



Article

Characterization of the Heat-Stable Proteome during Seed Germination in Arabidopsis with Special Focus on LEA Proteins

Orarat Ginsawaeng¹, Michal Gorka^{1,2}, Alexander Erban¹ , Carolin Heise^{1,3}, Franziska Brueckner¹, Rainer Hoefgen¹ , Joachim Kopka¹ , Aleksandra Skirydz^{1,4}, Dirk K. Hincha^{1,†} and Ellen Zuther^{1,*}

- ¹ Max Planck Institute of Molecular Plant Physiology, Am Muehlenberg 1, 14476 Potsdam, Germany; ginsawaeng@mpimp-golm.mpg.de (O.G.); michal.gorka@celonpharma.com (M.G.); Erban@mpimp-golm.mpg.de (A.E.); carolin.heise@uni-rostock.de (C.H.); brueckner@mpimp-golm.mpg.de (F.B.); hoefgen@mpimp-golm.mpg.de (R.H.); kopka@mpimp-golm.mpg.de (J.K.); skirydz@mpimp-golm.mpg.de (A.S.); hincha@mpimp-golm.mpg.de (D.K.H.)
- ² Celon Pharma S.A., Marymoncka 15, Kazun Nowy, 05-152 Czosnow, Poland
- ³ Institute of Life Science, University of Rostock, Albert-Einstein-Str. 3, 18059 Rostock, Germany
- ⁴ Boyce Thompson Institute, Cornell University, Ithaca, NY 14853, USA
- * Correspondence: zuