental lines between Huyou 15 as male parents and Zheshuang 6 as female parents (Hua et al., 2012, 2014), were grown in the field on the campus of Zhejiang Academy of Agricultural

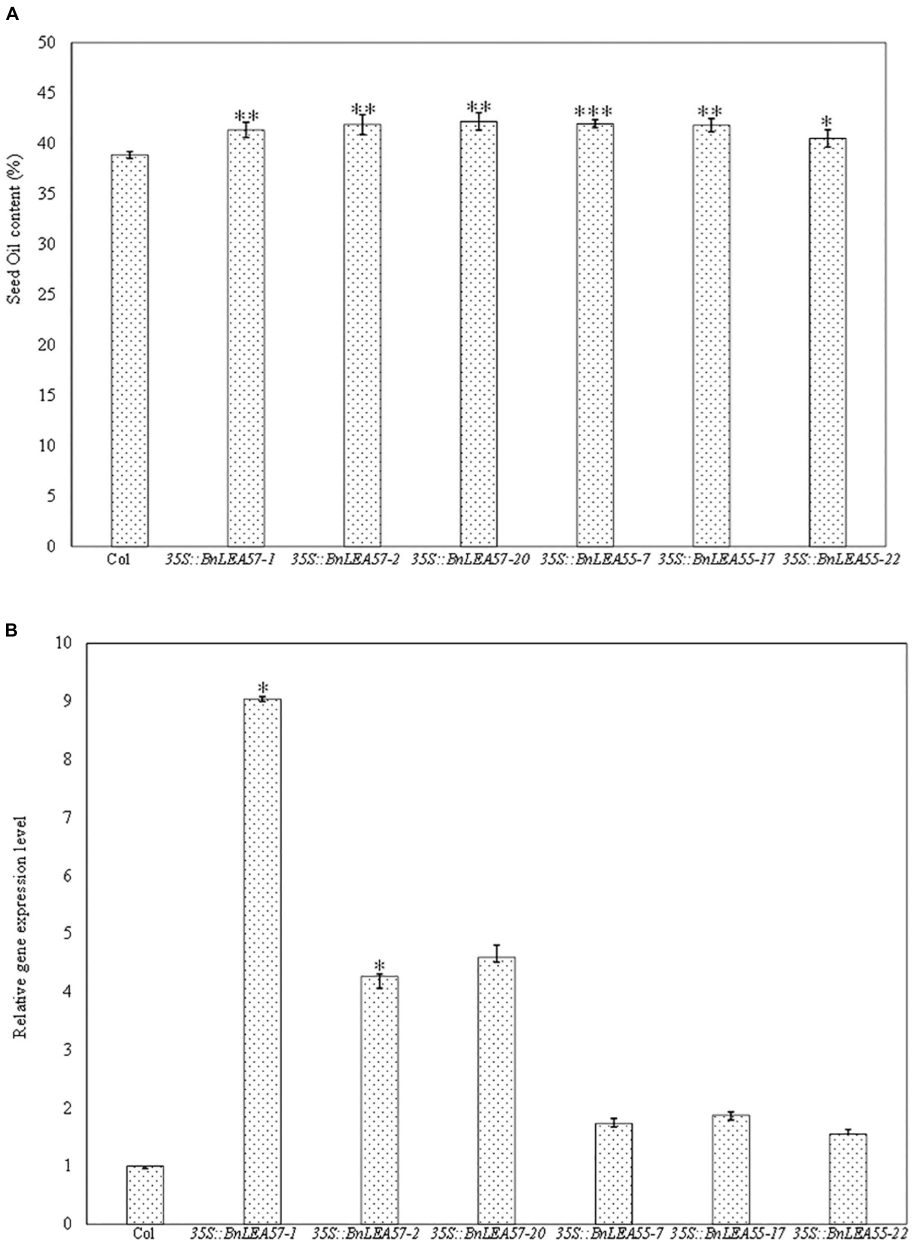


FIGURE 6 | Overexpression of *BnLEA57* or *BnLEA55* gene in *Arabidopsis* increased seed oil content. **(A)** Seed oil content in the Col-0 (WT) and different transgenic lines overexpressing *BnLEA57* or *BnLEA55 gene.* **(B)** Relative gene expression level of *BnLEA57* and *BnLEA55* in different *Arabidopsis* transgenic lines. Error bars indicate standard deviation ($n = 3$). Significant difference between the transgenic plants and wild type (Col-0) is indicated: * for $p < 0.05$, ** for $p < 0.01$, and *** for $p < 0.001$ for Student's t-test significance.

Sciences as described (Hua et al., 2014). The seeds for western blot assay were collected at 35, 45, and 55 days after anthesis on the same position of main inflorescence, the samples were immediately frozen in liquid nitrogen and kept at –80°C until use, and the samples collected at 35 DAA were also used for proteomic analysis.

Arabidopsis thaliana ecotype Columbia-0 was used for transformation. *Arabidopsis* seeds were surface-sterilized with 20 mg g^{-1} NaClO solution for 10 min and washed five times by sterilized distilled water, sown on agar medium (half-strength Murashige and Skoog basal salt mixture, 15 g kg^{-1} sucrose, pH 5.7), and vernalized at 4°C for 3 days in the dark. After germination, seedlings were transferred into soil and then grown in a growth room at 20–22°C, 16-h/8-h (day/night) photoperiod, and 60–70% humidity. The seeds were harvested as matured for oil content analysis. A total of three biological replicates for each sample were taken for analysis.

Protein Extraction and Digestion for Liquid Chromatography With Tandem Mass Spectrometry Analysis

Total proteins were extracted from the seeds of the *B. napus*, as previously described (Deng et al., 2007). Washed and pelleted protein sample was resuspended in 8 M urea solution, and the protein concentration was determined by the Bradford method.

For trypsin digestion, the samples were reduced with 10 mM DTT for 1 h at 56°C and alkylated with 55 mM iodoacetamide for 45 min at room temperature in darkness. Proteins were precipitated with 5 volumes of cold acetone overnight and then dissolved in 50 mM TEAB. Proteins were digested with trypsin overnight at 37°C in a 1:50 trypsin-to-protein mass ratio.

Tandem Mass Tag Labeling and High Performance Liquid Chromatography Fractionation

After trypsin digestion, peptides were desalted by Strata X C18 SPE column (Phenomenex) and vacuum-dried. Peptides were reconstituted in 50 mM TEAB and then labeled with respective 6-plex TMT reagents (LOCL: 126, 127, and 128 tags; HOCL: 129, 130, and 131 tags). The samples were pooled and fractionated by high pH reversed-phase high performance liquid chromatography (HPLC) using Agilent 300Extend C18 column (5-μm particles, 4.6 mm ID, 250 mm length). Peptides were separated with a gradient of 2–60% acetonitrile in 10 mM ammonium bicarbonate (pH 8) over 80 min. The peptides were then combined into 14 fractions and dried by vacuum centrifuging.

Liquid Chromatography With Tandem Mass Spectrometry Analysis

Multiplexed TMT-labeled samples were dissolved in 0.1% formic acid (solvent A), directly loaded onto a reversed-phase column (360 μm OD × 75 μm ID, 25 cm length) packed in-house with 3-μm C18 beads (ReproSil-Pur C18-AQ, Dr. Maisch) and eluted with a linear gradient of 6–25% solvent B (0.1% formic acid in 98% acetonitrile) for 28 min and 25–35% solvent B for 6 min at a constant flow rate of 300 nl/min on an EASY-nLC 1000 UPLC system. The resulting peptides were analyzed by Q Exactive Plus hybrid Quadrupole-Orbitrap mass spectrometer (Thermo Fisher Scientific, Waltham, MA, United States). A data-dependent procedure that alternated between one MS scan at a resolution of 70,000 and AGC target 3E6 followed by 20 MS/MS scans was applied for the top 20 precursor ions above a threshold ion count of 5E4 in the MS survey scan with 30.0-s dynamic exclusion. For MS scans, the m/z scan range was 350–1,600 Da. MS/MS scans were performed at the resolution at 17,500 and NCE at 30 with 3% stepped energy. The fixed first mass was set at 100 m/z for TMT quantification.

Data Analysis

The MS/MS raw data were analyzed using Proteome Discoverer (Version 2.4.0.305, Thermo Fisher Scientific, Waltham, MA, United States). Peptide and protein identifications were performed using both SEQUEST and MSFragger search engines against *Brassica napus* proteome from UniprotKB database (total 62,904 entries, as of 8 August 2021). Both searches were performed using the following settings: enzyme as trypsin/P, carbamidomethyl (C), TMT6plex (N-term), and TMT6plex (K) were selected as fixed modifications and oxidation (M) as variable modifications, 2 missed cleavage allowed and precursor error tolerance at 10 ppm, fragment deviation at 0.02 Da. Principal component analysis (PCA) of the samples was performed with R package PCAtools[1]. R package limma was used in screening the differentially accumulated proteins (DAPs) between HOCL and LOCL (Phipson et al., 2016). Adjusted p-value < 0.05 and fold change >1.20 or <0.83 were chosen as the cutoff threshold. The volcano plot was plotted using EnhancedVolcano package from R[2]. Gene Ontology (GO) enrichment for proteins that met differential expression criteria was performed using clusterProfiler in R environment (Wu et al., 2021).

Protein Immunoblot Analysis

Immunoblot analysis was performed as described (Deng et al., 2014). The proteins were extracted and dissolved in 2-D DIGE buffer (6 M urea, 2 M thiourea, and 4% CHAPS) and quantified using the Bio-Rad protein assay. Then, proteins were electrophoresed on 12% SDS-PAGE gels in a Tris-glycine buffer system at a constant 120 V/gel. Proteins were transferred to a nitrocellulose membrane at room temperature in a Bio-Rad semidry transfer System (Bio-Rad Laboratories, Hercules, CA, United States) according to the manufacturer's instructions. The transferred membrane was immunoblotted with anti-BnLEA57 antibody (Abmart, Shanghai, China) and then incubated with goat-anti-rabbit IgG-horseradish peroxidase (HRP) secondary antibody before development with ECL SuperSignal West Dura Extended Duration Substrate (Thermo Fisher Scientific, Waltham, MA, United States), and the intensities were captured by the ImageQuant LAS 4000 mini system (GE Healthcare Life Sciences, Piscataway, NJ, United States). Anti-BnLEA57 antibody was produced using the synthetic peptide C-TTQQKAQETAQATK (Abmart, Shanghai, China).

Vector Construction and Plant Transformation

cDNA sequences of *BnLEA57* (*BnaC05g37670D*) and *BnLEA55* (*BnaA05g23860D*) were downloaded from the *Brassica napus* genome database[3]. For both cloning constructs (*35S::LEA57* and *35S::LEA55*), target sequences were first cloned into the pENTR/D-TOPO vector (Invitrogen, Carlsbad, CA, United States) before being recombined into an destination vector pB7WGF2 using the gateway LR reaction (Invitrogen, Carlsbad, CA, United States). The following primers were used: LEA57-F,

[1]https://github.com/kevinblighe/PCAtools

[2]https://github.com/kevinblighe/EnhancedVolcano

[3]https://plants.ensembl.org/Brassica_napus/Info/Index

5′-CACCATGGCGTCTAACCAACAGAGC-3′; LEA57-R, 5′-CCTCTGATAAGTATGATGAGTAGGA-3′; LEA55-F, 5′-CACCATGGCGTCCAACCAACAA-3′; and LEA55-R, 5′-CTTCCTCTGATAAGTATGATGAGTCG-3′. Genetic transformation was performed as described (Zhou et al., 2011). The plasmids were transformed into *Agrobacterium tumefaciens* GV3101 strain and then transformed into Columbia-0 genotype using the floral-dip method (Clough and Bent, 1998). Transgenic seedlings were first selected on agar growth media with appropriate antibiotics and then confirmed by PCR using the corresponding primers.

RNA Extraction and Real-Time PCR Quantification

Total RNAs were extracted from the 10-day-old *Arabidopsis* seedlings with TRIzol Reagent (Invitrogen, Carlsbad, CA, United States), and cDNA was synthesized from 2 μg of total RNA using Moloney murine leukemia virus reverse transcriptase (Promega) and random primers in a 15 μl reaction according to the manufacturer's instructions. For real-time PCR, the cDNAs were diluted to 4 times, and 1 μl of the cDNAs was added to 5 μl of PowerUP SYBR Green PCR Master Mix (Thermo Fisher Scientific, Waltham, MA, United States) and 0.4 μl of each primer (100 nM final concentration) in 10 μl reactions. PCR amplification and detection were performed using a Bio-Rad CFX96 Tough Real-time PCR Detection System using the following cycling conditions: 50°C for 2 min, 95°C for 2 min, followed by 40 cycles of 95°C for 15 s, and 60°C for 1 min. Melting-curve analysis was used to confirm the absence of non-specific amplification products. *UBQ5* transcripts were used as an endogenous control to normalize expression of other genes. Relative expression levels were calculated by subtracting the threshold cycle (Ct) values for *UBQ5* from those of the target gene and then calculating $2^{-\Delta Ct}$ as described previously (Zhou et al., 2011). RT-PCR experiments were performed on at least three independent samples. The following primers were used: UBQ5-F, 5′-TCGACGCTTCATCTCGTCCT-3′; UBQ5-R, 5′-CGCTGAACCTTTCCAGATCC-3′; LEA57-QPCR-F, 5′-ACAAAGGAGACGGCTCAAGA-3′; LEA57-QPCR-R, 5′-TTCCAACAGTGTGCTTCACC-3′; LEA55-QPCR-F, 5′-ACAAAGGAGACGGCTCAAGA-3′; LEA55-QPCR-R, 5′-GGTCGTAGTTGTGCCTGGAT-3′.

Oil Content Measurement and Statistical Analysis

Oil content in *B. napus* seeds was determined as described (Hua et al., 2014). In this experiment, statistical analysis was carried out using two-tailed Student's *t*-test with two-samples unequal variance ("*," "**," and "***" represents as $p < 0.05$, $p < 0.01$, and $p < 0.001$) for significance.

DATA AVAILABILITY STATEMENT

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found below: ProteomeXchange, accession number PXD033185.

AUTHOR CONTRIBUTIONS

ZD, SH, and ZZ designed the experiment and wrote the manuscript. ZZ, ZD, and JT performed the proteomic analysis including data analysis. ZZ performed the western blotting and created the transgenic *Arabidopsis* lines. SH, BL, and PH prepared the *B. napus* samples and analyzed the seed oil content. All authors contributed to the article and approved the submitted version.

FUNDING

This work was supported by the National Key Research and Development Project (2018YFD1000900), Zhejiang Science and Technology Major Program on Agricultural New Variety Breeding (2021C02064), and Zhejiang Key Laboratory of Digital Dry Land Crops (2022E10012), and fundings from State Key Laboratory for Managing Biotic and Chemical Threats to the Quality and Safety of Agro-Products, and Zhejiang Academy of Agricultural Sciences.

ACKNOWLEDGMENTS

We thank the technical support of the proteomic analysis provided by Jingjie PTM Biolabs, Inc. (Hangzhou, China). The mass spectrometry proteomic data have been deposited to the ProteomeXchange Consortium (http://proteomecentral.proteomexchange.org) *via* the iProX partner repository (Ma et al., 2019) with the dataset identifier PXD033185.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpls.2022.907244/full#supplementary-material

Supplementary Table 1 | Lists of identified, quantified, and differentially accumulated *Brassica napus* proteins between seeds of HOCL and LOCL.

Supplementary Table 2 | Functional protein groups identified by GO enrichment analysis of DAPs between HOCL and LOCL seed proteome.

REFERENCES

Bai, S., Wallis, J. G., Denolf, P., Engelen, S., Bengtsson, J. D., Van Thournout, M., et al. (2020). The biochemistry of headgroup exchange during triacylglycerol synthesis in canola. *Plant J.* 103, 83–94. doi: 10.1111/tpj.14709

Chakrabortee, S., Tripathi, R., Watson, M., Schierle, G. S., Kurniawan, D. P., Kaminski, C. F., et al. (2012). Intrinsically disordered proteins as molecular shields. *Mol. biosyst.* 8, 210–219. doi: 10.1039/c1mb05263b

Chen, B., Zhang, G., Li, P., Yang, J., Guo, L., Benning, C., et al. (2020). Multiple GmWRI1s are redundantly involved in seed filling and nodulation by regulating

plastidic glycolysis, lipid biosynthesis and hormone signalling in soybean (*Glycine max*). *Plant Biotechnol. J.* 18, 155–171. doi: 10.1111/pbi.13183

Chen, M., Xuan, L., Wang, Z., Zhou, L., Li, Z., Du, X., et al. (2014). TRANSPARENT TESTA8 inhibits seed fatty acid accumulation by targeting several seed development regulators in *Arabidopsis*. *Plant Physiol.* 165, 905–916. doi: 10.1104/pp.114.235507

Clough, S. J., and Bent, A. F. (1998). Floral dip: a simplified method for agrobacterium-mediated transformation of *Arabidopsis thaliana*. *Plant J.* 16, 735–743. doi: 10.1046/j.1365-313x.1998.00343.x

Cuevas-Velazquez, C. L., Rendon-Luna, D. F., and Covarrubias, A. A. (2014). Dissecting the cryoprotection mechanisms for dehydrins. *Front. Plant Sci.* 5:583. doi: 10.3389/fpls.2014.00583

Dalal, M., Tayal, D., Chinnusamy, V., and Bansal, K. C. (2009). Abiotic stress and ABA-inducible group 4 LEA from *Brassica napus* plays a key role in salt and drought tolerance. *J. Biotechnol.* 139, 137–145. doi: 10.1016/j.jbiotec.2008.09.014

Deng, Z., Oses-Prieto, J. A., Kutschera, U., Tseng, T. S., Hao, L., Burlingame, A. L., et al. (2014). Blue light-induced proteomic changes in etiolated *Arabidopsis* seedlings. *J. Proteome Res.* 13, 2524–2533. doi: 10.1021/pr500010z

Deng, Z., Zhang, X., Tang, W., Oses-Prieto, J. A., Suzuki, N., Gendron, J. M., et al. (2007). A proteomics study of brassinosteroid response in *Arabidopsis*. *Mol. Cell Proteomics* 6, 2058–2071. doi: 10.1074/mcp.M700123-MCP200

Dupont, J., White, P. J., Johnston, K. M., Heggtveit, H. A., Mcdonald, B. E., Grundy, S. M., et al. (1989). Food safety and health effects of canola oil. *J. Am. Coll. Nutr.* 8, 360–375. doi: 10.1080/07315724.1989.10720311

Eastmond, P., Koláčá, L., and Rawsthorne, S. (1996). Photosynthesis by developing embryos of oilseed rape (*Brassica napus* L.). *J. Exp. Bot.* 47, 1763–1769. doi: 10.1093/jxb/47.11.1763

Giraudat, J., Hauge, B. M., Valon, C., Smalle, J., Parcy, F., and Goodman, H. M. (1992). Isolation of the *Arabidopsis* ABI3 gene by positional cloning. *Plant Cell* 4, 1251–1261. doi: 10.1105/tpc.4.10.1251

Goffman, F. D., Alonso, A. P., Schwender, J., Shachar-Hill, Y., and Ohlrogge, J. B. (2005). Light enables a very high efficiency of carbon storage in developing embryos of rapeseed. *Plant Physiol.* 138, 2269–2279. doi: 10.1104/pp.105.063628

Hay, J., and Schwender, J. (2011). Computational analysis of storage synthesis in developing *Brassica napus* L. (oilseed rape) embryos: flux variability analysis in relation to 13C metabolic flux analysis. *Plant J.* 67, 513–525. doi: 10.1111/j.1365-313X.2011.04611.x

He, Y. Q., and Wu, Y. (2009). Oil body biogenesis during *Brassica napus* embryogenesis. *J. Integr. Plant Biol.* 51, 792–799. doi: 10.1111/j.1744-7909.2009.00851.x

Herman, E. M. (2017). "Cell and molecular biology of seed oil bodies," in *Seed Development and Germination*, eds J. Kigel and G. Galili (Milton Park: Routledge), 195–214. doi: 10.1201/9780203740071

Houston, N. L., Hajduch, M., and Thelen, J. J. (2009). Quantitative proteomics of seed filling in castor: comparison with soybean and rapeseed reveals differences between photosynthetic and nonphotosynthetic seed metabolism. *Plant Physiol.* 151, 857–868. doi: 10.1104/pp.109.141622

Hu, J., Rampitsch, C., and Bykova, N. V. (2015). Advances in plant proteomics toward improvement of crop productivity and stress resistancex. *Front. Plant Sci.* 6:209. doi: 10.3389/fpls.2015.00209

Hua, S., Chen, Z. H., Zhang, Y., Yu, H., Lin, B., and Zhang, D. (2014). Chlorophyll and carbohydrate metabolism in developing silique and seed are prerequisite to seed oil content of *Brassica napus* L. *Bot. Stud.* 55:34. doi: 10.1186/1999-3110-55-34

Hua, S., Yu, H., Zhang, Y., Lin, B., and Chen, Z. (2012). Variation of carbohydrates and macronutrients during the flowering stage in canola (*Brasscia napus* L.) plants with contrasting seed oil content. *Aust. J. Crop Sci.* 6, 1275–1282. doi: 10.3316/informit.732097682474027

Hua, W., Liu, J., and Wang, H. (2016). Molecular regulation and genetic improvement of seed oil content in *Brassica napus* L. *Front. Agric. Sci. Eng.* 3, 186–194. doi: 10.15302/J-FASE-2016107

Huang, K. L., Zhang, M. L., Ma, G. J., Wu, H., Wu, X. M., Ren, F., et al. (2017). Transcriptome profiling analysis reveals the role of silique in controlling seed oil content in *Brassica napus*. *PLoS One* 12:e0179027. doi: 10.1371/journal.pone.0179027

Jofuku, K. D., Omidyar, P. K., Gee, Z., and Okamuro, J. K. (2005). Control of seed mass and seed yield by the floral homeotic gene APETALA2. *Proc. Natl. Acad. Sci. U.S.A.* 102, 3117–3122. doi: 10.1073/pnas.0409893102

King, S. P., Badger, M. R., and Furbank, R. T. (1998). CO_2 refixation characteristics of developing canola seeds and silique wall. *Funct. Plant Biol.* 25, 377–386. doi: 10.1071/PP97157

Li, T., Xu, S. L., Oses-Prieto, J. A., Putil, S., Xu, P., Wang, R. J., et al. (2011). Proteomics analysis reveals post-translational mechanisms for cold-induced metabolic changes in *Arabidopsis*. *Mol. Plant* 4, 361–374. doi: 10.1093/mp/ssq078

Liang, Y., Kang, K., Gan, L., Ning, S., Xiong, J., Song, S., et al. (2019). Drought-responsive genes, late embryogenesis abundant group3 (LEA3) and vicinal oxygen chelate, function in lipid accumulation in *Brassica napus* and *Arabidopsis* mainly via enhancing photosynthetic efficiency and reducing ROS. *Plant Biotechnol. J.* 17, 2123–2142. doi: 10.1111/pbi.13127

Liang, Y., Xiong, Z., Zheng, J., Xu, D., Zhu, Z., Xiang, J., et al. (2016). Genome-wide identification, structural analysis and new insights into late embryogenesis abundant (LEA) gene family formation pattern in *Brassica napus*. *Sci. Rep.* 6:24265. doi: 10.1038/srep24265

Liu, J., Hua, W., Zhan, G., Wei, F., Wang, X., Liu, G., et al. (2010). Increasing seed mass and oil content in transgenic *Arabidopsis* by the overexpression of wri1-like gene from *Brassica napus*. *Plant Physiol. Biochem.* 48, 9–15. doi: 10.1016/j.plaphy.2009.09.007

Lorenz, C., Rolletschek, H., Sunderhaus, S., and Braun, H. P. (2014). *Brassica napus* seed endosperm - metabolism and signaling in a dead end tissue. *J. Proteomics* 108, 382–426. doi: 10.1016/j.jprot.2014.05.024

Ma, J., Chen, T., Wu, S., Yang, C., Bai, M., Shu, K., et al. (2019). iProX: an integrated proteome resource. *Nucleic Acids Res.* 47, D1211–D1217. doi: 10.1093/nar/gky869

Maeo, K., Tokuda, T., Ayame, A., Mitsui, N., Kawai, T., Tsukagoshi, H., et al. (2009). An AP2-type transcription factor, WRINKLED1, of *Arabidopsis thaliana* binds to the AW-box sequence conserved among proximal upstream regions of genes involved in fatty acid synthesis. *Plant J.* 60, 476–487. doi: 10.1111/j.1365-313X.2009.03967.x

Martin, T., Oswald, O., and Graham, I. A. (2002). *Arabidopsis* seedling growth, storage lipid mobilization, and photosynthetic gene expression are regulated by carbon:nitrogen availability. *Plant Physiol.* 128, 472–481. doi: 10.1104/pp.010475

Mu, J., Tan, H., Zheng, Q., Fu, F., Liang, Y., Zhang, J., et al. (2008). LEAFY COTYLEDON1 is a key regulator of fatty acid biosynthesis in *Arabidopsis*. *Plant Physiol.* 148, 1042–1054. doi: 10.1104/pp.108.126342

Phipson, B., Lee, S., Majewski, I. J., Alexander, W. S., and Smyth, G. K. (2016). Robust hyperparameter estimation protects against hypervariable genes and improves power to detect differential expression. *Ann. Appl. Stat.* 10, 946–963. doi: 10.1214/16-AOAS920

Quettier, A. L., and Eastmond, P. J. (2009). Storage oil hydrolysis during early seedling growth. *Plant Physiol. Biochem.* 47, 485–490. doi: 10.1016/j.plaphy.2008.12.005

Rahman, H., Harwood, J., and Weselake, R. (2013). Increasing seed oil content in *Brassica* species through breeding and biotechnology. *Lipid Technol.* 25, 182–185. doi: 10.1002/lite.201300291

Ruuska, S. A., Girke, T., Benning, C., and Ohlrogge, J. B. (2002). Contrapuntal networks of gene expression during *Arabidopsis* seed filling. *Plant Cell* 14, 1191–1206. doi: 10.1105/tpc.000877

Ruuska, S. A., Schwender, J., and Ohlrogge, J. B. (2004). The capacity of green oilseeds to utilize photosynthesis to drive biosynthetic processes. *Plant Physiol.* 136, 2700–2709. doi: 10.1104/pp.104.047977

Schwender, J., Goffman, F., Ohlrogge, J. B., and Shachar-Hill, Y. (2004). Rubisco without the Calvin cycle improves the carbon efficiency of developing green seeds. *Nature* 432, 779–782. doi: 10.1038/nature03145

Schwender, J., and Hay, J. O. (2012). Predictive modeling of biomass component tradeoffs in *Brassica napus* developing oilseeds based on in silico manipulation of storage metabolism. *Plant Physiol.* 160, 1218–1236. doi: 10.1104/pp.112.203927

Schwender, J., and Ohlrogge, J. B. (2002). Probing in vivo metabolism by stable isotope labeling of storage lipids and proteins in developing *Brassica napus* embryos. *Plant Physiol.* 130, 347–361. doi: 10.1104/pp.004275

Sharma, R. V., Somidi, A. K., and Dalai, A. K. (2015). Preparation and properties evaluation of biolubricants derived from canola oil and canola biodiesel. *J. Agric. Food Chem.* 63, 3235–3242. doi: 10.1021/jf505825k

Stone, S. L., Braybrook, S. A., Paula, S. L., Kwong, L. W., Meuser, J., Pelletier, J., et al. (2008). *Arabidopsis* LEAFY COTYLEDON2 induces maturation traits and auxin activity: implications for somatic embryogenesis. *Proc. Natl. Acad. Sci. U.S.A.* 105, 3151–3156. doi: 10.1073/pnas.0712364105

Tan, H., Yang, X., Zhang, F., Zheng, X., Qu, C., Mu, J., et al. (2011). Enhanced seed oil production in canola by conditional expression of *Brassica napus* LEAFY COTYLEDON1 and LEC1-LIKE in developing seeds. *Plant Physiol.* 156, 1577–1588. doi: 10.1104/pp.111.175000

Tang, S., Zhao, H., Lu, S., Yu, L., Zhang, G., Zhang, Y., et al. (2021). Genome- and transcriptome-wide association studies provide insights into the genetic basis of natural variation of seed oil content in *Brassica napus*. *Mol. Plant* 14, 470–487. doi: 10.1016/j.molp.2020.12.003

Wang, F., and Perry, S. E. (2013). Identification of direct targets of FUSCA3, a key regulator of *Arabidopsis* seed development. *Plant Physiol.* 161, 1251–1264. doi: 10.1104/pp.112.212282

Wang, W., Vinocur, B., and Altman, A. (2003). Plant responses to drought, salinity and extreme temperatures: towards genetic engineering for stress tolerance. *Planta* 218, 1–14. doi: 10.1007/s00425-003-1105-5

Wang, Z., Chen, M., Chen, T., Xuan, L., Li, Z., Du, X., et al. (2014). TRANSPARENT TESTA2 regulates embryonic fatty acid biosynthesis by targeting FUSCA3 during the early developmental stage of *Arabidopsis* seeds. *Plant J.* 77, 757–769. doi: 10.1111/tpj.12426

Weselake, R. J., Taylor, D. C., Rahman, M. H., Shah, S., Laroche, A., Mcvetty, P. B. E., et al. (2009). Increasing the flow of carbon into seed oil. *Biotechnol. Adv.* 27, 866–878. doi: 10.1016/j.biotechadv.2009.07.001

Woodfield, H. K., Fenyk, S., Wallington, E., Bates, R. E., Brown, A., Guschina, I. A., et al. (2019). Increase in lysophosphatidate acyltransferase activity in oilseed rape (*Brassica napus*) increases seed triacylglycerol content despite its low intrinsic flux control coefficient. *New Phytol.* 224, 700–711.

Wu, T., Hu, E., Xu, S., Chen, M., Guo, P., Dai, Z., et al. (2021). clusterProfiler 4.0: a universal enrichment tool for interpreting omics data. *Innovation (N Y)* 2:100141. doi: 10.1016/j.xinn.2021.100141

Zhao, C., Wang, P., Si, T., Hsu, C. C., Wang, L., Zayed, O., et al. (2017). MAP kinase cascades regulate the cold response by modulating ICE1 protein stability. *Dev. Cell* 43, 618–629e5. doi: 10.1016/j.devcel.2017.09.024

Zhou, Z., An, L., Sun, L., Zhu, S., Xi, W., Broun, P., et al. (2011). Zinc finger protein5 is required for the control of trichome initiation by acting upstream of zinc finger protein8 in *Arabidopsis*. *Plant Physiol.* 157, 673–682. doi: 10.1104/pp.111.180281

Zhu, Y., Qiu, W., Li, Y., Tan, J., Han, X., Wu, L., et al. (2022). Quantitative proteome analysis reveals changes of membrane transport proteins in *Sedum* plumbizincicola under cadmium stress. *Chemosphere* 287:132302. doi: 10.1016/j.chemosphere.2021.132302

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

frontiers | Frontiers in Plant Science

ORIGINAL RESEARCH
published: 09 June 2022
doi: 10.3389/fpls.2022.867263

Multi-Omics Approaches Unravel Specific Features of Embryo and Endosperm in Rice Seed Germination

Naoto Sano[1], Imen Lounifi[1,2], Gwendal Cueff[1], Boris Collet[1], Gilles Clément[1], Sandrine Balzergue[3,4], Stéphanie Huguet[3], Benoît Valot[5,6], Marc Galland[1,7]* and Loïc Rajjou[1]**

[1] Université Paris-Saclay, INRAE, AgroParisTech, Institut Jean-Pierre Bourgin (IJPB), Versailles, France, [2] MBCC Group, Master Builders Construction Chemical, Singapore, Singapore, [3] Université Paris-Saclay, CNRS, INRAE, Univ Evry, Institute of Plant Sciences Paris-Saclay (IPS2), Orsay, France, [4] IRHS-UMR1345, Université d'Angers, INRAE, Institut Agro, SFR 4207 QuaSaV, Beaucouzé, France, [5] Université Paris-Saclay, INRAE, CNRS, AgroParisTech, GQE - Le Moulon, PAPPSO, Plateforme d'Analyse de Proteomique Paris-Sud-Ouest, Gif-sur-Yvette, France, [6] Chrono-Environnement Research Team UMR/CNRS-6249, Bourgogne-Franche-Comté University, Besançon, France, [7] Swammerdam Institute for Life Sciences, University of Amsterdam, Amsterdam, Netherlands

OPEN ACCESS

Edited by:
Andrej Frolov,
Leipzig University, Germany

Reviewed by:
Ravi Gupta,
Kookmin University, South Korea
Bing Bai,
University of Copenhagen, Denmark

***Correspondence:**
Naoto Sano
naoto.sano@inrae.fr
Marc Galland
m.d.galland@uva.nl
Loïc Rajjou
loic.rajjou@agroparistech.fr

Specialty section:
This article was submitted to Plant Proteomics and Protein Structural Biology, a section of the journal Frontiers in Plant Science

Received: *31 January 2022*
Accepted: *29 April 2022*
Published: *09 June 2022*

Citation:
Sano N, Lounifi I, Cueff G, Collet B, Clément G, Balzergue S, Huguet S, Valot B, Galland M and Rajjou L (2022) Multi-Omics Approaches Unravel Specific Features of Embryo and Endosperm in Rice Seed Germination. Front. Plant Sci. 13:867263. doi: 10.3389/fpls.2022.867263

Seed germination and subsequent seedling growth affect the final yield and quality of the crop. Seed germination is defined as a series of processes that begins with water uptake by a quiescent dry seed and ends with the elongation of embryonic axis. Rice is an important cereal crop species, and during seed germination, two tissues function in a different manner; the embryo grows into a seedling as the next generation and the endosperm is responsible for nutritional supply. Toward understanding the integrated roles of each tissue at the transcriptional, translational, and metabolic production levels during germination, an exhaustive "multi-omics" analysis was performed by combining transcriptomics, label-free shotgun proteomics, and metabolomics on rice germinating embryo and endosperm, independently. Time-course analyses of the transcriptome and metabolome in germinating seeds revealed a major turning point in the early phase of germination in both embryo and endosperm, suggesting that dramatic changes begin immediately after water imbibition in the rice germination program at least at the mRNA and metabolite levels. In endosperm, protein profiles mostly showed abundant decreases corresponding to 90% of the differentially accumulated proteins. An ontological classification revealed the shift from the maturation to the germination process where over-represented classes belonged to embryonic development and cellular amino acid biosynthetic processes. In the embryo, 19% of the detected proteins are differentially accumulated during germination. Stress response, carbohydrate, fatty acid metabolism, and transport are the main functional classes representing embryo proteome change. Moreover, proteins specific to the germinated state were detected by both transcriptomic and proteomic approaches and a major change in the network operating during rice germination was uncovered. In particular, concomitant changes of hormonal metabolism-related proteins (GID1L2 and CNX1) implicated in GAs and ABA metabolism, signaling proteins, and protein turnover events emphasized the importance of such biological networks in rice seeds. Using metabolomics, we highlighted the importance of an energetic supply in rice seeds during germination. In both embryo

and endosperm, starch degradation, glycolysis, and subsequent pathways related to these cascades, such as the aspartate-family pathway, are activated during germination. A relevant number of accumulated proteins and metabolites, especially in embryos, testifies the pivotal role of energetic supply in the preparation of plant growth. This article summarizes the key genetic pathways in embryo and endosperm during rice seed germination at the transcriptional, translational, and metabolite levels and thereby, emphasizes the value of combined multi-omics approaches to uncover the specific feature of tissues during germination.

Keywords: seed, multi-omics, germination, embryo, endosperm

INTRODUCTION

Rice is of major importance worldwide, economically, socially, and scientifically. As described for other species, rice germination implies a set of events that begins at the imbibition of the dry mature seed and finishes at the elongation of the embryonic axis (Bewley, 1997). Seed vigor and physiological performance are of paramount importance for proper seedling emergence, increase in crop yield, and reduce the cost of agriculture production (Rajjou et al., 2012; Finch-Savage and Bassel, 2016; Reed et al., 2022). Seed germination is usually defined as a sequential physiological process, including three phases of water uptake (Bradford, 1990). The first phase corresponding to rapid imbibition of dry mature seed, is critical to restore cell activities (*e.g.,* respiration, repair systems, translation) of the seed tissues, namely the embryo and the endosperm. The second phase is a lag period for the fully imbibed seed marked by an intensive metabolic activity with a very characteristic molecular regulation resulting in embryonic cell elongation. Germination is considered complete when the embryo emerges through the testa and cells start to divide in the third phase (Galland et al., 2014). This defined triphasic water uptake profile was also verified for rice seed germination and coleoptile growth (Yang et al., 2007). Upon imbibition, drastic biochemical changes occur within the rice seed tissues. Processes that take place during rice germination are chronologically and differentially regulated depending on imbibition conditions (Magneschi and Perata, 2009; Narsai and Whelan, 2013). For instance, under aerobic conditions, resumption of respiration and of the energy-producing related pathways, such as glycolysis, are activated rapidly upon imbibition (Howell et al., 2006, 2007, 2009). In contrast, under low oxygen conditions fermentative program is initiated to restrict energy consumption and thus produce ATP requisite for cell survival. Indeed, the carbohydrate metabolism can be redirected by transcriptional control into the fermentative branch mainly *via* the accumulation of sucrose synthase, pyruvate decarboxylase, lactate dehydrogenase, and alcohol dehydrogenase (Magneschi and Perata, 2009). Intermediate phases of rice germination include large transcriptional changes that are related to stress responses, energy production, protein degradation, and synthesis, as well as a large number of pathways that sustain elongation of the embryonic axes and the growth of the future seedling (Howell et al., 2009; Narsai et al., 2009). Previous studies highlighted transcriptome, proteome, and metabolome changes in rice during germination; however, these investigations focused mainly on the rice embryo or on the whole grain and the integrated functioning of the endospermic reserve tissue with the diploid embryo for the success of germination and seedling establishment has not yet been taken into consideration. A recent review underlined the coordination between embryo and endosperm and major interactions to ensure rice grain development (An et al., 2020). A multi-omics analysis of dry mature rice compartments emphasized molecular signature related to rice seed quality in a tissue-specific manner (Galland et al., 2017). Only very few studies have been based on detailed imbibition kinetics during the germination process considering both compartments in rice while some experimental evidence have shown in other species that the endosperm is capable to influence the growth of the germinating embryo (Yan et al., 2014). In this study, we deeply investigated the germination process by multi-omics of both the embryo and the endosperm of rice grain (*Oryza sativa* L. var. Nipponbare). We focused on the interpretation of the differential protein accumulation patterns as revealed by label-free shotgun proteomics. Indeed, proteins stored in dry mature seeds and proteins translated during seed imbibition from stored mRNAs have been shown to play a major role in seed germination, both in Arabidopsis (Rajjou et al., 2004) and rice (He et al., 2011; Sano et al., 2012; He and Yang, 2013). A comparison of the proteomic data with corresponding metabolic and transcriptomic data further helped unravel the main functional set of events occurring during rice germination and to have a better understanding of the embryo-endosperm interactions.

RESULTS

Rice Germination

Rice seed germination kinetics were studied by monitoring coleoptile protrusion that has previously been considered as the observable marker of this physiological process (Howell et al., 2009). Rice seeds used in this study (*Oryza sativa* L. var. Nipponbare) had high vigor and germinative capacity since T_{50} (*i.e.,* time to reach 50% germination) was equal to 16 h after imbibition (HAI) while G_{max} (maximum level of germination) ultimately reached 100% at 32 HAI. In addition, 91% of the seeds germinated within 24 HAI (**Figure 1A**). Germination can also be followed according to the three

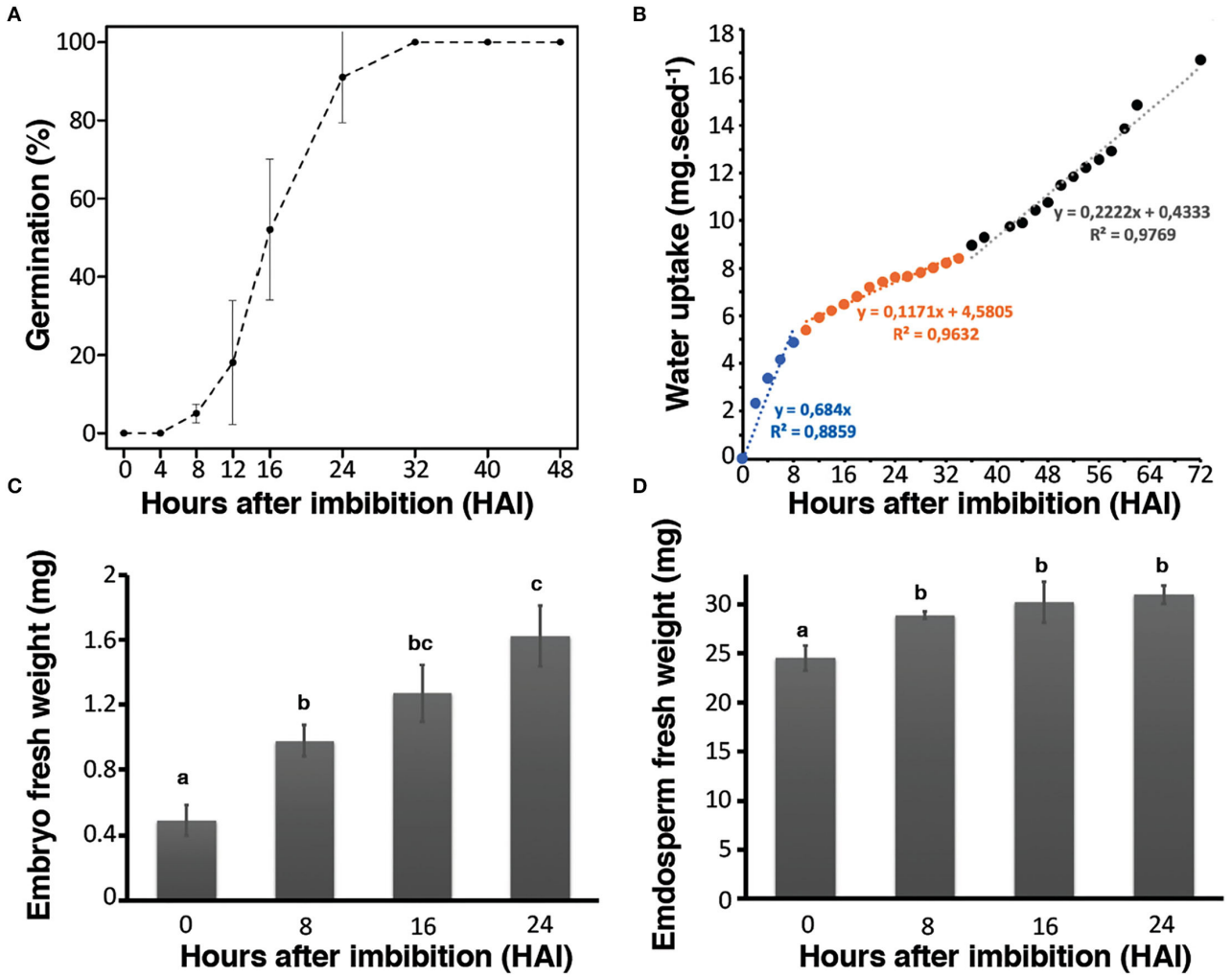


FIGURE 1 | Germination kinetics of Nipponbare rice seeds. Germination assays were carried out at 30°C under dark and continuous oxygenation conditions in triplicate. **(A)** Germination rate of seeds up to 48 h post-imbibition. **(B)** Water uptake in mg of whole rice seeds from 0 to 72 h post-imbibition. The fitted linear models of fresh weight uptake as a function of time are indicated on the graph to highlight the phase I (blue), phase II (orange), and phase III (black) of rice germination. **(C)** Water uptake of isolated embryos; **(D)** water uptake of isolated endosperms. Values are presented as the means ± SD of three replicates. Different letters in **(C)** and **(D)** indicate significant differences ($p < 0.05$, Tukey–Kramer tests).

phases of water uptake in each seed compartment (*i.e.,* embryo and endosperm) (Bradford, 1990). Accordingly, water uptake measurements were realized from the whole seeds and from both hand-dissected embryo and endosperm during the time course of germination. A very rapid increase in water uptake was observed in the first hours (8–10 HAI) corresponding to phase I of rice germination (**Figure 1B**). This first stage was followed by a slower period of water uptake corresponding to the phase II (lag period) of rice germination. The transition from phase II to phase III, corresponding to a strong restart of water uptake after germination completion and seedling establishment were visible from beyond 32 HAI (**Figure 1B**). Interestingly, the patterns of water uptake were remarkably different in the isolated embryo and endosperm (**Figures 1C,D**). The water content of the embryos steadily increased over the germination time course (**Figure 1C**), in contrast, the water content in the endosperm rapidly increased up to 8 HAI and then remained stable, consistent with the establishment of the lag period (phase II) (**Figure 1D**). Previously, water localization was characterized in germinating rice seeds by nuclear magnetic resonance (NMR) (Horigane et al., 2006). The results disclosed first rehydration occurring in the embryo at 2 HAI, which quickly reached the NMR saturation level. Past 8 HAI, the water diffused in the dorsal aleurone layer (transfer cells). Finally, from 8 to 15 HAI, the water slowly diffused into the endosperm. It appears that in the endosperm water distribution evolves convexly toward the center of the seed with a lower signal intensity indicating a lower rate of water penetration (Horigane et al., 2006). The present results are in accordance with these data. Besides physiological approaches, we used

combined "multi-omics" (*i.e.,* transcriptomics, proteomics, and metabolomics) approaches thus leading to the most exhaustive tissue-dependent data. Integrating these different levels revealed new and interesting factors impacting rice seed germination (**Supplementary Figure S1**).

Germination Phase Transitions in Both Compartments, Embryo and Endosperm, Based on Transcriptomics and Metabolomics

A transcriptomic approach was performed on both isolated embryos and endosperms during the time course of rice seed germination (0, 4, 8, 12, 16, 24 HAI). This transcriptomic analysis highlighted a significant signal for 22,343 (18,389 loci) and 20,052 locus-specific probes (16,657 loci) in the embryo (E) and endosperm (A), respectively (**Supplementary Tables S1A**,**B**). Differentially accumulated transcripts were detected in the embryo (7,665 up and 2,279 down) and endosperm (5,926 up and 2,103 down), respectively, after imbibition (**Supplementary Table S2**, **Supplementary Figure S1**). Moreover, during the germination time course, 15,926 rice loci could be detected in both compartments and only 2,463 and 731 transcripts were specifically detected in the embryo or the endosperm, respectively. This detailed transcriptome data was used to plot individual embryo and endosperm samples in a reduced dimension space using a Principal Component Analysis (**Figures 2A,B**). For both embryo and endosperm, the first component explained nearly half of the total variance (49.1 and 43.9%, respectively) and mainly distinguish early stages (embryo and endosperm at 0 HAI and 4 HAI). On the first component, embryo and endosperm samples at 8 HAI were nearly at the center suggesting that this time point might be the end of the first phase. Therefore, at the mRNA level, it can be thought that the end of phase I could end up as early as 8 HAI in a vigorous rice seed batch in accordance with the findings from the water uptake curve. Surprisingly, on the second component explaining 19.3 and 18.8% of the total variation in the embryo and endosperm transcriptomes (**Figures 2A,B**), the dry state is closest to the germinated state (E0 and E24, A0 and A24). This points to mRNAs whose levels might be comparable at the beginning and end of the germination process. As previous work did not examine metabolic changes occurring in the endosperm of germinating rice seeds (Galland et al., 2017), we performed a GC-MS based metabolite profiling on both embryo and endosperm during the germination time course. This allowed identifying 380 different metabolites of which 121 were matched to known compounds (**Supplementary Tables S3A**,**C**). Besides the fact that the present work is the first description of metabolite profiling in the endosperm of germinating rice seed, these data are currently the most detailed metabolomic study with a detailed time course from dry to germinated seeds. This allowed to highlight 104 and 62 metabolites differentially accumulated during the time course of rice germination in the embryo and the endosperm, respectively (**Supplementary Figure S2**, **Supplementary Tables S3B,D**, $p < 0.05$). A PCA analysis on the metabolomic data was also performed for both embryo and endosperm (**Figures 2C,D**). Contrarily to the transcriptome, here the percentage of variation explained by the first principal component was quite different between the embryo (57.1%) and the endosperm (24.8%). In the embryo, while it seems that PC1 was also related to time after imbibition, the dry state E0 was less well separated from early time points (4, 8, and 12 HAI) while it was easily isolated from later time points (16 and 24 HAI). Since the dry embryo can be distinguished from other time points on the second component (PC2, 26.1% total variation), it suggests that the major metabolic changes take place after 12 h in the embryo. In the endosperm, PC1 explains far less variation (24.8%) than the one in the embryo (57.1%). Yet, it also separates late time points (16 and 24 HAI) from early time points (4, 8, and to a lesser extent 12 HAI). Still, it is also remarkable that the dry endosperm (A0) can be easily separated on the second component PC2 (19.6% of total variation). Overall, in the transcriptome, the first 4 HAI has a dramatic effect on both embryo and endosperm transcriptomes and metabolomes as shown by the clear separation of E0 and A0 samples on PC1 and PC2 (**Figure 2**). Then we can distinguish a group of early time points (4 and 8 HAI). In the metabolome, the distinction is rather conserved but less sharp suggesting that transcriptome changes are more rapid and consistent than those at the metabolite level.

Proteome Analyses of Both Compartments, Embryo and Endosperm, in Dry and Germinating Seeds

To elucidate the developmental and metabolic changes accompanying rice seed germination in both isolated embryos and endosperms, a differential proteomic analysis was performed by LC-MS/MS shotgun approach allowing the comparison between dry vs. germinating (24 HAI) seeds (**Supplementary Table S4**). Thus, 2,419 non-redundant proteins (2,315 in the embryo, 832 in the endosperm, 725 common) with at least two locus-specific peptides in one biological replicate were identified (**Supplementary Tables S4A**,**C**). Furthermore, for each protein and in each compartment the protein abundance in dry (0 HAI) and germinating seeds (24 HAI) were compared by calculating the log2 ratio and a *p*-value based on a Student's *t*-test ($p < 0.05$, **Supplementary Tables S4B**,**D**). This statistical analysis yielded 417 (19% of the detected embryo proteome) and 120 (16% of the detected endosperm proteome) differentially accumulated proteins during germination in the embryo and the endosperm, respectively. Shotgun proteome analysis revealed 265 up accumulated and 152 down accumulated proteins in embryos during the period from 0 to 24 HAI (**Supplementary Tables S5A**,**C**). In the rice endosperm 11 up accumulated proteins and 109 down accumulated proteins were characterized between 0 and 24 HAI (**Supplementary Tables S5E**,**G**). To reveal the specific proteome features associated with the germination process, a gene ontology (GO) terms classification of the proteins differentially accumulated in rice embryo and endosperm during germination was also performed (**Figure 3**). Up accumulated proteins in the embryo during germination belonged mainly to the stress response (GOs:0046686, 0006979, 0009628, 0055114,

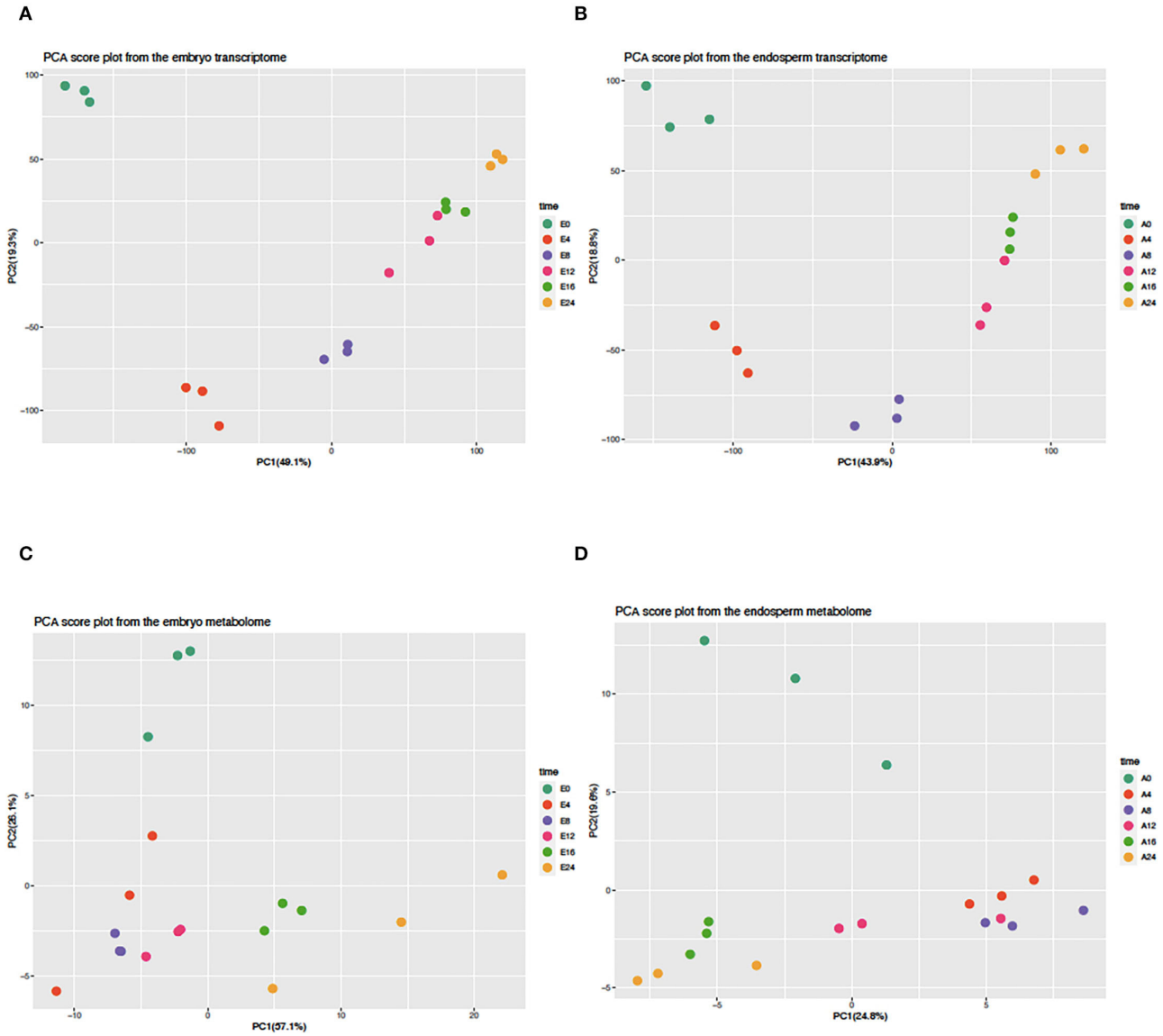


FIGURE 2 | Principal component analysis of embryo and endosperm transcriptomes and metabolomes. The normalized transcript **(A,B)** and metabolite **(C,D)** abundance of isolated embryos **(A,C)** and endosperms **(C,D)** were subjected to a principal component analysis (PCA) of which the first two components are shown in each graph. The percentage of explained variance is indicated along each axis. All values were centered and scaled before computing the PCA.

0009651, 0009737), carbohydrate metabolism (GOs:0005975, 0006096, 0009056), cytoskeleton (GOs:0051258, 0007017), and transport (GOs: 0015992, 0006886, 0015031) -related classes. These GO categories, characterized on the basis of proteomic analysis of germinating seeds, have been reviewed in detail (He and Yang, 2013; Tan et al., 2013; Czarna et al., 2016). Especially, stress response proteins might accumulate during germination to help the future seedling to cope with environmental conditions as well as the accumulation of proteins involved in carbohydrate metabolism is in accordance with the shift from the quiescent to active metabolic seed state upon imbibition. Down accumulated proteins in the embryo during germination also belonged mainly to the stress response and carbohydrate metabolism but showed specific enriched terms to the detailed categories, such as response to cold (GO:0009404), response to high light intensity (GO:0009644), and protein folding (GO:0006457). It is interesting to note that many of the 152 proteins that were down accumulated in the embryo during germination were previously associated with the maturation program (Zi et al., 2013). For instance, this concerns heat shock proteins (HSPs), LEAs (late embryogenesis abundant) proteins, and proteins related to reactive oxygen species (ROS) homeostasis (Zi et al., 2013). These proteins are definitely involved in desiccation tolerance allowing survival in a dry state. In the rice endosperm, only

11 proteins were up accumulated (**Supplementary Table S5E**). Functional ontological classification was not possible because of this low number and also because these proteins exhibited very diverse biological functions. However, it is noted that 10 of the 11 up accumulated proteins are specifically identified at 24 HAI as for example for alpha-amylases (AMY1A, LOC_Os02g52710, and AMY3E, LOC_Os08g36900), malate dehydrogenase (LOC_Os08g3372), or defensin DEF8 (LOC_Os03g03810). Ontological classification of the down-accumulated proteins in rice endosperm during germination was again enriched in stress response-related categories (GOs:0046686, 0006979, 0009628, 0009651, 0006950), while specifically over-represented in seed development, such as reproduction (GO:0000003), post-embryonic development (GO:0009791), and embryo development (GO:0009790), as well as cellular amino acid biosynthetic process (GO:0008652). This may highlight the shift from the maturation to the germination process in the endosperm. These GO analyses disclosed that proteins belonging to distinct ontology classes are up or down accumulated in embryos and endosperm during germination, while they also have a common ontology class, such as stress responses. This was also confirmed by mapping the identified proteins on seed-specific function with MapMan (**Supplementary Figure S3**). Some classes, such as fatty acid synthesis and proton transport, are clearly activated during embryo germination while heat stress-related proteins were down-accumulated in both germinating embryo and endosperm.

Combined Transcriptome and Proteome Changes in the Embryo During Rice Seed Germination

To further unravel the molecular functions involved during rice seed germination, we focused on the interpretation of the differential protein accumulation revealed by label-free shotgun proteomic in relation with the respective transcriptomic data. From the 417 differentially accumulated proteins in embryos during germination, 362 loci (genes with MSU locus IDs) matched with a detected transcript (**Supplementary Tables S5B**,**D**). Therefore, for 55 differentially accumulated proteins (36 up- and 19 down-accumulated), no detectable mRNAs were revealed by transcriptomic in both dry and germinating seeds (**Supplementary Tables S5B**,**D**). Most of these proteins (47 loci) have no corresponding probes on the chip. Only few of them (8 loci) have no detectable probe in all replicates. Thus, 265 up- and 152 down-accumulated proteins were considered for comparison with their corresponding transcripts (**Figure 4**; **Supplementary Tables S5B**,**D**). From this comparison, three clusters could be defined (**Figure 4A**). Cluster 1 corresponded to 11 transcripts whose accumulation profiles differed from those observed for the corresponding proteins during rice seed germination (**Figure 4A**). Thus, these transcripts were abundant in the dry embryos and their accumulation gradually decreased during germination while the corresponding proteins were up-accumulated. Cluster 2 corresponded to 128 transcripts whose accumulation profiles correlated well-with those of the corresponding proteins is an increased abundance over germination time (**Figure 4A**). Finally, Cluster 3 corresponded to 90 transcripts detected in the dry embryo for which the abundance did not change during germination while the corresponding proteins were up accumulated (**Figure 4A**). Similarly, from the comparison of down-accumulated proteins with their corresponding transcripts, three clusters were defined (**Figure 4B**). For the down-regulated transcripts during germination, Cluster 4 corresponded to 54 transcripts whose accumulation profiles correlated well with the observed down-accumulation of the corresponding proteins (**Figure 4B**, **Supplementary Table S5D**). Cluster 5 corresponded to 5 transcripts for which accumulation profiles during rice seed germination absolutely differed from the observed accumulation patterns of the corresponding proteins (**Supplementary Table S5D**). In Cluster 5, these transcripts were progressively accumulated during the time course of germination whereas the corresponding proteins were down accumulated (**Supplementary Table S5D**). Finally, Cluster 6 corresponded to 74 transcripts detected in the dry embryo and for which the abundance did not change during germination while the corresponding proteins were down-accumulated (**Figure 4B**, **Supplementary Table S5D**). Overall, a global comparison of mRNA and protein changes in the embryo and the endosperm shows that both tissues exhibit very poor correlations as previously shown in the dry rice seed (**Supplementary Figure S4**; Galland et al., 2017).

Combined Transcriptome and Proteome Changes in the Endosperm During Rice Seed Germination

The present proteome analysis highlighted 121 proteins that were differentially accumulated during germination in the rice seed endosperm (**Supplementary Tables S5E**,**G**). It is worth noting that out of them 109 proteins (90%) were down accumulated emphasizing the importance of protein degradation during the germination process. A likely explanation to account for this behavior might be that degradation of proteins accumulated in the mature endosperm is used for fuelling the embryo with free amino acids to boost metabolic restart and seedling establishment upon imbibition. Thus, considering the cereal endosperm as dead tissue, specific proteins may migrate from the aleurone layer to activate physiological processes leading to germination. Interestingly eight down accumulated proteins in the endosperm were up accumulated in the embryo during rice germination (**Supplementary Tables S5A**,**G**). All of them belong to the stress response protein family. This is for instance the case for the ATP-synthase (LOC_Os09g08910, LOC_Osg49190) implicated in the respiration pathway, the peptidyl-prolyl cis-trans isomerase (PPI, LOC_Osg01g18120) previously shown to be implicated in protein folding and in the control of the cellular redox state (Laxa et al., 2007; Bissoli et al., 2012), or the glycolytic enzymes, phosphoglucomutase (LOC_Os03g50480) and enolase (LOC_Os10g08550), that may be translocated from the endosperm to the embryo. From the transcriptomic data, 96 transcripts corresponded to the 109 down-accumulated proteins in the endosperm (**Supplementary Table S5H**). Three

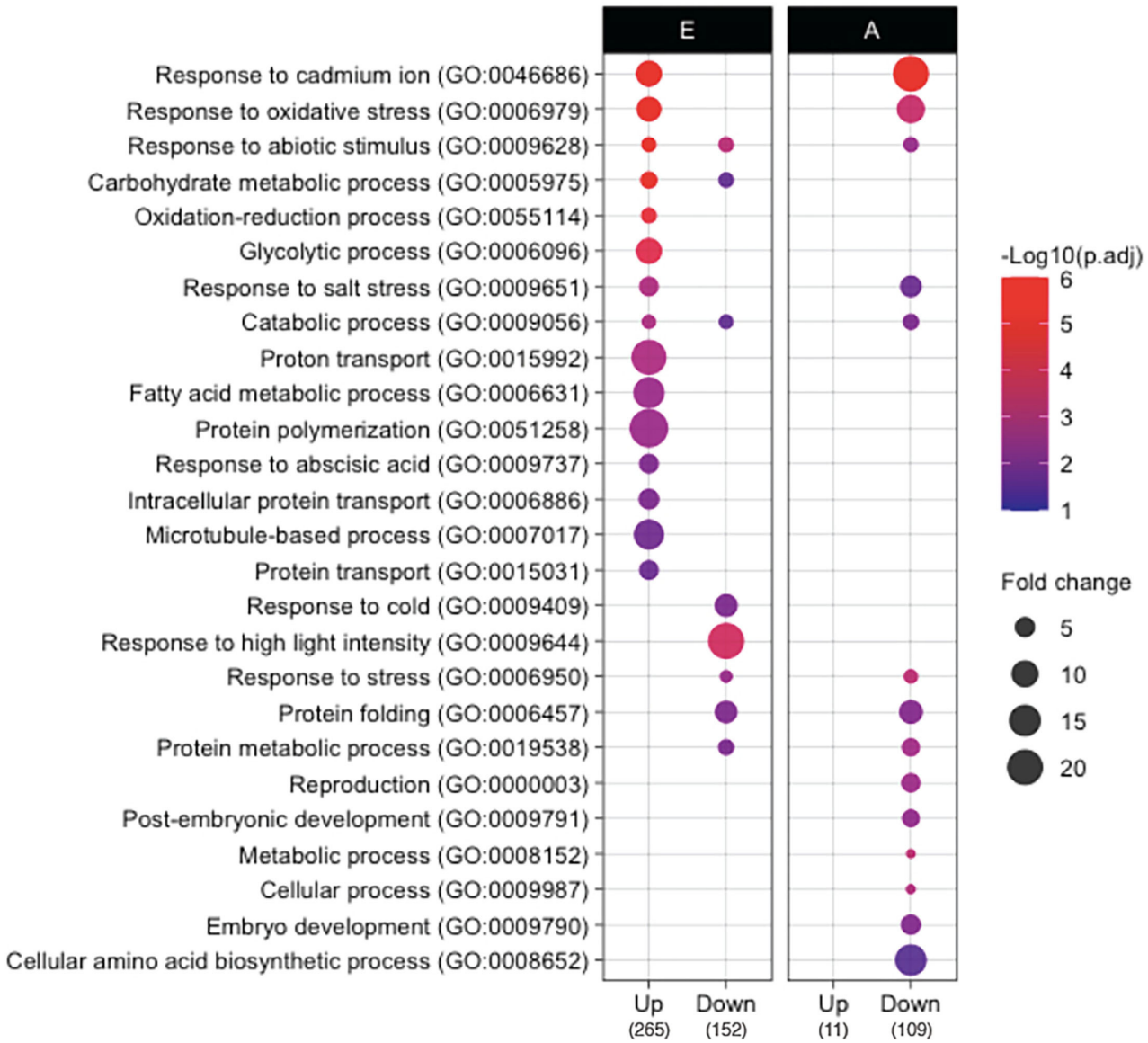


FIGURE 3 | Gene ontology (GO) comparison of proteins that accumulate differently in embryo and endosperm during germination. GO analysis was performed by using the CARMO (http://bioinfo.sibs.ac.cn/carmo/) for differentially accumulated proteins in the embryo (E) and endosperm (A) at 24 HAI compared to 0 HAI (the numbers in bottom brackets of the panel indicate the number of up or down-accumulated proteins). Significant GO terms (adjusted p-value < 0.05, Benjamini–Hochberg correction) in "biological process" were obtained from each protein group except for up-accumulated proteins in endosperm since too few proteins ($n = 11$) were up-accumulated in the endosperm. Fold change represents fold enrichment of genes with the given term as compared to ones in the rice genome background.

clusters could be defined (**Supplementary Table S5H** and **Figure 4C**). Cluster 9 corresponded to 20 down accumulated transcripts whose accumulation profiles during germination correlated well with a down-accumulation of their corresponding proteins (**Supplementary Table S5H**). Cluster 10 corresponded to 29 transcripts whose accumulation profiles completely differed compared to those of the corresponding proteins (**Supplementary Table S5H**); in cluster 10, these transcripts showed progressive accumulation during the time course of germination whereas the corresponding proteins were down accumulated (**Supplementary Table S5H**). Finally, Cluster 11 corresponded to 47 transcripts detected in the dry endosperm and for which abundance did not change during germination while corresponding proteins were down accumulated (**Supplementary Table S5H**). The shotgun proteomic data revealed only 11 proteins that were significantly up accumulated in the endosperm at 24 HAI, the time at which almost all seeds completed germination (**Supplementary Table S5E**).

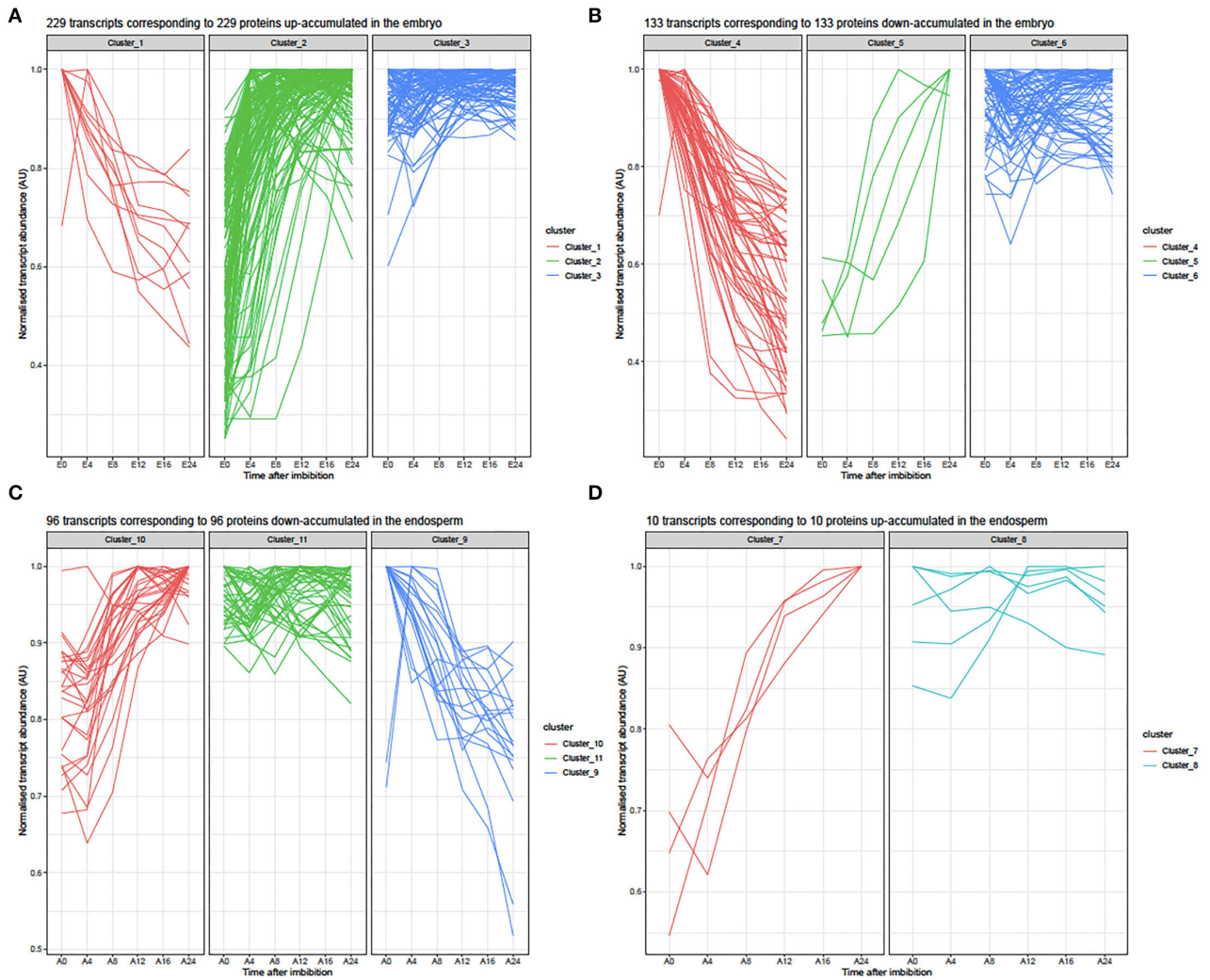


FIGURE 4 | Line plots of transcripts corresponding to embryo or endosperm with significant contrasted accumulations between 0 and 24 HAI. **(A)** Normalized abundances of 229 transcripts detected in the embryo during rice seed germination time course corresponding to 229 proteins up-accumulated at 24 HAI. From the shotgun proteome, 256 up-accumulated proteins were detected, of which only 229 showed changes in the corresponding mRNA abundances. In the embryo transcriptome data, the corresponding mRNA abundance was normalized by their corresponding maximum intensities reached during germination to yield accumulation values comprised between 0 and 1. E, Embryo; E0, dry embryos, E4-E24, dissected embryos from whole rice seed imbibed 4, 8, 12, 16, and 24 HAI. Clusters 1, 2, and 3 are described in **Supplementary Table S5B**. **(B)** Normalized abundances of 133 transcripts detected in the embryo during the rice seed germination time course corresponding to 133 proteins down-accumulated at 24 HAI. The shotgun proteome evidenced 152 down-accumulated proteins (**Supplementary Table S5C**). From the embryo transcriptome data, the corresponding mRNA abundances were normalized by their maximum intensities reached during germination to yield accumulation values comprised between 0 and 1. E, Embryo; E0, dry embryos; E4-E24, dissected embryos from whole rice seed imbibed for 4, 8, 12, 16, and 24 HAI. Clusters 4, 5, and 6 are described in **Supplementary Table S5D**. **(C)** Normalized abundances of 96 transcripts detected in the endosperm during rice seed germination time course corresponding to 96 proteins down-accumulated at 24 HAI. From the shotgun proteome, 109 were down-accumulated proteins at 24 HAI and corresponded to these 91 transcripts (**Supplementary Table S5G**). From the endosperm transcriptome data, corresponding mRNA abundances were normalized by their corresponding maximum intensities reached during germination to yield accumulation values comprised between 0 and 1. A, Endosperm; A0, dry endosperm; A4-A24, dissected endosperms from whole rice seed imbibed 4, 8, 12, 16, and 24 HAI. Clusters 9, 10, and 11 are described in **Supplementary Table S5H**. **(D)** Normalized abundances for 10 transcripts detected in the endosperm during rice seed germination time course corresponding to 10 proteins up-accumulated at 24HAI. From the shotgun proteome, 11 were up-accumulated proteins at 24 HAI and corresponded to these 10 transcripts detected in our transcriptome data (**Supplementary Table S5F**). From the endosperm transcriptome data, the corresponding mRNA abundances were normalized by their corresponding maximum intensity reached during germination to yield accumulation values comprised between 0 and 1. A, Endosperm; A0, dry endosperm; A4-A24, dissected endosperms from whole rice seeds imbibed for 4, 8, 12, 16, and 24 HAI. Clusters 7 and 8 are described in **Supplementary Table S5F**.

The corresponding transcripts were grouped according to two major clusters (**Supplementary Table S5F**, **Figure 4D**). All these transcripts except for LOC_Os05g49880 were detected in our transcriptome analyses and were strongly accumulated in the rice endosperm at 24 HAI in correlation with the accumulation of corresponding proteins. The main difference between Clusters

7 and 8 is the delay in the time of accumulation. Interestingly, out of the 11 up-accumulated endosperm proteins, 10 of them were germination-specific while the corresponding transcripts were detected both in dry and imbibed endosperm suggesting the occurrence of translational regulation. Altogether, these results emphasized the importance of post-transcriptional and translational controls in rice seed germination.

DISCUSSION

Seed germination is a complex process that includes, through imbibition, reinitiating cell metabolism from a quiescent to a highly active state, in which a large number of genes are known to be involved. However, there is no integrated analysis of gene expression during rice germination in the two tissues; embryo and endosperm. This study is the first systematic "multi-omics" (transcriptomics, proteomics, and metabolomics) study on both rice embryos and endosperm during germination. Here, we discuss the differences in embryos and endosperm at the multi-omics level, focusing on signaling and hormones, central metabolism, and protein turnover process to provide in-depth insight into the tissue-specific germination process.

Signaling and Hormones

The proteomic data corresponding to the rice embryo during germination revealed an up-accumulation of the GID1L2-gibberellin receptor (LOC_Os11g13670) and a down-accumulation of CNX1 (LOC_Os04g56620) (**Supplementary Table S4**), an enzyme involved in the biosynthesis of the molybdenum cofactor (MoCo) essential for ABA biosynthesis (Schwarz and Mendel, 2006), suggesting that as in other seed species these two phytohormones play a central role in rice seed germination. GAs are tetracyclic diterpenoid phytohormones implicated in a wide range of developmental processes, including seed germination (Olszewski et al., 2002; Hedden, 2020). The GID1 GA receptor was previously detected in rice as a soluble receptor implicated in GAs signaling. Mutants affected in the *GID1* gene produce dwarf rice plants (Ueguchi-Tanaka et al., 2005). It is admitted that GID1 action proceeds through active interactions between GID1 and GAs suppressors called DELLA proteins (Ueguchi-Tanaka et al., 2007). Experimental evidence showed that DELLA repression in rice is mainly achieved *via* its degradation as the GID1-GA-DELLA complex is recognized by the 26S proteasome (Ueguchi-Tanaka et al., 2005; Shimada et al., 2008; Sun, 2011). In Arabidopsis, DELLA repression was also shown to be effective through protein-protein interactions between GID1 and DELLA proteins that may be sufficient to block DELLA's repressing activity (Ariizumi et al., 2008). According to Rice Genome Annotation Project, 40 loci are annotated as putative gibberellin receptor GID1. An alignment of the amino acid sequences for these 40 genes, along with that for the three Arabidopsis GID1 orthologs (AtGID1a, b, and c), yields the phylogenetic tree presented in **Supplementary Figure S5**. All rice GID1 sequences (**Supplementary Figure S5**) possess a hydrolase domain that corresponds to a putative active GA-binding domain (Shimada et al., 2008). In Arabidopsis, GID1a and GID1c are surmised to be preferentially active during the vegetative development while GID1b is implicated in GA signaling during germination (Ariizumi et al., 2008; Hauvermale et al., 2014). According to amino acid sequence similarities, GID1L2 (LOC_Os11g13670) would be implicated in the GA metabolism during seed during germination. The detection of GID1L2 protein in rice embryos in a germination-specific manner is also supported by the transcriptomic data showing an up-accumulation of the *GID1L2* transcript abundance significantly until 12 HAI in the embryo (**Supplementary Table S1A**). A down accumulation of the MoCo biosynthesis enzyme CNX1 (LOC_Os04g56620) protein was observed in the imbibed embryo in comparison with the dry state (**Supplementary Table S4B**) with a decrease in its transcript level upon imbibition (**Supplementary Table S1A**). This MoCo cofactor is implicated in ABA biosynthesis and thereby its decreased abundance is in accordance with metabolism restart during rice germination as ABA inhibits germination. It is known that the MoCo interacts with the aldehyde oxidase that converts abscisic aldehyde to ABA. Thus, deficient mutants in MoCo biosynthesis do not accumulate ABA (Leydecker et al., 1995; Schwarz and Mendel, 2006). This strongly suggests that the down accumulation of CNX1 protein promotes rice seed germination. Our proteomic analysis also identifies other factors potentially implied in signaling in a germination specific manner, such as the transcription factor HBP-1b (LOC_Os01g06560, Histone Binding Protein-1b). This protein annotated as HBP-1b according to the RGAP (Rice Genome Annotation Project, 7th version) is referred to as OsDOG1L-2 (OsDOG1-like-2), thereby corresponding to an Arabidopsis DOG1 (Delay Of Germination) protein (At5g45830) homolog (Sugimoto et al., 2010). *OsDOG1L-2* gene is a target of the reported preharvest sprouting (PHS) resistance gene, namely Sdr4. In japonica rice cultivar Nipponbare, a highly dormant nearly isogenic line of *Sdr4* (*Seed dormancy 4*) gene containing the indica Kasalath allele (NIL[*Sdr4-k*]) over-express *OsDOGL1-1* but the *OsDOGL1-2* expression remained unaffected. *Sdr4*, is considered as a major regulator involved in seed dormancy and domestication of rice (Sugimoto et al., 2010). From our present omics data, HBP-1b/OsDOGL1-2 transcript and protein abundance increase significantly in the embryo during germination. Interestingly, HBP-1b was located inside of a QTL region for seed dormancy (Li et al., 2011) and or salinity tolerance (Lakra et al., 2015). The novel finding observed here showing OsHBP-1b accumulation in the embryo during germination might reveal a novel function, distinct from DOG1's, of this TF belonging to the bZIP family in rice germination. Recent work showed that transgenic rice over-expressing *OsHBP-1b* exhibits better germination capacity, higher shoot growth, and higher fresh weight under salinity stress than the wild type (Das et al., 2019). Further studies are needed to establish the function of DOG1-like genes in the modulation of rice seed germination and vigor.

The endothelial differentiation-related factor 1 (EDF1, LOC_Os06g39240) belonging to the MBF1 (multiprotein bridging factor 1) transcription factor family was also identified specifically in imbibed rice embryo, at the time of germination completion (24 HAI). OsEDF1 is orthologous to AtMBF1c (At3g24500) with 70% of amino acid sequence homology.

MBF1 is known as an evolutionary conserved transcriptional co-activator that mediates transcriptional activation by bridging between an activator and a TATA-box binding protein (Takemaru et al., 1997; Jaimes-Miranda and Chávez Montes, 2020). In Arabidopsis, *mbf1c* mutants were more sensitive to heat stress than are wild-type plants (Suzuki et al., 2008). In contrast, transgenic rice over-expressing wheat *MBF1c* was more tolerant to heat stress (Suzuki et al., 2005; Qin et al., 2015). In seeds, MBF1c is suggested to act as a positive regulator of ABA degradation during the early steps of germination (Di Mauro et al., 2012). To our knowledge, no data describing MBF1 protein action in rice seeds have been reported yet. We found that the EDF1/MBF1c protein was specifically accumulated in germinating embryos, although the transcriptomic data indicated a decreased level of the corresponding transcript **Supplementary Tables S1A**, **S4A**). This result suggested that the functional activation driven by *EDF1/MBF1* is under translational control during rice germination. The ABA content in rice seeds decreases rapidly after few hours of imbibition (Song et al., 2020) concomitantly with an increased expression of the ABA catabolism genes (*OsABA8ox2, OsABA8ox3*) (Zhu et al., 2009; Wang et al., 2021), supporting the finding that as in Arabidopsis, rice MBF1 acts as a positive regulator of ABA degradation (Di Mauro et al., 2012), thereby favoring germination. Interestingly, this gene is also described as an ethylene-response transcriptional co-activator (ERTCA) and could play a central role in the crosstalk between ABA and ethylene to control germination vigor and stress responses (Arc et al., 2013; Jaimes-Miranda and Chávez Montes, 2020). In Arabidopsis, ethylene boosts protein and mRNA levels of plant defensins to cope with pathogens (Penninckx et al., 1998). In the endosperm, the defensin DEF8 was up accumulated at 24 HAI. The OsDEF8 previously displayed antimicrobial activity against phytopathogens, such as *Xanthomonas oryzae* and *Fusarium oxysporum*, indicating a role in disease control (Tantong et al., 2016; Weerawanich et al., 2018). The specific accumulation of DEF8 in rice endosperm during germination suggests a protecting role of the embryo by this compartment through antimicrobial peptide production.

Central Metabolism

Tricarboxylic (TCA) and Glyoxylate Cycles for Energy Feeding

During rice seed germination, respiration starts just after 1 HAI and mitochondrial differentiation starts after 8 HAI (Howell et al., 2006). Plastidic and cytosolic glycolysis together with the mitochondrial TCA cycle are the main components of the respiratory metabolism that supply carbon and energy to cells (Fernie et al., 2004). The accumulation of AMY1A and AMY3E in rice endosperm during germination indicates the catalysis of starch breakdown providing substrate for the glycolysis and the pentose phosphate pathway and promoting energy production and reducing power for embryo germination. The TCA and glyoxylate cycle intermediates during rice germination were identified by the present metabolic analysis and all the identified metabolites were highly accumulated in the germinating embryos (**Supplementary Table S3A**, **Figure 5A**) in concordance with a previous study (Howell et al., 2009). In endosperm, 61 metabolites were identified as being significantly and differentially accumulated (**Supplementary Table S3C**). The highlighted metabolic pathways indicate an up accumulation of the glycolysis by-products in endosperms rapidly after imbibition; this may be correlated to starch degradation (major compound) (**Figure 5B**). Identification of the associated enzymes differentially accumulated between embryo and endosperm may pinpoint key steps involved in the activation and maintenance of the energetic flux for germination. NADP-malic enzymes (LOC_Os01g09320 and LOC_Os05g09440), citrate synthases (ACTS, LOC_Os11g47330, LOC_Os02g13840), and malate dehydrogenase (MDH, LOC_Os08g33720) specific isoforms were more accumulated in the germinated embryos than in the endosperm tissue. Interestingly, two plastidic isoforms of the NADP-malic enzyme (LOC_Os01g09320 and LOC_Os05g09440) were present specifically at the germinated state (**Supplementary Table S5A**; **Figure 5A**). In contrast, the abundance of the NADP-malic enzyme encoded by the *LOC_Os01g54030* gene significantly decreased at 24 HAI in both embryo and endosperm. NADP-malic enzymes catalyze the reversible and oxidative decarboxylation of malate to pyruvate. In the Arabidopsis genome, four genes encode the NADP-ME, namely three cytosolic NADP-ME 1,2,3 and a plastidic NADP-ME4, while according to the rice genome, seven isoforms are annotated. The Arabidopsis NADP-ME2 (AT5G11670) presents a protein sequence similarity of about 81% with the rice NADP-ME (LOC_Os05g09440). Transcript abundance of the NADP-malic enzyme isoform LOC_Os05g09440 was significantly up accumulated at the early stages of imbibition (0-4 HAI) (**Figure 5A**; **Supplementary Table S1A**), which may be correlated to the specific role of this enzyme in oxidative stress response, as previous studies reported an up accumulation of the *NADP-ME2* (AT5G11670) in Arabidopsis leaves submitted to oxidative stress, although accumulation of the enzyme was shown to be a non-limiting factor of the stress response (Li et al., 2013). Oxidative burst occurring during imbibition may be correlated to the increase of this *NADP*-ME transcript in the earliest steps of germination (Liu et al., 2007). The activation of the TCA cycle may explain the high accumulation of the NADP-ME that catalyzes the oxidative decarboxylation of malate to produce pyruvate and then feed the TCA cycle (Chang and Tong, 2003; Chen et al., 2019). The same behavior was observed for the malate dehydrogenase (MDH) where two isoforms (LOC_Os12g43630 and LOC_Os08g33720) were identified as being significantly up (LOC_Os08g33720) and down (LOC_Os12g43630) accumulated in the embryo (**Supplementary Tables S5A**,**C**; **Figure 5A**). Finally, it is admitted that the cyclic flux is maintained to ensure the respiratory metabolism *via* the acetyl CoA input where the oxaloacetate and acetyl CoA condensation is the starting reaction activating the TCA cycle flux (Sweetlove et al., 2010). However, a TCA cycle connection with the glyoxylate cycle and the aspartate family super pathway is not necessarily maintained as a circular flux especially under stress conditions to sustain adequate levels of ATP (Sweetlove et al., 2010). This observation may pinpoint a metabolic orientation in germinating seeds depending on seed storage compound composition and actual

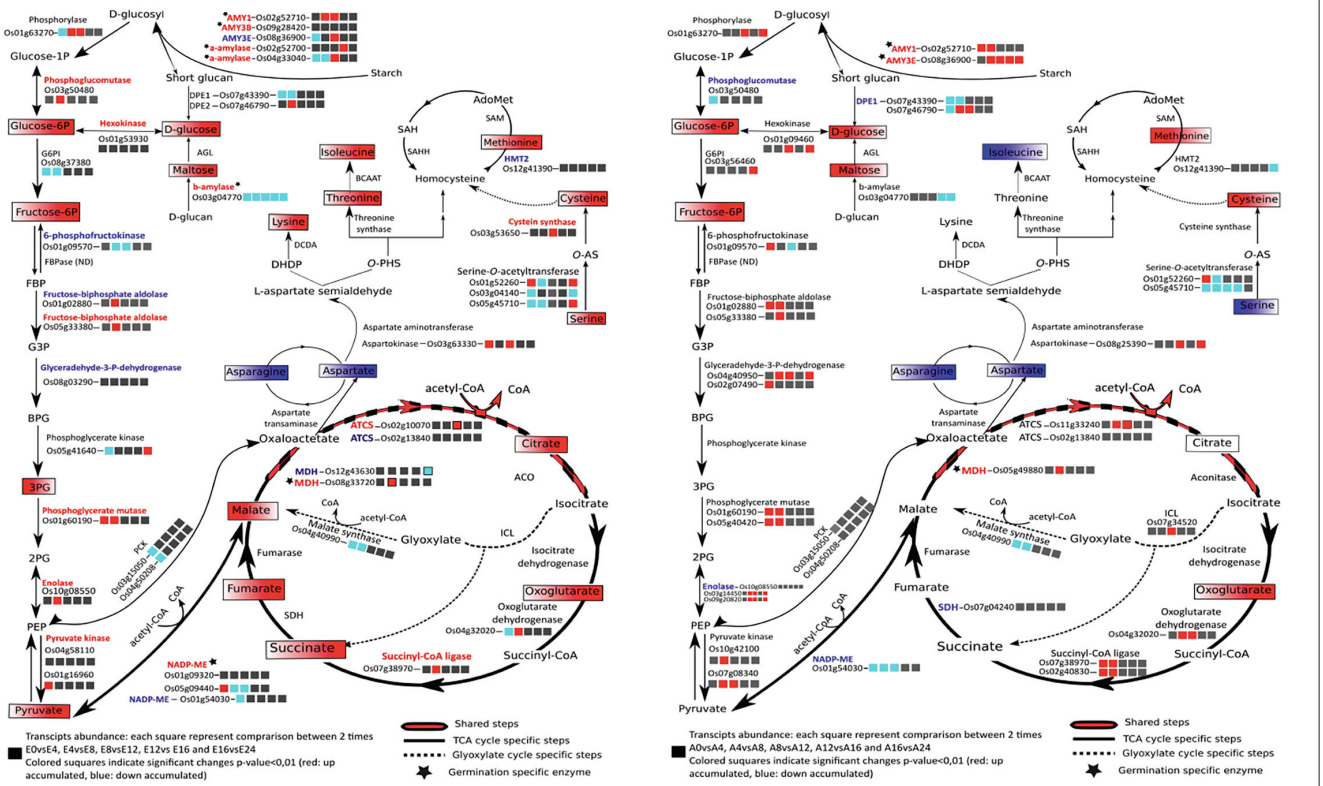


FIGURE 5 | Metabolic pathways related to seed glycolysis/neoglucogenesis, energy, and amino acid metabolisms in rice embryo **(A)** and rice endosperm **(B)** comparing mature dry and germinated seeds. Up and down-accumulated metabolites during germination are displayed, respectively, in red and blue boxes. Proteins present in the total embryo proteome are indicated in black if not detected or if their abundance change was not significant. ACO, aconitase; AGL, alpha-glucosidase; ATCS, citrate synthase; BCAAT, branched-chain-amino-acid aminotransferase; DCDA, diaminopimelate decarboxylase; DHDP, dihydrodipicolinate; DPE, depolymerizing enzyme; FBP, fructose-1,6-bisphosphate; G3P, glycerate-3-phosphate; HMT2, homocysteine *S*-methyltransferase; ICL, isocitrate lyase; MDH, malate dehydrogenase; NADP-ME, NADP-malic enzyme; *O*-PHS, *O*-phospho-L-homoserine; PCK, phosphoenolpyruvate carboxykinase; PEP, phosphoenolpyruvate; 3PG, 3-phosphosphoglycerate; *O*-AS-, *O*-acetylserine; SAH, *S*-adenosyl-L-homocysteine; SAHH, *S*-adenosyl-L-homocysteine hydrolase; SAM2, *S*-adenosylmethionine synthetase; SDH, succinate dehydrogenase; TCA, tricarboxylic acid.

metabolic requirements. In Arabidopsis seeds, glyoxylate and TCA cycle intermediates and enzymes are up accumulated during germination (Fait et al., 2006; Galland et al., 2014). In germinating rice seeds, as in Arabidopsis, fructose-6P, glucose 6P, and almost all the TCA and glyoxylate cycle intermediates presently identified are accumulated. The TCA and glyoxylate cycles exhibit a diverse set of features rendering this machinery much more able to deal with substrates supply (van Dongen et al., 2011).

Amino Acid Metabolism

In addition, to provide the monomeric precursors of proteins, amino acids are the precursor form of carbon and nitrogen and play a key role in the energy supply. Indeed, in non-photosynthetic organisms, amino acid catabolisms play a role to feed the tricarboxylic cycle (TCA) as an alternative way to maintain energy supply (Galili, 2011). Amino acid metabolism also plays as an important role during stress response where it is admitted that transcriptomic changes include the repression of amino acid biosynthesis genes and stimulation of genes implicated in the catabolism of some amino acids (Galili, 2011; Wang et al., 2018). From our GC-MS metabolomic data, 29 and 15 amino acids were detected in the rice embryo and endosperm, respectively, as being significantly and differentially accumulated during germination. In rice embryo, most of the amino acids were drastically accumulated at 24 HAI suggesting a burst of energy mobilization for post-germination phases and plant development (**Supplementary Tables S3A**,**C**, **Supplementary Figure S6**). In the endosperm, eight amino acids were down accumulated and seven were up accumulated during rice germination; however, fold changes are quite low varying from 0.3 to 1.7 (**Supplementary Table S3C**). This may indicate that most amino acid changes observed here are those of the aleurone layer, which contains more proteins than the starchy endosperm. Interestingly, aspartate abundance decreased very rapidly upon imbibition in both embryo and endosperm, suggesting consumption of this amino acid pool to feed the aspartate pathway (**Figure 5**, **Supplementary Figure S6**). This

observation is one of the main differences between the two plant models, Arabidopsis and rice, since in Arabidopsis the aspartate level increases during seed germination (Fait et al., 2006). The aspartate-degradation super pathway leads to the formation of methionine, lysine, and threonine. Each of these amino acids plays a key role in the cell metabolism and the organism surviving. From the present "omic" data, the aspartate pathway can be redrawn to show transcripts, proteins, and metabolites abundance in the embryo during germination (**Supplementary Figure S7**). As only few changes were observed in the endosperm, in the following, we will only focus on the germinating embryo aspartate pathway. Aspartate-derived amino acids were up accumulated in germinated embryos with ratios varying from 2 for the threonine to 5 for the isoleucine (**Supplementary Figures S6**, **S7**). It is noted that while lysine, threonine, and methionine drastically increase (**Figure 5A**; **Supplementary Table S3A**, **Supplementary Figure S6**), most enzymes of this pathway were not differentially accumulated during germination (**Figure 5** and **Supplementary Figure S7**; **Supplementary Table S4B**). Only three enzymes displayed significant changes at 24 HAI (**Supplementary Figure S7**; **Supplementary Table S4B**). Aspartate dehydrogenase (LOC_Os03g55280) and aminotransferase (LOC_Os10g25130) were significantly up accumulated during germination in accordance with the apparent activation of this pathway. Homocysteine-*S*-methyltransferase (methionine synthase) was significantly down accumulated (2-fold change) despite the increase in methionine level suggesting a regulated circular flux of synthesis/catabolism.

The sulfur amino acid metabolism is a major hub to the physiology of seed germination, allowing connecting housekeeping metabolic activity and hormonal regulation (Rajjou et al., 2012). Methionine is the initiator of protein elongation and is the precursor of *S*-adenosyl-methionine (SAM) the activated form of methionine and the universal methyl donor for a myriad of trans methylation reactions (Ravanel et al., 1998; Rahikainen et al., 2018). Methionine and SAM are involved in different central reaction pathways. Indeed, methionine biosynthesis is directly linked to the folate, cobalamin, and pyridoxine pathways and methionine-derived SAM is the substrate in most preponderant trans methylation reactions, including protein repair, as well as in DNA/RNA and histone methylation, Gas, ethylene, and biotin biosynthesis pathways rendering methionine a key regulator metabolite (Amir, 2010; Sauter et al., 2013; Huang et al., 2019). Methionine-derived ethylene is also a crucial hormone being involved in many pathways as plant growth and immune responses (Broekgaarden et al., 2015; Dubois et al., 2018). Methionine is required for protein synthesis in seed germination and seedling establishment both in Arabidopsis (Rajjou et al., 2004) and rice (Sano et al., 2012). It is noted that although chemical inhibition of methionine synthesis substantially delays seed germination in Arabidopsis but does not completely block this process while exogenous L-Met promotes germination (Gallardo et al., 2002; Ju et al., 2020). It is therefore possible that methionine accumulation through proteome renewal (*i.e.,* catabolism of proteins stored in the dry mature seeds) contributes significantly to this methionine requirement of rice embryo and therefore *de novo* methionine may not be the preferred mechanism, at least in the early phases of seed germination.

Protein Turnover

Protein Synthesis

During germination, the balance between synthesis and degradation of proteins is a highly regulated turnover process rendering possible the biological and metabolic switch from quiescence to active metabolism upon imbibition. The cellular proteome is established by the principal translation machinery; Cytosolic ribosomes consist of large 60S and small 40S subunits. During translation, recruitment of the 40S subunit starts at the 5′ end of mRNAs, which are bound to eukaryotic translation initiation factors, namely, a cap binding protein (eIF4E) and a helicase (eIF4A). A circle bridge is then formed between the 3′ and 5′ ends of the mRNA thus stimulating the translation process through the association of the eIF4G and eIF4B (Browning and Bailey-Serres, 2015). The eIF4B enhances the helicase activity of eIF4A, thereby determining the unwinding of 5′ mRNA cap's secondary structure (Bi et al., 2000). A total of 30 translation factors, including two eIF4B isoforms (LOC_Os02g24330 and LOC_Os02g38220), were identified by our proteomics, and LOC_Os02g24330 was up-accumulated at 24 HAI in rice embryos. These suggest that the translation activity of germinating rice embryos may be mainly managed through eIF4B (LOC_Os02g24330). Ribosomal proteins are the elementary components of ribosomes together with rRNA (Weis et al., 2015). Regulation of eukaryotic translation is also carried out by the interaction of the translation initiation factors with ribosomal proteins (Muench et al., 2012). In the present work, 13 ribosomal proteins were significantly up accumulated in rice embryos in which two (LOC_Os03g26860, LOC_Os05g48050) were strictly detected in the germinated embryos (24HAI) (**Supplementary Table S4A**). Interactions between ribosome and translation initiation factors orientate spatial translation and subsequent protein synthesis. Protein synthesis was detected in all rice compartments suggesting that translation could occur even in the isolated starchy endosperm (Galland et al., 2017). In agreement with this, the shotgun analysis data revealed an accumulation of two ribosomal proteins (40S ribosomal protein (LOC_Os02g48660) and a 60S ribosomal protein (LOC_Os02g48660) specifically in the endosperm at 24 HAI, although the most active compartment for protein synthesis is the embryo (Yang et al., 2007; Kim et al., 2009; He and Yang, 2013; Galland et al., 2017). Endosperms with conserved aleurone layer exhibited a slightly higher labeling signal than the starchy endosperm, suggesting an active protein synthesis from the aleurone layers during rice seed germination. Inhibition of translation is deleterious for rice germination while *de novo* transcription is not required (Sano et al., 2012, 2019). This suggests, as in Arabidopsis (Rajjou et al., 2004), a requirement for protein synthesis from stored mRNAs accumulated during seed development. The molecular mechanism of seed germination using stored mRNAs has been reviewed in detail recently (Sajeev et al., 2019; Sano et al., 2020). Particularly, the selective translation mechanisms of stored mRNAs could be one of

the reasons for those levels of transcript changes do not always correlate with levels of protein transitions during seed germination, which we also observed in the present study.

Protein Degradation

Protein degradation by proteolytic systems is an essential factor in many cellular processes, such as apoptosis, amino acid recycling and that avoids the detrimental effects (competition) associated with oxidized and misfolded protein accumulation. These proteins are degraded by the proteasome consisting of the 20S proteasome that degrades proteins in an ATP-independent manner and the ATP-dependent 26S proteasome that recognizes polyubiquinilated targeted proteins. This ATP-dependent reaction cascade involves the subsequent action of ubiquitin activating enzymes (E1s), ubiquitin conjugating enzymes (E2s), and ubiquitin protein ligase (E3s) enzymes (Vierstra, 2009; Marshall and Vierstra, 2019). In addition, the N-end rule pathway targets protein degradation and was shown to affect ABA sensitivity and promote seed germination (Holman et al., 2009; Zhang et al., 2018). In rice the RING Ub E3 ligase encoded by *OsCTR1* was shown to play an important role in drought tolerance; this study also showed that heterogeneous overexpression of *OsCTR1* in Arabidopsis entailed hypersensitive phenotypes with respect to ABA-responsive seed germination, seedling growth, and stomatal closure (Lim et al., 2014). Studies on endogenous plant proteases emerged and these enzymes are now admitted as taking part in the recognition and the subsequent degradation of targeted proteins (van der Hoorn, 2008). From our data, a cysteine endoproteinase was identified in the germinated rice embryo (**Supplementary Table S4A**). This protein is known to activate β-amylase through partial proteolysis and two cysteine endoproteinases, EP-A and EP-B, have been shown to digest hordein, the major barley seed storage protein (Guerin et al., 1992; Schmitt et al., 2013). Cysteine protease activities were related to the early remobilization of seed storage proteins during germination (Lu et al., 2015). In addition, we identified two isoforms of a proteasome subunit specifically at the germinated state (24 HAI) in rice embryos (**Supplementary Table S4A**). Likewise, RING-box (Really Interesting New Genes) protein 1a (RBX1A), is also germination specific (**Supplementary Table S4A**). These proteins are necessary for ubiquitin ligation activity of the multimeric cullin ring ubiquitin ligases. All RING finger proteins can bind to E2 ubiquitin-conjugating enzymes and possess E3 ubiquitin ligases, which promote targeted protein degradation. These enzymes are involved in many biological processes, such as DNA, RNA, and protein binding; nonetheless, germination specificity may pinpoint degradation machinery occurring during imbibition, necessary for proteome renewal from dry to imbibed seeds. Finally, LisH and RanBPM (Ran-binding protein in the microtubule-organizing center) domains containing protein (Complex GID, glucose-induced degradation) was also identified specifically at the germination stage (**Supplementary Table S4A**). This protein contains a LisH/CTHL (Lissencephaly type-1-like homology), CTHL being the (C-terminal domain of the LisH motif), and CRA domain. From previous studies in mammals and yeast, these domains demonstrated a proteosomal and/or lysosomal activity (Tomaru et al., 2010). CTLH identified in mammalian cells are homologous to glucose-induced degradation protein, which was previously identified as being involved in fructose biphosphatase degradation in both proteasome and vacuole in yeast (Regelmann et al., 2003; Tomaru et al., 2010). Studies on the Arabidopsis RanBPM containing LisH, CTHL, and CRA domains, disclosed a structural homology with CTHL complexes described in yeast and mammals and showed that the protein possesses a degradation activity (Tomaštíková et al., 2012). It will be important to specify the targets of these protein degradation-related enzymes for a better understanding of their roles in rice seed germination.

CONCLUSION

In the present study, we captured the multi-omes (transcriptome, proteome, and metabolome) of embryo and endosperm tissues during rice germination and identified characteristic molecular responses in each tissue by integrative analysis. These approaches are highly complementary and provide spatial and temporal atlas of molecular and biochemical transitions during rice germination. The dynamic changes observed in the endosperm during the time course of germination suggest a very important role of this compartment to both protect the embryo and promote its growth potential. These data not only summarize existing processes important for germination but also provide evidence for germination-specific molecular responses and candidate core genes in the embryo, and the endosperm contribution to seed vigor. Future detailed functional analysis of the relevant genes and metabolites as well as the development of molecular markers specific for each tissue will be promising for the control of rice germination.

METHODS

Biological Material

A total of 50 rice (*Oryza sativa* L. var. Nipponbare) seeds were used for germination assays at 30°C under dark and continuous oxygenation conditions in triplicate. Seeds with emerged coleoptiles were scored as germinated seeds at each time point. Water uptake values were calculated based on the dry and fresh weight of 10 bulks comprising 10 seeds. For multi-omics, three biological replicates of 150 whole seeds were used for each time point (0, 4, 8, 12, 16, 24 HAI). Rice seeds were dehulled and dissected with a sharp scalpel to separate isolated embryos (E0, E4, E8, E12, E16, E24) and isolated endosperms (A0, A4, A8, A12, A16, A24) during the time course of germination. The same samples were used for transcriptomic, proteomic, and metabolomic aiming to compare the multi-omics results as described below.

Transcriptome Analysis

Total mRNAs were isolated from three replicates of 100 embryos and 50 endosperms and hybridizations on the Affymetrix GeneChip® Rice Genome Array (Affymetrix, Santa Clara, CA, USA) were performed as previously described (Galland et al.,

2014, 2017). To obtain presence/absence calls for each probe, the CEL files were normalized by the MAS5 algorithm (Affymetrix). The CEL files were then normalized with the GC-RMA algorithm using the "gcrma" library available from the R Bioconductor suite of open-source software (Huber et al., 2015). Differentially expressed genes in the embryo and endosperm transcriptomes were detected by using a two-group *t*-test, followed by the Bonferroni method (adjusted *P*-value is <0.01). All raw CEL files are available from the Gene Expression Omnibus under the accession GSE43780.

Shotgun Proteomics

Protein Extraction and In-gel Digestion

For embryo protein extraction, three replicates of 50 embryos were ground in liquid nitrogen using a mortar and pestle. Total soluble proteins were extracted at room temperature in 400 µl thiourea/urea lysis buffer (7 M urea, 2 M thiourea, 6 mM Tris-HCl, 4.2 mM Trizma® base (Sigma-Aldrich, Lyon, France), 4% (w/v) CHAPS) supplemented with 50 µl of the protease inhibitor cocktail Complete Mini (Roche Diagnostics France, Meylan, France), 15 µl of dithiothreitol (DTT, 1 M, Sigma-Aldrich), 2 µl of DNase I (Roche Diagnostics), and 5 µl of RNase A (Sigma-Aldrich). For endosperm protein extraction, three replicates of 50 endosperms were ground in liquid nitrogen using a mortar and pestle and total soluble proteins were extracted at room temperature in 1 ml thiourea/urea lysis buffer (same composition as above) supplemented with 35 µl of DTT, 2 µl DNAse I, and 10 µl RNAse A. The protein extracts were left to agitate for 2 h at 4°C. All samples were then centrifuged at 20,000 g at 4°C for 15 min. The resulting supernatant was subjected to second clarifying centrifugation as above. The protein concentrations of the final supernatant were measured according to Bradford (1976) using Bovine Serum Albumin (BSA) as a standard. Twenty-five micrograms of embryo and endosperm soluble protein extracts (n = 3 biological replicates) were subjected to SDS-PAGE analysis with 10% acrylamide. Each lane was systematically cut into 16 slices and directly submitted to in-gel tryptic digestion with the Progest system (Genomic Solution) according to a standard trypsin protocol. Gel pieces were washed two times by successive separate baths of 10% acetic acid, 40% ethanol, and acetonitrile. They were then washed two times with successive baths of 25 mM NH4CO3 and ACN. Digestion was subsequently performed for 6 h at 37°C with 125 ng of modified trypsin (Promega) dissolved in 20% methanol and 20 mM NH4CO3. The peptides were extracted with 2% trifluoroacetic acid (TFA) and 50% ACN and then with ACN. Peptide extracts were dried in a vacuum centrifuge and suspended in 20 µl of 0.05% TFA, 0.05% HCOOH, and 2% ACN.

LC-MS/MS Analysis

Peptide separation by NanoLC was performed as described previously (Bonhomme et al., 2012). Eluted peptides were analyzed on-line with a Q-Exactive mass spectrometer (Thermo Electron) using a nano-electrospray interface. Peptide ions were analyzed using Xcalibur 2.1 with the following data-dependent acquisition parameters: a full MS scan covering 300–1,400 range of mass-to-charge ratio (m/z) with a resolution of 70,000 and an MS/MS step (normalized collision energy: 30%; resolution: 17,500). MS/MS Step was reiterated for the 8 major ions detected during the full MS scan. Dynamic exclusion was set to 45 s. For database searching, X!Tandem (Langella et al., 2017) was set up to search the 7th annotation of the Rice Genome Annotation Project database (Kawahara et al., 2013; Sakai et al., 2013) and a contaminant database (trypsin, keratins). Enzymatic cleavage was declared as a trypsin with two possible misscleavage. Cys carboxyamidomethylation was set to static modifications. Met oxydation was set as possible modifications. Precursor mass and fragment mass tolerance were 10 ppm and 0.02 Th, respectively. Only peptides with an E-value smaller than 0.1 were reported. Peptide quantification was performed by extracted ion current (XIC) using MassChroQ software (Valot et al., 2011). A 5 ppm precision windows was set for XIC extraction. The peptide ions not specific of a single protein were eliminated and reliably detectable ions that were detected at least twice out of the three biological replicates were used for the quantification. The Total Ionic Current (TIC) area under peak corresponding to the same ions was obtained for each peptide and then summed to get a protein abundance. The median normalization was used for the label-free data. The values of protein abundance were log2-transformed.

GO Enrichment Analysis for Differentially Accumulated Proteins

Differentially accumulated proteins during germination were determined by Student's *t*-test ($p < 0.05$) in embryos and endosperms, respectively. GO enrichment analysis for the up or down accumulated proteins was performed using the Functional Annotation tool in the CARMO (http://bioinfo.sibs.ac.cn/carmo/) with default parameters for thresholds ($p < 0.05$, fold enrichment > 2), followed by the Benjamini-Hochberg Procedure (adjusted $p < 0.05$). The values for -log10(adjusted *p*-values) and fold change of obtained GO terms in Biological Process were drawn using the ggplot2 package in R.

Metabolome Analysis by Gas Chromatography Coupled to Mass Spectrometry (GC-MS)

Metabolite samples were obtained starting from three replicates of 100 rice seeds manually dissected in embryo and endosperm. Embryos and endosperms were ground with a mortar and pestle under liquid nitrogen and with a Cyclotec™ 1093 Sample Mill (FOSS, Hillerød, Danemark), respectively. All samples were lyophilized and around 20 mg dry weight (DW) of each sample were placed in 2 ml Safelock Eppendorf tubes (Eppendorf AG, Hamburg, Germany). All analysis steps, including extraction, derivatization, analysis, and data processing, were adapted from the original protocol described by Fiehn (2008) and following the procedure described by Avila-Ospina et al. (2017). The extraction solvent was prepared by mixing water:acetonitrile:isopropanol at the volume ratio 2:3:3 allowing to extract metabolites with a broad range of polarities. For derivatization step, N-methyl-N-trimethylsilyl-trifluoroacetamide (MSTFA; Sigma-Aldrich) was used in silylation procedure of metabolites. Samples were

analyzed on an Agilent 7890A gas chromatograph coupled to an Agilent 5975C mass spectrometer. Raw Agilent datafiles were converted in NetCDF format and analyzed with AMDIS (Automated Mass Deconvolution and Identification System; http://chemdata.nist.gov/mass-spc/amdis/). A home retention indices/mass spectra library built from the NIST, Golm, and Fiehn databases and standard compounds was used for metabolites identification. Peak areas were then determined using the QuanLynx software (Waters, Milford, USA) after conversion of the NetCDF file in MassLynx format.

Data Analysis Using R and MapMan

Figures were created using R version 3.6.3 (R Core Team, 2020; https://www.R-project.org/) using the version 1.3.0 of the tidyverse package suite (Wickham et al., 2019), the version 1.1 of the RColorBrewer (by Erich Neuwirth, 2014; https://CRAN.R-project.org/package=RColorBrewer), the version 1.0.0 of the patchwork library (by Thomas Lin Pedersen, 2019; https://cran.r-project.org/web/packages/patchwork/index.html). **Supplementary Figure S3** was created using MapMap 3.6 (Schwacke et al., 2019) Code and data for **Figures 1**, **2**, **4** can be found on GitHub https://github.com/mgalland/ (https://github.com/mgalland/rice-germination).

DATA AVAILABILITY STATEMENT

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found below: Transcriptomic data are available on the National Center for Biotechnology Information (NCBI) BioProject database under accession number GSE43780; Proteomic data are available on ProticDB, http://proteus.moulon.inra.fr/w2dpage/proticdb/angular/#/projects/163.

AUTHOR CONTRIBUTIONS

MG, IL, and LR designed and performed the experimental work. GCl completed the metabolome analyses. SB and SH performed the transcriptomic analysis while BV gathered the proteomic data. GCu and BC were involved in the preparation of samples. NS, MG, and LR were involved in data analysis and wrote the manuscript. All authors contributed to the article and approved the submitted version.

FUNDING

This work was supported by the French Ministry of Industry (FUI, NUTRICE, agreement # 092906334). NS received funding from the European Union's Horizon 2020 research and innovation program under the Marie Skłodowska-Curie Grant Agreement No. [846387].

ACKNOWLEDGMENTS

This work has benefited from the support of IJPB's Plant Observatory technological platforms. The IJPB benefits from the support of Saclay Plant Sciences-SPS (ANR-17-EUR-0007).

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpls.2022.867263/full#supplementary-material

Supplementary Figure S1 | Summary of main results obtained by transcriptomic, proteomic, and metabolomic approaches performed on rice dry embryo and endosperm during germination.

Supplementary Figure S2 | Number of significant changes in metabolome during germination. **(A)** Embryo metabolites; **(B)** Endosperm metabolites. A one-way ANOVA was performed on each metabolite during the germination time course and significant ($p < 0.05$) metabolites were selected. The number of metabolites significantly up- (positive values) or down-accumulated (negative values) is shown for each pair of successive time points (*e.g.,* 0 HAI vs. 4 HAI, 4 HAI vs. 8 HAI, etc.) and for each tissue (embryo or endosperm).

Supplementary Figure S3 | Seed-specific overview of differentially accumulated proteins. MapMan overview of differentially accumulated proteins between 0 and 24 HAI in the embryo or endosperm with a preferential abundance in the endosperm or embryo. The 261 embryo and 96 endosperm proteins with measurable and significant ($p < 0.05$) $\log_2$ ratios were displayed on a seed-specific visualization (Joosen et al., 2011). A total of 177 (embryo) and 71 (endosperm) proteins are visible. Red and blue colors represent proteins favorably accumulated at 24 HAI and 0 HAI. respectively. Only proteins present in MapMan rice MSU7 mapping file are displayed.

Supplementary Figure S4 | Correlations between mRNA and protein accumulation profiles. Out of 1,659 and 597 embryo and endosperm genes for which a log2 fold change could be calculated between 24 and 0 HAI at both the transcript and protein level, the global Spearman correlation coefficient score was calculated. Both correlation coefficient and the number of genes are displayed for the embryo (left panel) and the endosperm (right panel).

Supplementary Figure S5 | Phylogenetic tree of Arabidopsis and rice GID1 proteins. The tree was built based on amino acid sequence similarities from the 40 annotated rice GID1 and the three Arabidopsis GID1 proteins (framed square). The red circle indicates the rice *GID1* gene mostly studied in the literature. The blue triangle indicates the rice GID1 identified in germinated embryos in the present work. This figure was generated using COBALT: Multiple Alignment Tool (http://www.ncbi.nlm.nih.gov/tools/cobalt/cobalt.cgi?CMD=Web).

Supplementary Figure S6 | Amino acid profiles in the embryo and endosperm during germination. The normalized median ($n = 3$) amino acid abundance was calculated for each time point and plotted on a relative (AU) scale for each tissue.

Supplementary Figure S7 | The aspartate super pathway in rice embryo during germination. Transcript abundances are represented in squares for each locus. Colored squares indicate significant changes ($p < 0.01$, red: up accumulated, blue: down accumulated) for comparison between 2 times of imbibition: 0 vs. S5, 4 vs. 8, 8 vs. 12, 12 vs. 16, and 16 vs. 24 HAI. Up and down accumulated metabolites during germination are displayed, respectively, in red and blue boxes. The same color code is applied for proteins present in the total embryo proteome. AHB, acetohydroxybutyric acid, BCCAT: branched-chain amino acid aminotransferase, DHMV: 2,3-dihyroxy-3-methylvalerate.

Supplementary Table S1 | Transcriptome of the germinating embryo and endosperm.

Supplementary Table S1A | Embryo germination transcriptome data. According to the MAS5 algorithm, we detected 27,503 Affymetrix probes called "present" in at least one imbibition time and in at least two replicates (out of three) during embryo germination. Affymetrix control probes and locus non-specific probesets were removed from the analysis yielding a final number of 22,343 "present" probes corresponding to 18,389 unique rice loci. Then, the GCRMA normalized abundance of the embryo germination CEL files was retrieved for these 22,343 probes. Probe: identifier of the Affymetrix probeset, MSU_id: locus number according to the RGAP 7th annotation, Gene_model: representative transcript for the locus, MSU7_annotation: gene function according to the RGAP 7th annotation, Uniprot: protein description according to Uniprot (accessed 17 November 2012), P_EX: presence (P) or absence (A) of the probe in the three replicates after X hours of imbibition, Embryo_specific: "TRUE" indicates genes

specifically detected in the embryo or "FALSE" if detected in both tissues (embryo and endosperm), $E_X\_R_Y$: GCRMA normalized probe intensity across the germination time course (X: hours after imbibition, Y: replicate number). The results of the differential analysis with log2 ratios (EYvsEX, Y > X) and associated Bonferroni-corrected *p*-value are indicated.

Supplementary Table S1B | Endosperm germination transcriptome data. According to the MAS5 algorithm, we detected 24,695 Affymetrix probes called "present" in at least one imbibition time and in at least two replicates (out of three) during embryo germination. Affymetrix control probes and locus non-specific probesets were removed from the analysis yielding a final number of 22,052 "present" probes corresponding to 16,657 unique rice loci. Then, the GCRMA normalized abundance of the embryo germination CEL files was retrieved for these 20,052 probes. Probe: identifier of the Affymetrix probeset, MSU_id: locus number according to the RGAP 7th annotation, Gene_model: representative transcript for the locus, MSU7_annotation: gene function according to the RGAP 7th annotation, Uniprot: protein description according to Uniprot (accessed 17 November 2012), P_AX: presence (P) or absence (A) of the probe in the three replicates after X hours of imbibition, Endosperm_specific: "TRUE" indicates genes specifically detected in the embryo or "FALSE" if detected in both tissues (embryo and endosperm), AX_RY: GCRMA normalized probe intensity across the germination time course (X: hours after imbibition, Y: replicate number).

Supplementary Table S2 | Differentially accumulated transcripts in the embryo and endosperm during rice seed germination.

Supplementary Table S2A | A total of 7,665 up-accumulated transcripts in embryo during rice germination. The results of the differential analysis with log2 ratios (EYvsEX, Y > X) and associated Bonferroni-corrected *p*-value are indicated.

Supplementary Table S2B | A total of 2,279 down-accumulated transcripts in embryo during rice germination. The results of the differential analysis with log2 ratios (EYvsEX, Y > X) and associated Bonferroni-corrected *p*-value are indicated.

Supplementary Table S2C | A total of 5,926 up-accumulated transcripts in endosperm during rice germination. The results of the differential analysis with log2 ratios (AYvsAX, Y > X) and associated Bonferroni-corrected *p*-value are indicated.

Supplementary Table S2D | A total of 2,103 down-accumulated transcripts in endosperm during rice germination. The results of the differential analysis with log2 ratios (AYvsAX, Y > X) and associated Bonferroni-corrected *p*-value are indicated.

Supplementary Table S3 | Metabolome of embryo and endosperm during rice seed germination.

Supplementary Table S3A | Single and averaged relative metabolite abundance data for the 121 detected metabolites in germinating embryos.

Supplementary Table S3B | Comparison of relative metabolite abundance detected in rice embryo during germination time course.

Supplementary Table S3C | Single and averaged relative metabolite abundance data for the 121 detected metabolites in endosperm during germination.

Supplementary Table S3D | Comparison of relative metabolite abundance detected in rice endosperm during germination time course.

Supplementary Table S4 | Shotgun proteomic data of the germinating embryo and endosperm.

Supplementary Table S4A | The 2,315 proteins identified in the dry (E0) or 24 h-imbibed embryo (E24) from 25 μg of proteins. The peptides identified in the embryo were filtered to keep only the peptides specific to a single protein and a protein was considered as "present" if detected by at least one specific peptide in one biological replicate. These proteins correspond to 2,306 unique rice loci (Rice Genome Annotation Project version 7.0). The specificity of the protein is indicated (e.g., only present in the dry embryo).

Supplementary Table S4B | Quantitative analysis of the 1,973 proteins common between the dry (E0) and 24h-imbibed embryo (E24) from 25 μg of proteins. The peptides identified in the embryo (ST1A) were filtered to keep only the peptides specific of a single protein and if the peptides were detected in at least 5 of the 6 samples (2 imbibition times x 3 replicates). MSU7_id: rice locus according to the RGAP 7.0 annotation, variance: result of the Fisher's test indicating if for each protein the variance is equal or different with $p < 0.05$, t_test: Student's *t*-test for each protein indicating significatively different mean with $p < 0.05$, $\log_2$ratio_E24vsE0: the $\log_2$ ratio between the protein mean abundance in E24 and E0 (a positive ratio significates that the protein is more abundant in 24 h-imbibed embryos as compared with dry embryos).

Supplementary Table S4C | The 832 proteins identified in the dry (A0) or 24 h-imbibed endosperm (A24) from 25 μg of proteins. The peptides identified in the endosperm (ST3D) were filtered to keep only the peptides specific of a single protein and a protein was considered as "present" if detected by at least one specific peptide in one biological replicate. These proteins correspond to 829 unique rice loci (Rice Genome Annotation Project version 7.0). The specificity of the protein is indicated (e.g., only present in the dry endosperm).

Supplementary Table S4D | Quantitative analysis of the 721 proteins common between the dry (A0) and 24h-imbibed endosperm (A24) from 25 μg of proteins. The peptides identified in the endosperm (ST1C) were filtered to keep only the peptides specific of a single protein and if the peptides were detected in at least 5 of the 6 samples (2 imbibition time x 3 replicates). MSU7_id: rice locus according to the RGAP 7.0 annotation, variance: result of the Fisher's test indicating if for each protein the variance is equal or different with $p < 0.05$, t_test: Student's *t*-test for each protein indicating significant different mean with $p < 0.05$, $\log_2$ratio_A24vsA0: the $\log_2$ ratio between the protein mean abundance in A24 and A0 (a positive ratio significates that the protein is more abundant in 24 h-imbibed endosperm as compared with dry endosperm).

Supplementary Table 5 | Transcript correspondence with differentially accumulated proteins in the embryo and endosperm during rice germination.

Supplementary Table S5A | The 265 significantly up-accumulated proteins in germinated rice embryos. Selection criteria: $p < 0.05$, $\log_2$ratio_E24vsE0 > 0. MSU7_id: rice locus according to the RGAP 7.0 annotation, Ontological class: obtained by a classification from agriGO analysis tool, t_test: Student's *t*-test for each protein indicating significant different mean with $p < 0.05$, $\log_2$ratio_E24vsE0: the $\log_2$ ratio between the protein mean abundance in E24 and E0.

Supplementary Table S5B | The 229 transcripts corresponding to 229 up-accumulated proteins in rice embryo during germination. According to the MAS5 algorithm, we detected 27,503 Affymetrix probes called "present" in at least one imbibition time and in at least two replicates (out of three) during embryo germination. Affymetrix control probes and locus non-specific probesets were removed from the analysis yielding a final number of 22,343 "present" probes corresponding to 18,389 unique rice loci. Then, the GCRMA normalized abundance of the embryo germination CEL files was retrieved for these 22,343 probes. mRNA GCRMA abundance were normalized for each probe intensity by its maximum intensity over germination to get a value between 0 and 1. MSU_id: locus number according to the RGAP 7th annotation, E_X: GCRMA normalized probe intensity across the germination time course with X standing for hours after imbibition (e.g., E_{24} stands for 24 HAI).

Supplementary Table S5C | The 152 significantly down-accumulated proteins in germinated rice embryos. Selection criteria: $p < 0.05$, $\log_2$ratio_E24vsE0 < 0. MSU7_id: rice locus according to the RGAP 7.0 annotation, Ontological class: obtained by a classification from agriGO analysis tool, t_test: Student's *t*-test for each protein indicating significant different mean with $p < 0.05$, $\log_2$ratio_E24vsE0: the $\log_2$ ratio between the protein mean abundance in E24 and E0.

Supplementary Table S5D | The 133 transcripts corresponding to 133 down-accumulated proteins in rice embryo during germination. According to the MAS5 algorithm, we detected 27,503 Affymetrix probes called "present" in at least one imbibition time and in at least two replicates (out of three) during embryo germination. Affymetrix control probes and locus non-specific probe sets were removed from the analysis yielding a final number of 22,343 "present" probes corresponding to 18,389 unique rice loci. Then, the GCRMA normalized abundance of the embryo germination CEL files was retrieved for these 22,343 probes. mRNA GCRMA abundance were normalized for each probe intensity by its maximum intensity over germination to get a value between 0 and 1. MSU_id: locus number according to the RGAP 7th annotation, E_X: GCRMA normalized probe intensity across the germination time course with X standing for hours after imbibition (e.g., E_{24} stands for 24 HAI).

Supplementary Table S5E | The 11 significantly up-accumulated proteins in germinated rice endosperm. Selection criteria: $p < 0.05$, $\log_2$ratio_A24vsA0 > 0. MSU7_id: rice locus according to the RGAP 7.0 annotation, Ontological class: obtained by a classification from agriGO analysis tool, t_test: Student's *t*-test for each protein indicating significant different mean with $p < 0.05$, $\log_2$ratio_A24vsA0: the $\log_2$ ratio between the protein mean abundance in A24 and A0.

Supplementary Table S5F | The 11 transcripts corresponding to 11 up-accumulated proteins in rice endosperm during germination. According to the MAS5 algorithm, we detected 24,695 Affymetrix probes called "present" in at least one imbibition time and in at least two replicates (out of three) during endosperm germination. Affymetrix control probes and locus non-specific probe sets were removed from the analysis yielding a final number of 22,052 "present" probes corresponding to 16,657 unique rice loci. Then, the GCRMA normalized abundance of the endosperm germination CEL files was retrieved for these 20,052 probes. mRNA GCRMA abundance was normalized for each probe intensity by its maximum intensity over germination to get a value between 0 and 1. MSU_id: locus number according to the RGAP 7th annotation, A_X: GCRMA normalized probe intensity across the germination time course with X standing for hours after imbibition (e.g., A_{24} stands for 24 HAI).

Supplementary Table S5G | The 109 significantly down-accumulated proteins in germinated rice endosperm. Selection criteria: $p < 0.05$, $\log_2$ratio_A24vsA0 < 0. MSU7_id: rice locus according to the RGAP 7.0 annotation, Ontological class: obtained by a classification from agriGO analysis tool, t_test: Student's *t*-test for each protein indicating significant different mean with $p < 0.05$, $\log_2$ratio_A24vsA0: the $\log_2$ ratio between the protein mean abundance in A24 and A0.

Supplementary Table S5H | The 96 transcripts corresponding to 96 down-accumulated proteins in rice endosperm during germination. According to the MAS5 algorithm, we detected 24,695 Affymetrix probes called "present" in at least one imbibition time and in at least two replicates (out of three) during endosperm germination. Affymetrix control probes and locus non-specific probesets were removed from the analysis yielding a final number of 22,052 "present" probes corresponding to 16,657 unique rice loci. Then, the GCRMA normalized abundance of the endosperm germination CEL files was retrieved for these 20,052 probes. mRNA GCRMA abundance were normalized for each probe intensity by its maximum intensity over germination to get a value between 0 and 1. MSU_id: locus number according to the RGAP 7th annotation, A_X: GCRMA normalized probe intensity across the germination time course with X standing for hours after imbibition (e.g., A_{24} stands for 24 HAI).

REFERENCES

Amir, R. (2010). Current understanding of the factors regulating methionine content in vegetative tissues of higher plants. *Amino Acids* 39, 917–931. doi: 10.1007/s00726-010-0482-x

An, L., Tao, Y., Chen, H., He, M., Xiao, F., Li, G., et al. (2020). Embryo-endosperm interaction and its agronomic relevance to rice quality. *Front. Plant Sci.* 11, 587641. doi: 10.3389/fpls.2020.587641

Arc, E., Sechet, J., Corbineau, F., Rajjou, L., and Marion-Poll, A. (2013). ABA crosstalk with ethylene and nitric oxide in seed dormancy and germination. *Front. Plant Sci.* 4, 63. doi: 10.3389/fpls.2013.00063

Ariizumi, T., Murase, K., Sun, T., and Steber, C. M. (2008). Proteolysis-independent downregulation of DELLA repression in *Arabidopsis* by the gibberellin receptor GIBBERELLIN INSENSITIVE DWARF1. *Plant Cell* 20, 2447–2459. doi: 10.1105/tpc.108.058487

Avila-Ospina, L., Clément, G., and Masclaux-Daubresse, C. (2017). Metabolite profiling for leaf senescence in barley reveals decreases in amino acids and glycolysis intermediates. *Agronomy* 7, 15. doi: 10.3390/agronomy7010015

Bewley, J. D. (1997). Seed germination and dormancy. *Plant Cell* 9, 1055–1066. doi: 10.1105/tpc.9.7.1055

Bi, X., Ren, J., and Goss, D. J. (2000). Wheat germ translation initiation factor eIF4B affects eIF4A and eIFiso4F helicase activity by increasing the ATP binding affinity of eIF4A. *Biochemistry* 39, 5758–5765. doi: 10.1021/bi992322p

Bissoli, G., Niñoles, R., Fresquet, S., Palombieri, S., Bueso, E., Rubio, L., et al. (2012). Peptidyl-prolyl cis-trans isomerase ROF2 modulates intracellular pH homeostasis in Arabidopsis: ROF2 modulates pH homeostasis. *Plant J.* 70, 704–716. doi: 10.1111/j.1365-313X.2012.04921.x

Bonhomme, L., Valot, B., Tardieu, F., and Zivy, M. (2012). Phosphoproteome dynamics upon changes in plant water status reveal early events associated with rapid growth adjustment in maize leaves. *Mol. Cell. Proteomics* 11, 957–972. doi: 10.1074/mcp.M111.015867

Bradford, K. J. (1990). A water relations analysis of seed germination rates. *Plant Physiol.* 94, 840–849. doi: 10.1104/pp.94.2.840

Bradford, M. M. (1976). A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding. *Anal. Biochem.* 72, 248–254. doi: 10.1016/0003-2697(76)90527-3

Broekgaarden, C., Caarls, L., Vos, I. A., Pieterse, C. M. J., and Van Wees, S. C. M. (2015). Ethylene: traffic controller on hormonal crossroads to defense. *Plant Physiol.* 169, 2371–2379. doi: 10.1104/pp.15.01020

Browning, K. S., and Bailey-Serres, J. (2015). Mechanism of cytoplasmic mRNA translation. *Arabidopsis Book* 13, e0176. doi: 10.1199/tab.0176

Chang, G.-G., and Tong, L. (2003). Structure and function of malic enzymes, a new class of oxidative decarboxylases. *Biochemistry* 42, 12721–12733. doi: 10.1021/bi035251+

Chen, Q., Wang, B., Ding, H., Zhang, J., and Li, S. (2019). The role of NADP-malic enzyme in plants under stress. *Plant Sci.* 281, 206–212. doi: 10.1016/j.plantsci.2019.01.010

Czarna, M., Kolodziejczak, M., and Janska, H. (2016). Mitochondrial proteome studies in seeds during germination. *Proteomes* 4, 19. doi: 10.3390/proteomes4020019

Das, P., Lakra, N., Nutan, K. K., Singla-Pareek, S. L., and Pareek, A. (2019). A unique bZIP transcription factor imparting multiple stress tolerance in rice. *Rice* 12, 58. doi: 10.1186/s12284-019-0316-8

Di Mauro, M. F., Iglesias, M. J., Arce, D. P., Valle, E. M., Arnold, R. B., Tsuda, K., et al. (2012). MBF1s regulate ABA-dependent germination of Arabidopsis seeds. *Plant Signal. Behav.* 7, 188–192. doi: 10.4161/psb.18843

Dubois, M., Van den Broeck, L., and Inzé, D. (2018). The pivotal role of ethylene in plant growth. *Trends Plant Sci.* 23, 311–323. doi: 10.1016/j.tplants.2018.01.003

Fait, A., Angelovici, R., Less, H., Ohad, I., Urbanczyk-Wochniak, E., Fernie, A. R., et al. (2006). Arabidopsis seed development and germination is associated with temporally distinct metabolic switches. *Plant Physiol.* 142, 839–854. doi: 10.1104/pp.106.086694

Fernie, A. R., Carrari, F., and Sweetlove, L. J. (2004). Respiratory metabolism: glycolysis, the TCA cycle and mitochondrial electron transport. *Curr. Opin. Plant Biol.* 7, 254–261. doi: 10.1016/j.pbi.2004.03.007

Fiehn, O. (2008). Extending the breadth of metabolite profiling by gas chromatography coupled to mass spectrometry. *Trends Anal. Chem.* 27, 261–269. doi: 10.1016/j.trac.2008.01.007

Finch-Savage, W. E., and Bassel, G. W. (2016). Seed vigour and crop establishment: extending performance beyond adaptation. *J. Exp. Bot.* 67, 567–591. doi: 10.1093/jxb/erv490

Galili, G. (2011). The aspartate-family pathway of plants: Linking production of essential amino acids with energy and stress regulation. *Plant Signal. Behav.* 6, 192–195. doi: 10.4161/psb.6.2.14425

Galland, M., He, D., Lounifi, I., Arc, E., Clément, G., Balzergue, S., et al. (2017). An integrated "multi-omics" comparison of embryo and endosperm tissue-specific features and their impact on rice seed quality. *Front. Plant Sci.* 8, 1984. doi: 10.3389/fpls.2017.01984

Galland, M., Huguet, R., Arc, E., Cueff, G., Job, D., and Rajjou, L. (2014). Dynamic proteomics emphasizes the importance of selective mRNA translation and protein turnover during Arabidopsis seed germination. *Mol. Cell. Proteomics* 13, 252–268. doi: 10.1074/mcp.M113.032227

Gallardo, K., Job, C., Groot, S. P. C., Puype, M., Demol, H., Vandekerckhove, J., et al. (2002). Importance of methionine biosynthesis for Arabidopsis seed germination and seedling growth. *Physiol. Plant.* 116, 238–247. doi: 10.1034/j.1399-3054.2002.1160214.x

Guerin, J. R., Lance, R. C. M., and Wallace, W. (1992). Release and activation of barley beta-amylase by malt endopeptidases. *J. Cereal Sci.* 15, 5–14. doi: 10.1016/S0733-5210(09)80052-0

Hauvermale, A. L., Ariizumi, T., and Steber, C. M. (2014). The roles of the GA receptors *GID1a*, *GID1b*, and *GID1c* in *sly1* -independent GA signaling. *Plant Signal. Behav.* 9, e28030. doi: 10.4161/psb.28030

He, D., Han, C., Yao, J., Shen, S., and Yang, P. (2011). Constructing the metabolic and regulatory pathways in germinating rice seeds through proteomic approach. *Proteomics* 11, 2693–2713. doi: 10.1002/pmic.201000598

He, D., and Yang, P. (2013). Proteomics of rice seed germination. *Front. Plant Sci.* 4, 246. doi: 10.3389/fpls.2013.00246

Hedden, P. (2020). The current status of research on gibberellin biosynthesis. *Plant Cell Physiol.* 61, 1832–1849. doi: 10.1093/pcp/pcaa092

Holman, T. J., Jones, P. D., Russell, L., Medhurst, A., Ubeda Tomas, S., Talloji, P., et al. (2009). The N-end rule pathway promotes seed germination and establishment through removal of ABA sensitivity in Arabidopsis. *Proc. Natl. Acad. Sci. U. S. A.* 106, 4549–4554. doi: 10.1073/pnas.0810280106

Horigane, A. K., Takahashi, H., Maruyama, S., Ohtsubo, K., and Yoshida, M. (2006). Water penetration into rice grains during soaking observed by gradient echo magnetic resonance imaging. *J. Cereal Sci.* 44, 307–316. doi: 10.1016/j.jcs.2006.07.014

Howell, K. A., Cheng, K., Murcha, M. W., Jenkin, L. E., Millar, A. H., and Whelan, J. (2007). Oxygen initiation of respiration and mitochondrial biogenesis in rice. *J. Biol. Chem.* 282, 15619–15631. doi: 10.1074/jbc.M609866200

Howell, K. A., Millar, A. H., and Whelan, J. (2006). Ordered assembly of mitochondria during rice germination begins with promitochondrial structures rich in components of the protein import apparatus. *Plant Mol. Biol.* 60, 201–223. doi: 10.1007/s11103-005-3688-7

Howell, K. A., Narsai, R., Carroll, A., Ivanova, A., Lohse, M., Usadel, B., et al. (2009). Mapping metabolic and transcript temporal switches during germination in rice highlights specific transcription factors and the role of RNA instability in the germination process. *Plant Physiol.* 149, 961–980. doi: 10.1104/pp.108.129874

Huang, X.-Y., Li, M., Luo, R., Zhao, F.-J., and Salt, D. E. (2019). Epigenetic regulation of sulfur homeostasis in plants. *J. Exp. Bot.* 70, 4171–4182. doi: 10.1093/jxb/erz218

Huber, W., Carey, V. J., Gentleman, R., Anders, S., Carlson, M., Carvalho, B. S., et al. (2015). Orchestrating high-throughput genomic analysis with Bioconductor. *Nat. Methods* 12, 115–121. doi: 10.1038/nmeth.3252

Jaimes-Miranda, F., and Chávez Montes, R. A. (2020). The plant MBF1 protein family: a bridge between stress and transcription. *J. Exp. Bot.* 71, 1782–1791. doi: 10.1093/jxb/erz525

Joosen, R. V. L., Ligterink, W., Dekkers, B. J. W., and Hilhorst, H. W. M. (2011). Visualization of molecular processes associated with seed dormancy and germination using MapMan. *Seed Sci. Res.* 21, 143–152. doi: 10.1017/S0960258510000449

Ju, C., Kong, D., Lee, Y., Ge, G., Song, Y., Liu, J., et al. (2020). Methionine synthase 1 provides methionine for activation of the GLR3.5 Ca2+ channel and regulation of germination in Arabidopsis. *J. Exp. Bot.* 71, 178–187. doi: 10.1093/jxb/erz431

Kawahara, Y., de la Bastide, M., Hamilton, J. P., Kanamori, H., McCombie, W. R., Ouyang, S., et al. (2013). Improvement of the *Oryza sativa* Nipponbare reference genome using next generation sequence and optical map data. *Rice* 6, 4. doi: 10.1186/1939-8433-6-4

Kim, S. T., Wang, Y., Kang, S. Y., Kim, S. G., Rakwal, R., Kim, Y. C., et al. (2009). Developing rice embryo proteomics reveals essential role for embryonic proteins in regulation of seed germination. *J. Proteome Res.* 8, 3598–3605. doi: 10.1021/pr900358s

Lakra, N., Nutan, K. K., Das, P., Anwar, K., Singla-Pareek, S. L., and Pareek, A. (2015). A nuclear-localized histone-gene binding protein from rice (OsHBP1b) functions in salinity and drought stress tolerance by maintaining chlorophyll content and improving the antioxidant machinery. *J. Plant Physiol.* 176, 36–46. doi: 10.1016/j.jplph.2014.11.005

Langella, O., Valot, B., Balliau, T., Blein-Nicolas, M., Bonhomme, L., and Zivy, M. (2017). X!TandemPipeline: A tool to manage sequence redundancy for protein inference and phosphosite identification. *J. Proteome Res.* 16, 494–503. doi: 10.1021/acs.jproteome.6b00632

Laxa, M., König, J., Dietz, K.-J., and Kandlbinder, A. (2007). Role of the cysteine residues in *Arabidopsis thaliana* cyclophilin CYP20-3 in peptidyl-prolyl *cis* – *trans* isomerase and redox-related functions. *Biochem. J.* 401, 287–297. doi: 10.1042/BJ20061092

Leydecker, M. T., Moureaux, T., Kraepiel, Y., Schnorr, K., and Caboche, M. (1995). Molybdenum cofactor mutants, specifically impaired in xanthine dehydrogenase activity and abscisic acid biosynthesis, simultaneously overexpress nitrate reductase. *Plant Physiol.* 107, 1427–1431. doi: 10.1104/pp.107.4.1427

Li, S., Mhamdi, A., Clement, C., Jolivet, Y., and Noctor, G. (2013). Analysis of knockout mutants suggests that Arabidopsis NADP-MALIC ENZYME2 does not play an essential role in responses to oxidative stress of intracellular or extracellular origin. *J. Exp. Bot.* 64, 3605–3614. doi: 10.1093/jxb/ert194

Li, W., Xu, L., Bai, X., and Xing, Y. (2011). Quantitative trait loci for seed dormancy in rice. *Euphytica* 178, 427–435. doi: 10.1007/s10681-010-0327-4

Lim, S. D., Lee, C., and Jang, C. S. (2014). The rice RING E3 ligase, OsCTR1, inhibits trafficking to the chloroplasts of OsCP12 and OsRP1, and its overexpression confers drought tolerance in *Arabidopsis*: OsCTR1 in trafficking inhibition of interactors. *Plant Cell Environ.* 37, 1097–1113. doi: 10.1111/pce.12219

Liu, S., Cheng, Y., Zhang, X., Guan, Q., Nishiuchi, S., Hase, K., et al. (2007). Expression of an NADP-malic enzyme gene in rice (*Oryza sativa* L.) is induced by environmental stresses; over-expression of the gene in Arabidopsis confers salt and osmotic stress tolerance. *Plant Mol. Biol.* 64, 49–58. doi: 10.1007/s11103-007-9133-3

Lu, H., Chandrasekar, B., Oeljeklaus, J., Misas-Villamil, J. C., Wang, Z., Shindo, T., et al. (2015). Subfamily-specific fluorescent probes for cysteine proteases display dynamic protease activities during seed germination. *Plant Physiol.* 168, 1462–1475. doi: 10.1104/pp.114.254466

Magneschi, L., and Perata, P. (2009). Rice germination and seedling growth in the absence of oxygen. *Ann. Bot.* 103, 181–196. doi: 10.1093/aob/mcn121

Marshall, R. S., and Vierstra, R. D. (2019). Dynamic regulation of the 26S proteasome: from synthesis to degradation. *Front. Mol. Biosci.* 6, 40. doi: 10.3389/fmolb.2019.00040

Muench, D. G., Zhang, C., and Dahodwala, M. (2012). Control of cytoplasmic translation in plants: control of cytoplasmic translation. *WIREs RNA* 3, 178–194. doi: 10.1002/wrna.1104

Narsai, R., Howell, K. A., Carroll, A., Ivanova, A., Millar, A. H., and Whelan, J. (2009). Defining core metabolic and transcriptomic responses to oxygen availability in rice embryos and young seedlings. *Plant Physiol.* 151, 306–322. doi: 10.1104/pp.109.142026

Narsai, R., and Whelan, J. (2013). How unique is the low oxygen response? An analysis of the anaerobic response during germination and comparison with abiotic stress in rice and Arabidopsis. *Front. Plant Sci.* 4, 349. doi: 10.3389/fpls.2013.00349

Olszewski, N., Sun, T., and Gubler, F. (2002). Gibberellin signaling: biosynthesis, catabolism, and response pathways. *Plant Cell* 14, S61–S80. doi: 10.1105/tpc.010476

Penninckx, I. A. M. A., Thomma, B. P. H. J., Buchala, A., Métraux, J.-P., and Broekaert, W. F. (1998). Concomitant activation of jasmonate and ethylene response pathways is required for induction of a plant defensin gene in Arabidopsis. *Plant Cell* 10, 2103–2113. doi: 10.1105/tpc.10.12.2103

Qin, D., Wang, F., Geng, X., Zhang, L., Yao, Y., Ni, Z., et al. (2015). Overexpression of heat stress-responsive TaMBF1c, a wheat (*Triticum aestivum* L.) multiprotein bridging factor, confers heat tolerance in both yeast and rice. *Plant Mol. Biol.* 87, 31–45. doi: 10.1007/s11103-014-0259-9

Rahikainen, M., Alegre, S., Trotta, A., Pascual, J., and Kangasjärvi, S. (2018). Trans-methylation reactions in plants: focus on the activated methyl cycle. *Physiol. Plantarum* 162, 162–176. doi: 10.1111/ppl.12619

Rajjou, L., Duval, M., Gallardo, K., Catusse, J., Bally, J., Job, C., et al. (2012). Seed germination and vigor. *Annu. Rev. Plant Biol.* 63, 507–533. doi: 10.1146/annurev-arplant-042811-105550

Rajjou, L., Gallardo, K., Debeaujon, I., Vandekerckhove, J., Job, C., and Job, D. (2004). The effect of α-amanitin on the Arabidopsis seed proteome highlights the distinct roles of stored and neosynthesized mRNAs during germination. *Plant Physiol.* 134, 1598–1613. doi: 10.1104/pp.103.036293

Ravanel, S., Gakiere, B., Job, D., and Douce, R. (1998). The specific features of methionine biosynthesis and metabolism in plants. *Proc. Natl. Acad. Sci. U. S. A.* 95, 7805–7812. doi: 10.1073/pnas.95.13.7805

Reed, R. C., Bradford, K. J., and Khanday, I. (2022). Seed germination and vigor: ensuring crop sustainability in a changing climate. *Heredity*. doi: 10.1038/s41437-022-00497-2. [Epub ahead of print].

Regelmann, J., Schüle, T., Josupeit, F. S., Horak, J., Rose, M., Entian, K.-D., et al. (2003). Catabolite degradation of fructose-1,6-bisphosphatase in the yeast *Saccharomyces cerevisiae*: a Genome-wide screen identifies eight novel *GID* genes and indicates the existence of two degradation pathways. *MBoC* 14, 1652–1663. doi: 10.1091/mbc.e02-08-0456

Sajeev, N., Bai, B., and Bentsink, L. (2019). Seeds: a unique system to study translational regulation. *Trends Plant Sci.* 24, 487–495. doi: 10.1016/j.tplants.2019.03.011

Sakai, H., Lee, S. S., Tanaka, T., Numa, H., Kim, J., Kawahara, Y., et al. (2013). Rice annotation project database (RAP-DB): an integrative and interactive database for rice genomics. *Plant Cell Physiol.* 54, e6–e6. doi: 10.1093/pcp/pcs183

Sano, N., Permana, H., Kumada, R., Shinozaki, Y., Tanabata, T., Yamada, T., et al. (2012). Proteomic analysis of embryonic proteins synthesized from long-lived mRNAs during germination of rice seeds. *Plant Cell Physiol.* 53, 687–698. doi: 10.1093/pcp/pcs024

Sano, N., Rajjou, L., and North, H. M. (2020). Lost in translation: physiological roles of stored mRNAs in seed germination. *Plants* 9, 347. doi: 10.3390/plants9030347

Sano, N., Takebayashi, Y., To, A., Mhiri, C., Rajjou, L., Nakagami, H., et al. (2019). Shotgun proteomic analysis highlights the roles of long-lived mRNAs and *de novo* transcribed mRNAs in rice seeds upon imbibition. *Plant Cell Physiol.* 60, 2584–2596. doi: 10.1093/pcp/pcz152

Sauter, M., Moffatt, B., Saechao, M. C., Hell, R., and Wirtz, M. (2013). Methionine salvage and *S* -adenosylmethionine: essential links between sulfur, ethylene and polyamine biosynthesis. *Biochem. J.* 451, 145–154. doi: 10.1042/BJ20121744

Schmitt, M. R., Skadsen, R. W., and Budde, A. D. (2013). Protein mobilization and malting-specific proteinase expression during barley germination. *J. Cereal Sci.* 58, 324–332. doi: 10.1016/j.jcs.2013.05.007

Schwacke, R., Ponce-Soto, G. Y., Krause, K., Bolger, A. M., Arsova, B., Hallab, A., et al. (2019). MapMan4: a refined protein classification and annotation framework applicable to multi-omics data analysis. *Mol. Plant* 12, 879–892. doi: 10.1016/j.molp.2019.01.003

Schwarz, G., and Mendel, R. R. (2006). Molybdenum cofactor biosynthesis and molybdenum enzymes. *Annu. Rev. Plant Biol.* 57, 623–647. doi: 10.1146/annurev.arplant.57.032905.105437

Shimada, A., Ueguchi-Tanaka, M., Nakatsu, T., Nakajima, M., Naoe, Y., Ohmiya, H., et al. (2008). Structural basis for gibberellin recognition by its receptor GID1. *Nature* 456, 520–523. doi: 10.1038/nature07546

Song, S., Wang, G., Wu, H., Fan, X., Liang, L., Zhao, H., et al. (2020). OsMFT2 is involved in the regulation of ABA signaling-mediated seed germination through interacting with OsbZIP23/66/72 in rice. *Plant J.* 103, 532–546. doi: 10.1111/tpj.14748

Sugimoto, K., Takeuchi, Y., Ebana, K., Miyao, A., Hirochika, H., Hara, N., et al. (2010). Molecular cloning of *Sdr4*, a regulator involved in seed dormancy and domestication of rice. *Proc. Natl. Acad. Sci. U. S. A.* 107, 5792–5797. doi: 10.1073/pnas.0911965107

Sun, T. (2011). The molecular mechanism and evolution of the GA–GID1–DELLA signaling module in plants. *Curr. Biol.* 21, R338–R345. doi: 10.1016/j.cub.2011.02.036

Suzuki, N., Bajad, S., Shuman, J., Shulaev, V., and Mittler, R. (2008). The transcriptional co-activator MBF1c is a key regulator of thermotolerance in *Arabidopsis thaliana*. *J. Biol. Chem.* 283, 9269–9275. doi: 10.1074/jbc.M709187200

Suzuki, N., Rizhsky, L., Liang, H., Shuman, J., Shulaev, V., and Mittler, R. (2005). Enhanced tolerance to environmental stress in transgenic plants expressing the transcriptional coactivator multiprotein bridging factor 1c. *Plant Physiol.* 139, 1313–1322. doi: 10.1104/pp.105.070110

Sweetlove, L. J., Beard, K. F. M., Nunes-Nesi, A., Fernie, A. R., and Ratcliffe, R. G. (2010). Not just a circle: flux modes in the plant TCA cycle. *Trends Plant Sci.* 15, 462–470. doi: 10.1016/j.tplants.2010.05.006

Takemaru, K.-I., Li, F.-Q., Ueda, H., and Hirose, S. (1997). Multiprotein bridging factor 1 (MBF1) is an evolutionarily conserved transcriptional coactivator that connects a regulatory factor and TATA element-binding protein. *Proc. Natl. Acad. Sci. U. S. A.* 94, 7251–7256. doi: 10.1073/pnas.94.14.7251

Tan, L., Chen, S., Wang, T., and Dai, S. (2013). Proteomic insights into seed germination in response to environmental factors. *Proteomics* 13, 1850–1870. doi: 10.1002/pmic.201200394

Tantong, S., Pringsulaka, O., Weerawanich, K., Meeprasert, A., Rungrotmongkol, T., Sarnthima, R., et al. (2016). Two novel antimicrobial defensins from rice identified by gene coexpression network analyses. *Peptides* 84, 7–16. doi: 10.1016/j.peptides.2016.07.005

Tomaru, K., Ueda, A., Suzuki, T., Kobayashi, N., Yang, J., Yamamoto, M., et al. (2010). Armadillo repeat containing 8α binds to HRS and promotes HRS interaction with ubiquitinated proteins. *Open Biochem. J.* 4, 1–8. doi: 10.2174/1874091X01004010001

Tomaštíková, E., Cenklová, V., Kohoutová, L., Petrovská, B., Váchová, L., Halada, P., et al. (2012). Interactions of an Arabidopsis RanBPM homologue with LisH-CTLH domain proteins revealed high conservation of CTLH complexes in eukaryotes. *BMC Plant Biol.* 12, 83. doi: 10.1186/1471-2229-12-83

Ueguchi-Tanaka, M., Ashikari, M., Nakajima, M., Itoh, H., Katoh, E., Kobayashi, M., et al. (2005). GIBBERELLIN INSENSITIVE DWARF1 encodes a soluble receptor for gibberellin. *Nature* 437, 693–698. doi: 10.1038/nature04028

Ueguchi-Tanaka, M., Nakajima, M., Motoyuki, A., and Matsuoka, M. (2007). Gibberellin receptor and its role in gibberellin signaling in plants. *Annu. Rev. Plant Biol.* 58, 183–198. doi: 10.1146/annurev.arplant.58.032806.103830

Valot, B., Langella, O., Nano, E., and Zivy, M. (2011). MassChroQ: A versatile tool for mass spectrometry quantification. *Proteomics* 11, 3572–3577. doi: 10.1002/pmic.201100120

van der Hoorn, R. A. L. (2008). Plant proteases: from phenotypes to molecular mechanisms. *Annu. Rev. Plant Biol.* 59, 191–223. doi: 10.1146/annurev.arplant.59.032607.092835

van Dongen, J. T., Gupta, K. J., Ramírez-Aguilar, S. J., Araújo, W. L., Nunes-Nesi, A., and Fernie, A. R. (2011). Regulation of respiration in plants: a role for alternative metabolic pathways. *J. Plant Physiol.* 168, 1434–1443. doi: 10.1016/j.jplph.2010.11.004

Vierstra, R. D. (2009). The ubiquitin—26S proteasome system at the nexus of plant biology. *Nat. Rev. Mol. Cell. Biol.* 10, 385–397. doi: 10.1038/nrm2688

Wang, G., Li, X., Ye, N., Huang, M., Feng, L., Li, H., et al. (2021). *OsTPP1* regulates seed germination through the crosstalk with abscisic acid in rice. *New Phytol.* 230, 1925–1939. doi: 10.1111/nph.17300

Wang, W., Xu, M., Wang, G., and Galili, G. (2018). New insights into the metabolism of aspartate-family amino acids in plant seeds. *Plant Reprod.* 31, 203–211. doi: 10.1007/s00497-018-0322-9

Weerawanich, K., Webster, G., Ma, J. K.-C., Phoolcharoen, W., and Sirikantaramas, S. (2018). Gene expression analysis, subcellular localization, and in planta antimicrobial activity of rice (Oryza sativa L.) defensin 7 and 8. *Plant Physiol. Biochem.* 124, 160–166. doi: 10.1016/j.plaphy.2018.01.011

Weis, B. L., Kovacevic, J., Missbach, S., and Schleiff, E. (2015). Plant-specific features of ribosome biogenesis. *Trends Plant Sci.* 20, 729–740. doi: 10.1016/j.tplants.2015.07.003

Wickham, H., Averick, M., Bryan, J., Chang, W., McGowan, L., François, R., et al. (2019). Welcome to the tidyverse. *JOSS* 4, 1686. doi: 10.21105/joss.01686

Yan, D., Duermeyer, L., Leoveanu, C., and Nambara, E. (2014). The functions of the endosperm during seed germination. *Plant Cell Physiol.* 55, 1521–1533. doi: 10.1093/pcp/pcu089

Yang, P., Li, X., Wang, X., Chen, H., Chen, F., and Shen, S. (2007). Proteomic analysis of rice (*Oryza sativa*) seeds during germination. *Proteomics* 7, 3358–3368. doi: 10.1002/pmic.200700207

Zhang, H., Gannon, L., Jones, P. D., Rundle, C. A., Hassall, K. L., Gibbs, D. J., et al. (2018). Genetic interactions between ABA signalling and the Arg/N-end rule pathway during Arabidopsis seedling establishment. *Sci. Rep.* 8, 15192. doi: 10.1038/s41598-018-33630-5

Zhu, G., Ye, N., and Zhang, J. (2009). Glucose-induced delay of seed germination in rice is mediated by the suppression of ABA catabolism rather than an enhancement of ABA biosynthesis. *Plant Cell Physiol.* 50, 644–651. doi: 10.1093/pcp/pcp022

Zi, J., Zhang, J., Wang, Q., Zhou, B., Zhong, J., Zhang, C., et al. (2013). Stress responsive proteins are actively regulated during rice (*Oryza sativa*) embryogenesis as indicated by quantitative proteomics analysis. *PLoS ONE* 8, e74229. doi: 10.1371/journal.pone.0074229

Conflict of Interest: IL was employed by company MBCC Group.

The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

frontiers | Frontiers in Plant Science

ORIGINAL RESEARCH
published: 27 June 2022
doi: 10.3389/fpls.2022.907042

Identification of Proteases and Protease Inhibitors in Seeds of the Recalcitrant Forest Tree Species *Quercus ilex*

Monica Escandón[1†‡], Ezequiel D. Bigatton[1,2‡], Victor M. Guerrero-Sánchez[1†], Tamara Hernández-Lao[1], Maria-Dolores Rey[1], Jesus V. Jorrín-Novo[1] and Maria Angeles Castillejo[1]**

[1]Agroforestry and Plant Biochemistry, Proteomics and Systems Biology, Department of Biochemistry and Molecular Biology, University of Córdoba, Córdoba, Spain, [2]Agricultural Microbiology, Faculty of Agricultural Science, National University of Córdoba, CONICET, Córdoba, Argentina

OPEN ACCESS

Edited by:
Pingfang Yang,
Hubei University, China

Reviewed by:
Zhenyu Wang,
Jiangxi Normal University, China
Wilco Ligterink,
KeyGene, Netherlands

***Correspondence:**
Maria Angeles Castillejo
bb2casam@uco.es
Jesus V. Jorrín-Novo
bf1jonoj@uco.es

†Present addresses:
Monica Escandon,
Plant Physiology, Department of Organisms and Systems Biology, Faculty of Biology and Biotechnology Institute of Asturias, University of Oviedo, Oviedo, Spain
Victor M. Guerrero-Sánchez,
Vascular Pathophysiology Area, Cardiovascular Proteomics Laboratory, Centro Nacional de Investigaciones Cardiovasculares Carlos III (CNIC), Madrid, Spain

‡These authors have contributed equally to this work

Specialty section:
This article was submitted to Plant Proteomics and Protein Structural Biology, a section of the journal Frontiers in Plant Science

Received: 29 March 2022
Accepted: 06 June 2022
Published: 27 June 2022

Citation:
Escandón M, Bigatton ED, Guerrero-Sánchez VM, Hernández-Lao T, Rey M-D, Jorrín-Novo JV and Castillejo MA (2022) Identification of Proteases and Protease Inhibitors in Seeds of the Recalcitrant Forest Tree Species Quercus ilex. Front. Plant Sci. 13:907042. doi: 10.3389/fpls.2022.907042

Proteases and protease inhibitors have been identified in the recalcitrant species *Quercus ilex* using *in silico* and wet methods, with focus on those present in seeds during germination. *In silico* analyses showed that the *Q. ilex* transcriptome database contained 2,240 and 97 transcripts annotated as proteases and protease inhibitors, respectively. They belonged to the different families according to MEROPS,[1] being the serine and metallo ones the most represented. The data were compared with those previously reported for other *Quercus* species, including *Q. suber, Q. lobata,* and *Q. robur*. Changes in proteases and protease inhibitors alongside seed germination in cotyledon and embryo axis tissues were assessed using proteomics and *in vitro* and in gel activity assays. Shotgun (LC–MSMS) analysis of embryo axes and cotyledons in nonviable (NV), mature (T1) and germinated (T3) seeds allowed the identification of 177 proteases and 12 protease inhibitors, mostly represented by serine and metallo types. Total protease activity, as determined by *in vitro* assays using azocasein as substrate, was higher in cotyledons than in embryo axes. There were not differences in activity among cotyledon samples, while embryo axis peaked at germinated T4 stage. Gel assays revealed the presence of protease activities in at least 10 resolved bands, in the *Mr* range of 60–260 kDa, being some of them common to cotyledons and embryo axes in either nonviable, mature, and germinated seeds. Bands showing quantitative or qualitative changes upon germination were observed in embryo axes but not in cotyledons at *Mr* values of 60–140 kDa. Proteomics shotgun analysis of the 10 bands with protease activity supported the results obtained in the overall proteome analysis, with 227 proteases and 3 protease inhibitors identified mostly represented by the serine, cysteine, and metallo families. The combined use of shotgun proteomics and protease activity measurements allowed the identification of tissue-specific (e.g., cysteine protease inhibitors in embryo axes of mature acorns) and stage-specific proteins (e.g., those associated with mobilization of storage proteins accumulated in T3 stage). Those proteins showing differences between nonviable and

[1]http://www.merops.ac.uk

viable seeds could be related to viability, and those variables between mature and germinated could be associated with the germination process. These differences are observed mostly in embryo axes but not in cotyledons. Among them, those implicated in mobilization of reserve proteins, such as the cathepsin H cysteine protease and Clp proteases, and also the large number of subunits of the CNS and 26S proteasome complex differentially identified in embryos of the several stages suggests that protein degradation *via* CNS/26S plays a major role early in germination. Conversely, aspartic proteases such as nepenthesins were exclusively identified in NV seeds, so their presence could be used as indicator of nonviability.

Keywords: protease, protease inhibitors, nonorthodox seeds, germination, *Quercus ilex*, proteomics, protease activity

INTRODUCTION

Germination is a complex process by which a seed embryo develops into a seedling. It is also a critical stage for plant development, survival, propagation and reproduction. Seeds can be classified as orthodox and nonorthodox (recalcitrant) according mainly to their tolerance to dehydration (Berjak and Pammenter, 2008). Orthodox seeds generally retain viability and germinability even after storage over long periods under suitable dry, cool conditions. On the other hand, nonorthodox seeds are damaged by, and cannot survive, dehydration (Roberts, 1973; Wyse and Dickie, 2017). *Quercus* species belong to the latter, recalcitrant group, so they must be shed and germinate immediately after maturation, when their moisture content is still relatively high (Connor and Sowa, 2002; Pasquini et al., 2012; Joët et al., 2013; Sghaier-Hammami et al., 2016). Therefore, after harvesting and processing recalcitrant seeds, storing them over long periods in unsuitable conditions can cause loss of their viability and pose serious difficulties to conservation and propagation of the species (Pasquini et al., 2011, 2012).

Seed germination has been widely studied in orthodox seeds. The process involves major changes including reprogramming of gene expression under hormonal control (Angelovici et al., 2010; Wang et al., 2015). By contrast, available knowledge of germination on the nonorthodox seeds such as those of *Q. ilex* is limited (Pammenter and Berjak, 2014) despite the fact that the storage, desiccation sensitivity, germination and chemical composition of *Q. ilex* acorns have been the subject of some study (Valero-Galván et al., 2011, 2021a; Pasquini et al., 2012; Joët et al., 2013; Romero-Rodríguez et al., 2014, 2015, 2018, 2019; Sghaier-Hammami et al., 2016, 2021; López-Hidalgo et al., 2021). Despite its significance to seed propagation and conservation, germination at the molecular level in recalcitrant seeds remains poorly understood. In this work, molecular events occurring during germination of seeds of the recalcitrant species *Q. ilex* and, specifically, in proteases and protease inhibitors (PIs), were examined with *in silico* and wet methods.

Maturation and germination in *Q. ilex* acorns have been the subjects of morphometric, physiological, transcriptomic, proteomic and metabolomic analyses (Valero-Galván et al., 2011, 2021b; Romero-Rodríguez et al., 2015, 2018, 2019; Sghaier-Hammami et al., 2016, 2021; Rey et al., 2019; Escandón et al., 2021a). Multiomic approaches have allowed differences in phytohormone, sugar and phenolic metabolism, proteases, ROS and storage proteins, and late embryogenesis proteins during seed germination to be detected (Romero-Rodríguez et al., 2018, 2019; Sghaier-Hammami et al., 2021). Also, regulation through post-translational modifications such as phosphorylation has been reported (Romero-Rodríguez et al., 2015). Romero-Rodríguez et al. (2019) observed accumulation of proteases and enzymes of amino acid metabolism in mature seeds. Their results support the hypothesis that mature nonorthodox *Q. ilex* seeds possess the machinery needed to rapidly resume metabolic activities and start germination.

Proteases are involved in almost all aspects of plant growth and development including germination, circadian rhythms, senescence, programmed cell death and responses to environmental changes (García-Lorenzo et al., 2006). Together with specific endogenous inhibitors that regulate their activities, proteases are in fact the main actors in effecting and regulating protein breakdown (Vaseva et al., 2011). During seed germination, much of the amino acid supply needed for emerging seedlings to grow comes from degradation of seed storage proteins. Proteases are not only important for storing hydrolysis proteins, but also responsible for other processes such as protein turnover, post-translational modifications, enzyme activation and inactivation, and plant defense (Schaller, 2004). Most proteolytic enzymes degrading seed storage proteins during seed germination are cysteine proteases, but others such as serine, aspartic and metalloproteases have also been reported (Tan-Wilson and Wilson, 2012). Protease activity is regulated at the transcriptional and translational levels but, most importantly, at the protein level. PIs are proteinaceous molecules that operate by regulating peptidase activity (Santamaría et al., 2014) and can have two different functions. Thus, they act as inhibitors of endogenous peptidases to regulate the activity of plant peptidases in order to avoid indiscriminate degradation when not needed (Martínez et al., 2012; Volpicella et al., 2012). Also, they regulate the activity of exogenous peptidases such as those used by various pests and pathogens to feed on and attack plants (Haq et al., 2004; Santamaría et al., 2012; Hörger and van der Hoorn, 2013).

Proteolytic enzymes have been classified according to catalytic mechanism, substrate specificity, cell locus, structure and function (Rawlings and Barrett, 1993; Van der Hoorn, 2008). The proteases

and PIs in the MEROPS database[2] contains a hierarchical classification where homologous sets of peptidases and protein inhibitors are grouped into families and clans (Rawlings et al., 2010). Complete sequence analysis of various genomes has shown that approximately 2% of the genetic information corresponds to proteases, which thus constitute one of the largest and best characterized functional groups of enzymes (Barrett et al., 1998). Plant genomes encode hundreds of proteases involved in a number of cellular processes; however, their regulation and subsequent actions are still poorly understood (van der Hoorn and Klemenčič, 2021).

Characterizing and comparing proteases and PIs in nonorthodox seeds such as those of *Q. ilex* is expected to improve our understanding of the role of these proteins in maturation and germination. In this work we conducted a comparative *in silico* analysis at the transcriptome level of *Q. ilex* and other *Quercus* species. Analyses were performed by using wet methods whereby protease and PI forms were examined at the proteomic level, and compared among *Q. ilex* organs. Changes in protease and PI profiles at the protein and activity levels were examined by using shotgun (LC–MSMS) proteomic analysis, and *in vitro* and *in gel* activity assays.

MATERIALS AND METHODS

Comparative *in silico* Analysis of Protease and Protease Inhibitors in the Gene Product Databases for *Quercus* spp.

Proteases and PIs in the species-specific *Q. ilex* database (Guerrero-Sánchez et al., 2017, 2019, 2021) were validated and classified against the Merops (see Footnote 2), Gene Ontology[3] and Panther[4] databases. The transcriptome *Q. ilex* DB is deposited at NCBI GEO (Gene Expression Omnibus) repository with the accession GSE145009.[5] In addition, the numbers of proteases and inhibitors found in the transcriptome of *Q. ilex* DB were compared with those of other *Quercus* spp. For this purpose, we used the transcriptome or gene product databases for *Q. suber* (Ramos et al., 2018),[6] *Q. robur* (Plomión et al., 2018)[7] and *Q. lobata* (Sork et al., 2016, 2022).[8] The previous annotated databases were searched with the following key words: protease, proteinase, peptidase, proteasome and protease inhibitor.

Plant Material and Experimental Germination Design

Mature acorns were harvested from a *Q. ilex* tree located at 38° 19′ 46″ N, 5° 33′ 15″ W in Aldea de Cuenca (Córdoba, southern Spain). Once in the lab, acorns were sterilized in a solution of commercial bleach (1%), washed with abundant tap water and dried with filter paper. Then, they were allowed to germinate in the dark for 4 weeks at a mean temperature of 22°C and (50 ± 10)% relative humidity according to Simova-Stoilova et al. (2015). Viable and nonviable seeds—the latter including seeds that failed to germinate within 4 weeks—from mature acorns at stage T1 and NV, respectively, as well as germinated seeds at stages T2–T4, were also collected for analysis (Romero-Rodríguez et al., 2018; Sghaier-Hammami et al., 2020; **Supplementary Figure S1**). After sampling, embryonic axes and cotyledons were separated and ground under liquid nitrogen in a mortar for storage at −70°C. Three replicates per stage and type of tissue (embryo axis or cotyledon), and one pool of 20 acorns per replicate, were used.

Protein Extraction, Identification and Quantification by Gel-nLC-Orbitrap/MS Analysis

Two different protein extractions protocols were followed for shotgun protein profiling and protease activity analysis. With the first, an amount of 100 mg of embryo axes or cotyledons was extracted by using the TCA–acetone/phenol protocol (Wang et al., 2006; Maldonado et al., 2008). For activity assays, 100 mg of each type of tissue was extracted by following a slightly modified version of the protocol of Amaral et al. (2020). First, 600 µl of extraction buffer [50 mm Tris–HCl, pH 7.4, containing 10% (v/v) glycerol, 0.25% (v/v) Triton X-100, 1 mm dithiothreitol-DTT and 3% (w/v) polyvinylpolypyrrolidone (PVPP)] was added to ground tissue and mixed by inversion. After centrifugation at 17000 *g* at 4°C for 30 min, the supernatant was collected for protein and activity determinations. Proteins were quantified by following the Bradford protocol according to the manufacturer's (Sigma-Aldrich) instructions.

Samples containing 70 µg of protein were subjected to sodium dodecyl sulfate–polyacrylamide gel electrophoresis (SDS-PAGE) on 12% acrylamide for protein cleaning according to Romero-Rodríguez et al. (2019). The only resulting band was excised and digested with trypsin and the peptides thus obtained were dissolved in 50 µl of a 4% acetonitrile/0.25% formic acid mixture for loading into an Ultimate 3,000 nano-LC-MS-UHPLC-Orbitrap Fusion instrument from Thermo Fisher Scientific (Waltham, MA, United States; Gómez-Gálvez et al., 2020). MS/MS data were processed with the software Proteome Discoverer v. 2.3, also from Thermo Fischer Scientific, and identified by using the SEQUEST algorithm against the species-specific database of *Q. ilex* developed from the transcriptome (Guerrero-Sánchez et al., 2017, 2019, 2021), using the setting reported by San-Eufrasio et al. (2021). Proteins were quantified in relative form from the peak areas for the precursor ions, using the three strongest peptide ion signals for each protein (Al Shweiki et al., 2017). Raw data were deposited in the ProteomeXchange Consortium *via* PRIDE (Perez-Riverol et al., 2018), using the identifier PXD032845. The following criteria were used to deem identifications confident: score ≥ 2, coverage ≥15% and at least two different peptides for each protein.

[2]https://www.ebi.ac.uk/merops/
[3]http://geneontology.org/
[4]http://www.pantherdb.org/
[5]https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE145009
[6]https://corkoakdb.org/
[7]https://www.oakgenome.fr/
[8]https://valleyoak.ucla.edu/genomicresources/

In vitro and *in gel* Proteolytic Activity Assays

Proteolytic activity was assessed by following the protocol of Coêlho et al. (2016) in slightly modified form. For this purpose, we used a 1.5% (w/v) azocasein solution in protease extraction buffer [50 mM Tris–HCl, pH 7.4, containing 10% (v/v) glycerol and 0.25% (v/v) Triton X-100]. The enzyme extract solution was prepared at a total protein concentration of 1 mg/ml in extraction buffer. The reaction mixture contained identical amounts of enzyme extract and azocasein solution (125 µl each) and was incubated at 35°C for 6 h. The reaction was stopped by adding 700 µl of 5% trichloroacetic acid (TCA). Then, the mixture was allowed to stand at room temperature for 10 min and centrifuged at 2000 *g* for 10 min before the supernatant was collected and supplied with 0.5 N sodium hydroxide in a 1:1 ratio for absorbance measurements at 440 nm. A blank prepared by mixing 5% TCA with the substrate before adding the enzyme extract was also measured in parallel.

In gel protease activity was determined according to Heussen and Dowdle (1980). Protein extracts (40 µg per sample) were separated on 9% acrylamide–0.1% gelatin copolymerized gels. Samples were loaded with nondenaturing buffer [62.5 mm Tris–HCl containing 10% (v/v) glycerol and 0.001% (w/v) Bromophenol Blue]. Electrophoresis runs were performed at 4°C, using a voltage of 50 V until the proteins entered the resolving gel and then raising it to 80 V through the end. Then, gels were incubated under agitation at room temperature in 2.5% (v/v) Triton X-100 for 30 min, washed 3 times with distilled water and incubated overnight in proteolysis buffer (100 mm citrate buffer, Na_2HPO_4/citric acid pH 6.8, 4 mm DTT and 10 mM cysteine) under continuous shaking at 35°C. Proteolysis was stopped by transferring gels to a solution containing 0.1% (w/v) Coomassie Brilliant Blue R-250 (Neuhoff et al., 1988). Gel images were acquired with a calibrated GS-900 densitometer from Bio-Rad (Hercules, CA, United States). The effect of the PIs ethylenediaminetetraacetic acid (EDTA) and phenylmethylsulfonyl fluoride (PMSF) was evaluated by following the above-described protocol except that the sample was supplied with the corresponding inhibitor at a 100 mm concentration prior to use. Molecular weight of bands was calculated by mobility comparisons with protein standards markers (Thermo scientific Spectra Multicolor Broad Range). Protease activity bands were excised from the gels for shotgun analysis. Ten bands were cut from protease activity gels corresponded to the higher activity. Then, the bands were digested and analyzed by shotgun proteomic analysis such as described above.

Statistical and Bioinformatic Analyses

Statistical analyses of shotgun and *in vitro* protease activity were performed with the core functions in the software R v. 3.6.2 (R Core Team, 2018) and the package pRocessomics v.1.8.[9] Following Valledor et al. (2014), proteomic data were preprocessed with the Random Forest algorithm for assignation of missing values, using a threshold of 0.34 and a consistency-based criterion of 0.2. The abundance of each variable was normalized by following a sample centric approach. Preprocessed, filtered data were z-centered to ensure normality and homoscedasticity. Scaled and centered values (*z*-scores) were subjected to multivariate analysis, using the Manhattan distance method for heatmapping with the pheatmap library (Kolde, 2015). Venn diagrams were constructed and Principal Component Analysis (PCA) was done with pRocessomic and UpsetR v.1.4. (Lex et al., 2014). Univariate analyses (one-way ANOVA) and Tukey's HSD *post-hoc* test were conducted on proteins and activities with $p \leq 0.05$ and FDR (5%).

RESULTS

In silico Analysis of Proteases and Protease Inhibitors in the *Quercus ilex* Transcriptome Database, and Comparison With Other *Quercus* spp.

Proteases and PIs in the species-specific *Q. ilex* transcriptome database (Guerrero-Sánchez et al., 2017, 2019, 2021) were subjected to *in silico* analysis (**Supplementary Table S1**). This database contains 45,815 transcripts 2,240 and 97 of which are annotated as proteases and protease inhibitors, respectively. These transcripts belong to the different protease families in MEROPS[10] and were classified according to GO into biological process, molecular function and cellular component categories (**Supplementary Table S1**). Serine proteases and metalloproteases accounted for one-half of all proteases in the *Q. ilex* database, and degradation and metabolic proteins were highly represented. Also, one-half of the proteins in this group were predicted to be extracellular or membrane proteins. In addition, 50% of PIs were of the serine family, and 40% had defense and stress response functions.

Similarly, large numbers of proteases and PIs were previously found in other *Quercus* spp. Thus, 40,599 *Q. suber* transcripts have been listed, 974 being proteases and 17 PIs (Ramos et al., 2018).[11] The *Q. lobata* database (Sork et al., 2022; see Footnote 8) contains 39,373 protein-coding genes with 721 corresponding to proteases and 44 to PIs. The figures for *Q. ilex* are similar to those for *Q. robur*, with 43,240 protein-coding genes (Plomión et al., 2018; see Footnote 7) 1815 of which are annotated as proteases and 173 as PIs.

Shotgun Proteomic Analysis of Proteases and Protease Inhibitors in *Quercus ilex* Seeds: A Comparison of Tissues

A shotgun proteomics strategy was used to analyze the protein profile of seed cotyledons and embryo axes tissues from viable (T1 and T3) and nonviable (NV) seeds. In total, 3,512 proteoforms were identified in *Q. ilex* embryo axes and cotyledons. Editing and normalizing the data yielded a total of 1926 confident

[9]https://github.com/Valledor/pRocessomics

[10]https://www.ebi.ac.uk/merops/cgi-bin/family_index?type=P

[11]https://corkoakdb.org/home

proteins (**Supplementary Table S2a**). PCA of the raw data (**Supplementary Table S3**; **Supplementary Figure S2**) revealed that the first two components accounted for 41% of total variability, PC1 mostly separating stages and PC2 tissues.

Proteins were then analyzed and classified in terms of protease activity as verified against the GO, Merops and Panther databases. This allowed a total of 177 proteases (**Supplementary Table S2b**) and 12 PIs (**Supplementary Table S2c**) to be identified. Most proteases (149) and PIs (8) were present in both embryo axes and cotyledon tissues (**Figures 1A1**,**B1**), 2 proteases and 1 PI being unique to cotyledons, and 26 proteases and 3 PIs to embryo axes.

In terms of germination stage, NV and T1 embryo axes, and NV cotyledons, contained the largest numbers of proteases and PIs (**Figures 1A2**,**B2**). These embryo axes and cotyledons shared 38 proteases and 2 PIs. More than one-half of all proteases identified belonged to the serine (28.7%), and metalloprotease (23.6%) families, followed by the threonine (13.5%), cysteine (11.2%) and aspartyl (10.1%) families (**Figure 1A3**). The remaining proteases (12.4%), which belonged largely to the 26S proteasome, could be assigned to no particular family. The vast majority of PIs (66.7%) were cysteine protease inhibitors, followed by an unassigned group (25%) and serine proteases (8.3%; **Figure 1B3**).

A plot of total protein abundance for the identified proteases and PIs based on their combined peak areas (**Figures 1A4**,**B4**) revealed differences that were significant in NV seeds only, where total proteases were better represented in embryo axes. Proteases were well represented by the metallo and serine families in both types of tissues at all stages (**Figure 1A4**). Although the most common PIs were from the cysteine family, the serine family was also well represented in embryo axes (**Figure 1B4**).

A comparison of the results of our shotgun proteomic analysis of *Q. ilex* seeds with those of previous studies (Sghaier-Hammami et al., 2020, 2021; Escandón et al., 2021b; Guerrero-Sánchez et al., 2021) allowed the identification of 400 proteases and 23 PIs differentially accumulating in various tissues (embryo axis, cotyledon, leaf and root) of *Q. ilex*. The largest number of them (354) was found in leaves, although a high number (129) were also present in all tissues. Several proteins were specifically detected depending on samples: 133 in leaves, 8 in a mix composed by acorn, leaf and roots, and 1 in acorns (**Supplementary Table S2d**). The best represented among them belonged to the serine and metalloprotease families, which together accounted for more than one-half of all identified proteins.

Differential Proteases and Protease Inhibitors Between Embryo Axes and Cotyledons, and Seed Developmental Stages

Based on the PCA results, those proteases and PIs most markedly contributing to variability were selected for further investigation (**Supplementary Table S3**). For that we established a cut off at the 300 proteins contributing with higher loadings (absolute values) to the PC1 and PC2. A total of 55 proteases and 3 PIs were thus identified in this group, of which 3 proteases (*viz.*, carboxypeptidase scpl44, qilexprot_44690; Xaa-Pro aminopeptidase 2, qilexprot_22106; and acetylornithine deacetylase aodD, qilexprot_78419) were present in both components. A heatmap for the quantities of PCA-selected proteins that were differentially present in tissues or stages (**Figure 2**) exhibited two clusters. One included highly represented proteins in mainly ET1 and ENV; the other, encompassed two subclusters containing proteins that were more abundant in embryo axes (mainly in ET3) or cotyledons (mainly in CT1 and CT3). As can be seen from the heatmap, proteases accumulated more markedly in embryo axes than they did in cotyledons.

A number of RPN proteins (26S proteasome) fell in the highly represented group ET3, namely: qilexprot_8321, qilexprot_55116, qilexprot_26142, qilexprot_12738, and the 26S protease regulatory subunits qilexprot_51657, qilexprot_59275 and qilexprot_78318. Interestingly, 3 cysteine proteinase inhibitors (qilexprot_51394, qilexprot_15999, qilexprot_28688) were ET1-specific, and 8 proteases including the aspartic proteinases qilexprot_63251, qilexprot_78560 and qilexprot_60038 were NV-specific in both types of tissues.

T3 stage showed a marked increase of the proteases and PIs analyzed. Among them, 5 serine carboxypeptidases (*viz.*, qilexprot_44690 in CT3, qilexprot_16893 and qilexprot_52755 in ET3, and qilexprot_48367 and qilexprot_40121 in both CT3 and ET3) and 4 cysteine proteases (*viz.*, qilexprot_78020 and qilexprot_56434 in both tissues, and qilexprot_77592 and qilexprot_15272 in cotyledons). Embryo axes of the three stages differentially accumulated 2 signalosome (CSN) proteins (qilexprot_12955, qilexprot_63361) and 2 neddylation proteins (qilexprot_26596, qilexprot_46052).

In vitro and *in gel* Protease Activity

In vitro protease activity as determined according to Heussen and Dowdle (1980) differed between embryo axes and cotyledons, being significant at T3 stage. Cotyledon tissues exhibited higher baseline activity at all germination stages except T4 (**Figure 3**). However, only embryo axis tissues differed throughout the germination process, with a decrease at T3 followed by an increase at T4.

As can be seen from **Figure 4**, *in gel* activity profiles differed between cotyledons and embryo axes. Both types of tissue gave a band at *ca.* 260 kDa irrespective of developmental stages (**Figures 4A1**,**B1**). Such a band, however, was broader in embryo axes, where it ranged from 260 to 140 kDa at NV and T1, and from 260 to 100 kDa at T2, T3 and T4. A thin band was also seen at 70 kDa in cotyledons, and another at about 60 kDa in embryo axes, both of which increased at later stages.

Treating samples with the serine protease inhibitor PMSF (**Figures 4A2**,**B2**) considerably reduced band strength, which suggests the presence of abundant serine proteases in the extracts from both types of tissues. This finding is consistent with the number of serine proteases identified in the shotgun analyses. The metalloprotease inhibitor EDTA (**Figures 4A3**,**B3**) inhibited protease activity less markedly—so much so that the

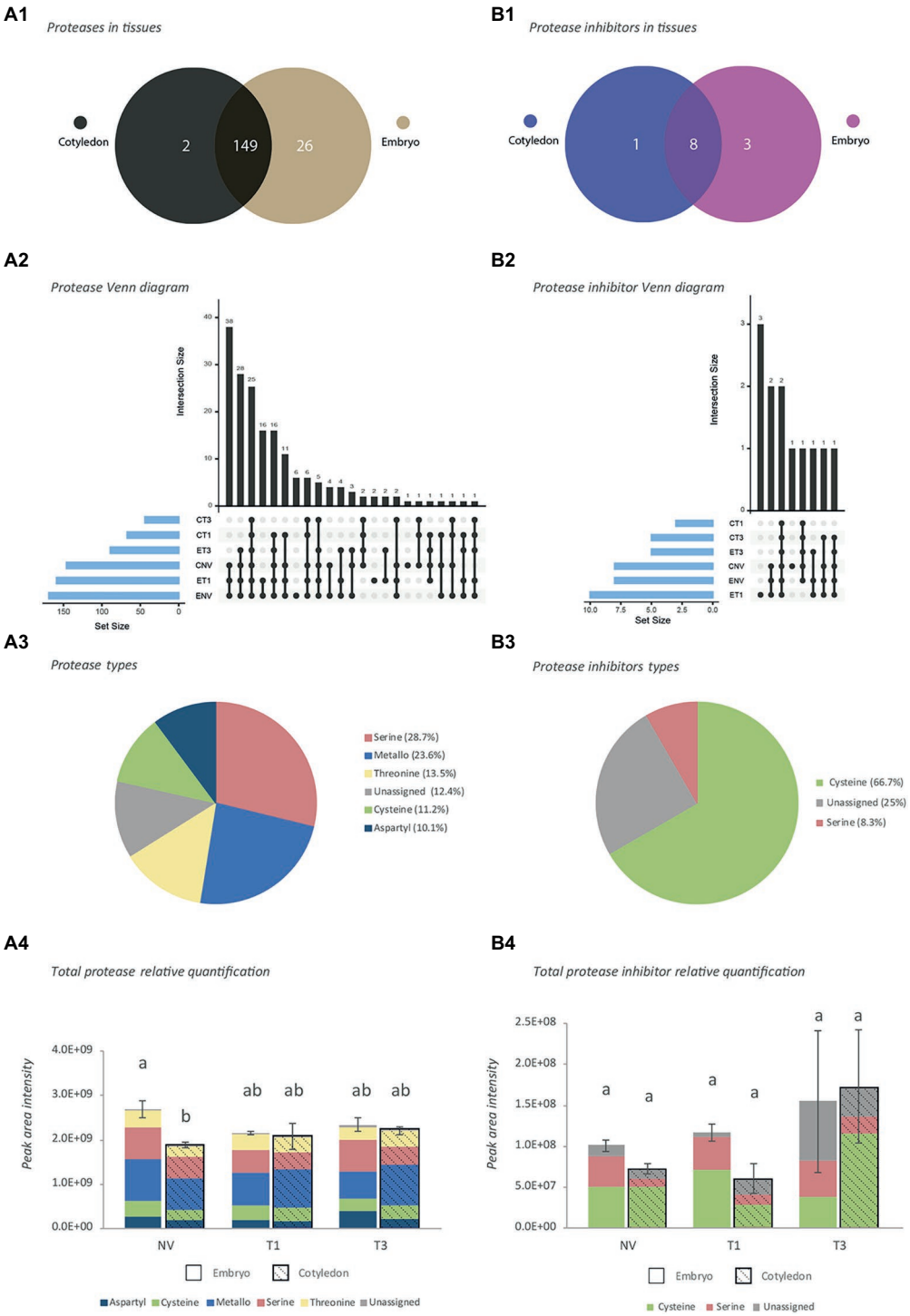


FIGURE 1 | Characterization of proteases and protease inhibitors identified by shotgun analysis. **(A)** Proteases. **(B)** Protease inhibitors. 1: Venn diagram at tissue level, 2: Venn diagram at stage level, 3: Activity-type classification, 4: Total protein abundance based on combined peak areas. NV: Nongerminated acorns after 4 weeks of germination. T1: mature acorns prior to germination. T3: germinated acorns when root tip size reached 6.5 mm. The same letter indicates that there is no statistical difference between tissues or stages.

band of 70 kDa was even stronger in both embryo axes and cotyledons, possibly as a result of EDTA inhibiting certain PIs of the metalloprotease family.

Shotgun analysis of gel activity bands allowed the proteases and PIs potentially contributing to the greatest extent to some functions to be identified (**Figure 4**). A total of 224 proteases

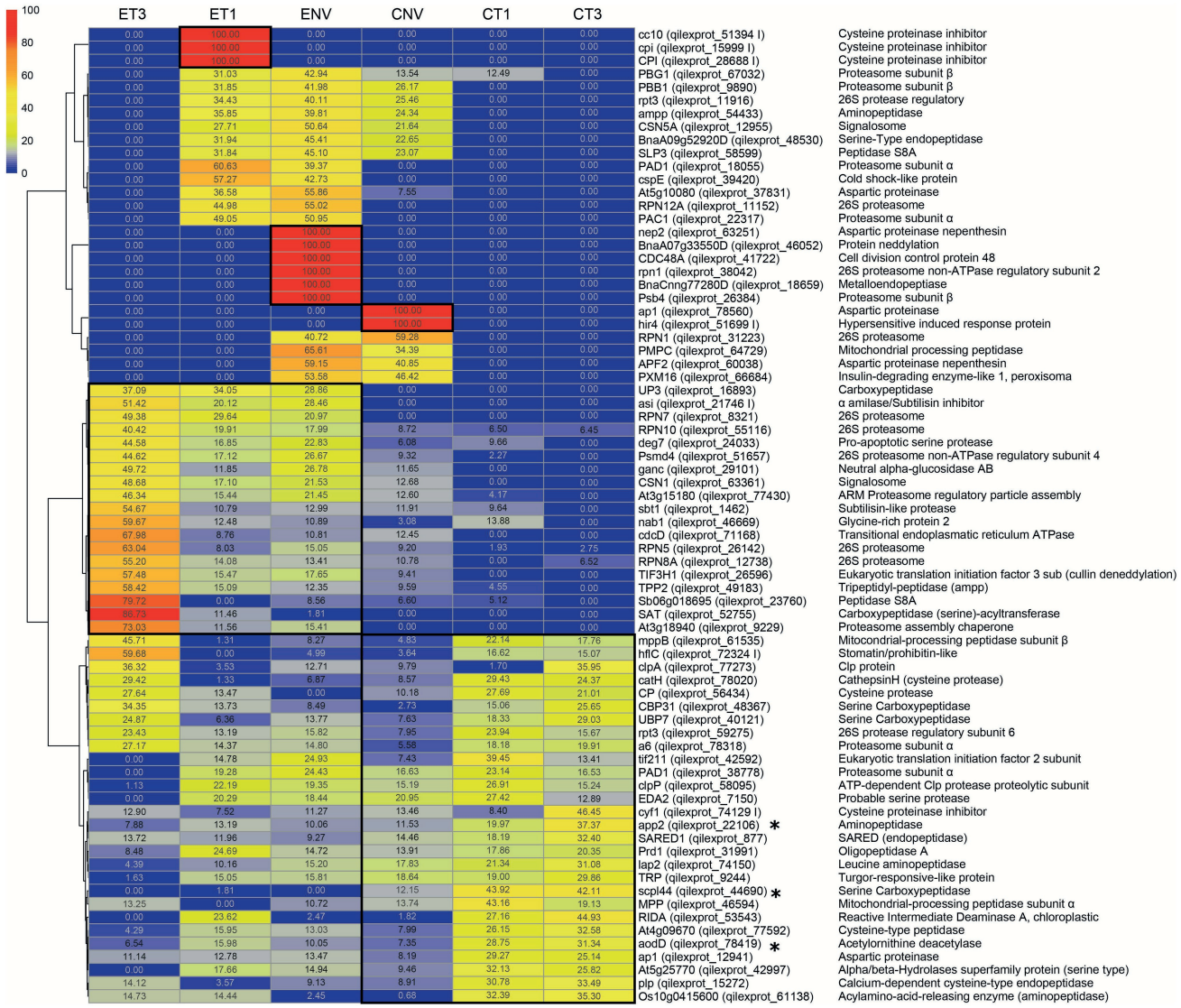


FIGURE 2 | Proteases and protease inhibitors heatmap included in the top300 loadings of PCA and Venn specific tissue/stage protease. E, embryo axis. C, cotyledon. NV: Nongerminated acorns after 4 weeks of germination. T1: mature acorns prior to germination. T3: germinated acorns when root tip size reached 6.5 mm. * proteases present in both the first and second PC. Numbers in each cell represent the normalized abundance of the proteins found in each treatment. This abundance is reflected in the color scale shown.

and 3 PIs were identified, mostly of serine, cysteine and metallo types. Band 1 (~260 kDa) was cut from cotyledons at T3 and allowed a leucine aminopeptidase (qilexprot_74150) to be identified with a high score. Bands 2–4 from embryo axes at T1 (260–140 kDa) revealed the presence of two cysteine protease inhibitors (qilexprot_28688 and qilexprot_15999) that were consistently represented in the three fractions. The serine peptidase qilexprot_24033 and the oligopeptidase qilexprot_31991 were also represented in them. The most salient proteins identified from bands 5–8 in T3 embryo axes (260–120 kDa) were the 26S proteasome regulatory subunit qilexprot_31223, clpA (qilexprot_77273), transitional endoplasmic reticulum ATPase (qilexprot_71168) and serine protease qilexprot_24033. Band 9 (~60 kDa) corresponded to T4 embryo axes, stage not examined in the whole shotgun analysis. Some proteins identified in embryo axes of T3 stage have been found in band 9. Such was the case with mitochondrial processing peptidase subunit beta (qilexprot_61535), transitional endoplasmic reticulum ATPase (qilexprot_71168), ubiquitin carboxyl-terminal hydrolase (qilexprot_40121), SAT1 (qilexprot_52755), serine protease (qilexprot_24033), ARM proteasome regulatory particle assembly (qilexprot_77430) and clpA (qilexprot_77273). Finally, band 10 (70 kDa), which was extracted from EDTA treated gel containing T1 embryo axis, allowed 5 aminopeptidases (qilexprot_74150, qilexprot_22106, qilexprot_49183, qilexprot_54433 and qilexprot_61138), 3 serine endopeptidases (qilexprot_48530, qilexprot_24033 and qilexprot_1,462) and 1 metalloprotease (qilexprot_31991) to be identified. The proteases and PIs identified

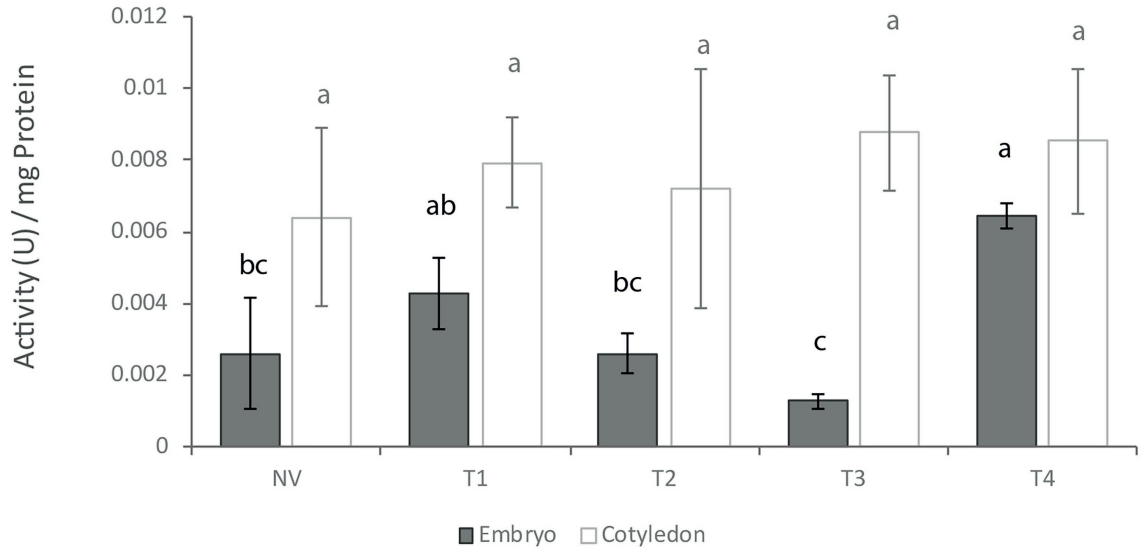


FIGURE 3 | *In vitro* protease activity at different germination stages. NV: Nongerminated acorns after 4 weeks of germination. T1: mature acorns prior to germination. T2, T3, T4: Germinated acorns when root tips were visibly emerging from cotyledons (T2), their size reached 6.5 mm (T3) and it exceeded 20 mm (T4). Different letters denote significant differences between germination stages in each tissue.

with high scores (≥4) in gel activity bands are listed in **Supplementary Tables S2b**,**c**, respectively. Some proteases and PIs identified using the whole proteome were also identified in the bands from activity gels in the same experimental system.

DISCUSSION

Eukaryotes have a high number of protease-coding genes. More than 641 and 677 protease genes have been identified in the human and mouse genome, respectively (Bond, 2019). Numbers are even greater in plants, with 723 in *Arabidopsis* and 955 in *Populus*, for example (García-Lorenzo et al., 2006). The numbers for *Q. ilex* found here, and those previously reported for other *Quercus* species, are roughly similar if one considers post-transcriptional variants, the high level of heterozygosity in the genus *Quercus* and genome assembly errors (Sork et al., 2016, 2022; Plomión et al., 2018; Ramos et al., 2018). Thus, *Q. ilex* has 2,240 transcripts for proteases (Guerrero-Sánchez et al., 2017, 2019, 2021), *Q. suber* 974 (Ramos et al., 2018), *Q. lobata* 721 (Sork et al., 2022) and *Q. robur* 1815 (Plomión et al., 2018). The number of PI genes in these species is much smaller, with 97 transcripts in *Q. ilex* (this work), 17 in *Q. suber* (Ramos et al., 2018), 44 in *Q. lobata* (Sork et al., 2022) and 173 in *Q. robur* (Plomión et al., 2018). Also, the number of PIs in *Q. ilex*, 97, is much greater than those reported for other plant species such as tomato, with 55 (Fan et al., 2019).

Proteases in *Q. ilex* are complex in specificity, structure and catalytic properties, with representatives of the different families in MEROPS classification involved in various biological processes and belonging to the different molecular and cell locus groups in GO. The serine and metallo families are the best represented protease groups (especially those in the protein degradation, protein metabolic processes, and extracellular and membrane locus groups). As noted by Puente and López-Otín (2004), the views on proteases have changed since early reports from their being deemed nonspecific enzymes involved in proteolysis and protein catabolism to their enacting specific, limited selective cleavage of proteins, and regulating proteostasis and developmental processes ranging from growth and development to responses to environmental conditions. Seed germination is one of the biological processes where proteases play a key role; thus, they hydrolyze and mobilize reserve proteins (Martinez et al., 2019). Based on protein annotations in *Q. ilex* DB, 6% of the total proteases could be related to seed germination or embryo development, and 40% of the PIs with defense and stress response. The MEROPS database contains 83 family members of PIs, the serine family being the best represented in *Q. ilex* DB—mostly annotated with defense and response to stress roles. While PIs are present in all plant tissues, they are constitutive components of seeds and storage organs (McManus et al., 1999). Ever since they were assumed to be induced in response to pathogen and herbivore attacks (Ryan, 1990), PIs have been associated with defense responses. Interest in plant PIs has grown since the outburst of the COVID-2019 pandemic by effect of their potential against viral proteases (Hellinger and Gruber, 2019; Adhikari et al., 2021). Based on their abundance in some seeds, a twofold role can be envisaged, namely: protection against proteases from other organisms, and regulation of endogenous enzymes during mobilization of reserve proteins.

Out of the total 2,240 and 97 transcripts annotated as proteases and PIs, respectively, in the *Q. ilex* transcriptome, only 8% (177) and 12% (12) were detected by shotgun analysis of embryo axes and cotyledons here, 28 and 5, respectively, being tissue-specific. The differences between the two sets of data may be due to analytical or biological reasons such as

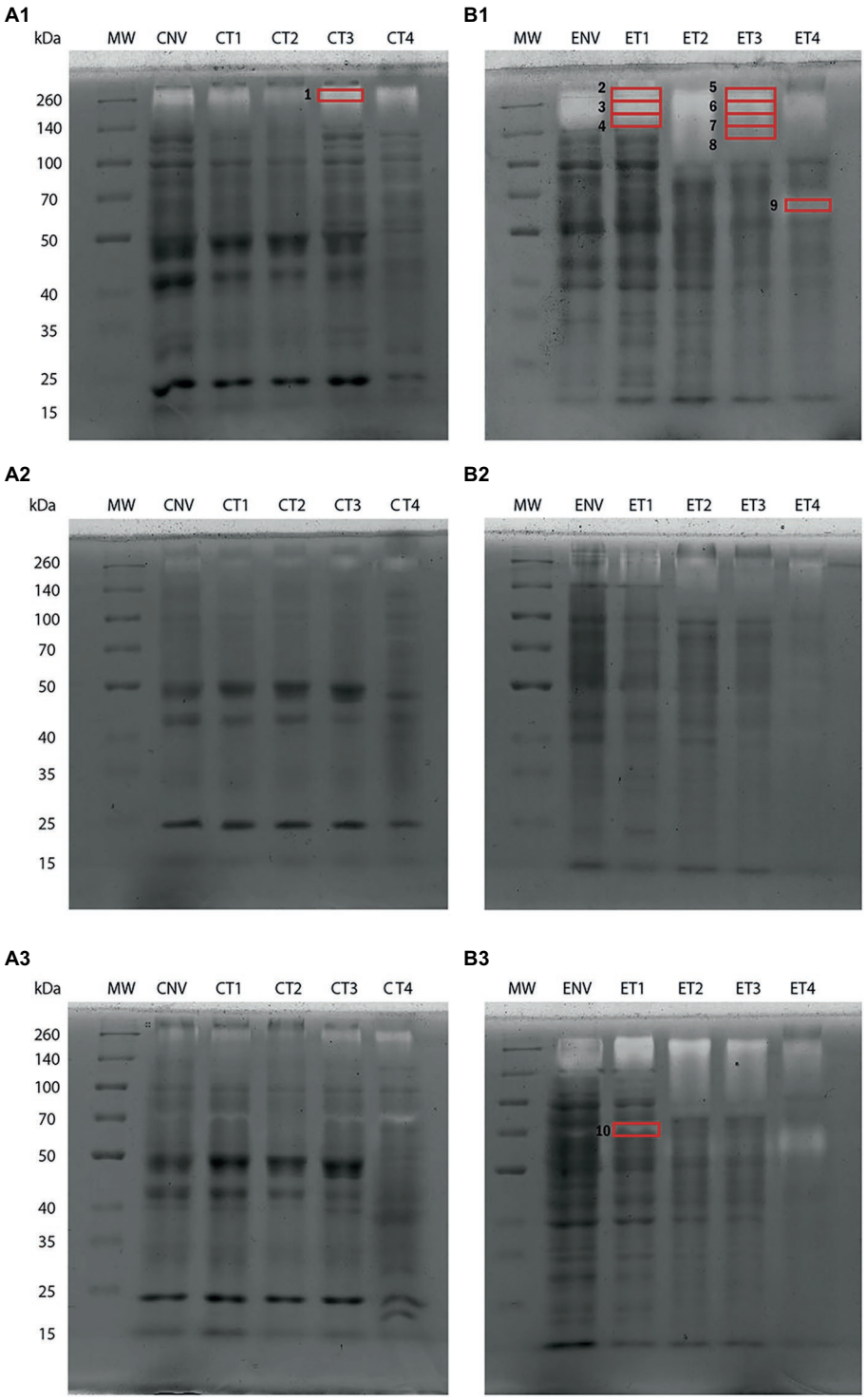


FIGURE 4 | Protease activity in 9% polyacrylamide gel slabs containing gelatin at different germination stages. Cotyledon samples **(A)**. Embryo axis samples **(B)**: Gels containing no inhibitor (1); an inhibitor of serine protease activity (PMSF; 2); and one of metalloprotease activity (EDTA; 3). NV: Nongerminated acorns after 4 weeks of germination. T1: mature acorns prior to germination. T2, T3, T4: Germinated acorns when root tips were visibly emerging from cotyledons (T2), their size reached 6.5 mm (T3) and it exceeded 20 mm (T4). Ten bands were cut from protease activity gels for shotgun analysis (red boxes).

differences in analytical potential between the two omic tools or in cell and temporal specific transcription and translation of genes. Including previously reported data for other organs increased the total number of proteoforms corresponding to proteases and PIs to 400 and 23, respectively, 129 being present in all organs (Sghaier-Hammami et al., 2020, 2021; Escandón

et al., 2021b; Guerrero-Sánchez et al., 2021). The organ specific profile of proteases was reported by Hüynck et al. (2019). That for *Q. ilex* seeds is rich in serine and metallo proteases, and also in cysteine PIs, whereas that for cereal seeds is especially rich in proteases of the cysteine family, which mobilize and hydrolyze storage proteins (Martinez et al., 2019).

Comparing the results of the shotgun analyses among samples suggested that some of the proteins identified were also developmentally regulated. Thus, 11 proteases associated with mobilization of storage proteins were up-accumulated in T3 cotyledons and embryo axes, namely: 3 aminopeptidases (app2-qilexprot_22106, acylamino-acid-releasing enzyme-qilexprot_61138 and lap2-qilexprot_74150), 6 serine carboxypeptidases (scpl44-qilexprot_44690, EDA2-qilexprot_7150, CBP31-qilexprot_48367, UBP7-qilexprot_40121, UP3-qilexprot_16893 and SAT-qilexprot_52755) and 2 cysteine proteases (cathepsin H-qilexprot_78020 and CP protein-qilexprot_56434). Only after the function of a protein has been elucidated, can its role be interpreted in biological terms; in fact, transcriptomic and proteomic analyses by themselves do not suffice for this purpose since post-transcriptional and post-translational events rendering proteins inactive should always be considered. Enzymes should be assessed for activity. In this work, *in vitro* and in gel activity were assessed in extracts from embryo axes and cotyledons at a mature stage and upon germination. Gel activity tests provided few bands compared to the number of proteases identified by shotgun, revealing differences in the analytical potential of the used approaches and the complexity of the mechanisms regulating protein synthesis and activation. The differences in gel activity patterns, and those in shotgun analysis results, between samples confirmed that some identified proteins were tissue-specific or stage-specific.

Let us now discuss the organ and developmental pattern of some identified proteases and their putative biological role in terms of shotgun and gel activity data. Shotgun and protease activity analyses revealed more marked differences in embryo axes than there were in cotyledons. In gel activity tests, proteases spanned the 260–60 kDa *Mr* range in embryo axes, and the 260–70 kDa range in cotyledons. Nonspecific exo- and endopeptidases (*viz.*, amino and carboxyl isoforms) involved in storage protein mobilization were identified including leucine aminopeptidases and cathepsin H. A leucine aminopeptidase (LAP) was previously identified in mature, nongerminated *Q. ilex* seeds by Romero-Rodríguez et al. (2019). LAPs are highly conserved exopeptidases and among the most frequently used enzymes as gene markers in forest genetics (Rudin, 1976). LAP in *Picea abies* and other conifers exhibits high variability and tissue-unspecific activity in seeds, needles and pollen (Müller-Starck and Hüttermann, 1981). In this work, lap2 proteins were also abundant in band 1 from T3 cotyledons of gels activity. Also, cathepsin H-like protein was identified in viable *Q. ilex* seeds but not in nonviable seeds; as a result, this protein could be used as a putative marker of seed quality. Cathepsin H-like is an aleurain isolated from the aleurone of barley seeds (Rogers et al., 1985; Holwerda and Rogers, 1992). By using fluorescence activity-based probes on germinated *Arabidopsis* seeds, Lu et al. (2015) demonstrated dynamic activity in aleurain-like proteases, cathepsin B-like protein and vacuolar processing enzymes concomitantly with remobilization of seed storage proteins.

Some protein targeted proteases such as those that are ATP-dependent (Adam, 2007) were identified here and in previous studies (Balbuena et al., 2011; Romero-Rodríguez et al., 2019). Such proteases include members of the Clp family. For example, ClpA and ClpP are proteolytic subunits of the ATP-dependent Clp protease, which is found in the chloroplasts of higher plants, and plays essential roles in modulating the availability of short-lived regulatory proteins and in removing abnormal or damaged proteins (Shikanai et al., 2001). Besides its proteolytic role, ClpA functions as a chaperone independently of ClpP (Wickner et al., 1994). ATP-dependent ClpC proteases were specifically identified in germinated embryos of *Araucaria angustifolia*, indicating that this enzyme is synthesized during germination (Balbuena et al., 2011). Several Clp proteases were identified by shotgun in T3 embryo axes and cotyledons in this study (*viz.*, ClpA, qilexprot_77273; ClpP, qilexprot_58095; and ClpB, qilexprot_8964), as well as in the activity gels: band 1 (qilexprot_8964), bands 7 and 8 (qilexprot_8964, qilexprot_77273, qilexprot_50467, qilexprot_50), and band 9 (qilexprot_8964, qilexprot_77273). Overall, accumulation of these proteins at T3 indicates active mobilization of reserve proteins, some of which were present in mature (T1) fruits, albeit to a lesser extent. As previously was postulated by Romero-Rodríguez et al. (2019), in contrast to that occurs in orthodox seeds, in which all metabolic activity ceases in mature dry seeds, mature *Q. ilex* seeds have the machinery necessary for rapidly resuming metabolic activities and start the germination process.

Seed quality and viability can be assumed to depend on seed storage age and conditions, so assessing seed viability after sourcing may be useful. Despite great progress in seed biology, the sequence of cellular events that dictate whether seeds can germinate is poorly understood. Chaitanya et al. (2000) observed a gradual decline in total protein content due to corresponding increase in protease activity preceding loss of viability in *Shorea robusta* seeds after 6 days of harvest. In addition, substantially higher amounts of protease activity in embryonic axes than in the cotyledons might underline a special role played by embryonic axes in inducing protease activity in storage tissues (Chaitanya et al., 2000). In this work, the differential accumulation in tissues of proteins of the 26S proteasome and the Constitutive Photomorphogenesis 9 (COP9) signalosome (CSN) in embryo axes was striking (especially at T3). The protease complex encompassed by the 26S proteasome, and various types of E3 ligases ubiquitinating target proteins, are two key actors in the mechanism governing accumulation of regulatory proteins involved in phytohormone and light signaling pathways, and ultimately determining seed germination potential (Oracz and Stawska, 2016). CSN was originally identified in plants (Wei et al., 1994) and subsequently in all eukaryotic organisms. It is highly homologous to the lid sub-complex of the 26S proteasome, which is the major proteolysis machinery in eukaryotic cells, and involved in various cellular and developmental processes where it is believed to regulate ubiquitin proteasome-mediated protein degradation (Wei and Deng, 2003).

CSN regulates the Cullin–Ring ubiquitin ligase (CRL) complex, which is the main class of E3 ligase complexes in eukaryotes (Lydeard et al., 2013; Teixeira and Reed, 2013). The existence of super-complexes consisting of CSN, the 26S proteasome and cullin-based Ub ligases has been suggested (Peng et al., 2003; Huang et al., 2005). CSN5 subunit catalyzes the hydrolysis of NEDD8 proteins from CRL, so it is responsible for CRL deneddylation acting as a deactivator (Cope and Deshaies, 2003; Wei and Deng, 2003). According to Franciosini et al. (2015), COP9 is deactivated during the maturation of *Arabidopsis thaliana* embryos and subsequently reactivated at germination.

Interestingly, two CSN proteins (CSN5A, qilexprot_12955; and CSN1, qilexprot_63361), a cullin deneddylation protein (qilexprot_26596) and a neddylation protein (qilexprot_46052) were differentially identified between *Q. ilex* embryo axes stages, with the neddylation protein exclusively found in ENV and the deneddylation protein highly represented in ET3. In addition to the above-described proteins, a number of others belonging to the 26S proteasome complex—6 up-represented at ET3, and 4 at ENV and ET1—were identified. Seven other proteins were identified as proteasome subunit alpha (4) and proteasome subunit beta (3) accumulating in embryo axes many of which were also identified in gel activity bands from T3 embryo axes. Although no categorical explanation for the differential accumulation of proteins from the COP9 signalosome and the 26S proteasome at different developmental stages can be put forward, a clear connection exists with embryonic axes development in germinated seeds.

This suggests that degradation *via* CNS/26S proteasome systems early during *Q. ilex* seed germination is crucial for adequate seedling development. Subtilisin-like endopeptidase (qilexprot_1,462) was also highly accumulated in T3 embryo axes. This protein performs specific functions in plant development and in signaling cascades. For example, subtilisins play a major role in cuticle development during embryogenesis (Van der Hoorn, 2008).

A few other proteins were tissue-specific. Such was the case with the cysteine PIs qilexprot_51394, qilexprot_15999 and qilexprot_28688, which were exclusively identified in ET1 seeds—and also consistently identified in gel activity bands from ET1 seeds. Proteinase inhibitors may play a role as reserve proteins in plants (Richardson, 1991; Hansen et al., 2007) and be involved in plant defense mechanisms. Cysteine PIs were previously isolated from seeds of the *Enterolobium contortisiliquum* tree targeting larval growth inhibition in the pest *Collasobruchus maculatus* (Nunes et al., 2021). Several studies have shown feeding insects with transgenic plants that express proteinase inhibitors to delay insect growth and development, and to cause starvation and death (Koiwa et al., 1998; Schuler et al., 1998; Mosolov et al., 2001). These proteins may thus not only play a defensive role against a potential attack by pathogens at a maturate stage but also act as inhibitors of endogenous peptidases to regulate their activity. Three other proteins were specifically identified in NV seeds (*viz.*, the aspartic proteinases nepenthesin-qilexprot_63251, qilexprot_78560 and nepenthesin-qilexprot_60038). Aspartic proteases are known to play major roles in storage protein processing, nitrogen remobilization, biotic and abiotic stress responses, and senescence and programmed cell death (PCD; Ge et al., 2005; reviewed in Bekalu et al., 2020a). Nepenthesin was initially reported in pitcher fluid from the carnivorous plant *Nepenthes* (Vines, 1901). Apart from carnivorous plants, nepenthesins have also been found in *A. thaliana* leaves, stems, seeds and pods (Takahashi et al., 2008) and in 24 tissues during the life cycle of *Oryza sativa* (Chen et al., 2009) suggesting ubiquitous occurrence and multiple functions. Overexpression of the nepenthesin-1 gene in the endosperm of barley grains significantly reduces infection and disease progression of Fusarium head blight disease, and also accumulation of toxins from the fungus (Bekalu et al., 2020a,b). Although it is difficult to ascertain whether these proteins are involved in defense or senescence processes, or both, their absence from viable seeds could be used as an indicator of nonviability. Also, the number of identified proteins related to proteasome: 26S proteasome regulatory proteins, proteasome subunit alpha, proteasome subunit beta and neddylation, highly accumulated in NV could be indicating a senescence process.

CONCLUSION

Analyzing proteases and protease inhibitors (PIs) at different developmental stages of *Q. ilex* seeds by using a combination of shotgun proteomics and protease activity tests allowed the identification of a number of tissue- and stage-specific proteins in addition to others that were present in all systems. Comparing the omic and enzymatic activity results revealed differences in analytical potential between the two approaches but also useful complementariness. Accumulation of some proteases at the T3 germination stage such as the cathepsin H cysteine protease and Clp proteases among others, suggests active remobilization of reserve proteins; some proteases, however, were also active during maturation, which confirms the hypothesis that nonorthodox seeds such as *Q. ilex* acorns continue to be metabolically active at this stage. The large number of subunits of the CNS and 26S proteasome complex differentially identified in embryo axes of the several stages suggests that protein degradation *via* CNS/26S plays a major role early in germination, so it could be a useful indicator of seed viability. On the other hand, aspartic proteases such as nepenthesins were exclusively identified in NV seeds, so their presence could be used as indicator of nonviability. In addition, the specific accumulation of three cysteine protease inhibitors in mature embryo axes raises the possibility of their playing a protective role against potential attacks by pathogens during maturation. These results are important with a view to conserving recalcitrant native seeds as they reveal that some proteases and PIs can be useful as indicators of seed viability and quality, and also for biotechnological purposes.

DATA AVAILABILITY STATEMENT

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found in the article/ **Supplementary Material**.

AUTHOR CONTRIBUTIONS

MAC and JJ-N: conceptualization, supervision, and funding acquisition. ME, EB, TH-L, and MAC: methodology development and experimental design. ME, VG-S, and MAC: bioinformatics and statistical analysis. MAC and ME: writing—original draft preparation. MAC, ME, M-DR, and JJ-N: writing—review and editing. All authors have read and agreed to the published version of the manuscript.

FUNDING

This research was funded by the Spanish Ministry of Economy and Competitiveness in the framework of Projects BIO2015-64737-R and PID2019-109038RB-I00.

ACKNOWLEDGMENTS

MAC, MDR, and ME are grateful for award of a Ramón y Cajal (RYC-2017-23706), Juan de la Cierva-Incorporación (IJC2018-035272-I), and Juan de la Cierva-Formación (FJCI-2017-31613) contracts, respectively, by the Spanish Ministry of Science, Innovation and Universities. EB is grateful for a fellowship of the Argentinian National Research Council (CONICET) and the National University of Córdoba. TH-L thanks the University of Córdoba (Spain) for award of a contract under Project RYC-2017-23706.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpls.2022.907042/full#supplementary-material

Supplementary Figure S1 | Germination stages. T1: mature acorns prior to germination. T2, T3, T4: Germinated acorns when root tips were visibly emerging from the cotyledons (T2), their size reached 6.5 mm (T3) and it exceeded 20 mm (T4). The bar corresponds to 1 cm.

Supplementary Figure S2 | Classification of germination stages from a PCA plot of shotgun data. The PCA encompassed all proteins. E, embryo axis. C, cotyledon. NV: Nongerminated acorns after 4 weeks of germination. T1: mature acorns prior to germination. T3: germinated acorns when root tip size reached 6.5 mm.

REFERENCES

Adam, Z. (2007). "Protein stability and degradation in plastids in cell and molecular biology of plastids," in *Current Genetics*. ed. R. Bock (Berlin: springer), 315–338.

Adhikari, B., Marasini, B. P., Rayamajhee, B., Bhattarai, B. R., Lamichhane, G., Khadayat, K., et al. (2021). Potential roles of medicinal plants for the treatment of viral diseases focusing on COVID-19: A review. *Phytother. Res.* 35, 1298–1312. doi: 10.1002/ptr.6893

Al Shweiki, M. R., Mönchgesang, S., Majovsky, P., Thieme, D., Trutschel, D., and Hoehenwarter, W. (2017). Assessment of label-free quantification in discovery proteomics and impact of technological factors and natural variability of protein abundance. *J. Proteome Res.* 16, 1410–1424. doi: 10.1021/acs.jproteome.6b00645

Amaral, J., Correia, B., Escandón, M., Jesus, C., Serôdio, J., Valledor, L., et al. (2020). Temporal physiological response of pine to *Fusarium circinatum* infection is dependent on host susceptibility level: the role of ABA catabolism. *Tree Physiol.* 41, 801–816. doi: 10.1093/treephys/tpaa143

Angelovici, R., Galili, G., Fernie, A. R., and Fait, A. (2010). Seed desiccation: a bridge between maturation and germination. *Trends Plant Sci.* 15, 211–218. doi: 10.1016/j.tplants.2010.01.003

Balbuena, T., Jo, L., Pieruzzi, F. P., Dias, L. L. C., Silveira, V., Santa-Catarina, C., et al. (2011). Differential proteome analysis of mature and germinated embryos of *Araucaria angustifolia*. *Phytochemistry* 72, 302–311. doi: 10.1016/j.phytochem.2010.12.007

Barrett, A. J., Rawlings, N. D., and Woessner, J. F. (eds.) (1998). *Handbook of Proteolytic Enzymes*. London: Academic Press.

Bekalu, Z. E., Dionisio, G., and Brinch-Pedersen, A. H. (2020a). Molecular properties and new potentials of plant nepenthesins. *Plants (Basel, Switzerland)* 9:570. doi: 10.3390/plants9050570

Bekalu, Z. E., Krogh Madsen, C., Dionisio, G., Bæksted Holme, I., Jørgensen, L. N., S Fomsgaard, I., et al. (2020b). Overexpression of Nepenthesin HvNEP-1 in barley endosperm reduces Fusarium head blight and mycotoxin accumulation. *Agronomy* 10:0203. doi: 10.3390/agronomy10020203

Berjak, P., and Pammenter, N. W. (2008). From avicennia to zizania: seed recalcitrance in perspective. *Ann. Bot.* 101, 213–228.

Bond, J. S. (2019). Proteases: history, discovery, and roles in health and disease. *J. Biol. Chem.* 294, 1643–1651. doi: 10.1074/jbc.TM118.004156

Chaitanya, K. S. K., Keshavkant, S., and Naithani, S. C. (2000). Changes in total protein and protease activity in dehydrating recalcitrant Sal (*Shorea robusta*) seeds. *Silva Fennica* 34, 71–77. doi: 10.14214/sf.646

Chen, J., Ouyang, Y., Wang, L., Xie, W., and Zhang, Q. (2009). Aspartic proteases gene family in rice: gene structure and expression, predicted protein features and phylogenetic relation. *Gene* 442, 108–118. doi: 10.1016/j.gene.2009.04.021

Coêlho, D. F., Saturnino, T. P., Fernandes, F. F., Mazzola, P. G., Silveira, E., and Tambourgi, E. B. (2016). Azocasein substrate for determination of Proteolytic activity: reexamining a traditional method using Bromelain samples. *Biomed. Res. Int.* 2016, 1–6. doi: 10.1155/2016/8409183

Connor, K. F., and Sowa, S. (2002). "Recalcitrant behavior of temperate forest tree seeds: Storage, biochemistry, and physiology." in *Proceedings of the Biennial Southern Silvicultural Research Conference. Vol. 48*. (Asheville, NC: U.S. Department of Agriculture, Forest Service), 47–50.

Cope, G. A., and Deshaies, R. J. (2003). COP9 signalosome: a multifunctional regulator of SCF and other cullin-based ubiquitin ligases. *Cell* 114, 663–671. doi: 10.1016/s0092-8674(03)00722-0

Escandón, M., Castillejo, M. Á., Jorrín-Novo, J. V., and Rey, M.-D. (2021a). Molecular research on stress responses in *Quercus* spp.: From classical biochemistry to systems biology through Omics analysis. *Forests* 12:0364. doi: 10.3390/f12030364

Escandón, M., Jorrín-Novo, J. V., and Castillejo, M. Á. (2021b). Application and optimization of label-free shotgun approaches in the study of *Quercus ilex*. *J. Proteome* 233:104082. doi: 10.1016/j.jprot.2020.104082

Fan, Y., Yang, W., Yan, Q., Chen, C., and Li, J. (2019). Genome-wide identification and expression analysis of the protease inhibitor gene families in tomato. *Genes (Basel)* 11:1. doi: 10.3390/genes11010001

Franciosini, A., Moubayidin, L., Du, K., Matari, N. H., Boccaccini, A., Butera, S., et al. (2015). The COP9 SIGNALOSOME is required for postembryonic meristem maintenance in *Arabidopsis thaliana*. *Mol. Plant* 8, 1623–1634. doi: 10.1016/j.molp.2015.08.003

García-Lorenzo, M., Sjödin, A., Jansson, S., and Funk, C. (2006). Protease gene families in *Populus* and *Arabidopsis*. *BMC Plant Biol.* 6:30. doi: 10.1186/1471-2229-6-30

Ge, X., Dietrich, C., Matsuno, M., Li, G., Berg, H., and Xia, Y. (2005). An Arabidopsis aspartic protease functions as an anti-cell-death component in reproduction and embryogenesis. *EMBO Rep.* 6, 282–288. doi: 10.1038/sj.embor.7400357

Gómez-Gálvez, I., Sánchez-Lucas, R., San-Eufrasio, B., Rodríguez de Francisco, L. E., Maldonado-Alconada, A. M., Fuentes-Almagro, C., et al. (2020). "Optimizing shotgun proteomics analysis for a confident protein identification and quantitation in orphan plant species: the case of Holm oak (*Quercus ilex*)," in *Plant Proteomics. Vol. 2139*. eds. J. V. Jorrin-Novo, L. Valledor, M. A. Castillejo and M. D. Rey (New York, NY: Springer), 157–168.

Guerrero-Sánchez, V. M., Castillejo, M. Á., López-Hidalgo, C., Alconada, A. M. M., Jorrín-Novo, J. V., and Rey, M.-D. (2021). Changes in the transcript and protein profiles of *Quercus ilex* seedlings in response to drought stress. *J. Proteome* 243:104263. doi: 10.1016/j.jprot.2021.104263

Guerrero-Sánchez, V. M., Maldonado-Alconada, A. M., Amil-Ruiz, F., and Jorrin-Novo, J. V. (2017). Holm oak (*Quercus ilex*) transcriptome. De novo sequencing and assembly analysis. *Front. Mol. Biosci.* 4:70. doi: 10.3389/fmolb.2017.00070

Guerrero-Sánchez, V. M., Maldonado-Alconada, A. M., Amil-Ruiz, F., Verardi, A., Jorrín-Novo, J. V., and Rey, M.-D. (2019). Ion torrent and lllumina, two complementary RNA-seq platforms for constructing the holm oak (*Quercus ilex*) transcriptome. *PLoS One* 14:e0210356. doi: 10.1371/journal.pone.0210356

Hansen, D., Macedo-Ribeiro, S., Veríssimo, P., Yoo Im, S., Sampaio, M. U., and Oliva, M. L. V. (2007). Crystal structure of a novel cysteinless plant Kunitz-type protease inhibitor. *Biochem. Biophys. Res. Commun.* 360, 735–740. doi: 10.1016/j.bbrc.2007.06.144

Haq, S. K., Atif, S. M., and Khan, R. H. (2004). Protein proteinase inhibitor genes in combat against insects, pests, and pathogens: natural and engineered phytoprotection. *Arch. Biochem. Biophys.* 431, 145–159. doi: 10.1016/j.abb.2004.07.022

Hellinger, R., and Gruber, C. W. (2019). Peptide-based protease inhibitors from plants. *Drug Discov. Today* 24, 1877–1889. doi: 10.1016/j.drudis.2019.05.026

Heussen, C., and Dowdle, E. B. (1980). Electrophoretic analysis of plasminogen activators in polyacrylamide gels containing sodium dodecyl sulfate and copolymerized substrates. *Anal. Biochem.* 102, 196–202. doi: 10.1016/0003-2697(80)90338-3

Holwerda, B. C., and Rogers, J. C. (1992). Purification and characterization of aleurain: a plant thiol protease functionally homologous to mammalian cathepsin h. *Plant Physiol.* 99, 848–855. doi: 10.1104/pp.99.3.848

Hörger, A. C., and van der Hoorn, R. A. L. (2013). The structural basis of specific protease-inhibitor interactions at the plant-pathogen interface. *Curr. Opin. Struct. Biol.* 23, 842–850. doi: 10.1016/j.sbi.2013.07.013

Huang, X., Hetfeld, B. K. J., Seifert, U., Kähne, T., Kloetzed, P. M., Naumann, M., et al. (2005). Consequences of COP9 signalosome and 26S proteasome interaction. *FEBS J.* 272, 3909–3917. doi: 10.1111/j.1742-4658.2005.04807.x

Hüynck, J., Kaschani, F., van der Linde, K., Ziemann, S., Müller, A. N., Colby, T., et al. (2019). Proteases underground: analysis of the maize root Apoplast identifies organ specific papain-Like cysteine protease activity. *Front. Plant Sci.* 10:473. doi: 10.3389/fpls.2019.00473

Joët, T., Ourcival, J. M., and Dussert, S. (2013). Ecological significance of seed desiccation sensitivity in *Quercus ilex*. *Ann. Bot.* 111, 693–701. doi: 10.1093/aob/mct025

Koiwa, H., Shade, R. E., Zhu-Salzman, K., Subramanian, L., Murdock, L. L., Nielsen, S. S., et al. (1998). Phage display selection can differentiate insecticidal activity of soybean cystatins. *Plant J.* 14, 371–379. doi: 10.1046/j.1365-313x.1998.00119.x

Kolde, R. (2015). Pheatmap: Pretty Heatmaps. *R package version 0.7.7*. Available at: ftp://cran.r-project.org/pub/R/web/packages/pheatmap/ (Accessed September, 2021).

Lex, A., Gehlenborg, N., Strobelt, H., Vuillemot, R., and Pfister, H. (2014). UpSet: visualization of intersecting sets. *IEEE Trans. Vis. Comput. Graph.* 20, 1983–1992. doi: 10.1109/TVCG.2014.2346248

López-Hidalgo, C., Trigueros, M., Menéndez, M., and Jorrin-Novo, J. V. (2021). Phytochemical composition and variability in *Quercus ilex* acorn morphotypes as determined by NIRS and MS-based approaches. *Food Chem.* 338:127803. doi: 10.1016/j.foodchem.2020.127803

Lu, H., Chandrasekar, B., Oeljeklaus, J., Misas-Villamil, J. C., Wang, Z., Shindo, T., et al. (2015). Subfamily-specific fluorescent probes for cysteine proteases display dynamic protease activities during seed germination. *Plant Physiol.* 168, 1462–1475. doi: 10.1104/pp.114.254466

Lydeard, J. R., Schulman, B. A., and Harper, J. W. (2013). Building and remodelling Cullin-RING E3 ubiquitin ligases. *EMBO Rep.* 14, 1050–1061. doi: 10.1038/embor.2013.173

Maldonado, A. M., Echevarría-Zomeño, S., Jean-Baptiste, S., Hernández, M., and Jorrín-Novo, J. V. (2008). Evaluation of three different protocols of protein extraction for *Arabidopsis thaliana* leaf proteome analysis by two-dimensional electrophoresis. *J. Proteome* 71, 461–472. doi: 10.1016/j.jprot.2008.06.012

Martínez, M., Cambra, I., González-Melendi, P., Santamaría, M. E., and Díaz, I. (2012). C1A cysteine-proteases and their inhibitors in plants. *Physiol. Plant.* 145, 85–94. doi: 10.1111/j.1399-3054.2012.01569.x

Martinez, M., Gómez-Cabellos, S., Giménez, M. J., Barro, F., Diaz, I., and Diaz-Mendoza, M. (2019). Plant proteases: From key enzymes in germination to allies for fighting human gluten-related disorders. *Front. Plant Sci.* 10:721. doi: 10.3389/fpls.2019.00721

McManus, M. T., Ryan, S. N., and Laing, W. A. (1999). The functions of proteinase inhibitors in seeds. *Seed Symposium* 1999, 3–13.

Mosolov, V. V., Grigor'eva, L. I., and Valueva, T. A. (2001). Plant proteinase inhibitors as multifunctional proteins. *Appl. Biochem. Microbiol.* 37, 545–551. doi: 10.1023/A:1012352914306

Müller-Starck, G., and Hüttermann, A. (1981). Aminopeptidases in seeds of *picea abies* (L.) karst.: characterization of leucine aminopeptidase by molecular properties and inhibitors. *Biochem. Genet.* 19, 1247–1259. doi: 10.1007/BF00484577

Neuhoff, V., Arold, N., Taube, D., and Ehrhardt, W. (1988). Improved staining of proteins in polyacrylamide gels including isoelectric focusing gels with clear background at nanogram sensitivity using Coomassie brilliant blue G-250 and R-250. *Electrophoresis* 9, 255–262. doi: 10.1002/elps.1150090603

Nunes, N. N. S., Ferreira, R. S., de Sá, L. F. R., de Oliveira, A. E. A., and Oliva, M. L. V. (2021). A novel cysteine proteinase inhibitor from seeds of *Enterolobium contortisiliquum* and its effect on *Callosobruchus maculatus* larvae. *Biochem. Biophys. Reports* 25:100876. doi: 10.1016/j.bbrep.2020.100876

Oracz, K., and Stawska, M. (2016). Cellular recycling of proteins in seed dormancy alleviation and germination. *Front. Plant Sci.* 7:1128. doi: 10.3389/fpls.2016.01128

Pammenter, N. W., and Berjak, P. (2014). Physiology of desiccation-sensitive (recalcitrant) seeds and the implications for cryopreservation. *Int. J. Plant Sci.* 175, 21–28. doi: 10.1086/673302

Pasquini, S., Braidot, E., Petrussa, E., and Vianello, A. (2011). Effect of different storage conditions in recalcitrant seeds of holm oak (*Quercus ilex* L.) during germination. *Seed Sci. Technol.* 39, 165–177. doi: 10.15258/sst.2011.39.1.14

Pasquini, S., Mizzau, M., Petrussa, E., Braidot, E., Patui, S., Gorian, F., et al. (2012). Seed storage in polyethylene bags of a recalcitrant species (Quercus ilex): analysis of some bio-energetic and oxidative parameters. *Acta Physiol Plant.* 34, 1963–1974. doi: 10.1007/s11738-012-0996-9

Peng, Z., Shen, Y., Feng, S., Wang, X., Chitteti, B. N., Vierstra, R. D., et al. (2003). Evidence for a physical association of the COP9 signalosome, the proteasome, and specific SCF E3 ligases in vivo. *Curr. Biol.* 13, R504–R505. doi: 10.1016/S0960-9822(03)00439-1

Perez-Riverol, Y., Csordas, A., Bai, J., Bernal-Llinares, M., Hewapathirana, S., Kundu, D. J., et al. (2018). The PRIDE database and related tools and resources in 2019: improving support for quantification data. *Nucleic Acids Res.* 47, D442–D450. doi: 10.1093/nar/gky1106

Plomión, C., Aury, J., Amselem, J., Leroy, T., Murat, F., Duplessis, S., et al. (2018). Oak genome reveals facets of long lifespan. *Nat. Plants* 4, 440–452. doi: 10.1038/s41477-018-0172-3

Puente, X. S., and López-Otín, C. (2004). A genomic analysis of rat proteases and protease inhibitors. *Genome Res.* 14, 609–622. doi: 10.1101/gr.1946304

Ramos, A. M., Usie, A., Barbosa, P., Barros, P. M., Capote, T., Chaves, I., et al. (2018). The draft genome sequence of cork oak. *Sci. Data* 5:180069. doi: 10.1038/sdata.2018.69

Rawlings, N. D., and Barrett, A. J. (1993). Evolutionary families of peptidases. *Biochem. J.* 290, 205–218. doi: 10.1042/bj2900205

Rawlings, N. D., Barrett, A. J., and Bateman, A. (2010). MEROPS: the peptidase database. *Nucleic Acids Res.* 38, D227–D233. doi: 10.1093/nar/gkp971

R Core Team (2018). R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna. Available at: https://www.R-project.org

Rey, M.-D., Castillejo, M. Á., Sánchez-Lucas, R., Guerrero-Sanchez, V. M., López-Hidalgo, C., Romero-Rodríguez, C., et al. (2019). Proteomics, holm oak (*Quercus ilex* L.) and other recalcitrant and orphan forest tree species: how do they see each other? *Int. J. Mol. Sci.* 20:692. doi: 10.3390/ijms20030692

Richardson, M. (1991). "Seed storage proteins: The enzyme inhibitors," in *Methods in Plant Biochemistry*. ed. L. J. Rogers (New York, NY: Academic Press), 259–305.

Roberts, E. H. (1973). Predicting the storage life of seeds. *Seed Sci. Technol.* 1, 499–514.

Rogers, J. C., Dean, D. A., and Heck, G. R. (1985). Aleurain: a barley thiol protease closely related to mammalian cathepsin H. *Proc. Natl. Acad. Sci. U. S. A.* 82, 6512–6516. doi: 10.1073/pnas.82.19.6512

Romero-Rodríguez, M. C., Abril, N., Sánchez-Lucas, R., and Jorrín-Novo, J. V. (2015). Multiplex staining of 2-DE gels for an initial phosphoproteome analysis of germinating seeds and early grown seedlings from a non-orthodox specie: *Quercus ilex* L. subsp. ballota [Desf.] Samp. *Front. Plant Sci.* 6:620. doi: 10.3389/fpls.2015.00620

Romero-Rodríguez, C., Archidona-Yuste, A., Abril, N., Gil-Serrano, A. M., Meijón, M., and Jorrín-Novo, J. V. (2018). Germination and early seedling development in *Quercus ilex* recalcitrant and non-dormant seeds: targeted transcriptional, hormonal, and sugar analysis. *Front. Plant Sci.* 9:1508. doi: 10.3389/fpls.2018.01508

Romero-Rodríguez, M. C., Jorrín-Novo, J. V., and Castillejo, M. A. (2019). Toward characterizing germination and early growth in the non-orthodox forest tree species *Quercus ilex* through complementary gel and gel-free proteomic analysis of embryo and seedlings. *J. Proteome* 197, 60–70. doi: 10.1016/j.jprot.2018.11.003

Romero-Rodríguez, M., Maldonado-Alconada, A., Valledor, L., and Jorrin-Novo, J. (2014). Back to Osborne. Sequential protein extraction and LC-MS analysis for the characterization of the holm oak seed proteome. *Methods Mol. Biol.* 1072, 379–389. doi: 10.1007/978-1-62703-631-3_27

Rudin, D. (1976). "Biochemical genetics and selection, application of isozymes in tree breeding." *Proceedings of IUFRO Joint Meeting on Advanced Generation Breeding*. 145–164.

Ryan, C. A. (1990). Protease inhibitors in plants: genes for improving defenses Against insects and pathogens. *Annu. Rev. Phytopathol.* 28, 425–449. doi: 10.1146/annurev.py.28.090190.002233

San-Eufrasio, B., Bigatton, E., Guerrero-Sanchez, V., Chaturvedi, P., Jorrin-Novo, J., Rey, M.-D., et al. (2021). Proteomics data analysis for the identification of proteins and derived proteotypic peptides of potential use as putative drought tolerance markers for *Quercus ilex*. *Int. J. Mol. Sci.* 22:3191. doi: 10.3390/ijms22063191

Santamaría, M. E., Diaz-Mendoza, M., Diaz, I., and Martinez, M. (2014). Plant protein peptidase inhibitors: an evolutionary overview based on comparative genomics. *BMC Genomics* 15:812. doi: 10.1186/1471-2164-15-812

Santamaría, M. E., Hernández-Crespo, P., Ortego, F., Grbić, V., Grbic, M., Díaz, I., et al. (2012). Cysteine peptidases and their inhibitors in Tetranychus urticae: a comparative genomic approach. *BMC Genomics* 13:307. doi: 10.1186/1471-2164-13-307

Schaller, A. (2004). A cut above the rest: the regulatory function of plant proteases. *Planta* 220, 183–197. doi: 10.1007/s00425-004-1407-2

Schuler, T. H., Poppy, G. M., Kerry, B. R., and Denholm, I. (1998). Insect-resistant transgenic plants. *Trends Biotechnol.* 16, 168–175. doi: 10.1016/S0167-7799(97)01171-2

Sghaier-Hammami, B., Hammami, S. B. M., Baazaoui, N., Gómez-Díaz, C., and Jorrín-Novo, J. V. (2020). Dissecting the seed maturation and germination processes in the non-orthodox *Quercus ilex* species based on protein signatures as revealed by 2-DE coupled to MALDI-TOF/TOF proteomics strategy. *Int. J. Mol. Sci.* 21:4870. doi: 10.3390/ijms21144870

Sghaier-Hammami, B., Castillejo, M. Á., Baazaoui, N., Jorrín-Novo, J. V., and Escandón, M. (2021). GeLC-Orbitrap/MS and 2-DE-MALDI-TOF/TOF comparative proteomics analysis of seed cotyledons from the non-orthodox *Quercus ilex* tree species. *J. Proteome* 233:104087. doi: 10.1016/j.jprot.2020.104087

Sghaier-Hammami, B., Redondo-López, I., Valero-Galvàn, J., and Jorrín-Novo, J. V. (2016). Protein profile of cotyledon, tegument, and embryonic axis of mature acorns from a non-orthodox plant species: *Quercus ilex*. *Planta* 243, 369–396. doi: 10.1007/s00425-015-2404-3

Shikanai, T., Shimizu, K., Ueda, K., Nishimura, Y., Kuroiwa, T., and Hashimoto, T. (2001). The chloroplast clpP gene, encoding a proteolytic subunit of ATP-dependent protease, is indispensable for chloroplast development in tobacco. *Plant Cell Physiol.* 42, 264–273. doi: 10.1093/pcp/pce031

Simova-Stoilova, L. P., Romero-Rodriguez, M. C., Sanchez-Lucas, R., Navarro-Cerrillo, R. M., Medina-Aunon, J. A., and Jorrin-Novo, J. V. (2015). 2-DE proteomics analysis of drought treated seedlings of *Quercus ilex* support a root active strategy for metabolic adaptation in response to water shortage. *Front. Plant Sci.* 6:627. doi: 10.3389/fpls.2015.00627

Sork, V. L., Cokus, S. J., Fitz-Gibbon, S. T., Zimin, A. V., Puiu, D., Garcia, J. A., et al. (2022). High-quality genome and methylomes illustrate features underlying evolutionary success of oaks. *Nat. Commun.* 13:2047. doi: 10.1038/s41467-022-29584-y

Sork, V., Fitz-Gibbon, S., Puiu, D., Crepeau, M., Gugger, P. F., Sherman, R. M., et al. (2016). First draft assembly and annotation of the genome of a California endemic oak *Quercus lobata* Née (Fagaceae). *G3 Genes|Genomes|Genetics* 6, 3485–3495. doi: 10.1534/g3.116.030411

Takahashi, K., Niwa, H., Yokota, N., Kubota, K., and Inoue, H. (2008). Widespread tissue expression of nepenthesin-like aspartic protease genes in *Arabidopsis thaliana*. *Plant Physiol. Biochem.* 46, 724–729. doi: 10.1016/j.plaphy.2008.04.007

Tan-Wilson, A., and Wilson, K. A. (2012). Mobilization of seed protein reserves. *Physiol. Plant.* 145, 140–153. doi: 10.1111/j.1399-3054.2011.01535.x

Teixeira, L. K., and Reed, S. I. (2013). Ubiquitin ligases and cell cycle control. *Annu. Rev. Biochem.* 82, 387–414. doi: 10.1146/annurev-biochem-060410-105307

Valero-Galván, J., González-Fernández, R., and Jorrin-Novo, J. V. (2021a). Interspecific variation between the American *Quercus virginiana* and Mediterranean *Quercus* species in terms of seed nutritional composition, phytochemical content, and antioxidant activity. *Molecules* 26:2351. doi: 10.3390/molecules26082351

Valero-Galván, J., Jorrin-Novo, J. V., Gómez Cabrera, A., Ariza, D., García-Olmo, J., and Navarro-Cerrillo, R. M. (2021b). Population variability based on the morphometry and chemical composition of the acorn in holm oak (*Quercus ilex* subsp. *ballota* [Desf.] Samp.). *Eur J Forest Res* 131, 893–904. doi: 10.1007/s10342-011-0563-8

Valero-Galván, J., Valledor, L., Navarro-Cerrillo, R. M., Gil Pelegrín, E., and Jorrín-Novo, J. V. (2011). Studies of variability in holm oak (*Quercus ilex* subsp. *ballota* [Desf.] Samp.) through acorn protein profile analysis. *J. Proteome* 74, 1244–1255. doi: 10.1016/j.jprot.2011.05.003

Valledor, L., Romero-Rodriguez, M. C., and Jorrin-Novo, J. V. (2014). Standardization of data processing and statistical analysis in comparative plant proteomics experiment. *Methods Mol. Biol.* 1072, 51–60. doi: 10.1007/978-1-62703-631-3_5

Van der Hoorn, R. A. L. (2008). Plant proteases: From phenotypes to molecular mechanisms. *Annu. Rev. Plant Biol.* 59, 191–223. doi: 10.1146/annurev.arplant.59.032607.092835

van der Hoorn, R. A. L., and Klemenčič, M. (2021). Plant proteases: from molecular mechanisms to functions in development and immunity. *J. Exp. Bot.* 72, 3337–3339. doi: 10.1093/jxb/erab129

Vaseva, I., Sabotič, J., Sustar-Vozlic, J., Meglic, V., Kidrič, M., Demirevska, K., et al. (2011). "The response of plants to drought stress: The role of dehydrins, chaperones, proteases and protease inhibitors in maintaining cellular protein function," in *Droughts: New Research*. eds. D. F. Neves and D. Sanz (Nova Science), 1–46.

Vines, S. H. (1901). The Proteolytic enzyme of Nepenthes (III). *Ann. Bot.* 15, 563–573. doi: 10.1093/oxfordjournals.aob.a088838

Volpicella, M., Leoni, C., Costanza, A., Leo, F.De, Gallerani, R., and Ceci, L. R. (2012). Cystatins, serpins and other families of protease inhibitors in plants. *Curr. Protein Pept. Sci.* 12, 386–398. doi: 10.2174/138920311796391098

Wang, W.-Q., Liu, S.-J., Song, S.-Q., and Møller, I. M. (2015). Proteomics of seed development, desiccation tolerance, germination and vigor. *Plant Physiol. Biochem. PPB* 86, 1–15. doi: 10.1016/j.plaphy.2014.11.003

Wang, W., Vignani, R., Scali, M., and Cresti, M. (2006). A universal and rapid protocol for protein extraction from recalcitrant plant tissues for proteomic analysis. *Electrophoresis* 27, 2782–2786. doi: 10.1002/elps.200500722

Wei, N., Chamovitz, D. A., and Deng, X. W. (1994). Arabidopsis COP9 is a component of a novel signaling complex mediating light control of development. *Cell* 78, 117–124. doi: 10.1016/0092-8674(94)90578-9

Wei, N., and Deng, X. W. (2003). The COP9 signalosome. *Annu. Rev. Cell Dev. Biol.* 19, 261–286. doi: 10.1146/annurev.cellbio.19.111301.112449

Wickner, S., Gottesman, S., Skowyra, D., Hoskins, J., McKenney, K., and Maurizi, M. R. (1994). A molecular chaperone, ClpA, functions like DnaK and DnaJ. *Proc. Natl. Acad. Sci. U. S. A.* 91, 12218–12222. doi: 10.1073/pnas.91.25.12218

Wyse, S. V., and Dickie, J. B. (2017). Predicting the global incidence of seed desiccation sensitivity. *J. Ecol.* 105, 1082–1093. doi: 10.1111/1365-2745.12725

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may

be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

frontiers | Frontiers in Plant Science

ORIGINAL RESEARCH
published: 07 July 2022
doi: 10.3389/fpls.2022.937392

Comparative Proteome and Phosphoproteome Analyses Reveal Different Molecular Mechanism Between Stone Planting Under the Forest and Greenhouse Planting of *Dendrobium huoshanense*

Liping Wu[1], Xiaoxi Meng[2], Huizhen Huang[3], Yingying Liu[4], Weimin Jiang[3], Xinglong Su[1], Zhaojian Wang[1], Fei Meng[1], Longhai Wang[5], Daiyin Peng[1,6,7] and Shihai Xing[1,6,8]**

[1] College of Pharmacy, Anhui University of Chinese Medicine, Hefei, China, [2] Department of Horticultural Science, University of Minnesota, St. Paul, MN, United States, [3] Hunan Key Laboratory for Conservation and Utilization of Biological Resources in the Nanyue Mountainous Region, College of Life Sciences and Environment, Hengyang Normal University, Hengyang, China, [4] College of Humanities and International Education Exchange, Anhui University of Chinese Medicine, Hefei, China, [5] School of Integrated Chinese and Western Medicine, Anhui University of Chinese Medicine, Hefei, China, [6] Institute of Traditional Chinese Medicine Resources Protection and Development, Anhui Academy of Chinese Medicine, Hefei, China, [7] MOE-Anhui Joint Collaborative Innovation Center for Quality Improvement of Anhui Genuine Chinese Medicinal Materials, Hefei, China, [8] Anhui Province Key Laboratory of Research and Development of Chinese Medicine, Hefei, China

The highly esteemed Chinese herb, *Dendrobium huoshanense*, whose major metabolites are polysaccharides and alkaloids, is on the verge of extinction. The stone planting under the forest (SPUF) and greenhouse planting (GP) of *D. huoshanense* are two different cultivation methods of pharmaceutical Dendrobium with significantly differences in morphology, metabolites content and composition, and medication efficacy. Here, we conducted proteomics and phosphoproteomics analyses to reveal differences in molecular mechanisms between SPUF and GP. We identified 237 differentially expressed proteins (DEPs) between the two proteomes, and 291 modification sites belonging to 215 phosphoproteins with a phosphorylation level significantly changed (PLSC) were observed. GO, KEGG pathway, protein domain, and cluster analyses revealed that these DEPs were mainly localized in the chloroplast; involved in processes such as posttranslational modification, carbohydrate transport and metabolism, and secondary metabolite biosynthesis; and enriched in pathways mainly including linoleic acid metabolism, plant-pathogen interactions, and phenylpropanoid, cutin, suberin, and wax biosynthesis. PLSC phosphoproteins were mainly located in the chloroplast, and highly enriched in responses to different stresses and signal transduction mechanisms through protein kinase and phosphotransferase activities. Significant differences between SPUF and GP were observed by mapping the DEPs and phosphorylated proteins to photosynthesis and polysaccharide and

OPEN ACCESS

Edited by:
Pingfang Yang,
Hubei University, China

Reviewed by:
Xiaojian Yin,
China Pharmaceutical University, China
Mingquan Guo,
Wuhan Botanical Garden (CAS), China

***Correspondence:**
Daiyin Peng
pengdy@ahtcm.edu.cn
Shihai Xing
xshshihai@163.com

Specialty section:
This article was submitted to Plant Proteomics and Protein Structural Biology, a section of the journal Frontiers in Plant Science

Received: *06 May 2022*
Accepted: *20 June 2022*
Published: *07 July 2022*

Citation:
Wu L, Meng X, Huang H, Liu Y, Jiang W, Su X, Wang Z, Meng F, Wang L, Peng D and Xing S (2022) Comparative Proteome and Phosphoproteome Analyses Reveal Different Molecular Mechanism Between Stone Planting Under the Forest and Greenhouse Planting of Dendrobium huoshanense. Front. Plant Sci. 13:937392. doi: 10.3389/fpls.2022.937392

alkaloid biosynthesis pathways. Phosphorylation characteristics and kinase categories in *D. huoshanense* were also clarified in this study. We analyzed different molecular mechanisms between SPUF and GP at proteomic and phosphoproteomic levels, providing valuable information for the development and utilization of *D. huoshanense*.

Keywords: proteome, phosphoproteome, stone planting under the forest, greenhouse planting, *Dendrobium huoshanense*

INTRODUCTION

Dendrobium huoshanense C.Z. Tang et S.J. Cheng (Huoshan Shihu, 霍山石斛, "石" means stone) (**Figure 1A**), commonly known as "Mihu," is a perennial, aerial, and epiphytic herb that grows on stone (**Figure 1B**) and belongs to the family Orchidaceae (Wang et al., 2019). It is commonly used as a valuable high-end medicinal herb and is only found in the northern mountain regions of the Yangtze River in China. The *D. huoshanense* stem has medicinal properties (**Figure 1C**) and fresh stems are often processed into "Huoshan Fengdou" (**Figure 1D**; Niu et al., 2018), the main commodity in markets. The annual production of *D. huoshanense* exceeded 40 tons in 2020 in China with a total output value of 2 billion Chinese Yuan (Chen and Xia, 2020). Increasing studies have shown that *D. huoshanense* possesses considerable pharmacological activities, involved in cataract prevention (Ge et al., 2018), immunomodulation, tumor proliferation inhibition, decreases in blood lipid and glucose levels (Wang et al., 2014), and hepatoprotection (Tian et al., 2015). Moreover, *D. huoshanense* has been used as a functional food ingredient to produce tea drinks, porridges, and soups, *D. huoshanense* comprises medicinal, nutritional, and ornamental functions and has great prospects for further development (Hao et al., 2018).

The active medicinal components of *D. huoshanense* are very complex, including polysaccharides (Deng et al., 2016), alkaloids (Liang et al., 2019), flavonoids (Chang et al., 2010), and other compounds. Polysaccharide is one of the most important active components of medicinal plants and its pharmacological activities were influenced greatly by its content and monosaccharide compositions (Yuan et al., 2018). Therefore, the content of total polysaccharide and main monosaccharides are the essential quality evaluation standards of plant medicines, especially for plants with polysaccharides as the main active components, such as *D. huoshanense* (Tian et al., 2013). Studies have shown that the major monosaccharide components of polysaccharides in *D. huoshanense* and *Dendrobium officinale* are glucose and mannose (Xing et al., 2015; Hao et al., 2018). The mannose-containing polysaccharides obtained from natural plants exhibit higher bioactivities than those of glucose-containing polysaccharides and other monosaccharides (Yu et al., 2018). Studies have shown that alkaloids are also important bioactive compounds in *D. huoshanense* (Jin et al., 2016).

The wild *D. huoshanense* is often epiphytic to rocks and tree trunks under forest at an altitude of 200–1,200 m (Yuan et al., 2018; Chen et al., 2020), can grow well on wet stone walls facing northwest and northeast or cliff stones with waterfalls, and often forms small communities with mosses and lichens (Chen et al., 2020). Due to the unique growth environment, slow growth rates, and excessive manual collection, the wild resources of *D. huoshanense* are seriously threatened and cannot meet market demand (Jin et al., 2016). Thus, the stone planting under the forest (SPUF) and greenhouse planting (GP) of *D. huoshanense* have already become its major pharmaceutical source. SPUF restored the wild growth environment of *Dendrobium* to a certain extent, broad-leaved forest, coniferous forest or coniferous broad-leaved mixed forest with good ventilation, moderate density and thick trunk in the hillside were selected, and the canopy closure is 0.4–0.6, stones and moss are used as substrates (**Figure 1E**). GP was evolved by controlling the temperature (room temperature), humidity (80%), ventilation in a greenhouse. The substrate is a mixture of stone and bark in a certain proportion. In summer, a sunshade is used to prevent direct sunlight. In winter, a plastic film is installed for thermal insulation. Spray and ventilation and cooling facilities are set inside to control the external environment to make it suitable for the growth of *Dendrobium* (**Figure 1F**). The two cultivation modes have different environmental conditions including light, temperature, soil conditions, pH, humidity, etc., which consequently affect the accumulation of primary and secondary metabolites, and ultimately lead to a different chemical composition and efficacy (Yang et al., 2018). Studies had established the characteristic chromatogram of flavonoids in wild and cultivated *D. huoshanense* by High performance liquid chromatography (HPLC) characteristic chromatogram analysis technology, and found that there was difference in flavonoid content under different cultivation modes (Wei et al., 2014). The results also showed significant differences in the content of polysaccharides and ethanol extracts under different cultivation patterns (Chen et al., 2015). Moreover, long-term exposure to a harsh environment might contribute to the accumulation

Abbreviations: SPUF, stone planting under the forest; GP, greenhouse planting; DEPs, differentially expressed proteins; PLSC, phosphorylation level significantly changed; PTM, post-translational modification; RACK1A, receptor for activated C kinase 1A; MAPK, mitogen-activated protein kinase; ROS, reactive oxygen species; Thr, threonine; Ser, serine; Tyr, tyrosine; DTT, dithiothreitol; PMP, 1-phenyl-3-methyl-5-pyrazolone; TBST, tris-buffered saline with Tween®; TEAB, triethylammonium bicarbonate; IMAC, immobilized metal affinity chromatography; AGC, automatic gain control; FDR, false discovery rate; GO, gene ontology; KEGG, Kyoto Encyclopedia of Genes and Genomes; FC, fold change; START, star-related lipid-transfer; PLAT/LH2, lipase/lipooxigenase domain; BP, biological process; CC, cellular component; MF, molecular function; RT-qPCR, real-time fluorescent quantitative PCR; FBP, fructose-1,6-bisphosphatase; petH, ferredoxin–NADP reductase; PPC, phosphoenolpyruvate carboxylase; GAPA, glyceraldehyde-3-phosphate dehydrogenase A; GAPDH, glyceraldehyde-3-phosphate dehydrogenase; BPGA, 1,3-bisphosphoglyceric acid; UXS1, UDP-arabinose 4-epimerase 3 isoform; AXS, UDP-D-xylose synthase; GPI, glucose-6-phosphate isomerase; Sus, sucrose synthase; CMK, 4-diphosphocytidyl-2-C-methyl-D-erythritol kinase; 10HGO, mannitol dehydrogenase.

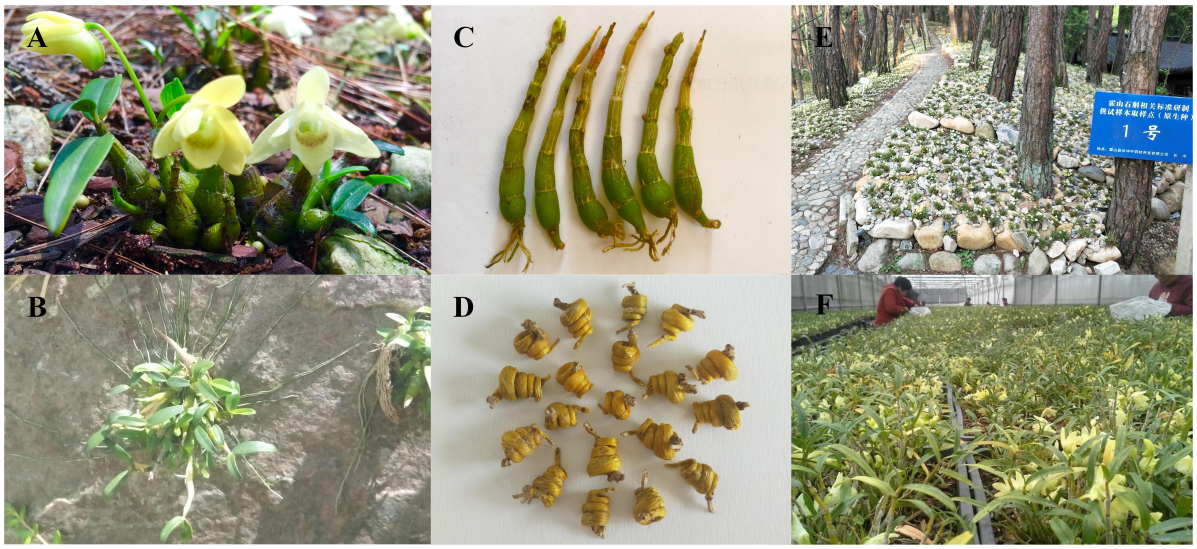


FIGURE 1 | Figures of the of *Dendrobium huoshanense* original plant, its habitat, medicinal part, processing products, and cultivating pattern. **(A)** The original plant of *D. huoshanense*. **(B)** *D. huoshanense* can live on the stone, indicating the derivation of its Chinese name. **(C)** The stem is the medicinal part of *D. huoshanense*. **(D)** "Huoshan Fengdou," the processed product of *D. huoshanense*. **(E)** Stone planting under the forest (SPUF). **(F)** Greenhouse planting (GP).

of metabolites, which can elevate the quality of traditional Chinese medicine (Bowne et al., 2012; Brunetti et al., 2013). A similar study has shown that *Dendrobium moniliforme* (L.) Sw. can alleviate drought stress by changing its physiological mechanism. Additionally, the primary and secondary metabolites of *D. moniliforme* were improved at the stage of antagonism with the environment (Wu et al., 2016). However, the molecular mechanism of the differences in *D. huoshanense* composition under different planting modes, such as SPUF and GP, has not been reported, especially at proteomic and phosphoproteomic level. Based on the differences in morphology, physiology and metabolites, especially the content of polysaccharides and alkaloids, and composition of polysaccharides, we inferred there were significant discrepancy on protein abundances and protein post-translational modification levels between GP and SPUF of *D. huoshanense*.

Protein phosphorylation/dephosphorylation, catalyzed by kinases and phosphatases, was first discovered (Levene and Alsberg, 1996). This post-translational modification (PTM) has become one of the most widely studied reversible protein modifications. It is involved in the regulation of many cellular functions through activation/inhibition of enzymes and transporters, functional protein localization, and signal transduction (Ghelis, 2011). Studies have shown that protein phosphorylation is involved in the signal recognition and transduction of plants under abiotic stress. For example, tyrosine phosphorylation of Arabidopsis RACK1A (Receptor for Activated C Kinase 1A) protein regulates multiple environmental stress signaling pathways, including drought and salt stress resistance pathways (Sabila et al., 2016). The classical mitogen-activated protein kinase (MAPK) signaling pathway is activated through phytohormones, reactive oxygen species (ROS), and calcium ions, and the phosphorylation process transfers the upstream signal to downstream response molecules, finally inducing the stress response (Zhao et al., 2017). In eukaryotes, phosphorylation mostly occurs on hydroxylated amino acid residues such as threonine (Thr), serine (Ser), and tyrosine (Tyr) (Pearlman et al., 2011).

In this study, comparative proteome and phosphoproteome analyses were performed to quantitatively display the protein abundance and phosphorylation status between SPUF and GP, and attempt to reveal the mechanisms for the different accumulation of metabolites and biological adaptability between them.

MATERIALS AND METHODS

Plant Materials, Morphology Measurements, and Reagents

The greenhouse planting (GP) and stone planting under the forest (SPUF) materials were collected from *D. huoshanense* cultivated on a hill under a forest of Taipingfan town Huoshan County, Anhui, China and whose man-made environment imitated a wild-type one, and three independent 3-year-old whole plants had been collected for the samples. All samples used for proteome and transcriptome analysis between SPUF and GP were whole plants with the same batch cultivated in the same year. The morphology index such as stem length and stem diameter were measured with vernier caliper.

The TMT 10plex^{TM} Isobaric Mass Tag Labeling Kit, trypsin, and BCA Protein Assay Kit were purchased from ThermoFisher Scientific, Promega Corporation, and Beyotime Biotechnology, respectively. Acetonitrile was obtained from Fisher Chemical, and other chemicals [trifluoroacetic acid, formic acid, iodoacetamide, dithiothreitol (DTT), urea, trichloroacetic

acid, protease inhibitor, EDTA, and TEAB] were all purchased from Sigma-Aldrich. Standards of (+)-glucose, Dendrobine were purchased from Chengdu Push Bio-technology Co., Ltd., Chengdu, China.

Extraction and Determination of Water-Soluble Polysaccharides and Alkaloids

Polysaccharides in *D. huoshanense* were extracted, and polysaccharides content was determined by the phenol-sulfuric acid method described by Chinese Pharmacopoeia (National Pharmacopoeia Committee, 2020). The polysaccharides content was calculated using glucose as a standard. There were three biological replicates for each sample in this experiment.

Bush et al.'s method was modified to analyze alkaloid contents (Bush et al., 1997; Yuan et al., 2019). The total alkaloid content was calculated using Dendrobine as a standard. There were three biological replicates for each sample in this experiment.

Polysaccharide Hydrolysis and Monosaccharide Determination

Dried *D. huoshanense* powder (0.12 g) was defatted with 80% alcohol at 80°C for 4 h and then extracted with 100 mL of distilled water at 100°C for 1 h, as previously described (Mohan et al., 2020). The filtered supernatant (1.0 mL) was hydrolyzed with 0.5 mL of 3.0 M HCl at 110°C for 1 h in a sealed glass and neutralized to pH 7.0 with 3.0 M NaOH after being cooled to room temperature. The hydrolyzed products (0.4 mL) were added to a solution of 1-phenyl-3-methyl-5-pyrazolone (PMP) consisting of 0.3 M aqueous NaOH (0.4 mL) and 0.5 M methanol (0.4 mL), and then kept at 70°C for 100 min in an oven. After neutralization with 0.3 M HCl, the PMP derivatives were determined in an Agilent Series 1260 system (Agilent Technologies, Shanghai, China) equipped with a Topsil C18 HPLC column (4.6 mm × 250 mm, 5 μm particle size) and monitored by UV absorbance at 250 nm. Elution was carried out using (A) 0.02 M amine acetate solution and (B) acetonitrile as a mobile phase. The ratio of A to B was 80:20. The flow rate was 1.0 mL/min, the sample injection volume was 10 μL, and the column temperature was 30°C. Glucose (100 μg/ml) and mannose (100 μg/ml) were used as internal standards, and there were three biological replicates for each sample in this experiment were used.

Proteome Analysis

Protein Extraction and Western Blot Analysis

Proteins were extracted according to the procedures of previous study (Dong et al., 2021), and separated using 15.0% SDS-PAGE gel and electro-transferred to a PVDF membrane (Millipore) for Western blots. After blocking for 1 h with 5% skimmed milk in Tris-buffered saline with Tween® (TBST), the membrane was incubated overnight at 4°C in a 1/1000 dilution of rabbit-derived pan anti-phosphotyrosine antibody (Thermo Pierce). The unbound primary antibody was removed by washing thrice with TBST buffer (10 min each). Then, the membrane was incubated in a 1/10000 dilution of Goat Anti-Mouse IgG (H + L), Peroxidase Conjugate (Thermo Pierce) at 25°C. The unbound secondary antibody was also removed by washing thrice with TBST buffer (10 min per wash). Enhanced chemiluminescence reagent (LABLEAD, Beijing, China) was used for the detection of proteins in the Western blot.

Trypsin Digestion of Proteins

For protein digestion, proteins extracted from all samples in three duplicates were reduced with 5 mM dithiothreitol for 30 min at 56°C, and then were alkylated with 11 mM iodoacetamide for 15 min at room temperature in the dark. Then, 100 mM TEAB (triethylammonium bicarbonate) was added to dilute the urea concentration to less than 2 M. Trypsin in sequencing grade was added in a 1:50 (trypsin: protein, w/w) ratio for the first digestion overnight. Finally, a second 4-h digestion with trypsin was performed at a 1:100 (w/w) ratio for complete protein digestion.

Tandem Mass Tag Labeling of Peptides

Peptides were desalted with a Strata X C18 SPE column (Phenomenex), dried under vacuum, and re-dissolved with 0.5 M TEAB following the manufacturer's protocol for the TMT labeling kit. The specific steps were as follows: after thawing, the labeled reagent was dissolved in acetonitrile, mixed with the peptide and incubated at room temperature for 2 h. After mixing, the labeled peptide was desalted and vacuum freeze-dried.

High Performance Liquid Chromatography Fractionation of Peptides

High pH reverse-phase HPLC fractionation of the peptide samples after digestion was performed with an Agilent 300 Extend C18 column (5-μm particles, 10 mm ID, 250 mm length). In short, a gradient of 8% to 32% acetonitrile (pH 9.0) over 60 min was used to fractionate the sample into 60 fractions. Then, peptides were further combined into 18 fractions and dry-centrifuged under vacuum.

LC-MS/MS Analysis

Formic acid (0.1%), as solvent A, was used to desalt the peptides, while a reversed-phase analytical column (15 cm length, 75 μm inner diameter) was used to load the peptides instantly. The concentration of solvent B (0.1% formic acid in 98% acetonitrile) in the gradient increased from 6 to 23% for 26 min, increased from 23 to 35% for 8 min, and increased to 80% in 3 min, and then reached 80%, being held for the last 3 min. All steps were performed on the EASY-nLC 1000 UPLC system at a constant flow rate of 400 nL/min. Peptides were processed with an NSI source and then subjected to tandem mass spectrometry (MS/MS) analysis in Q ExactiveTM Plus (ThermoFisher Scientific, Guangzhou, China) connected to UPLC. The applied electrospray voltage was 2.0 kV, and the m/z scan range of the full MS scan was 350–1800. The integral peptides were observed at a resolution of 70,000 in Orbitrap, and the peptides were selected as MS/MS using standardized collision energy, the number was set to 28, and the ion fragment was detected resolution at 17,500 in Orbitrap. A data related process alternates between 1 MS scan and 20 MS/MS scans, and the dynamic exclusion was 15.0 s. Automatic gain control (AGC) was set to 5E4. The first mass setting was kept at 100 m/Z.

Database Search

We used the Maxquant search engine (v.1.5.2.8) to handle the obtained MS/MS data. Tandem mass spectra of the data were searched against a transcriptome database uploaded by us,[1] combined with a reverse decoy database. Trypsin/P was consigned as the cleavage enzyme, and there were up to 2 missing cleavages; the minimum length of the peptide segment was set to 7 amino acid residues. For precursor ions, the mass error was set to 20 ppm in the first search and 5 ppm in the main search, followed by 0.02 Da for ion fragments. The quantitative method was set to TMT-6plex, and the FDR of protein identification and PSM (peptide-spectrum matches) identification were both set to 1%.

Bioinformatics Analysis

Gene Ontology (GO) annotations of phosphorylated proteins were compared against the UniProt-GOA database.[2] By GO annotation, proteins are divided into three categories: biological process, cellular compartment, and molecular function.

Pathways associated with the phosphorylated proteins were annotated based on the Kyoto Encyclopedia of Genes and Genomes (KEGG) database. Wolfpsort software was applied to predict the subcellular localization of the phosphorylated proteins in *D. huoshanense*.

For GO, KEGG pathway, and protein domain enrichment analyses, a two-tailed Fisher's exact test was used to examine the differentially expressed proteins and phosphorylated proteins against all identified proteins. Multiple testing corrections were performed utilizing standard false discovery rate control methods, and the terms with a corrected p-value < 0.05 were considered significantly enriched. For further hierarchical clustering, we first collated all the categories and p-values. P-values were log10 (p-value) transformed and then z-transformed for each category ($P < 0.05$). The one-way hierarchical clustering method was utilized to cluster z scores. Cluster membership was visualized by heat map utilizing the "heatmap.2" function from the "gplots" R-package.

Subcellular Localization

Prediction of subcellular localization of proteins by using WoLF PSORT (v.0.25).

Phosphorylation Protein Analysis

The method of protein extraction, western blot analysis and trypsin digestion was the same as that described in proteome analysis.

Immobilized Metal Affinity Chromatography Enrichment for Phosphorylation

Firstly, separated peptides were incubated with vibrating Ti-IMAC (immobilized metal affinity chromatography) microsphere suspensions in a loading buffer containing 50% acetonitrile and 6% trifluoroacetic acid. Then, we collected IMAC microspheres enriched in phosphopeptides and discarded the supernatant after centrifugation. Next, aiming at removing non-specific adsorbed peptides, IMAC microspheres were sequentially rinsed with 50% acetonitrile, 6% trifluoroacetic acid, 30% acetonitrile, and 0.1% trifluoroacetic acid. Enriched phosphopeptides were vibrationally diluted with an elution buffer containing 10% NH_4OH, and collected and lyophilized for LC-MS/MS analysis.

LC-MS/MS Analysis

Solvent A and Solvent B, the analytical column used in UPLC and the instrument of tandem mass spectrometry (MS/MS) are the same as above. The concentration of solvent B in the gradient increased from 4 to 23% for 40 min, increased from 23 to 35% for 12 min, and increased to 80% in 4 min, and then reached 80%, being held for the last 4 min. The constant flow rate of 300 nL/min was used in UPLC system.

For MS analysis, the m/z scan range of the full MS scan was 350–1600, the integral peptides were observed at a resolution of 60,000, and the ion fragment was detected resolution at 15,000 in Orbitrap. AGC was set to 1E5, and the first mass setting was kept at 100 m/Z.

Database Search

Cysteine alkylation was set as fixed modification and variable modification as oxidation of methionine, acetylation of protein N-terminus, and phosphorylation of serine, threonine, and tyrosine. The minimum score for modified or non-modified peptides was set to > 40. All of the other steps and parameters are the same as shown in proteome analysis.

Gene Ontology, Kyoto Encyclopedia of Genes and Genomes Annotation, and Enrichment Analysis

For Gene Ontology, KEGG annotation, and enrichment analysis, the method was also same as that described in proteome analysis.

Motif Analysis

MoMo Modification Motifs (version5.3.3[3]) (Bailey et al., 2009) software was applied to analyze the conserved amino acid sequence motifs. The Motif-x algorithm (Chou and Schwartz, 2011)[4] was used to analyze the model of conserved motifs of phosphoryl-13-mers (six amino acids upstream and downstream of the phosphorylation site). All the database protein sequences were used as the background database, other parameters were set to default. Only when the number of peptides of a specific sequence is more than 20 and the p-value ≤ 0.000001, the specific amino acid sequence was considered as a conserved characteristic motif.

Kinase and Phosphatase Classification

According to the downloaded genome release pathway (Zm_B73_5b_FGS_cds_2012), kinases and phosphatases

[1]https://pan.baidu.com/s/1bdA4oE_LjkhC41eePsv0pg. Access code: 2290.

[2]http://www.ebi.ac.uk/GOA/

[3]https://meme-suite.org/meme/meme_5.3.3/

[4]http://motif-x.med.harvard.edu/

were identified by MapMan software and categorized referring to the maize databases of ProFITS, P3DB,[5] and ITAK.[6] We used the P3DB database to perform BLAST searches on the kinase sequences of Arabidopsis and rice.

GPS5.0 software[7] (Wang et al., 2020) was used for predicting the activity sites of upstream phosphorylation kinases. There is a hypothesis that similar short peptides may act as alike roles and similar protein kinases may regulate alike short-term motifs. Thus, software predictions may affect the kinase protein families by regulating specific phosphorylation sites, then gains the kinases by comparing with iEKPD database (Guo et al., 2019). At last, the upstream kinase family corresponding to the phosphorylation sites and the kinase IDs corresponding to the kinase family in the database were obtained.

Total RNA Extraction, Construction of cDNA Libraries, and RNA-Seq

The ethanol precipitation protocol and CTAB-pBIOZOL reagent were used to purify total RNA from the whole plants of SPUF and GP according to the manufacturer's instructions (Bioflux), and each sample was prepared with three biological replicates. RNA concentration and quality were measured using a Nano Drop and Agilent 2100 bioanalyzer (Thermo Fisher Scientific, Waltham, MA, United States). After the RNA sample was qualified, fragment buffer was added to the enriched RNA to break the RNA into small fragments. Then, the first strand of cDNA was synthesized by reverse transcription with 6 bp random primers, and then the second strand was synthesized by adding buffer, dNTPs, DNA polymerase I and RNase H. Finally, cDNA library construction was performed by high-throughput sequencing using the Illumina HiSeq platforms at Novogene.

Real-Time Fluorescent Quantitative PCR Analysis

RT-qPCR analysis was performed using a SuperReal PreMix Plus SyBr Green PCR kit [Tiangen Biotech (Beijing) Co., Ltd., Beijing, China] on a Cobas z480 Real-Time PCR System, following the method previously described (Wang et al., 2021). Candidate primers were designed using Primer 5.0 (**Supplementary Table 1**). Reactions contained 2.0 μL of diluted cDNA, 0.6 μL of each primer, 10 μL of 2 × SuperReal PreMix Plus, and 6.8 μL of RNase-free double-distilled water (ddH_2O). All RT-qPCRs were performed as follows: denaturation at 95°C for 15 min, followed by 45 cycles of 95°C for 10 s, 58°C for 20 s, and 72°C for 30 s. Each sample contained three biological replicates, a housekeeping gene (*18S rRNA*) was used as a reference, and the relative expression level of each gene was calculated using the $2^{-\Delta\Delta Ct}$ approach.

Statistical Analysis

At least three independent biological replicates were used for physiological and biochemical analyses. The unpaired two-tailed student's *t*-test was used to determine statistically significant *p*-values in the mean. A significance level of p-value < 0.05 and p-value < 0.01 were considered as statistically difference and significant difference, respectively. For protein abundance and phosphorylation level analysis, the relative quantitative values were log2 transformed to make the data conform to the normal distribution, and then a two-sample two-tailed *T*-test method was performed. When p-value < 0.05 and protein abundance/modification ratio $>$ FC (fold change) was regarded as up-regulation/hyperphosphorylated. When p-value < 0.05 and protein ratio $<$ FC was regarded as down-regulation/hypophosphorylated.

RESULTS

Differences in Morphology, Polysaccharides, and Alkaloids Between Stone Planting Under the Forest and Greenhouse Planting

Significant differences in the morphology index and main metabolites content of SPUF and GP were revealed (**Figure 2A**). The stem length of GP was much higher than that of SPUF, while SPUF had a larger stem diameter than that of GP (**Figures 2A,B** and **Supplementary Table 2**). The leaf to total plant mass ratio in GP was higher than that in SPUF. However, the reverse relationship was found for the root to total plant mass ratio (**Supplementary Table 2**). Total alkaloids and total polysaccharides content in SPUF were higher than those in GP (**Figures 2C,D** and **Supplementary Table 3**). Moreover, regarding polysaccharide composition, mannose and glucose levels in SPUF were higher than those in GP (**Figures 2E,F** and **Supplementary Table 3**). These results indicated that the cultivation modes of SPUF was more conducive to the accumulation of active constituents of polysaccharide and alkaloid.

Global Proteome and Phosphoproteome Profiling in *Dendrobium huoshanense*

Aiming at establishing a comprehensive database specifically for the proteome and phosphoproteome of *D. huoshanense*, we extracted the proteins from SPUF and GP in three independent biological replicates (**Figure 3A**). A western blot experiment was performed to verify the phosphorylation of proteins in SPUF and GP by Pan anti-phosphotrysine antibody (**Figure 3B**). As a result, different patterns were revealed in proteome and phosphoproteome by SDS-PAGE and Western blot, indicating different protein composition and protein phosphorylation status between SPUF and GP (**Figures 3A,B**). Subsequently, the digested peptides of proteome and enriched peptides of phosphoproteome were identified by liquid chromatography followed by mass spectrometry (LC/MS) analysis. The mass spectrometry proteomics data of this study have been deposited into the ProteomeXchange Consortium *via* the PRIDE partner repository with the dataset identifier PXD022337.

A total of 4,820 *D. huoshanense* proteins were identified in this study by proteome and phosphoproteome analysis

[5]http://p3db.org/

[6]http://bioinfo.bti.cornell.edu/cgi-bin/itak/index.cgi

[7]http://gps.biocuckoo.cn/

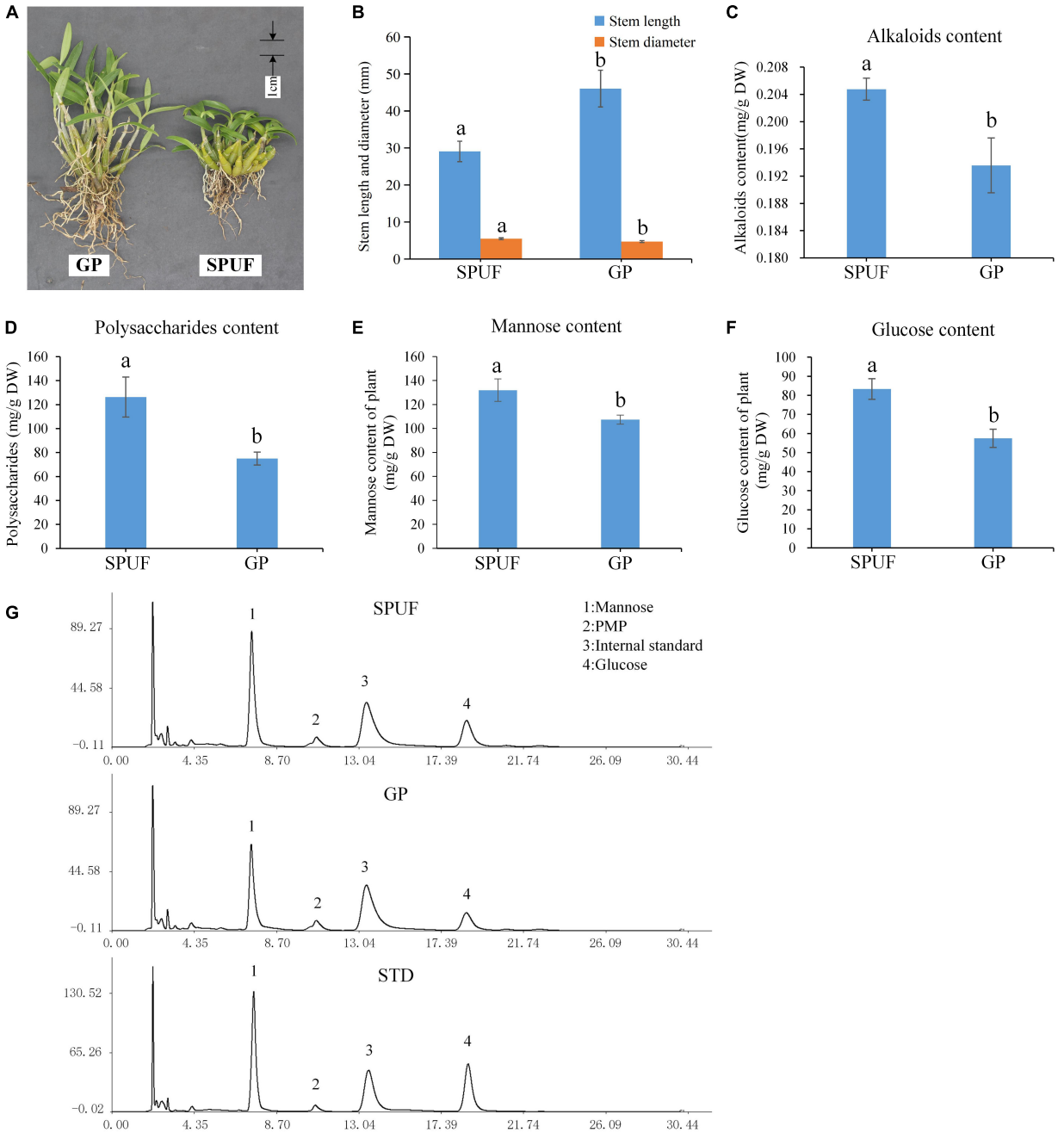


FIGURE 2 | Differences in phenotype and main metabolites content between SPUF and GP. A significance level of p-value < 0.05 was considered as statistically significant. **(A)** Phenotype of SPUF and GP. **(B)** Stem length and diameter of SPUF and GP. **(C–F)** Alkaloid, polysaccharide, mannose, and glucose content in SPUF and GP. **(G)** Chromatogram of the main monosaccharide composition of SPUF and GP in *D. huoshanense* derived by PMP (1-phenyl-3-methyl-5-pyrazolone). Glucosamine hydrochloride was used as internal standard. Different letters above bars indicate significant differences ($P < 0.01$).

(**Figure 3C** and **Supplementary Table 4**). Among them, 3,504 proteins for which quantitative information was available were identified in a comparative proteomics study (**Figure 3C**), and 1,376 proteins containing 2,113 phosphorylated sites were obtained by screening based on localization probability ≥ 0.75 in a phosphoproteomic study (**Figure 3C** and **Supplementary Table 4**). Regardless of the localization information, a total of 1,535 proteins containing 2,872 phosphorylated sites were identified (**Supplementary Table 5**).

Approximately 85.79% of the detected phosphorylated sites were pSer (pS), 13.47% pThr (pT), and 0.73% pTyr (pY) (**Figure 3D**), more than half of the phosphopeptides were singly phosphorylated (**Figure 3E**). There were 42 proteins with more than six phosphorylation sites, and the one with the highest

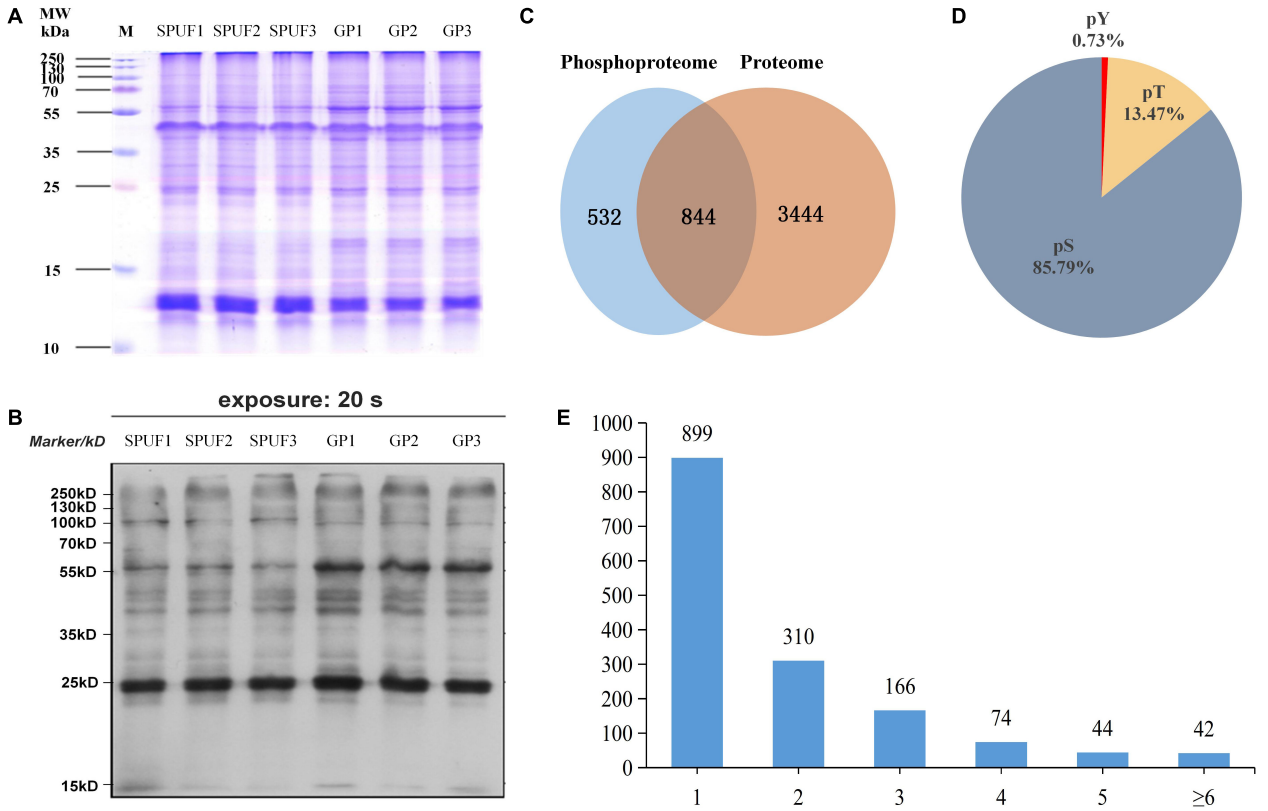


FIGURE 3 | Estimation and determination of global protein phosphorylation site occupancy in *D. huoshanense*. **(A)** SDS-PAGE image of total protein stained with Coomassie Brilliant Blue G-250. Molecular weight is labeled in the left, samples and their duplicates are labeled on the top. M, marker; SPUF1, SPUF2, and SPUF3 are duplicates of stone planting under the forest; GP1, GP2, and GP3 are duplicates of the greenhouse planting. **(B)** Western blot image of protein phosphorylation. The primary antibodies used were rabbit-derived pan anti-phosphotyrosine antibodies (Jingjie PTM Biolab, Hangzhou, China). **(C)** Venn diagram of proteins identified between the proteome and phosphoproteome analysis. **(D)** Frequency of identified pS, pT, pY phosphosites. **(E)** Distribution of phosphopeptides. More than half of the phosphorylated sites were confidently singly phosphorylated.

degree of phosphorylation was annotated as "PREDICTED: RNA-binding motif protein, X chromosome-like isoform X2 [Phoenix dactylifera]" and contained 12 pS (**Supplementary Table 4**). Among these identified proteins, a total of 1,400 proteins with 2,329 phosphorylated sites were quantifiable (**Supplementary Table 5**).

Characterization of Serine-, Threonine-, and Tyrosine-Phosphorylated Peptides

To understand the regulation and amino acid residue preference around serine, threonine, and tyrosine phosphorylation sites, motif analysis was performed for the sequence from sites −6 to +6 of all the 2,329 phosphorylation sites. From the 1,673 conserved peptides, 24 conserved motifs were identified, of which 21 contained serine residues, 3 contained threonine residues, and none contained tyrosine residues (**Supplementary Table 6**). Eight distinguished motifs were identified (marked as Motif Logo): 5 conserved serine motifs (with a high value fold increase) such as SPK, SPR, RXXSP, SPXK, SPXXXR (S indicates the phosphorylated serine, R, P, K indicate arginine, proline, and lysine, respectively, and X indicates a random amino acid residue) and 3 conserved threonine motifs such as PXTP, TP, and RXXT (T indicates the phosphorylated threonine), as shown in **Figure 4A**. Additionally, the frequencies of amino acid residues flanking the phosphorylated serine and threonine were analyzed to demonstrate the enrichment or depletion of various amino acids (**Figure 4B** and **Supplementary Table 7**).

It is intriguing that a significant preference for motifs was discovered in phosphorylated proteins associated with different domains, pathways, biological processes, and cellular components (**Figures 4C–F**). For example, peptides involved in biosynthetic and metabolic processes preferentially harbor the motif TphP (**Figure 4C**). Endoplasmic reticulum-related cellular components preferentially harbor the motif SphPR, and chromosome-related cellular components preferentially harbor the motif PXTphP (**Figure 4E**). Hydrolase- and phosphatase-related molecular functions preferentially harbor the motif PXTphP (**Figure 4F**). The identification of conserved motifs and their preferences in different protein clusters verified that, instead of a non-specific modification, protein phosphorylation is a highly regulated modification in *D. huoshanense*, whose function is affected by the composition of neighboring amino acid residues.

FIGURE 4 | Motif and logo-based clustering analysis of the phosphorylation sites. **(A)** Eight representative conserved motifs of phosphoryl-21-mers flanking phosphorylation sites ("S" and "T," which represent serine and threonine, respectively, are at position 0). The size of each letter correlates to the frequency of that amino acid residue occurring in that position. **(B)** Heat map of the amino acid compositions around the phosphorylation sites. The –log10 (Fisher's exact test *p*-value) for every amino acid in each position (from –6 to +6) is shown, phosphorylation sites in serine are shown on the left, and phosphorylation sites in threonine on the right. Motif logo-based clustering analyses: **(C)** protein domain enrichment analysis, **(D)** KEGG pathway enrichment analysis, **(E)** biological process enrichment analysis, and **(F)** molecular function analysis.

Differentially Expressed Proteome Analysis

Differentially expressed proteomes between GP and SPUF were profiled by a tandem mass tag (TMT)-labeled LC-MS quantitative proteome method with three biological replicates. Among the identified proteins, 3,504 proteins containing quantitative information were identified in the comparative proteomic study (**Figure 3C** and **Supplementary Table 4**). A total of 611 proteins were revealed to be differentially expressed between SPUF and GP with a fold change (FC) $\geq$ 1.5 (upregulated) or $\leq$ 0.667 (downregulated) and *p*-value $\leq$ 0.05. Among these 611 differentially expressed proteins (DEPs), 310 proteins were upregulated and 301 were downregulated in GP compared to their levels in SPUF. The protein with the highest FC value was

predicted as a NUCLEAR FUSION DEFECTIVE 4-like protein, with a FC of 26.477 (*p*-value: 2.19E-6) (**Supplementary Table 8**).

Functional Annotation of Differentially Expressed Proteins

To thoroughly understand the function and characteristics of the DEPs, subcellular localization, functional classification, functional enrichment, cluster of orthologous groups, and protein domain enrichment analyses of these proteins were performed.

Subcellular localization analysis of these DEPs (**Supplementary Table 9**) revealed that most (54.6%) were in chloroplasts, including 279 proteins (176 upregulated and 103 downregulated in GP vs. SPUF), indicating that photosynthesis is significantly different between SPUF and GP. The second and third largest classes for DEPs were in the cytoplasm and nucleus (**Supplementary Table 9**).

Gene ontology (GO) analysis of the 611 DEPs suggested that metabolic and cellular process, as well as proteins with catalytic and binding activity, were significantly divergent between SPUF and GP (**Supplementary Table 10**).

From GO enrichment analysis, these DEPs were mainly enriched in biological processes such as responses to biotic and abiotic processes (GO:0009628 and GO:0009607) and other defense processes (GO:0006950, GO:0001101, GO:0009415, and GO:0006952) (**Figure 5A** and **Supplementary Table 11**). Among these DEPs, proteins responding to abiotic stimuli were enriched with the highest −log10 (*p*-value), both upregulated and downregulated DEPs between GP and SPUF. In GP, most of those DEPs were enriched in responses to water (GO:0009415), heat (GO:0009408), acid (GO:0001101), and inorganic substances (GO:0010035), as well as enriched in biosynthesis (GO:0072525, GO:0009108, GO:0051188, and GO:0044085). Proteins pertaining to ion homeostasis (GO:0055082, GO:0048878, GO:0006879, GO:0055076, and GO:0055072) and catabolic processes (GO:0042744, GO:0006308, and GO:1901136) were enriched in SPUF (**Supplementary Table 11**). These enrichments of DEPs are better explained by differences in growth conditions and morphology between SPUF and GP. Many enriched proteins involved in biotic and abiotic stress and ion homeostasis grant SPUF an increased tolerance to harsh habitat.

Based on KOG/COG (cluster of orthologous groups) functional analysis, SPUF and GP proteins showed differences in posttranslational modifications, protein turnover, chaperones, and carbohydrate transport and metabolism (**Figure 5B**). Additionally, many DEPs participated in the biosynthesis, transport, and catabolism of secondary metabolites, as well as energy production and conversion (**Figure 5B**). KOG/COG analysis showed that DEPs upregulated in SPUF mainly belonged to pathways such as plant-pathogen interactions, alpha-linolenic acid metabolism, and linoleic acid metabolism, phenylpropanoid biosynthesis (**Supplementary Figure 1A**), which were mainly localized in the nucleus, cytoskeleton, and non-membrane-bounded organelles (**Supplementary Figure 1B**). Proteins with a higher expression in GP were mainly categorized into pathways associated with amino acid metabolism, carbon fixation in photosynthesis, carbon metabolism, zeatin biosynthesis, etc. (**Supplementary Figure 1A**), which were mainly localized in thylakoids and the photosynthetic membrane (**Supplementary Figure 1B**).

Based on the Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway enrichment analysis, proteins in GP were enriched in pathways such as glyoxylate and dicarboxylate metabolism; glycine, serine, and threonine metabolism; carbon metabolism; zeatin biosynthesis; carbon fixation in photosynthetic organisms; flavone and flavonol biosynthesis; tropane, piperidine, and pyridine alkaloid biosynthesis; and thiamine metabolism. Proteins abundant in SPUF were involved in pathways such as linoleic acid metabolism; plant-pathogen interactions; fatty acid elongation; ribosome biosynthesis; alpha-linolenic acid metabolism; phenylpropanoid biosynthesis; cutin, suberin, wax, and unsaturated fatty acid biosynthesis (**Supplementary Table 12**).

Protein domain enrichment analysis showed that proteins with aquaporin-like, bet v I/Major latex, START (star-related lipid-transfer)-like, secretory peroxidase, PLAT/LH2 (lipase/lipooxigenase domain), and heat shock protein Hsp90 (N terminal) domains were enriched in SPUF compared to their enrichment in GP (**Supplementary Table 13**), which indicated that these proteins may increase the resistance of SPUF to water, heat, oxidation, and other stresses.

Differential Phosphorylation Analysis

Phosphoproteome analysis between GP and SPUF was carried out by phosphorylation affinity enrichment technology and TMT-labeled high-resolution LC-MS. Among the identified phosphoproteins, a total of 1,400 with 2,329 phosphorylated sites were quantifiable (**Supplementary Table 5**). The phosphorylation level significantly changed (PLSC) on 291 modified sites belonging to 215 phosphoproteins with a FC $\geq$ 1.5 (upregulated) or $\leq$ 0.667 (downregulated) and *p*-value $\leq$ 0.05 based on three biological replicates (**Supplementary Table 14**). Among them, 56 hypophosphorylated and 235 hyperphosphorylated sites were detected in GP compared to their levels in SPUF.

Functional Annotation of Phosphorylation Level Significantly Changed Phosphoproteins

The 215 PLSC phosphoproteins containing 291 phosphorylated modification sites underwent GO annotation and GO enrichment analyses. Distribution bar charts of biological process (BP), cellular component (CC), and molecular function (MF) are shown in **Figure 5C** and **Supplementary Table 15**. From the BP perspective, single-organism metabolic process (GO:0044710), oxidation-reduction process (GO:0055114), alpha-amino acid metabolic process (GO:1901605), oxoacid metabolic process (GO:0043436), etc., were significantly enriched. In terms of CC, ribosome (GO:0005840), ribonucleoprotein complex (GO:1990904), intracellular ribonucleoprotein complex (GO:0030529), proton-transporting two-sector ATPase complex, catalytic domain (GO:0033178), etc., were highly enriched.

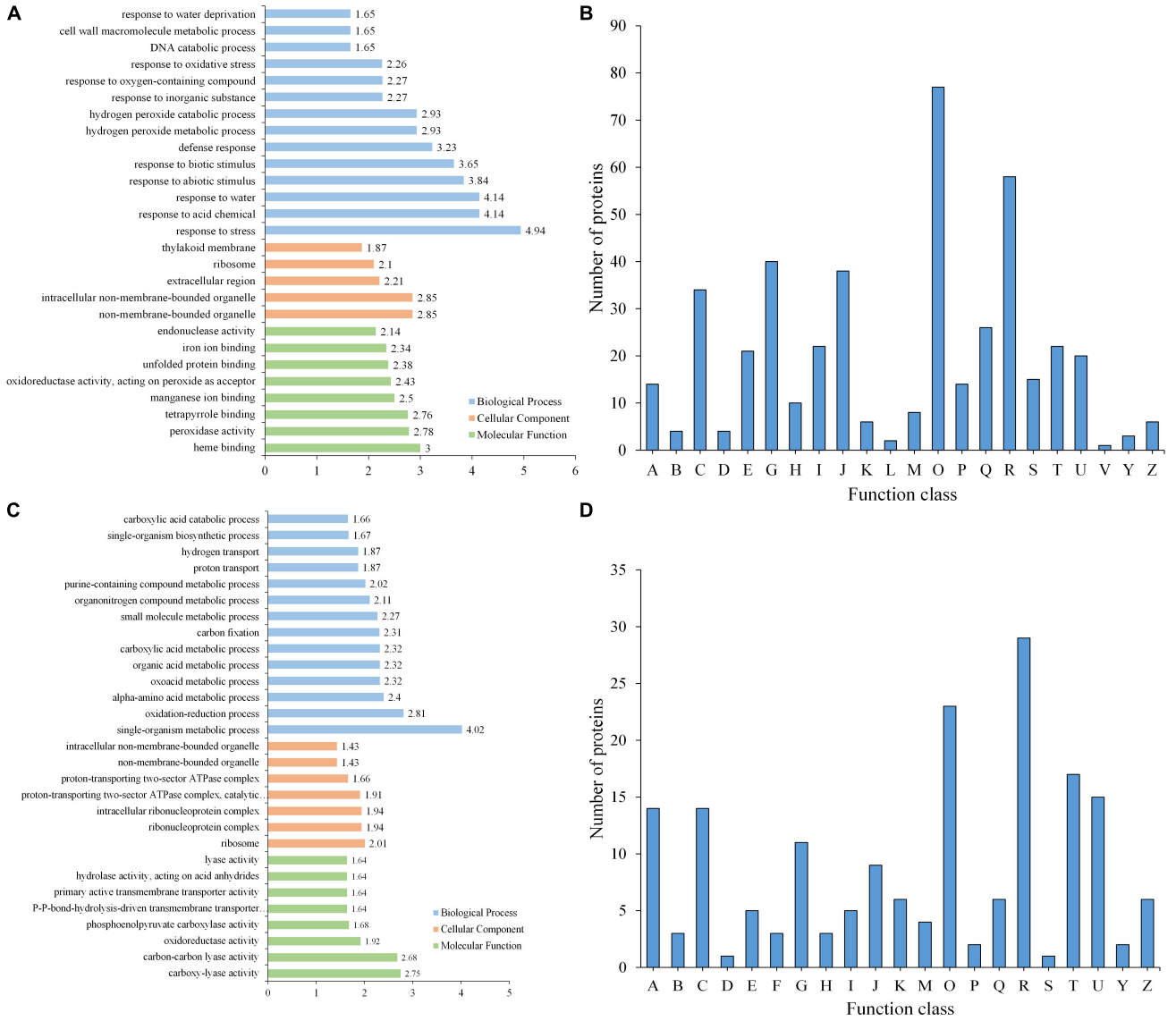


FIGURE 5 | GO enrichment and KOG functional analyses of differentially expressed proteins. **(A)** GO enrichment analysis of DEPs. **(B)** KOG analysis of DEPs. **(C)** GO enrichment analysis of PLSC phosphoproteins. **(D)** KOG analysis of PLSC phosphoproteins. A, RNA processing and modification; B, Chromatin structure and dynamics; C, Energy production and conversion; D, Cell cycle control, cell division, chromosome partitioning; E, Amino acid transport and metabolism; F, Nucleotide transport and metabolism; G, Carbohydrate transport and metabolism; H, Coenzyme transport and metabolism; I, Lipid transport and metabolism; J, Translation, ribosomal structure and biogenesis; K, Transcription; L, Replication, recombination, and repair; M, Cell wall/membrane/envelope biogenesis; O, Posttranslational modification, protein turnover, chaperones; P, Inorganic ion transport and metabolism; R, General function prediction only; S, Function unknown; T, Signal transduction mechanisms; U, Intracellular trafficking, secretion, and vesicular transport; Z, Cytoskeleton.

Regarding MF, carboxy-lyase activity (GO:0016831), carbon-carbon lyase activity (GO:0016830), oxidoreductase activity, acting on the aldehyde or oxo group of donors, NAD or NADP as acceptor (GO:0016620), phosphoenolpyruvate carboxylase activity (GO:0008964), etc., were significantly enriched relative to the background (**Figure 5C** and **Supplementary Table 16**).

Kyoto Encyclopedia of Genes and Genomes enrichment analysis of PLSC proteins in SPUF revealed that the pathways of glyoxylate and dicarboxylate metabolism, endocytosis, and oxidative phosphorylation were enriched. The pathways of pyruvate metabolism, taurine and hypotaurine metabolism, histidine metabolism, and glycerolipid metabolism were enriched in GP. The pathway of beta-alanine metabolism was enriched in both GP and SPUF (**Supplementary Table 17**). The pathways with PLSC phosphorylated protein enrichment in SPUF were closely associated to plant resistance and secondary metabolism, while those in GP were related to plant growth and development. These results reasonably account for the higher metabolite content and better adaptation to worse growth conditions of SPUF despite the larger biomass of GP.

KOG analysis suggested that a substantial number of PLSC phosphoproteins were involved in post-translational modification, signal transduction, and different carbohydrate transport and metabolism for adaptation to different growth environments (**Figure 5D**). Differential transport and metabolism of carbohydrates results in different contents of carbohydrate-associated metabolites between GP and SPUF.

Moreover, subcellular localization analysis suggested that PLSC phosphoproteins are largely located in chloroplasts, with 72 (33%) proteins (57 hyperphosphorylated and 15 hypophosphorylated in GP vs. SPUF), followed by the nucleus and cytoplasm (**Supplementary Table 18**). Since several PLSC phosphoproteins located in chloroplast are involved in metabolic processes and have catalytic and binding activity, phosphorylation likely plays a role in photosynthesis regulation in SPUF and GP and may result in different morphology traits and carbohydrate accumulation.

These results implied that these PLSC proteins were highly enriched in response to different stresses and signal transduction mechanism processes in GP and SPUF and caused different carbohydrate transport and metabolism, which might associate with differences in polysaccharide content and composition.

Protein Kinases and Phosphatases in *Dendrobium huoshanense*

We further exploited our extensive phosphoproteomics dataset and predicted the potential enzymes responsible for phosphorylation. Since there are no related phosphatase databases for the phosphatase analysis of *D. huoshanense* at present, only kinases were analyzed in this study. Altogether, 677 phosphosites catalyzed by kinases were predicted from the 2,329 identified phosphosites. These kinases catalyzing phosphoproteins were assigned to eight categories involving CAMK, TKL, CMGC histone kinase, STE, Atypical, AGC, CK1, and others in *D. huoshanense* (**Supplementary Table 19**). The percentages of phosphosites in *D. huoshanense* catalyzed by major kinases such as CAMK, TKL, and CMGC were 15.84, 26.63, and 28.65%, respectively (**Supplementary Table 19**). From bioinformatics analysis and prediction, 911 putative kinases in *D. huoshanense* may be involved in catalyzing these phosphosites. For instance, ATPase family AAA domain-containing protein 1-A isoform X3 is a kinase that might be involved in the modification of 5 phosphosites, and its expression level in GP is 1.71 times higher than that in SPUF (**Supplementary Table 20**).

Validation by Real-Time Fluorescent Quantitative PCR

We verified the transcript level of relevant genes which encode enzymes involved in the Calvin cycle and polysaccharide and alkaloid biosynthesis pathways, such as glpX-SEBP, FBP, GAPA, DXS, CS, SCS, GCK, and MVD. Transcriptional levels of glpX-SEBP, FBP, GAPA, DXS, CS, and SCS in GP were higher than those in SPUF. Meanwhile, the expression levels of GCK and MVD in SPUF were higher than those in GP (**Figure 6**). These qRT-PCR results were consistent with data from the proteome analysis.

DISCUSSION

SPUF *Dendrobium huoshanense* Contains Higher Active Constituents

Different cultivation modes lead to great changes in traits, quality, and effects. To reveal differences between SPUF and GP, we conducted a systematic comparison of their appearance and quality. The stems of SPUF are short and thick, resembling the wild *D. huoshanense*. Our study found that the content of major secondary metabolites in SPUF is significantly higher than that in GP (**Figure 2**), which is consistent with previous reports (Yi et al., 2021). Since mannose-containing polysaccharides exhibit higher bioactivities than those of glucose-containing polysaccharides and other monosaccharides (Yu et al., 2018), it is conceivable that SPUF, with a higher content of mannose and polysaccharides, is more medicinally effective than GP. It had been reported that the protective effect against mouse liver injury of SPUF was better than that of GP (Li et al., 2020). And the study had verified that SPUF was more similar to wild *D. huoshanense* and the quality of its medicinal materials was considerably better than that of GP's medicinal materials (Yi et al., 2021). The fresh and dried stem of SPUF tastes bitterer, which might be due to its higher content of alkaloids compared to GP. From the literature and our results, SPUF is more suitable for medicine than GP.

Protein Phosphorylation in *Dendrobium huoshanense*

Various cellular activities are controlled by protein PTMs, and more than 200 types of PTMs have been identified, such as phosphorylation, methylation, acetylation, and ubiquitination (Krishna and Wold, 1993). Protein kinase cascades play essential roles in diverse intracellular signaling processes in animals and yeast (Tetlow et al., 2004). In plants, phosphorylation is involved in signaling pathways triggered by abiotic stress, pathogen invasion, and plant hormones (Knetsch et al., 1996; Sheen, 1996; Zhang et al., 1998). Additionally, protein phosphorylation is also important in metabolism, transcription, translation, proteolysis, homeostasis, and signaling (Thingholm et al., 2009). Despite the importance and widespread occurrence of protein phosphorylation in plants, this is the first report to date on the proteome and phosphoproteome profile of *D. huoshanense* to the best of our knowledge. In this study, a total of 532 specific proteins were identified in the phosphoproteome by an immobilized metal affinity chromatography (IMAC) sphere phosphopeptide enrichment approach, which were not identified in the proteome (**Figure 3C**). Protein phosphorylation is the most abundant and extensively studied PTM (Ptacek and Snyder, 2006), with approximately 30% of proteins being phosphorylated (Hauser et al., 2017). Here, the percentage of protein phosphorylation was 28.55% (**Figure 3C**). The occurrence of identified phosphosites were pSer > pThr > pTyr in our study (**Figure 3D**), which was consistent with most

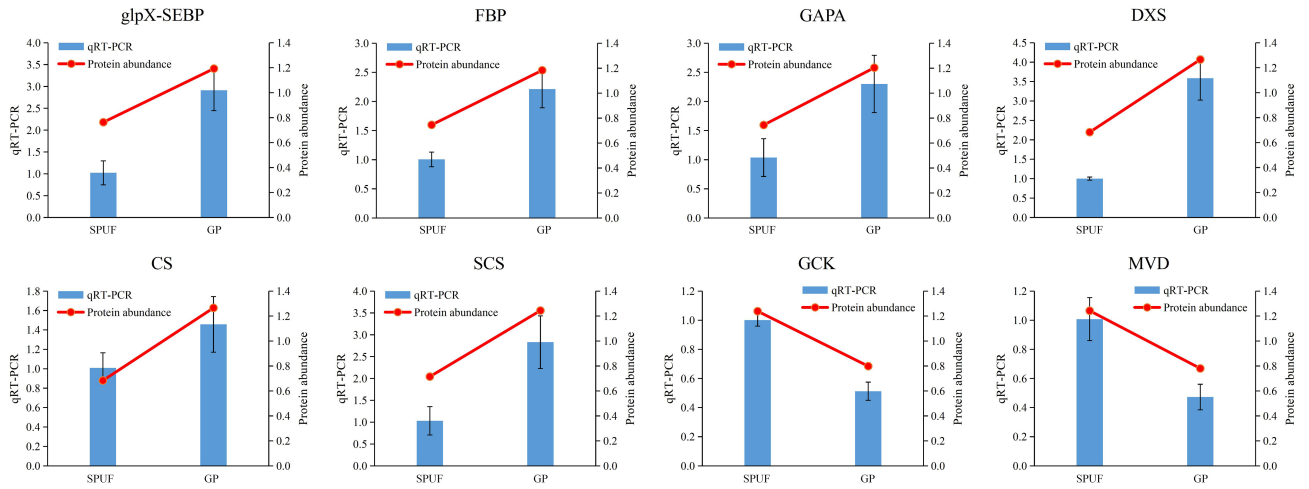


FIGURE 6 | qRT-PCR analysis of genes encoding enzymes differentially expressed in major metabolites biosynthesis. The *18S rRNA* was used as an internal reference gene for normalization. Blue bars represent qRT-PCR data and red lines indicate protein abundance from proteomic analyses. Data represent mean ± standard error of three replicates. The left Y-axis denotes relative expression levels of genes determined by qRT-PCR, and the right Y-axis denotes protein abundance data.

plants (Rattanakan et al., 2016; Chen et al., 2017). The less frequent nature of tyrosine phosphorylation in plants might due to their lack of receptor tyrosine kinases (Luan, 2002). Studies have shown that mono-phosphosites in phosphoproteins are dominant (Rattanakan et al., 2016; Chen et al., 2017). In our study, the percentage of mono-phosphorylated peptides was 58.57% (**Figure 3E**). The above results demonstrate the reliability of this study's results.

Analysis of Differentially Expressed Kinases and Phosphorylation Level Significantly Changed Kinases

Significant differences in conditions such as temperature, light, water, soil composition, soil pathogens, pH (**Figures 1E,F**) may affect the accumulation of metabolites in plants and ultimately lead to different chemical compositions and efficacies between SPUF and GP. These environmental conditions influence signal transduction, and the widespread nature of phosphorylation in plants renders it an ideal model to explore the molecular basis of differences between SPUF and GP. There were 19 differentially expressed kinases and 12 PLSC kinases identified in the proteomics and phosphoproteomics studies, respectively. We also found 14 phosphatases differentially expressed and 2 PLSC phosphatases between GP and SPUF, among these DEPs, the levels of 8 were upregulated and 6 were downregulated in GP compared with those in SPUF (**Supplementary Table 21**). Overall, 9 upregulated and 10 downregulated kinases were differentially expressed in GP. A D-glycerate 3-kinase involved in the reaction of the photorespiratory C2 cycle, an indispensable ancillary metabolic pathway to the photosynthetic C3 cycle (Boldt et al., 2005), was a differentially expressed kinase with the highest fold difference in GP compared with SPUF; while the expression of CBL-interacting serine/threonine-protein kinase 24, a kinase conferring salt tolerance in plants (Yang et al., 2019), in SPUF was much higher than that in GP. The expression of all PLSC kinases generating pSer in GP was higher than that in SPUF, while most of them were located in the chloroplast (**Supplementary Table 21**). It is suggested that SPUF requires more functional kinases than GP for signal transduction to adapt to harsh environment.

Difference Expressed and Phosphorylation Level Significantly Changed Enzymes Involved in Photosynthesis

Photosynthesis involves enzymes that are highly heat-sensitive, and previous studies have shown that the phosphorylation of photosynthesis proteins plays an important role in photosynthesis regulation (Boex-Fontvieille et al., 2014; Pan et al., 2018). There were 21 enzymes involved in the photosynthesis pathway identified in this study. The expression of two putative enzymes (FBP: fructose-1,6-bisphosphatase with EC: 3.1.3.11, and petH: ferredoxin-NADP reductase with EC: 1.18.1.2) in GP was higher than that in SPUF, while the PPC candidate enzyme (EC: 4.1.1.31, phosphoenolpyruvate carboxylase 2) had a higher expression in SPUF (**Supplementary Table 22**). We identified 11 enzymes with 39 phosphorylated sites involved in photosynthesis between SPUF and GP. Glyceraldehyde-3-phosphate dehydrogenase A (GAPA, EC: 1.2.1.13) is an essential enzyme for glycolysis/gluconeogenesis and mediates carbon fixation in photosynthetic organisms. Five sites in GAPA were phosphorylated, among them, Ser120 and Thr121 were PLSC sites (**Supplementary Table 22**). Glyceraldehyde-3-phosphate dehydrogenase (GAPDH, EC 1.2.1.12) is a key enzyme in the Calvin cycle that catalyzes the reductive dephosphorylation of 1,3-bisphosphoglyceric acid (BPGA) to inorganic phosphate and glyceraladehyde-3-phosphate (Trost et al., 2006; Mekhalfi et al., 2014). Six serine

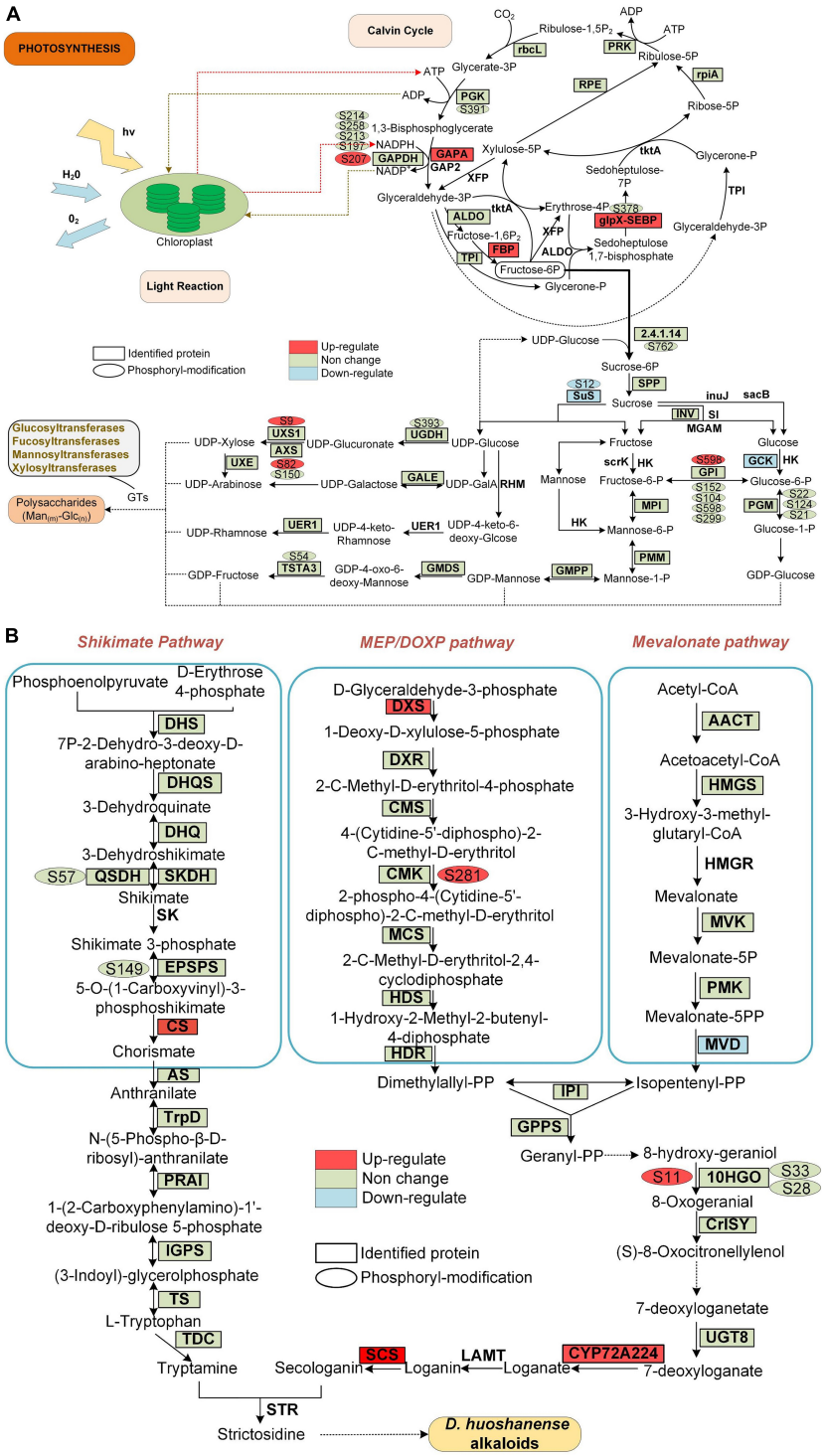


FIGURE 7 | Proteins and phosphoproteins involved in photosynthesis and alkaloid biosynthesis pathways in *D. huoshanense*. Proteins identified in proteome analysis are surrounded by a square frame, and phosphorylated proteins are shown with an oval frame near them. Upregulated and downregulated proteins are shown in GP compared to SPUF. **(A)** Proteins and phosphoproteins involved in photosynthesis. **(B)** Proteins and phosphoproteins involved in alkaloid biosynthesis pathways.

and three threonine sites in GAPDH were phosphorylated in both GP and SPUF (**Figure 7A** and **Supplementary Table 22**), but the phosphorylation level of Ser207 changed significantly between them. The phosphorylation of Ser120 and Thr121 of GAPA and Ser207 of GAPDH may be related to the biomass and polysaccharide content of GP and SPUF.

Differentially Expressed Proteins and Phosphorylation Level Significantly Changed Proteins Involved in Polysaccharides and Alkaloids Biosynthesis

The main bioactive metabolites in *D. huoshanense* are polysaccharides and alkaloids (Hsieh et al., 2008; Matus et al., 2008). Previous studies showed that the whole crotonylation proteome of *D. huoshanense* was analyzed, and the enzyme genes of Calvin cycle, polysaccharide and alkaloid pathway were also crotonylated (Wu et al., 2022). In our study, we found a significant difference in the content of both polysaccharides and alkaloids between GP and SPUF (**Figure 2**). We found that the protein abundance and phosphorylation level in chloroplasts are significantly different between GP and SPUF. We mapped the proteins identified in our proteome and phosphoproteome analyses to the Calvin cycle and polysaccharide and alkaloid biosynthesis pathways (**Figure 7**). We found that most of the enzymes involved in these pathways had been identified in our research, and many of them were phosphorylated. We also found that several enzymes were significantly differentially expressed and the phosphorylation level of some phosphorylated sites was significantly changed between GP and SPUF. GAPA, FBP, and glpX-SEBP in the Calvin Cycle (**Figure 7A**) and CS, DXS, SCS, and CYP72A224 in alkaloid biosynthesis (**Figure 7B**) were highly expressed in GP. The expression of Sus and GCK, involved in polysaccharide biosynthesis, and MVD, involved in alkaloid biosynthesis, in SPUF was higher than that in GP. NADPH had the most abundant modified sites, with five sites among proteins involved in these pathways, and the phosphorylation level of Ser207 in GP was higher than that in SPUF. Other sites also had higher phosphorylation levels, such as Ser9 of UXS1, Ser82 of AXS, and Ser598 of GP1 in polysaccharides synthesis and Ser281 in CMK and S11 in 10HGO in GP. Only the phosphorylation level of Ser12 of Sus was higher in SPUF.

There is no report on the correlation between protein phosphorylation and polysaccharides or alkaloids biosynthesis. In this study, we found that there may be a correlation between phosphorylation levels of different enzyme sites and the polysaccharide content, including Ser9 of UDP-arabinose 4-epimerase 3 isoform (UXS1), Ser82 of UDP-D-xylose synthase (AXS), Ser598 of glucose-6-phosphate isomerase (GPI), Ser12 of sucrose synthase 1 (Sus) (**Figure 7A**). There may also be a correlation between the phosphorylation levels and alkaloid content at various enzyme sites, including Ser281 of 4-diphosphocytidyl-2-C-methyl-D-erythritol kinase (CMK) and Ser11 of mannitol dehydrogenase (10HGO) (**Figure 7B**).

CONCLUSION

Altogether, 4,820 proteins were identified by proteome and phosphoproteome analyses in this study from *D. huoshanense*. Among them, 237 were differentially expressed between SPUF and GP, and 215 PLSC phosphoproteins containing 291 modification sites were identified in the phosphoproteome analysis. Gene ontology, KEGG pathway, protein domain, and cluster analyses revealed that these DEPs were mainly localized in the chloroplasts, involved in processes such as posttranslational modification, carbohydrate transport and metabolism, and secondary metabolite biosynthesis, and enriched in pathways mainly including linoleic acid metabolism, plant-pathogen interactions, and phenylpropanoid, cutin, suberin, and wax biosynthesis. PLSC proteins were also mainly located in chloroplasts, and highly enriched in different stress responses and signal transduction mechanisms processes through protein kinase and phosphotransferase activities. These differences accounted for the much stronger tolerance of SPUF to harsh environments, and likely for the obvious differences in morphology between SPUF and GP. Significant differences between SPUF and GP were observed by mapping DEPs and PLSC proteins to photosynthesis and polysaccharide and alkaloid biosynthesis pathways. The phosphorylation characteristics and kinase categories of *D. huoshanense* were also clarified in this study, which constitutes the first proteomic and phosphoproteomic report on *D. huoshanense*. The results might be helpful to understand the mechanism of these two different varieties and improve the levels of cultivation, management and breeding of *D. huoshanense*.

DATA AVAILABILITY STATEMENT

The proteome data can be found below: PRIDE with ProteomeXchange, PXD022337 (https://www.ebi.ac.uk/pride/archive/projects/PXD022337).

AUTHOR CONTRIBUTIONS

SX, LWu, and DP conceived, designed, and implemented the study. LWu and SX collected specimens and prepared samples for sequencing. LWu, XS, DP, ZW, FM, HH, YL, LWa, and WJ statistics analyzed and interpreted. SX, DP, ZW, and YL reagents, materials, and analyzed tools. SX, LWu, and XM drafted the manuscript. All authors edited the manuscript and approved the final version.

FUNDING

This work was supported by National Natural Science Foundation of China (No. U19A2009), NSF of Anhui Province (No. 1908085MH268), Key Natural Science Research Projects in Anhui Universities (Nos. KJ2019A0453 and KJ2018A0275), Anhui Collaborative Innovation Project of Universities (No. GXXT-2019-043), and the Open Fund Project of Key Laboratory in Hunan Universities (No. NY20K04).

ACKNOWLEDGMENTS

We would like to thank Dr. Zhang Zhang from Jingjie PTM BioLab (Hangzhou) Co. Ltd. in China for MS analysis and

some bioinformatics of proteome. We also thank Shutang Tan from University of Science and Technology of China and Lei Zhang from Naval Medical University in China for revising the manuscript.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpls.2022.937392/full#supplementary-material

The raw data and annotated MS spectra of *Dendrobium huoshanese* are available in the proteomics repository PRIDE with ProteomeXchange data set accession number of PXD022337 (for reviewing, the username is reviewer_022337@ebi.ac.uk and password is fHJF6HMM). The transcriptome data generated during the current study as reference genome information are available in the Baidu Cloud repository: https://pan.baidu.com/s/1bdA4oE_LjkhC41eePsv0pg, and access code is 2290.

Supplementary Figure 1 | Cluster analysis of the DEPs from proteome.

Supplementary Table 1 | Gene descriptions and primers used for qRT-PCR.

Supplementary Table 2 | Morphology characteristics of SPUF and GP of *D. huoshanense*.

Supplementary Table 3 | Major metabolites and their composition comparison between SPUF and GP of *D. huoshanense*.

Supplementary Table 4 | Proteins identified in proteome and phosphoproteome between SPUF and GP of *D. huoshanense*.

Supplementary Table 5 | All phosphoproteins were identified without considering the localization probability ≥ 0.75.

Supplementary Table 6 | Conserved motif analysis on conserved phosphorylated peptides.

Supplementary Table 7 | The −log10 (Fisher's exact test *p*-value) of the amino acids around the phosphorylation sites in each position (from −6 to +6). Phosphorylation sites in serine are shown on the left and those in threonine are shown on the right.

Supplementary Table 8 | Data matrix of differentially expressed proteins from proteome study.

Supplementary Table 9 | Subcellular localization categories of the DEPs from proteome of *D. huoshanense*.

Supplementary Table 10 | GO classification analysis of the DEPs from proteome of *D. huoshanense*.

Supplementary Table 11 | GO enrichment analysis of the DEPs from proteome.

Supplementary Table 12 | KEGG pathway enrichment analysis of the DEPs from proteome.

Supplementary Table 13 | Proteins domain analysis of the DEPs from proteome.

Supplementary Table 14 | Data matrix of phosphorylation level significantly changes proteins from phosphoproteome study.

Supplementary Table 15 | GO annotation of the PLSC proteins from phosphoproteome of *D. huoshanense*.

Supplementary Table 16 | GO enrichment analysis of the PLSC proteins.

Supplementary Table 17 | KEGG pathway enrichment analysis of the PLSC proteins from phosphoproteome of *D. huoshanense*.

Supplementary Table 18 | Subcellular localization analysis of the PLSC proteins.

Supplementary Table 19 | Kinases predicted in phosphosites identified of *D. huoshanense*.

Supplementary Table 20 | Category of the predicted protein kinases of *D. huoshanense*.

Supplementary Table 21 | Differentially expressed kinases and phosphatases in proteome and PLSC kinases and phosphatases in phosphoproteome.

Supplementary Table 22 | Protein enzymes involved in photosynthesis of *D. huoshanense* identified in proteome and phosphoproteome.

REFERENCES

Bailey, T. L., Boden, M., Buske, F. A., Frith, M., Grant, C. E., Clementi, L., et al. (2009). MEME SUITE: tools for motif discovery and searching. *Nucleic Acids Res.* 37, W202–W208. doi: 10.1093/nar/gkp335

Boex-Fontvieille, E., Daventure, M., Jossier, M., Hodges, M., Zivy, M., and Tcherkez, G. (2014). Phosphorylation pattern of Rubisco activase in *Arabidopsis* leaves. *Plant Biol.* 16, 550–557. doi: 10.1111/plb.12100

Boldt, R., Edner, C., Kolukisaoglu, U., Hagemann, M., Weckwerth, W., Wienkoop, S., et al. (2005). D-GLYCERATE 3-KINASE, the last unknown enzyme in the photorespiratory cycle in *Arabidopsis*, belongs to a novel kinase family. *Plant Cell* 17, 2413–2420. doi: 10.1105/tpc.105.033993

Bowne, J. B., Erwin, T. A., Juttner, J., Schnurbusch, T., Langridge, P., Bacic, A., et al. (2012). Drought responses of leaf tissues from wheat cultivars of differing drought tolerance at the metabolite level. *Mol. Plant*. 5, 418–429. doi: 10.1093/mp/ssr114

Brunetti, C., George, R. M., Tattini, M., Field, K., and Davey, M. P. (2013). Metabolomics in plant environmental physiology. *J. Exp. Bot*. 64, 4011–4020. doi: 10.1093/jxb/ert244

Bush, L. P., Wilkinson, H. H., and Schardl, C. L. (1997). Bioprotective alkaloids of grass-fungal endophyte symbioses. *Plant Physiol*. 114, 1–7. doi: 10.1104/pp.114.1.1

Chang, C. C., Ku, A. F., Tseng, Y. Y., Yang, W. B., Fang, J. M., and Wong, C. H. (2010). 6,8-Di-C-glycosyl flavonoids from *Dendrobium huoshanense*. *J. Nat. Prod*. 73, 229–232. doi: 10.1021/np900252f

Chen, G. X., Zhen, S. M., Liu, Y. L., Yan, X., Zhang, M., and Yan, Y. M. (2017). In vivo phosphoproteome characterization reveals key starch granule-binding phosphoproteins involved in wheat water-deficit response. *BMC Plant Biol*. 17:168. doi: 10.1186/s12870-017-1118-z

Chen, N. D., Chen, N. F., Wang, T. T., and Gu, Z. (2015). Study on the dynamic accumulation of polysaccharides and ethanol leachables of *Dendrobium huoshanense*, wild *Dendrobium Huoshanense* and *Dendrobium henan*. *Nat. Prod. Res. Dev*. 27, 2090–2094.

Chen, S., Dai, J., Song, X., Jiang, X., Zhao, Q., Sun, C., et al. (2020). Endophytic microbiota comparison of *Dendrobium huoshanense* root and stem in different growth years. *Planta Med*. 86, 967–975. doi: 10.1055/a-1046-1022

Chen, Z. Q., and Xia, L. P. (2020). Exploration of industrial poverty alleviation on the old revolutionary base area of Dabie Mountain: a study on poverty alleviation in the industry of *Dendrobium Huoshanese*. *J. Anhui Acad. Gov*. 2, 76–80.

Chou, M. F., and Schwartz, D. (2011). Biological sequence motif discovery using motif-x. *Curr. Protoc. Bioinformatics* Chapter 13, Unit 13.15–13.24. doi: 10.1002/0471250953.bi1315s35

Deng, Y., Chen, L. X., Han, B. X., Wu, D. T., Cheong, K. L., Chen, N. F., et al. (2016). Qualitative and quantitative analysis of specific polysaccharides in *Dendrobium huoshanense* by using saccharide mapping and chromatographic methods. *J. Pharm. Biomed. Anal*. 129, 163–171. doi: 10.1016/j.jpba.2016.06.051

Dong, Y., Li, P., Li, P., and Chen, C. (2021). First comprehensive analysis of lysine succinylation in paper mulberry (*Broussonetia papyrifera*). *BMC Genomics* 22:255. doi: 10.1186/s12864-021-07567-5

Ge, J. C., Zha, X. Q., Nie, C. Y., Yu, N. J., Li, Q. M., Peng, D. Y., et al. (2018). Polysaccharides from *Dendrobium huoshanense* stems alleviates lung inflammation in cigarette smoke-induced mice. *Carbohydr. Polym*. 189, 289–295. doi: 10.1016/j.carbpol.2018.02.054

Ghelis, T. (2011). Signal processing by protein tyrosine phosphorylation in plants. *Plant Signal. Behav*. 6, 942–951. doi: 10.4161/psb.6.7.15261

Guo, Y., Peng, D., Zhou, J., Lin, S., Wang, C., Ning, W., et al. (2019). iEKPD 2.0: an update with rich annotations for eukaryotic protein kinases, protein phosphatases and proteins containing phosphoprotein-binding domains. *Nucleic Acids Res*. 47, D344–D350. doi: 10.1093/nar/gky1063

Hao, J. W., Chen, N. D., Chen, C. W., Zhu, F. C., Qiao, D. L., Zang, Y. J., et al. (2018). Rapid quantification of polysaccharide and the main monosaccharides in *Dendrobium huoshanense* by near-infrared attenuated total reflectance spectroscopy. *J. Pharm. Biomed. Anal*. 151, 331–338. doi: 10.1016/j.jpba.2018.01.027

Hauser, A., Penkert, M., and Hackenberger, C. (2017). Chemical approaches to investigate labile peptide and protein phosphorylation. *Acc. Chem. Res*. 50, 1883–1893. doi: 10.1021/acs.accounts.7b00170

Hsieh, Y. S., Chien, C., Liao, S. K., Liao, S. F., Hung, W. T., Yang, W. B., et al. (2008). Structure and bioactivity of the polysaccharides in medicinal plant *Dendrobium huoshanense*. *Bioorg. Med. Chem*. 16, 6054–6068. doi: 10.1016/j.bmc.2008.04.042

Jin, Q., Jiao, C., Sun, S., Song, C., Cai, Y., Lin, Y., et al. (2016). Metabolic analysis of medicinal *Dendrobium officinale* and *Dendrobium huoshanense* during different growth years. *PLoS One* 11:e0146607. doi: 10.1371/journal.pone.0146607

Knetsch, M., Wang, M., Snaar-Jagalska, B. E., and Heimovaara-Dijkstra, S. (1996). Abscisic acid induces mitogen-activated protein kinase activation in barley aleurone protoplasts. *Plant Cell* 8, 1061–1067. doi: 10.1105/tpc.8.6.1061

Krishna, R. G., and Wold, F. (1993). Post-translational modification of proteins. *Adv. Enzymol. Relat. Areas Mol. Biol*. 67, 265–298. doi: 10.1002/9780470123133.ch3

Levene, P. A., and Alsberg, C. L. (1996). The cleavage products of vitellin. *J. Biol. Chem*. 2, 127–133. doi: 10.1016/s0021-9258(17)46054-6

Li, Z. Q., Zhou, H. Q., Ouyang, Z., Dai, J., Wei, Y., and Han, B. X. (2020). Protective effects of differently cultivated *Dendrobium huoshanense* on acute liver injury in mice. *Chin. Trad. Pat. Med*. 42, 1155–1162.

Liang, Z. Y., Zhang, J. Y., Huang, Y. C., Zhou, C. J., Wang, Y. W., Zhou, C. H., et al. (2019). Identification of flavonoids in *Dendrobium huoshanense* and comparison with those in allied species of *Dendrobium* by TLC, HPLC and HPLC coupled with electrospray ionization multi-stage tandem MS analyses. *J. Sep. Sci*. 42, 1088–1104. doi: 10.1002/jssc.201801021

Luan, S. (2002). Tyrosine phosphorylation in plant cell signaling. *Proc. Natl. Acad. Sci. U.S.A*. 99, 11567–11569. doi: 10.1073/pnas.182417599

Matus, J. T., Aquea, F., and Arce-Johnson, P. (2008). Analysis of the grape MYB R2R3 subfamily reveals expanded wine quality-related clades and conserved gene structure organization across *Vitis* and *Arabidopsis* genomes. *BMC Plant Biol*. 8:83. doi: 10.1186/1471-2229-8-83

Mekhalfi, M., Puppo, C., Avilan, L., Lebrun, R., Mansuelle, P., Maberly, S. C., et al. (2014). Glyceraldehyde-3-phosphate dehydrogenase is regulated by ferredoxin-NADP reductase in the diatom *Asterionella formosa*. *New Phytol*. 203, 414–423. doi: 10.1111/nph.12820

Mohan, K., Muralisankar, T., Uthayakumar, V., Chandirasekar, R., Revathi, N., Ramu Ganesan, A., et al. (2020). Trends in the extraction, purification, characterization and biological activities of polysaccharides from tropical and sub-tropical fruits - A comprehensive review. *Carbohydr. Polym*. 238:116185. doi: 10.1016/j.carbpol.2020.116185

National Pharmacopoeia Committee (2020). *The Pharmacopoeia of the People's Republic of China*. Beijing: China Medical Science and Technology Press.

Niu, Z., Pan, J., Xue, Q., Zhu, S., Liu, W., and Ding, X. (2018). Plastome-wide comparison reveals new SNV resources for the authentication of *Dendrobium huoshanense* and its corresponding medicinal slice (Huoshan Fengdou). *Acta Pharm. Sin. B* 8, 466–477. doi: 10.1016/j.apsb.2017.12.004

Pan, D., Wang, L., Tan, F., Lu, S., Lv, X., Zaynab, M., et al. (2018). Phosphoproteomics unveils stable energy supply as key to flooding tolerance in *Kandelia candel*. *J. Proteomics* 176, 1–12. doi: 10.1016/j.jprot.2018.01.008

Pearlman, S. M., Serber, Z., and Ferrell, J. E. Jr. (2011). A mechanism for the evolution of phosphorylation sites. *Cell* 147, 934–946. doi: 10.1016/j.cell.2011.08.052

Ptacek, J., and Snyder, M. (2006). Charging it up: global analysis of protein phosphorylation. *Trends Genet*. 22, 545–554. doi: 10.1016/j.tig.2006.08.005

Rattanakan, S., George, I., Haynes, P. A., and Cramer, G. R. (2016). Relative quantification of phosphoproteomic changes in grapevine (*Vitis vinifera* L.) leaves in response to abscisic acid. *Hortic. Res*. 3:16029. doi: 10.1038/hortres.2016.29

Sabila, M., Kundu, N., Smalls, D., and Ullah, H. (2016). Tyrosine phosphorylation based homo-dimerization of *Arabidopsis* RACK1A proteins regulates oxidative stress signaling pathways in yeast. *Front. Plant Sci*. 7:176. doi: 10.3389/fpls.2016.00176

Sheen, J. (1996). Ca^{2+}-dependent protein kinases and stress signal transduction in plants. *Science* 274, 1900–1902. doi: 10.1126/science.274.5294.1900

Tetlow, I. J., Wait, R., Lu, Z., Akkasaeng, R., Bowsher, C. G., Esposito, S., et al. (2004). Protein phosphorylation in amyloplasts regulates starch branching enzyme activity and protein-protein interactions. *Plant Cell* 16, 694–708. doi: 10.1105/tpc.017400

Thingholm, T. E., Jensen, O. N., and Larsen, M. R. (2009). Analytical strategies for phosphoproteomics. *Proteomics* 9, 1451–1468. doi: 10.1002/pmic.200800454

Tian, C. C., Zha, X. Q., and Luo, J. P. (2015). A polysaccharide from *Dendrobium huoshanense* prevents hepatic inflammatory response caused by carbon tetrachloride. *Biotechnol. Biotechnol. Equip*. 29, 132–138. doi: 10.1080/13102818.2014.987514

Tian, C. C., Zha, X. Q., Pan, L. H., and Luo, J. P. (2013). Structural characterization and antioxidant activity of a low-molecular polysaccharide from *Dendrobium huoshanense*. *Fitoterapia* 91, 247–255. doi: 10.1016/j.fitote.2013.09.018

Trost, P., Fermani, S., Marri, L., Zaffagnini, M., Falini, G., Scagliarini, S., et al. (2006). Thioredoxin-dependent regulation of photosynthetic glyceraldehyde-3-phosphate dehydrogenase: autonomous vs. CP12-dependent mechanisms. *Photosynth. Res*. 89, 263–275. doi: 10.1007/s11120-006-9099-z

Wang, C., Xu, H., Lin, S., Deng, W., Zhou, J., Zhang, Y., et al. (2020). GPS 5.0: an update on the prediction of kinase-specific phosphorylation sites in proteins. *Genomics Proteomics Bioinformatics* 18, 72–80. doi: 10.1016/j.gpb.2020.01.001

Wang, H. Y., Li, Q. M., Yu, N. J., Chen, W. D., Zha, X. Q., Wu, D. L., et al. (2019). *Dendrobium huoshanense* polysaccharide regulates hepatic glucose homeostasis and pancreatic β-cell function in type 2 diabetic mice. *Carbohydr. Polym*. 211, 39–48. doi: 10.1016/j.carbpol.2019.01.101

Wang, X. Y., Luo, J. P., Chen, R., Zha, X. Q., and Wang, H. (2014). The effects of daily supplementation of *Dendrobium huoshanense* polysaccharide on ethanol-induced subacute liver injury in mice by proteomic analysis. *Food Funct*. 5, 2020–2035. doi: 10.1039/c3fo60629e

Wang, Z. J., Jiang, W. M., Liu, Y. Y., Meng, X. X., Su, X. L., Cao, M. Y., et al. (2021). Putative genes in alkaloid biosynthesis identified in *Dendrobium officinale* by correlating the contents of major bioactive metabolites with genes expression between Protocorm-like bodies and leaves. *BMC Genomics* 22:579. doi: 10.1186/s12864-021-07887-6

Wei, G., Shun, Q. S., Dai, Y. F., Ye, J. H., Li, B., Huang, Y. C., et al. (2014). Study on HPLC characteristic chromatogram of *Dendrobium huoshanense*. *Chin. Pat. Med*. 36, 2642–2644.

Wu, J., Meng, X. X., Jiang, W. M., Wang, Z. J., Zhang, J., Meng, F., et al. (2022). Qualitative Proteome-wide analysis reveal the diverse functions of lysine crotonylation in *Dendrobium huoshanense*. *Front. Plant Sci.* 13:822374. doi: 10.3389/fpls.2022.822374

Wu, X., Yuan, J., Luo, A. X., Chen, Y., and Fan, Y. J. (2016). Drought stress and re-watering increase secondary metabolites and enzyme activity in *Dendrobium moniliforme*. *Ind. Crops Prod*. 94, 385–393. doi: 10.1016/j.indcrop.2016.08.041

Xing, X., Cui, S. W., Nie, S., Phillips, G. O., Goff, H. D., and Wang, Q. (2015). Study on *Dendrobium officinale* O-acetyl-glucomannan (Dendronan®): part II. fine structures of *O*-acetylated residues. *Carbohydr. Polym*. 117, 422–433. doi: 10.1016/j.carbpol.2014.08.121

Yang, L., Wen, K. S., Ruan, X., Zhao, Y. X., Wei, F., and Wang, Q. (2018). Response of plant secondary metabolites to environmental factors. *Molecules* 23:762. doi: 10.3390/molecules23040762

Yang, Y., Zhang, C., Tang, R. J., Xu, H. X., Lan, W. Z., Zhao, F., et al. (2019). Calcineurin B-like proteins CBL4 and CBL10 mediate two independent salt

tolerance pathways in *Arabidopsis*. *Int. J. Mol. Sci.* 20:2421. doi: 10.3390/ijms20102421

Yi, S. Y., Kang, C. Z., Wang, W., Song, X. W., Xu, T., Lu, H. B., et al. (2021). Comparison of planting modes of *Dendrobium huoshanense* and analysis of advantages of simulated cultivation. *China J. Chin. Mater. Med.* 46, 1864–1868.

Yu, Z., He, C., Teixeira da Silva, J. A., Luo, J., Yang, Z., and Duan, J. (2018). The GDP-mannose transporter gene *(DoGMT*) from *Dendrobium officinale* is critical for mannan biosynthesis in plant growth and development. *Plant Sci.* 277, 43–54. doi: 10.1016/j.plantsci.2018.07.021

Yuan, Y., Yu, M., Jia, Z., Song, X., Liang, Y., and Zhang, J. (2018). Analysis of *Dendrobium huoshanense* transcriptome unveils putative genes associated with active ingredients synthesis. *BMC Genomics* 19:978. doi: 10.1186/s12864-018-5305-6

Yuan, Y., Yu, M., Zhang, B., Liu, X., and Zhang, J. (2019). Comparative nutritional characteristics of the three major Chinese *Dendrobium* species with different growth years. *PLoS One* 14:e0222666. doi: 10.1371/journal.pone.0222666

Zhang, S., Du, H., and Klessig, D. F. (1998). Activation of the tobacco SIP kinase by both a cell wall-derived carbohydrate elicitor and purified proteinaceous elicitins from *Phytophthora* spp. *Plant Cell* 10, 435–450. doi: 10.1105/tpc.10.3.435

Zhao, C., Wang, P., Si, T., Hsu, C. C., Wang, L., Zayed, O., et al. (2017). MAP kinase cascades regulate the cold response by modulating ICE1 protein stability. *Dev. Cell* 43, 618–629. doi: 10.1016/j.devcel.2017.09.024

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

frontiers | Frontiers in Plant Science

TYPE Original Research
PUBLISHED 23 November 2022
DOI 10.3389/fpls.2022.874761

OPEN ACCESS

EDITED BY
Klára Kosová,
Crop Research Institute (CRI), Czechia

REVIEWED BY
Ghazala Mustafa,
Quaid-i-Azam University, Pakistan
Xia Wu,
University of Washington,
United States

*CORRESPONDENCE
Andrej Frolov
afrolov@ipb-halle.de

SPECIALTY SECTION
This article was submitted to Plant Proteomics and Protein Structural Biology, a section of the journal Frontiers in Plant Science

RECEIVED 12 February 2022
ACCEPTED 26 October 2022
PUBLISHED 23 November 2022

CITATION
Leonova T, Ihling C, Saoud M, Frolova N, Rennert R, Wessjohann LA and Frolov A (2022) Does filter-aided sample preparation provide sufficient method linearity for quantitative plant shotgun proteomics? *Front. Plant Sci.* 13:874761. doi: 10.3389/fpls.2022.874761

Does filter-aided sample preparation provide sufficient method linearity for quantitative plant shotgun proteomics?

Tatiana Leonova[1,2], Christian Ihling[3], Mohamad Saoud[1], Nadezhda Frolova[1,2], Robert Rennert[1], Ludger A. Wessjohann[1] and Andrej Frolov[1,2]*

[1]Department of Bioorganic Chemistry, Leibniz Institute of Plant Biochemistry, Halle (Saale), Germany, [2]Department of Biochemistry, St Petersburg State University, St Petersburg, Russia, [3]Institute of Pharmacy, Department of Pharmaceutical Chemistry and Bioanalytics, Martin-Luther Universität Halle-Wittenberg, Halle (Saale), Germany

Due to its outstanding throughput and analytical resolution, gel-free LC-based shotgun proteomics represents the gold standard of proteome analysis. Thereby, the efficiency of sample preparation dramatically affects the correctness and reliability of protein quantification. Thus, the steps of protein isolation, solubilization, and proteolysis represent the principal bottleneck of shotgun proteomics. The desired performance of the sample preparation protocols can be achieved by the application of detergents. However, these compounds ultimately compromise reverse-phase chromatographic separation and disrupt electrospray ionization. Filter-aided sample preparation (FASP) represents an elegant approach to overcome these limitations. Although this method is comprehensively validated for cell proteomics, its applicability to plants and compatibility with plant-specific protein isolation protocols remain to be confirmed. Thereby, the most important gap is the absence of the data on the linearity of underlying protein quantification methods for plant matrices. To fill this gap, we address here the potential of FASP in combination with two protein isolation protocols for quantitative analysis of pea (*Pisum sativum*) seed and *Arabidopsis thaliana* leaf proteomes by the shotgun approach. For this aim, in comprehensive spiking experiments with bovine serum albumin (BSA), we evaluated the linear dynamic range (LDR) of protein quantification in the presence of plant matrices. Furthermore, we addressed the interference of two different plant matrices in quantitative experiments, accomplished with two alternative sample preparation workflows in comparison to conventional FASP-based digestion of cell lysates, considered here as a reference. The spiking experiments revealed high sensitivities (LODs of up to 4 fmol) for spiked BSA and LDRs of at least 0.6×10^2. Thereby, phenol extraction yielded slightly better

recoveries, whereas the detergent-based method showed better linearity. Thus, our results indicate the very good applicability of FASP to quantitative plant proteomics with only limited impact of the protein isolation technique on the method's overall performance.

KEYWORDS

detergent-assisted proteolysis, filter aided sample preparation (FASP), label-free quantification, LC-MS, phenol extraction, plant proteomics, shotgun proteomics, sodium dodecyl sulfate

Introduction

To date, bottom-up proteomics represents one of the most established methodological platforms for post-genomic research (Zhang et al., 2013). During the last decade, gel-free LC-based shotgun proteomics became a gold standard of proteome analysis due to its higher throughput, superior proteome coverage, better analytical resolution, and reproducibility (Smolikova et al., 2020). However, because of the mechanistic limitations of electrospray ionization (ESI), shotgun proteomics is critically sensitive to detergents (Frolov et al., 2018). On the other hand, in contrast to in-gel proteolysis (which can be quantitatively accomplished in ammonium bicarbonate buffer), in-solution digestion for shotgun proteomics ultimately requires supplementation of detergents to ensure quantitative solubilization of protein isolates and their efficient proteolysis (Wiśniewski et al., 2009). Therefore, a broad range of protocols for detergent-assisted proteolysis, employing degradable or selectively removable detergents, were successfully established to date (Waas et al., 2014). One of the most widely spread detergent-based digestion techniques is the filter-aided sample preparation (FASP)—an elegant approach efficiently combining the advantages of in-gel and in-solution digestion protocols (Wiśniewski et al., 2009). This method allows complete solubilization of dried protein isolates in sodium dodecyl sulfate (SDS) aqueous solutions, centrifugal concentration of reconstituted proteins, efficient reduction, and alkylation, followed by detergent removal and digestion of proteins in one centrifugal filter device. After the introduction of FASP by Mann's group in 2009 (Wiśniewski et al., 2009), the original protocol was subjected to various modifications to improve recovery of proteolytic peptides (Wiśniewski, 2019). Thus, binding of peptides to the membrane could be reduced by conditioning of filters with 5% (*v*/*v*) Tween® 20, whereas supplementation of 0.2% (*w*/*v*) deoxycholic acid prior to proteolysis resulted in enhancement of digestion efficiency (Erde et al., 2014). Furthermore, implementation of a multienzyme digestion FASP provided improved protein identification rates and sequence coverage of individual species (Wiśniewski and Mann, 2012).

The principal advantage of FASP technology is its ability to couple uniform and efficient protein extraction and/or reconstitution protocols with powerful proteolysis techniques. In this context, due to the tremendous variation in the properties of biological matrices from different (plant) species, the entire sample preparation protocol typically requires intensive optimization for each of them to ensure the best possible performance (Wang et al., 2018). For this reason, integration of FASP with protein extraction techniques still requires validation, which typically relies on the assessment of proteome coverage in comparison to other sample preparation pipelines (Wiśniewski, 2019). In these experiments, FASP [which was originally proposed as the method for digestion of cell lysates (Wiśniewski et al., 2009)] was shown to be widely applicable to human and animal samples (Wiśniewski and Mann, 2012). Moreover, this method was successfully employed in proteomics analyses of plant organs—leaves (Wang et al., 2018), seeds (Min et al., 2019), roots (Jiang et al., 2017), and fruits (Szymanski et al., 2017). Thereby, FASP proved to be compatible with all three major protein isolation strategies —phenol extraction, precipitation with TCA/acetone, and their combination (Wang et al., 2018; Heyer et al., 2019).

Although the efficiency of FASP for plant samples was characterized in terms of protein identification rates and sequence coverage, the potential of this technique for quantitative plant proteomics remains completely unknown. Indeed, high contents of proteases, carbohydrates, and secondary metabolites, including protein-binding polyphenols, characteristic for recalcitrant plant tissues, might interfere with protein extraction and MS analysis, dramatically affecting, thereby, the linear dynamic range (LDR) of protein quantification. However, the information about the LDR of protein quantification by FASP-based bottom-up shotgun proteomics is still missing. Moreover, the impact of specific protein isolation techniques in the overall result of such experiments is also unknown. Hence, the FASP-based sample preparation methods still require validation in terms of their applicability for quantitative assessments. Therefore, to fill this gap, we addressed the potential of FASP in combination with

two protein isolation protocols for quantification of *Arabidopsis thaliana* leaf and pea (*Pisum sativum* L.) seed proteins by LC-MS-based bottom-up shotgun proteomics. Due to the presence of strongly dominating major proteins (RuBisCO in leaves and storage proteins in seeds), these organs belong to the most difficult plant matrices, i.e., the most strongly affecting protein quantitation. Therefore, we evaluated the LDR of protein quantification in the presence of these complex plant matrices processed by two alternative sample preparation workflows, each in comparison to conventional FASP-based digestion protocols of cell lysates. Our results indicate the applicability of FASP for quantitative plant proteomics with a limited impact of protein isolation technique used on the overall method performance.

Materials and methods

Materials and reagents

Materials were obtained from the following manufacturers: Biowest (Nuaillé, France): fetal bovine serum, South America; Capricorn Scientific GmbH (Ebsdorfergrund, Germany): RPMI 1640, Dulbecco's PBS (1×), penicillin/streptomycin (100×), L-glutamine solution (200 mmol/L), and trypsin-EDTA (0.05%) in DPBS (1×); Carl Roth GmbH and Co (Karlsruhe, Germany): tris (hydroxymethyl)aminomethane (tris, ultra-pure grade), tetramethylethylenediamine (TMED, p.a.), ammonium persulfate (ACS grade), glycerol (p.a.), and bovine serum albumin (BSA); CDS Analytical, LLC (Oxford, PA, USA): Empore Extraction C18 Disks; Honeywell (Charlotte, NC, USA): acetonitrile (LC-MS grade) and methanol (LC-MS grade); PanReac AppliChem (Darmstadt, Germany): glycerol (ACS grade), phenylmethylsulfonyl fluoride (PMSF), and polysorbate 20 (Tween® 20); SERVA Electrophoresis GmbH (Heidelberg, Germany): 2-mercaptoethanol (research grade), acrylamide/bis-acrylamide solution [37.5/1, 30% (*w/v*), 2.6% C], NB sequencing grade modified trypsin from porcine pancreas, and sodium dodecyl sulfate (SDS, electrophoresis grade); Thermo Fisher Scientific (Bremen, Germany): PageRuler™ Prestained Protein Ladder #26616 (10–180 kDa); VWR Chemicals, LLC (Solon, OH, USA): phenol (ultra-pure). Amicon® Ultra-0.5 Centrifugal Filter Unit of 30 kDa molecular weight cutoff (MWCO) and all other chemicals were purchased from Merck KGaA (Darmstadt, Germany).

The prostate cancer cell line PC-3 was obtained from German Collection of Microorganisms and Cell Cultures GmbH and maintained routinely in complete RPMI medium 1640 supplemented with 10% of heat-inactivated FBS, 1% glutamine, and 1% penicillin/streptomycin at 37°C in a humidified atmosphere with 5% CO_2 to reach subconfluency (~80%) prior to subsequent usage or subculturing (Smolko et al., 2020). For protein isolation, cells were washed three times with ice-cold PBS solution and harvested by adding the detergent-containing extraction buffer.

Pea seeds of the cultivar "Millennium" were obtained from the Research and Practical Center of the National Academy of Science of the Republic of Belarus for Arable Farming (Zhodino, Belarus, harvested in the year 2015 and stored at 4°C). *Arabidopsis thaliana* (Columbia 1092) seeds were planted in wet soil–sand mixture, and the plants were grown in a phytotron MLR-351H (Sanyo Electric Co., Ltd., Moriguchi, Japan) under a short day with an 8-h light (150 ± 2.5 µmol photons m^{-2} s^{-1})/16-h dark cycle at 23/18°C, respectively, and 60% humidity. The plants were harvested after 6 weeks of growth, and the leaves were ground in liquid nitrogen using a Mixer Mill MM 400 ball mill with 3-mm-diameter stainless steel balls (Retsch, Haan, Germany) at a vibration frequency of 30 Hz for 2 min. The plants were characterized elsewhere (Bilova et al., 2016).

Protein isolation

Protein isolation relied on two approaches: (i) the phenol extraction method described in detail previously (Antonova et al., 2019) and (ii) treatment with an SDS-based solution according to the procedure recently introduced by Bassal et al. (2020) with minor modifications. All experiments were performed in triplicate. In detail, protein isolation from PC-3 cells and a mixture of pea seed powder and Arabidopsis leaf material was accomplished by treatment with extraction buffer [4% (*w/v*) SDS, 10 mmol/L dithiothreitol (DTT), 10 mmol/L EDTA, and 1 mmol/L PMSF in 50 mmol/L Tris-HCl, pH 8.0], followed by two steps of incubation (900 rpm; at 95°C for 10 min, at room temperature for 20 min) and centrifugation (25,000 *g*, 30 min, 10°C) after each incubation. The resulting supernatants were transferred to new tubes.

The total protein fractions from pea seeds and mixtures of pea seed powder and Arabidopsis leaf material were isolated by the phenol extraction procedure described in detail previously (Antonova et al., 2019), and dry protein pellets were reconstituted in 10% (*w/v*) SDS solution. The protein contents were determined by the 2D Quant Kit according to the manufacturer's instructions.

FASP protocol

The Amicon® Ultra 30K filter units were conditioned (passivated) on the day prior to digestion with 5% (*v/v*) aqueous Tween® 20 under continuous shaking overnight. Afterwards, filters were washed three times for 10 min with distilled water. The protein aliquots (50 µg) were adjusted to the total volume of 200 µl with urea solution (8 mol/L urea in 50 mmol/L Tris-HCl, pH 8.0), applied to the filter unit and centrifuged (here and below—14,000 *g*, 10 min). The concentrated samples were washed three times with 200 µl of urea solution followed each time by centrifugation. Disulfide

bonds were reduced by the addition of 100 µl of a solution of 100 mmol/L DTT and 8 mol/L urea in 50 mmol/L Tris-HCl, pH 8.0, and incubation at 22°C for 1 h under continuous shaking (450 rpm) followed by centrifugation. Alkylation of sulfhydryl groups was accomplished by the addition of 100 µl of 50 mmol/L iodoacetamide in a solution of 8 mol/L urea in 50 mmol/L Tris-HCl, pH 8.0, and incubation at 22°C for 1 h in the dark under continuous shaking followed by centrifugation. The resulting concentrated samples were washed three times with 200 µl of urea solution and three times with 100 µl of 50 mmol/L aq. NH_4HCO_3 followed each time by centrifugation. Afterwards, the proteins were digested by sequential addition of two aliquots of 2.5 and 1 µg trypsin, reconstituted in 50 µl of 50 mmol/L aq. NH_4HCO_3 (enzyme-to-protein ratio 1:20 and 1:50 *w/w*, respectively) and incubation at 37°C for 4 and 12 h under continuous shaking (450 rpm), respectively. The digests were collected by centrifugation, and the filter units were rinsed with 40 µl of 50 mmol/L aq. NH_4HCO_3 three times followed by centrifugation. The resulting filtrates were desalted by solid phase extraction (SPE) as described by Mamontova et al. (2019). The completeness of tryptic digestion was verified by SDS-PAGE as described by Greifenhagen et al. (2016).

LC-MS/MS

All samples were analyzed by nanoRP-HPLC-ESI-MS/MS using an Orbitrap XL hybrid mass spectrometer equipped with a NanoFlex source, coupled online to an Ultimate 3000 nano-HPLC system (Thermo Fisher Scientific, Bremen, Germany). Proteolytic peptides (1 µg) were loaded on a trap column (PepMap 100 C18, 300 µm × 5 mm, particle size 5 µm) during 15 min with 0.1% (*v/v*) TFA at a flow rate of 30 µl/min and resolved on a separation column (PepMap 100 C18, 75 µm × 150 mm, particle size 3 µm, Thermo Fisher Scientific, Bremen, Germany). The peptides were separated with the linear gradient from 3% to 35% eluent B over 90 min (A: 0.1% *v/v* aqueous formic acid, B: 0.08% *v/v* formic acid in acetonitrile) at a flow rate of 300 nl/min. The raw files were acquired as data-dependent acquisition (DDA) experiments accomplished in positive ion mode. Dependent tandem mass spectrometric (MS/MS) experiments relied on higher-energy collision-induced dissociation (HCD) at 27% normalized collision energy (NCE). MS data (*m/z* range 300–1500) were recorded with $R = 60{,}000$, the target of the automated gain control (AGC) was set to 2×10^5, and the maximum injection time was 50 ms. Each full scan was followed by high-resolution HCD product ion scans within 5 s, starting with the most intense signal in the mass spectrum, with charge states ranging from 2 to 6. For MS/MS scans, the following parameters were applied: a resolution (R) of 15,000, an AGC of 5×10^4, and a maximum injection time of 200 ms. Dynamic exclusion of multi-charged peptide ions was set to 60 s. Targeted analyses relied on a nonscheduled SRM acquisition method to quantitate up to 10 peptides in one run (three transitions per peptide). Mass tolerances for searching precursor and fragment ions (defined as trap isolation window) were ± 0.5 and 1.5 Da, respectively. The mass spectrometry proteomics data were deposited to the ProteomeXchange Consortium *via* the PRIDE (Perez-Riverol et al., 2019) partner repository with the dataset identifier PXD025897 and 10.6019/PXD025897.

Data analysis

Identification of peptides relied on Proteome Discoverer software (version 2.2.0.388, Thermo Fisher Scientific, Bremen, Germany) and Sequest HT search engine. For database search, the following UniProt reference FASTA files were used: *A. thaliana* (39,299 entries, downloaded 7 September 2019) and BSA (entry P02769, downloaded 29 October 2019). The enzyme was set to trypsin, tolerating two missed cleavages. The precursor and fragment mass tolerance were set to 10 ppm and 0.8 Da, respectively. Carbamidomethylation of cysteine was employed as a fixed modification, and oxidation of methionine was specified as a variable one. False discovery rate (FDR) was set to 0.05. Proteotypic peptides were selected for integration if they were confidently (XCorr ≥ 2.20) annotated as $[M+2H]^{2+}$ ions, and did not contain missed cleavage sites, modifications, and methionine or cysteine amino acid residues. The peak integration was accomplished in the Quan Browser application of the Xcalibur software (Thermo Fisher Scientific). Detection and integration of the peptide-specific peaks in corresponding extracted ion chromatograms (XICs, m/z ± 0.02) were accomplished by ICIS algorithm with the following parameters: Smoothing Points 15, Baseline Window 40–80, Area Noise Factor 5, and Peak Noise Factor 10. The linearity of protein quantification was assessed for several proteotypic peptides by Xcalibur software (version 2.0.7, Thermo Fisher Scientific, Bremen, Germany). For relative quantification of SRM data, the peak areas were determined using Skyline software (version 22.2.0.255, MacCoss Lab Software, USA, https://skyline.ms/project/home/software/Skyline/begin.view). The default transition settings were applied except for method match tolerance *m/z* that was set to 0.6. Intensity of each proteotypic peptide was calculated as a sum of chromatographic area of each fragment ion.

Results and discussion

Quantitative analysis of BSA spiked in cell lysates

Originally, FASP was proposed for (human/animal) cell lysates (Wiśniewski et al., 2009), and its performance was comprehensively characterized with various cell lines and experimental setups (Wiśniewski et al., 2009; Ni et al., 2017; Potriquet et al., 2017). Therefore, here we decided for cultured cells as a reference to estimate the linear performance of this

method. Following the previously established methodology, prostate cancer (PC-3) cells were lysed with extraction buffer containing 4% (*w/v*) SDS, incubated at 95°C, and centrifuged; i.e., the procedure reproduced the classical workflow of Wiśniewski et al. (2009). Of course, the total cell lysate represents a complex system, highly prone to matrix effects (most probably, and predominantly, ion suppression), (Mitchell, 2010), which can affect the LDRs of individual tryptic peptides.

To get access to the LDRs in the easiest and most straightforward way, we applied a spiking approach, which represents a well-established normalization strategy in label-free quantification (LFQ) experiments (Tuli et al., 2012). Specifically, after determination of protein concentrations, aliquots of cell lysates were spiked with BSA at the percentage concentration ratios of 3.125%, 6.25%, 12.5%, 25%, 50%, and 100% (*w/w*) that corresponded to 0.47, 0.94, 1.88, 3.76, 7.5, and 15 pmol of BSA loads, respectively. This range of ratios covered two orders of magnitude that in most of the cases is sufficient for characterization of typical protein expression responses, i.e., allows monitoring up to 100-fold alteration in expression of individual proteins. Thereby, based on earlier performed linearity tests, the point 15 pmol is above the LDR (Zauber et al., 2013). As the protein expression dynamic range in both cell lysates and plant protein extracts accounts at least seven orders of magnitude (Geiger et al., 2012), specific groups of individual proteins can be addressed by optimization of the protein extraction scale and by a broad selection of enrichment, depletion, and fractionation methods (Ly and Wasinger, 2011).

Although spiking with mixtures of standard peptides represents an adequate approach to estimate method linearity performance (Beri et al., 2015), we decided here for the spike with allogenic protein, as this approach might consider not only matrix effects, but also the contribution of factors related to proteolysis. Bovine serum albumin seems to be a suitable protein for this purpose. Indeed, it has 58 lysyl and 24 arginyl residues, which are relatively homogeneously distributed in the protein sequence. It gives a rich selection of potential proteotypic peptides covering the whole range of peptide-specific retention times. Indeed, serum albumin was established as a tool for reference protein normalization in both cell (Chang et al., 2012) and plant (Zauber et al., 2013) proteomics. Thus, the behavior of its proteolytic peptides in plant protein hydrolysates is well-characterized and is in agreement with dynamic ranges of cell- and plant-derived tryptic peptides. This, in turn, will ensure comparability of the results obtained with cell lysates and plant protein extracts. In agreement with earlier reports (Ni et al., 2017; Potriquet et al., 2017), FASP proved to be an efficient tool for the analysis of cell lysates. The completeness of tryptic digestion was verified by SDS-PAGE and also was in agreement with the results of previous studies (Wiśniewski et al., 2009).

After the nanoRP-HPLC-ESI-LIT-Orbitrap-MS analysis, the raw files were searched by SEQUEST engine against the BSA sequence database. Sequence coverage of BSA spiked to cell lysates corresponded to 83% with 84 identified unique peptides (for the sequences, see Supplementary information 2, Tables S2-1). Based on this search, three proteotypic BSA peptides, namely, LVNELTEFAK, AEFVEVTK, and YLYEIAR, with *m/z* 582.3190 ± 0.02, 464.2504 ± 0.02, and 756.4250 ± 0.02 at t_R 52.4, 45.6, and 50.2, corresponding to the $[M+2H]^{2+}$ ions, respectively, were used to assess the linearity of the sample preparation method. The method delivered acceptable linear correlation (R^2 = 0.975) over the half of the assessed dynamic range (Figure 1A) based on the relative intensity of the overall BSA abundance response (calculated as an integrated sum of corresponding peptide signal intensities) and its contents spiked to aliquots of cell lysates. Accordingly, each of the individual peptides demonstrated an excellent linearity of response R^2 from 0.91 to 0.99 (Figures 1B–D).

Thus, in our hands, the FASP-based quantitation of BSA spiked to the cell lysate proved to be reliable at the low pmol levels. This result was in agreement with the classical works published before (Wiśniewski, 2019). Based on this finding, we assume that the PC-3 cell lysate spiked with BSA is an appropriate reference for our plant shotgun proteomics experiments.

Quantitative analysis of BSA spiked to the protein extracts of the pea seeds

Having the verified FASP method in hand, we extended our spike approach to pea seed total protein fraction. For this aim, proteins were isolated from the plant tissues by phenol extraction, dry protein pellets were reconstituted in 10% (*w/v*) SDS solution, and protein concentrations were determined. Based on these values, a 1 g/L BSA solution in 10% SDS was spiked to protein solution aliquots using the scheme described above for the cell lysates. Thus, the digestion protocol of Mann's group was transferred to the appropriate protein extraction method, which is currently considered as the most efficient one in terms of sample quality and protein identification rates (Saravanan and Rose, 2004; Carpentier et al., 2005).

The FASP approach proved to be an appropriate method for digestion of the pea seed proteins; i.e., it was ideally compatible with the phenol extraction method. Indeed, most of the total seed protein fraction (97.7%) was successfully digested and transferred through the cellulose membrane of the filter units (Figure 2A), whereas only 2.3% of the protein aliquot, subjected to proteolysis, remained on the membrane after 3× washing with 50 mmol/L aq. NH_4HCO_3 (as the averaged total lane abundance of the non-filtered fraction constituted 23% of the ND reference, which, in turn, corresponded to 10% of the digested aliquot, Figure 2B). These assessments relied on our quantitative approach, assuming sensitivity of colloidal Coomassie of about 30 ng/band (Frolov et al., 2014).

To address the efficiency of protein quantification with the FASP approach, the integrated peak areas were calculated for each peptide signal in corresponding extracted ion

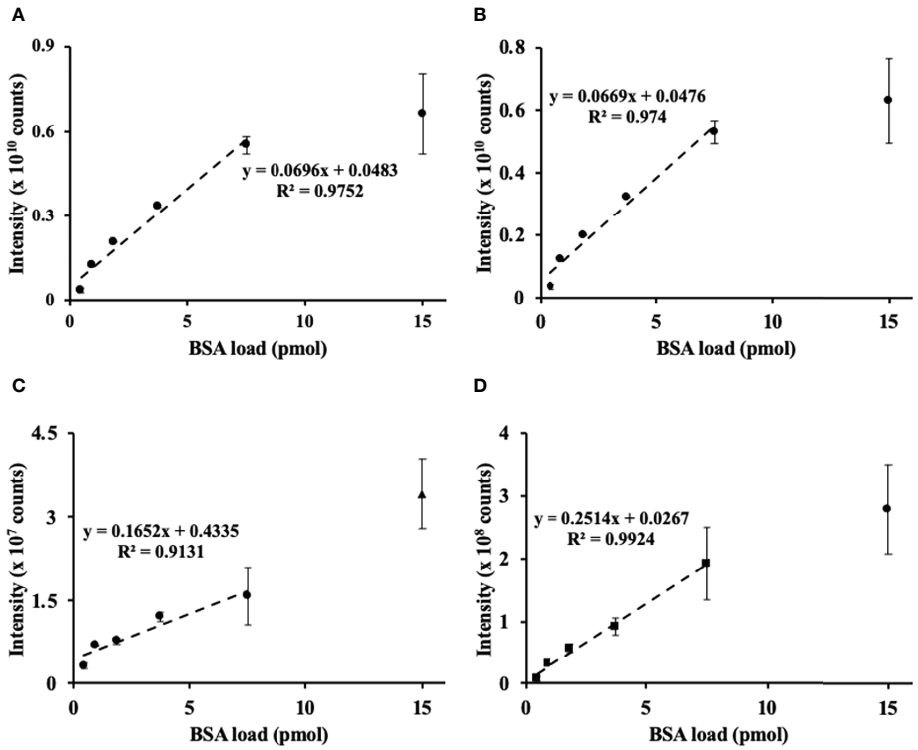


FIGURE 1
Assessment of method linearity for the quantification of bovine serum albumin (BSA) spiked to the prostate cancer (PC-3) cell lysate to obtain 0.47, 0.94, 1.88, 3.76, 7.5, and 15 pmol of BSA loads. Quantification of BSA relied on the sum **(A)** of integrated and individual peak areas obtained for *m/z* 582.3190 ± 0.02, 464.2504 ± 0.02, and 756.4250 ± 0.02 at t_R 78.4, 65.9, and 75.9, corresponding to the $[M+2H]^{2+}$ ions of the proteotypic tryptic peptides LVNELTEFAK **(B)**, YLYEIAR **(C)**, and VPQVSTPTLVEVSR **(D)**, respectively. The peak integration was accomplished in the Quan Browser application of the Xcalibur software (Thermo Fisher Scientific).

chromatograms (XICs, *m/z* ± 0.02) and summed for each concentration point. This approach yielded the same LDR as was observed above for the PC-3 cells (Figure 3A). For all three selected proteotypic BSA peptides, similar concentration–signal intensity curves were observed (Figure 3B). Thereby, the response was linear for up to 50% of spiked BSA (R^2 = 0.97–0.99).

As the next logical step, implementation of the selected reaction monitoring (SRM) as a quantitation method allowed extending the LDR of our approach down to the femtomole level. For this aim, the pea seed proteins were spiked with BSA amounts corresponding to 1, 4, 16, 62.5, 125, 500, and 2000 fmol of BSA column loads (i.e., we extended the dynamic range of our experiment three orders of magnitude down). The quantification relied on five proteotypic BSA peptides (7–20 residues, not containing methionine and internal trypsin cleavage sites) and the SRM acquisition method with three specific combinations of precursor and product *m/z* ranges per peptide (transitions) yielding the highest signal intensity in the MS/MS spectra of the DDA experiment (Table S1-1). Quantitative analysis was performed using Skyline software (Figure S1-1). The analysis yielded limits of detection (LODs) for the peptides DAFLGSFLYEYSR and LVNELTEFAK as low as 4 fmol and limits of quantification (LOQs) of 125 fmol (Table S1-2; Figures S1-2A, B), which corresponded to LDRs of 0.6 × 10^2. These values corresponded well to the published data for LTQ instruments and allowed comparable sensitivity of BSA quantification (Ishihama et al., 2005). The other selected proteotypic peptides obviously had lower ionization efficiency and showed, therefore, less favorable LODs and LOQs (Table S1-2 and Figures S1-2C–E).

This result highlights the general applicability of FASP in quantitative plant proteomics, i.e., its combination with phenol extraction gave access to acceptable linearity for quantification of BSA spiked to plant protein isolates (Figure 3). Moreover, BSA spiked to pea seed protein yielded superior sequence coverage (92%) and a higher number of identified unique peptides (93) in comparison to BSA spiked to PC-3 cell

lysate. This fact can be explained by the higher efficiency of phenol extraction with respect to discrimination from non-protein contaminations (Isaacson et al., 2006). On the other hand, this observation can be attributed to different relative abundances of individual proteins in these matrices. Indeed, pea seeds contain several strongly dominant proteins; most proteins are much less abundant (Mamontova et al., 2018). This might result in lower ion suppression at most of the chromatogram span.

Quantitative analysis of Arabidopsis proteins in pea seed protein matrix

At the next step, we addressed the impact of the protein isolation method on the performance of FASP-based quantitative proteome analysis. To get access to this information, we compared the linearity of two protein isolation methods: (i) phenol extraction and (ii) treatment with a detergent-containing solution followed by incubation at

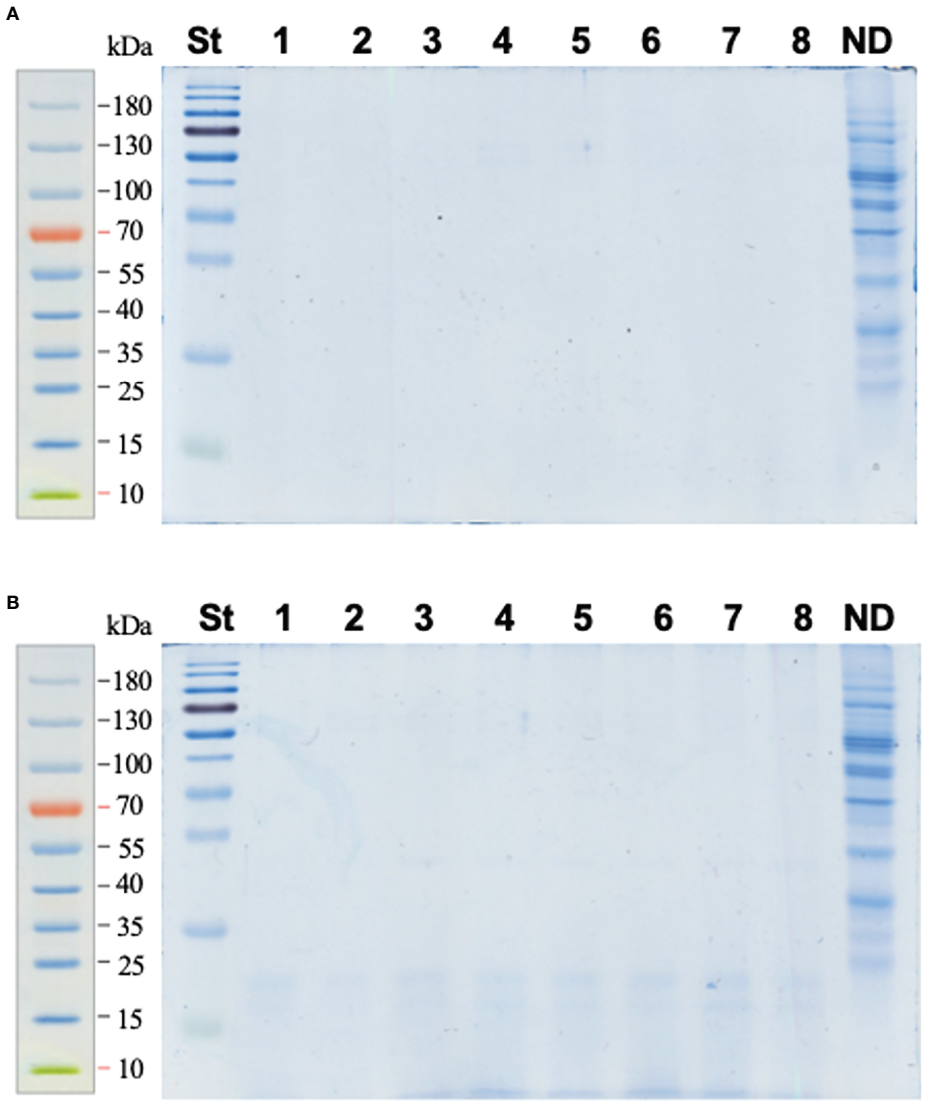


FIGURE 2
SDS-PAGE electrophoreograms acquired for the pea seed protein digested with trypsin using the FASP approach (a total of 50 μg of protein was applied to each filter unit). Lines 1–8 correspond to samples of digested pea seed protein with the spiked BSA at the percentage concentration ratios of 12.5 (bands 1–3), 6.25 (bands 4–6), and 3.125 (bands 7–8). The analyses were performed with filtrate **(A)** and the fraction retained on the filter **(B)** after 3× washing with 40 μl of 50 mmol/L aq. NH_4HCO_3 and centrifugal filtration (14,000 *g*, 10 min). To assess the completeness of hydrolysis, 5 μg of each digest was applied on the gel. The whole retained fraction (corresponding to 50 μg of digested protein) was completely transferred to a polypropylene tube, lyophilized, reconstituted in SDS-PAGE sample buffer, and loaded on the gel. The overall lane densities were compared to those of non-digested (ND) protein (5 μg) applied to a separate lane. St – Page Ruler Prestained Protein Ladder.

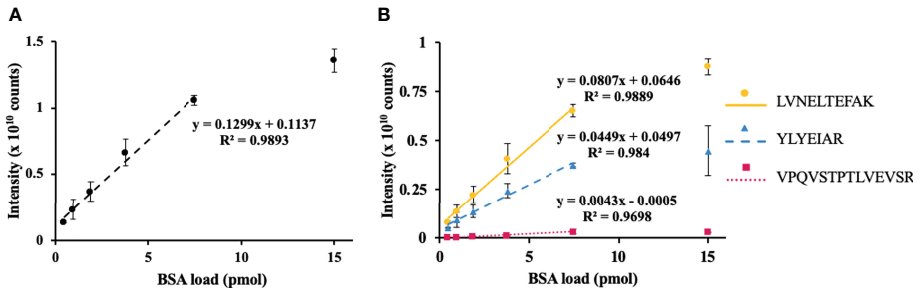


FIGURE 3
Assessment of method linearity for the quantification of bovine serum albumin (BSA) spiked to the total pea seed protein to obtain 0.47, 0.94, 1.88, 3.76, 7.5, and 15 pmol of BSA loads. Seed proteins were isolated by phenol extraction. Quantification of BSA relied on the sum **(A)** of integrated peak areas obtained for *m/z* 582.3190 ± 0.02, 464.2504 ± 0.02, and 756.4250 ± 0.02 at t_R 78.1, 66.0, and 75.7, corresponding to the $[M+2H]^{2+}$ ions of the proteotypic tryptic peptides LVNELTEFAK (circles, solid line), YLYEIAR (triangles, dashed line), and VPQVSTPTLVEVSR (squares, dotted line), respectively. Quantification of individual peptides is shown on panel **(B)** The peak integration was accomplished in the Quan Browser application of the Xcalibur software (Thermo Fisher Scientific).

95°C. For this purpose, to simulate different relative representation of proteins in the proteome, frozen milled Arabidopsis leaf material was added to the pea seed powder at the percentage concentration ratios 10%, 25%, 50%, 75%, 90%, and 100% (*w/w*). The total protein of the resulting mixtures was co-extracted using the phenol extraction procedure or treatment with SDS-containing extraction solution. The completeness of tryptic digestion was verified by SDS-PAGE (Figure S1-3).

As can be seen from Table S1-3, the efficiency of the two applied digestion protocols clearly differed—the detergent-based protocol yielded 11% more identified peptides, 17% more possible proteins, and 24% more identified non-redundant proteins in comparison to the phenol extraction. However, the phenol-based protocol yielded 12% more identified membrane protein than the former procedure (1,078 vs. 945, Table S1-4), although it was less specific for trans-membrane domains (Table S1-5). For the complete lists of the Arabidopsis proteins identified by both isolation protocols, see Tables S2-2 and S2-3.

As protein yields from pea seeds were eightfold higher in comparison to Arabidopsis leaf (5.4 vs. 42.2 mg/g fresh weight), even the most abundant Arabidopsis proteins acted as minor components of the mixed protein isolates. Therefore, we followed the abundance of two proteins, characteristic for Arabidopsis leaves—large subunit of chloroplastic ribulose bisphosphate carboxylase/oxygenase (RuBisCO, the most abundant leaf polypeptide, Figure 4A) and RuBisCO activase, which is less abundant (Figure 4B). The first protein was quantified with *m/z* 511.2693 ± 0.02, 614.8302 ± 0.02, and 704.3376 ± 0.02 at t_R 60.0, 49.4, and 44.6, respectively. These *m/z* values corresponded to $[M+H]^{2+}$ ions of proteotypic peptides DTDILAAFR, DLAVEGNEIIR, and LTYYTPEYETK, respectively, whereas the signals at *m/z* 504.2741 ± 0.02, 849.3843 ± 0.02, and 576.8606 ± 0.02 with t_R 42.2, 40.9, and 67.9, corresponded to the $[M+2H]^{2+}$ ions of the proteotypic tryptic peptides FVESLGVEK, GLAYDTSDDQQDITR, and VPLILGIWGGK, respectively.

As can be seen from Figure 4 (and its more detailed presentation in Figure S1-4), direct extraction with SDS-containing solution provided better linearity (higher maximal point of the LDR) of the integrated response of Arabidopsis proteins, each based on the sum of the selected three proteotypic peptides. Furthermore, we implemented SRM-MS for deeper investigation of the linearity behavior of Arabidopsis proteins. The quantification relied on 20 proteotypic peptides (7–20 residues, not containing methionine and internal trypsin cleavage sites) representing six proteins of different abundance and SRM acquisition method with three specific combinations of precursor and product *m/z* ranges per peptide (transitions) yielding the highest signal intensity in the MS/MS spectra of the DDA experiment (Table S1-6). Quantitative analysis was performed using Skyline software (Figure S1-5). Based on the obtained data, linear regression curves could be built and the sensitivity and linearity parameters could be assessed (Figures S1-6 and S1-7). The analysis yielded the LODs and LOQ at the level of 10% of Arabidopsis material for peptides DTDILAAFR and ESTLGFVDLLR of RuBisCO large chain, as well as for VPLILGIWGGK and NILLNEGIR of RuBisCO activase and Photosystem II D2 protein, respectively. The other selected proteotypic peptides obviously had lower ionization efficiency and showed, therefore, less favorable LODs and LOQs (Tables S1-7 and S1-8). For all analyzed proteins, the detergent-based extraction showed compromised (in comparison to the phenol extraction) recovery of the corresponding proteins in SDS-containing extraction buffer when Arabidopsis leaf material contributed more than 50% in the plant material mix (Figures S1–4). Indeed, when the detergent solution was used for extraction, linearity was superior (R^2 values better than 0.95) up to a contribution of Arabidopsis material accounting 90%,

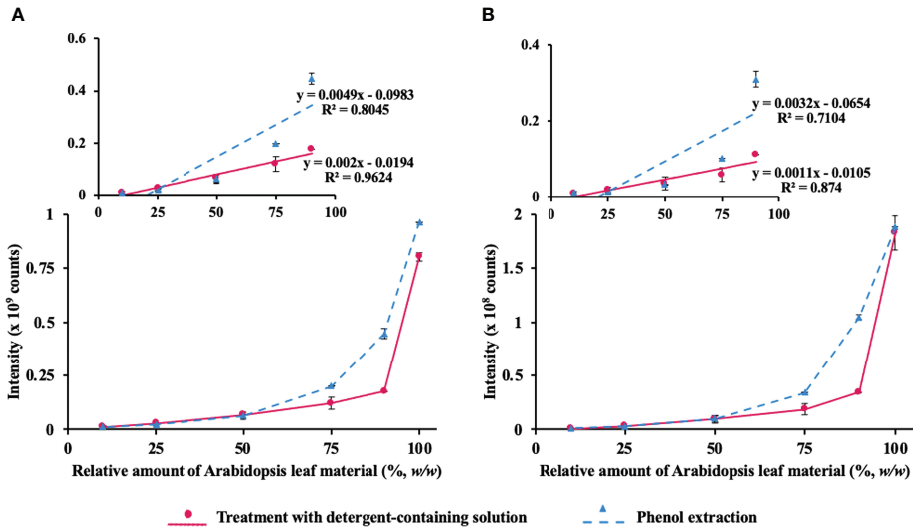


FIGURE 4
Assessment of the method linearity for quantification of *Arabidopsis thaliana* leaf proteins ribulose-1,5-bisphosphate carboxylase-oxygenase (RuBisCO) and RuBisCO activase after co-extraction from Arabidopsis leaf material added to pea seed powder at the different percentage concentrations [10%, 25%, 50%, 75%, 90%, and 100% (*w/w*)] using the phenol extraction procedure (triangles, dashed line) or treatment with SDS-containing extraction solution (circles, solid line). Quantification of RuBisCO **(A)** relied on the sum of integrated peak areas obtained for *m/z* 511.2693 ± 0.02, 614.8302 ± 0.02, and 704.3376 ± 0.02 at t_R 60.0, 49.4, and 44.6, corresponding to the $[M+2H]^{2+}$ ions of the proteotypic tryptic peptides DTDILAAFR, DLAVEGNEIIR, and LTYYTPEYETK, respectively, whereas RuBisCO activase **(B)** was quantified with *m/z* 504.2741 ± 0.02, 849.3843 ± 0.02, and 576.8606 ± 0.02 at t_R 42.2, 40.9, and 67.9, corresponding to the $[M+2H]^{2+}$ ions of the proteotypic tryptic peptides FVESLGVEK, GLAYDTSDDQQDITR, and VPLILGIWGGK, respectively. The peak integration was accomplished in the Quan Browser application of the Xcalibur software (Thermo Fisher Scientific).

whereas for phenol extraction, R^2 values typically were 0.85 or lower. However, when the quantification dynamics range was reduced to a contribution of Arabidopsis material accounting 50%, both detergent and phenol extraction methods yielded similar linearity (R^2 values of 0.99 and 0.97, respectively, data not shown).

Most likely, the observed differences were attributed to stronger matrix effects, which might accompany detergent-based extraction. Indeed, although all reagents used in the digestion are quantitatively removed during sample preparation and by online trapping in terms of the nano-LC setup, secondary metabolites (which are co-extracted with proteins in this design) cannot be quantitatively removed by both these steps and might cause inhibition of trypsin activity and ion suppression *via* co-elution with individual peptides in RP-HPLC experiments (Wang et al., 2018). On the other hand, phenol isolates are free from non-protein contaminants, and pea proteolytic peptides represented the only factor of ion suppression in our experimental system. Accordingly, the signal intensity of Arabidopsis tryptic peptides increased when the contribution of pea seed protein in the total isolate decreased.

It is worth noting that the presented approach has limitations. The spike-in experiments are quite convenient and, certainly, suitable for validation of label-free proteomics protocol, as the differences between samples are known, and method performance can be reliably characterized by the ability to identify the true differences (Riter et al., 2011). However, testing the protocol in a case study representing typical real research is another important part of the validation procedure (Riter et al., 2011; Välikangas et al., 2018) and needs to be its next step. Indeed, even though the presented FASP methodology proved to be efficient in the spike-in experiments, its performance in identifying the biologically relevant alterations in protein dynamics can be different.

Moreover, our experimental setup relied on the whole plant organs, whereas currently single-cell proteomics becomes the main road of the state-of-the-art proteomics, as it gives much better insight into biological processes without averaging effects (Kelly, 2020). However, unfortunately, plant systems bring unique challenges for single-omics experiments such as optimization of individual cell isolation from different plant species and plant organs, determining and detecting cell type-specific marker genes as well as data analysis methods (Clark et al., 2022). For this reason, the protocol validation for whole organ lysates represents a critical, absolutely mandatory, and not avoidable step in obtaining biologically relevant information.

Thus, protein isolation protocol affects the LDR of FASP-based quantitative proteomics techniques. However, this difference in LDRs can be considered during data interpretation and appropriately corrected by experimental design.

Conclusions

FASP represents a powerful and versatile technique to access quantitative and reproducible protein solubilization and digestion for shotgun proteomics. It was originally proposed for cell lysates in the mid-2000s by Mann's group. Since that time, it was comprehensively optimized and validated for cells, blood plasma, and homogenates of animal organs. Finally, during recent years, FASP was employed in plant proteomics as well. However, in our opinion, this step requires a comprehensive estimation of the methods' behavior with respect to plant matrix, which is known to be a much more complex biological material that is difficult to handle when compared to mammalian cells. The most critical aspect here is the effect of the plant matrix on LDRs of individual proteins. This knowledge is critically important for a correct assessment of quantitative alterations. Here, to the best of our knowledge, for the first time, we provide data on the linearity of FASP in plant matrix. Surprisingly, when coupled to plant-specific protein isolation protocols, this method demonstrates even better performance in comparison to mammalian matrices. The selection of the protein isolation protocol for plant FASP assumes a compromise between recovery (which is more favored by the phenol extraction method) and LDR (which is better when direct detergent treatment is applied). Therefore, a linearity/recovery test prior to working with a new plant matrix is mandatory for obtaining adequate quantitative information.

Data availability statement

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found below: http://www.proteomexchange.org/, PXD025897.

Author contributions

TL performed protein isolation, FASP of pea seed and Arabidopsis leaf proteins, and data analysis and wrote the manuscript draft. CI performed proteomics analysis. MS performed FASP of cell proteins. RR supervised cell experiment and contributed to the writing of the final draft of the manuscript. LW contributed to the writing of the final draft of the manuscript. AF proposed the idea of the manuscript, supervised the work, and wrote the final draft of the manuscript. All authors contributed to the article and approved the submitted version.

Funding

The financial support from the Russian Foundation of Basic Research (#20-54-00044 and #20-34-90160), Deutsche Forschungsgemeinschaft (DFG, project #FR3117/2-3), and Leibniz Society is gratefully acknowledged.

Acknowledgments

We thank Dr. Wolfgang Höhenwarter for fruitful discussions.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpls.2022.874761/full#supplementary-material

References

Antonova, K., Vikhnina, M., Soboleva, A., Mehmood, T., Heymich, M.-L., Leonova, T., et al. (2019). Analysis of chemically labile glycation adducts in seed proteins: case study of methylglyoxal-derived hydroimidazolone 1 (MG-H1). *Int. J. Mol. Sci.* 20 (15), 3659. doi: 10.3390/ijms20153659

Bassal, M., Abukhalaf, M., Majovsky, P., Thieme, D., Herr, T., Ayash, M., et al. (2020). Reshaping of the arabidopsis thaliana proteome landscape and Co-regulation of proteins in development and immunity. *Mol. Plant* 13, 1709–1732. doi: 10.1016/j.molp.2020.09.024

Beri, J., Rosenblatt, M. M., Strauss, E., Urh, M., and Bereman, M. S. (2015). Reagent for evaluating liquid chromatography-tandem mass spectrometry (LC-MS/MS) performance in bottom-up proteomic experiments. *Anal. Chem.* 87, 11635–11640. doi: 10.1021/acs.analchem.5b04121

Bilova, T., Lukasheva, E., Brauch, D., Greifenhagen, U., Paudel, G., Tarakhovskaya, E., et al. (2016). A snapshot of the plant glycated proteome: Structural, functional, and mechanistic aspects. *J. Biol. Chem.* 291, 7621–7636. doi: 10.1074/jbc.M115.678581

Carpentier, S. C., Witters, E., Laukens, K., Deckers, P., Swennen, R., and Panis, B. (2005). Preparation of protein extracts from recalcitrant plant tissues: An evaluation of different methods for two-dimensional gel electrophoresis analysis. *Proteomics* 5, 2497–2507. doi: 10.1002/PMIC.200401222

Chang, Y. H., Lee, S. H., Liao, I. C., Huang, S. H., Cheng, H. C., and Liao, P. C. (2012). Secretomic analysis identifies alpha-1 antitrypsin (A1AT) as a required protein in cancer cell migration, invasion, and pericellular fibronectin assembly for facilitating lung colonization of lung adenocarcinoma cells. *Mol. Cell. Proteomics* 11, 1320–1339. doi: 10.1074/mcp.M112.017384

Clark, N. M., Elmore, J. M., and Walley, J. W. (2022). To the proteome and beyond: advances in single-cell omics profiling for plant systems. *Plant Physiol.* 188, 726–737. doi: 10.1093/PLPHYS/KIAB429

Erde, J., Loo, R. R. O., and Loo, J. A. (2014). Enhanced FASP (eFASP) to increase proteome coverage and sample recovery for quantitative proteomic experiments. *J. Proteome Res.* 13, 1885–1895. doi: 10.1021/pr4010019

Frolov, A., Blüher, M., and Hoffmann, R. (2014). Glycation sites of human plasma proteins are affected to different extents by hyperglycemic conditions in type 2 diabetes mellitus. *Anal. Bioanal Chem.* 406, 5755–5763. doi: 10.1007/s00216-014-8018-y

Frolov, A., Mamontova, T., Ihling, C., Lukasheva, E., Bankin, M., Chantseva, V., et al. (2018). Mining seed proteome: from protein dynamics to modification profiles. *Biol. Commun.* 63, 43–58. doi: 10.21638/spbu03.2018.106

Geiger, T., Wehner, A., Schaab, C., Cox, J., and Mann, M. (2012). Comparative proteomic analysis of eleven common cell lines reveals ubiquitous but varying expression of most proteins. *Mol. Cell. Proteomics* 11 (3), M111.014050. doi: 10.1074/mcp.M111.014050

Greifenhagen, U., Frolov, A., Blüher, M., and Hoffmann, R. (2016). Plasma proteins modified by advanced glycation end products (AGEs) reveal site-specific susceptibilities to glycemic control in patients with type 2 diabetes. *J. Biol. Chem.* 291, 9610–9616. doi: 10.1074/jbc.M115.702860

Heyer, R., Schallert, K., Büdel, A., Zoun, R., Dorl, S., Behne, A., et al. (2019). A robust and universal metaproteomics workflow for research studies and routine diagnostics within 24 h using phenol extraction, fasp digest, and the metaproteomeanalyzer. *Front. Microbiol.* 10, 1883. doi: 10.3389/fmicb.2019.01883

Isaacson, T., Damasceno, C. M. B., Saravanan, R. S., He, Y., Catala, C., and Rose, J. K. C. (2006). Sample extraction techniques for enhanced proteomic analysis of plant tissues. *Nat. Protoc.* 1, 769–774. doi: 10.1038/nprot.2006.102

Ishihama, Y., Oda, Y., Tabata, T., Sato, T., Nagasu, T., Rappsilber, J., et al. (2005). Exponentially modified protein abundance index (emPAI) for estimation of absolute protein amount in proteomics by the number of sequenced peptides per protein. *Mol. Cell. Proteomics* 4, 1265–1272. doi: 10.1074/mcp.M500061-MCP200

Jiang, Q., Li, X., Niu, F., Sun, X., Hu, Z., and Zhang, H. (2017). iTRAQ-based quantitative proteomic analysis of wheat roots in response to salt stress. *Proteomics* 17, 1600265. doi: 10.1002/pmic.201600265

Kelly, R. T. (2020). Single-cell proteomics: Progress and prospects. *Mol. Cell. Proteomics* 19, 1739–1748. doi: 10.1074/mcp.R120.002234

Ly, L., and Wasinger, V. C. (2011). Protein and peptide fractionation, enrichment and depletion: Tools for the complex proteome. *Proteomics* 11, 513–534. doi: 10.1002/PMIC.201000394

Mamontova, T., Afonin, A. M., Ihling, C., Soboleva, A., Lukasheva, E., Sulima, A. S., et al. (2019). Profiling of seed proteome in pea (pisum sativum l.) lines characterized with high and low responsivity to combined inoculation with nodule bacteria and arbuscular mycorrhizal fungi. *Molecules* 24 (8), 1603. doi: 10.3390/molecules24081603

Mamontova, T., Lukasheva, E., Mavropolo-Stolyarenko, G., Proksch, C., Bilova, T., Kim, A., et al. (2018). Proteome map of pea (Pisum sativum l.) embryos containing different amounts of residual chlorophylls. *Int. J. Mol. Sci.* 19 (20), 4066. doi: 10.3390/ijms19124066

Min, C. W., Gupta, R., Agrawal, G. K., Rakwal, R., and Kim, S. T. (2019). Concepts and strategies of soybean seed proteomics using the shotgun proteomics approach. *Expert Rev. Proteomics* 16, 795–804. doi: 10.1080/14789450.2019.1654860

Mitchell, P. (2010). Proteomics retrenches. *Nat. Biotechnol.* 28, 665–670. doi: 10.1038/nbt0710-665

Ni, M., Wang, L., Chen, W., Mou, H., Zhou, J., and Zheng, Z. (2017). Modified filter-aided sample preparation (FASP) method increases peptide and protein identifications for shotgun proteomics. *Rapid Commun. Mass Spectrom* 31, 171–178. doi: 10.1002/rcm.7779

Perez-Riverol, Y., Csordas, A., Bai, J., Bernal-Llinares, M., Hewapathirana, S., Kundu, D. J., et al. (2019). The PRIDE database and related tools and resources in 2019: Improving support for quantification data. *Nucleic Acids Res.* 47, D442–D450. doi: 10.1093/nar/gky1106

Potriquet, J., Laohaviroj, M., Bethony, J. M., and Mulvenna, J. (2017). A modified FASP protocol for high-throughput preparation of protein samples for mass spectrometry. *PloS One* 12, e0175967. doi: 10.1371/journal.pone.0175967

Riter, L. S., Jensen, P. K., Ballam, J. M., Urbanczyk-Wochniak, E., Clough, T., Vitek, O., et al. (2011). Evaluation of label-free quantitative proteomics in a plant matrix: A case study of the night-to-day transition in corn leaf. *Anal. Methods* 3, 2733–2739. doi: 10.1039/C1AY05473B

Saravanan, R. S., and Rose, J. K. C. (2004). A critical evaluation of sample extraction techniques for enhanced proteomic analysis of recalcitrant plant tissues. *Proteomics* 4, 2522–2532. doi: 10.1002/PMIC.200300789

Smolikova, G., Gorbach, D., Lukasheva, E., Mavropolo-Stolyarenko, G., Bilova, T., Soboleva, A., et al. (2020). Bringing new methods to the seed proteomics platform: Challenges and perspectives. *Int. J. Mol. Sci.* 21 (23), 9162. doi: 10.3390/ijms21239162

Smolko, L., Smolková, R., Samoľová, E., Morgan, I., Saoud, M., and Kaluđerović, G. N. (2020). Two isostructural Co(II) flufenamato and niflumato complexes with bathocuproine: Analogues with a different cytotoxic activity. *J. Inorg. Biochem.* 210, 111160. doi: 10.1016/j.jinorgbio.2020.111160

Szymanski, J., Levin, Y., Savidor, A., Breitel, D., Chappell-Maor, L., Heinig, U., et al. (2017). Label-free deep shotgun proteomics reveals protein dynamics during tomato fruit tissues development. *Plant J.* 90, 396–417. doi: 10.1111/tpj.13490

Tuli, L., Tsai, T. H., Varghese, R. S., Xiao, J. F., Cheema, A., Ressom,, et al. (2012). Using a spike-in experiment to evaluate analysis of LC-MS data. *Proteome Sci.* 10, 13. doi: 10.1186/1477-5956-10-13

Välikangas, T., Suomi, T., and Elo, L. L. (2018). A systematic evaluation of normalization methods in quantitative label-free proteomics. *Brief Bioinform.* 19, 1–11. doi: 10.1093/BIB/BBW095

Waas, M., Bhattacharya, S., Chuppa, S., Wu, X., Jensen, D. R., Omasits, U., et al. (2014). Combine and conquer: Surfactants, solvents, and chaotropes for robust mass spectrometry based analyses of membrane proteins. *Anal. Chem.* 86, 1551–1559. doi: 10.1021/ac403185a

Wang, W. Q., Jensen, O. N., Møller, I. M., Hebelstrup, K. H., and Rogowska-Wrzesinska, A. (2018). Evaluation of sample preparation methods for mass spectrometry-based proteomic analysis of barley leaves. *Plant Methods* 14, 72. doi: 10.1186/s13007-018-0341-4

Wiśniewski, J. R. (2019). Filter aided sample preparation – a tutorial. *Anal. Chim. Acta* 1090, 23–30. doi: 10.1016/j.aca.2019.08.032

Wiśniewski, J. R., and Mann, M. (2012). Consecutive proteolytic digestion in an enzyme reactor increases depth of proteomic and phosphoproteomic analysis. *Anal. Chem.* 84, 2631–2637. doi: 10.1021/ac300006b

Wiśniewski, J. R., Zougman, A., Nagaraj, N., and Mann, M. (2009). Universal sample preparation method for proteome analysis. *Nat. Methods* 6, 359–362. doi: 10.1038/nmeth.1322

Zauber, H., Schüler, V., and Schulze, W. (2013). Systematic evaluation of reference protein normalization in proteomic experiments. *Front. Plant Sci.* 4, 25. doi: 10.3389/fpls.2013.00025

Zhang, Y., Fonslow, B. R., Shan, B., Baek, M. C., and Yates, J. R. (2013). Protein analysis by shotgun/bottom-up proteomics. *Chem. Rev.* 113, 2343–2394. doi: 10.1021/cr3003533

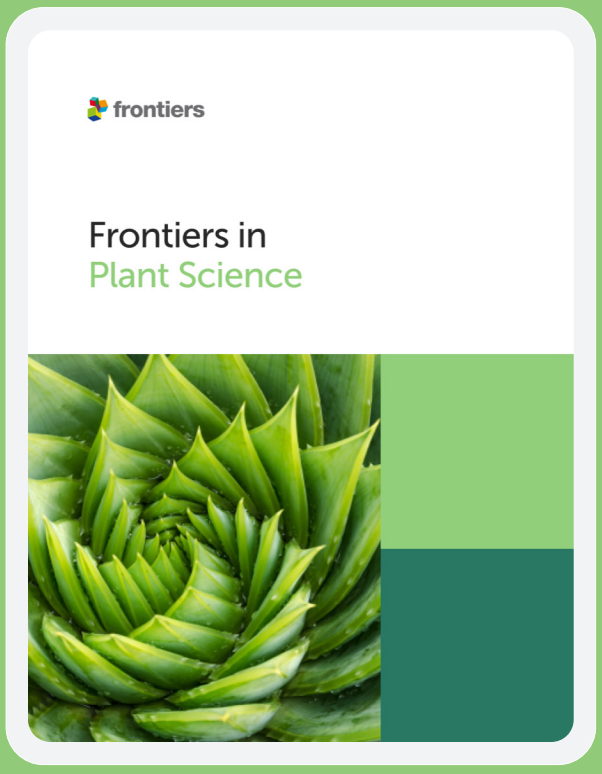
Frontiers in Plant Science

Cultivates the science of plant biology and its applications

The most cited plant science journal, which advances our understanding of plant biology for sustainable food security, functional ecosystems and human health.

Discover the latest Research Topics

See more →



Frontiers

Avenue du Tribunal-Fédéral 34
1005 Lausanne, Switzerland
frontiersin.org

Contact us

+41 (0)21 510 17 00
frontiersin.org/about/contact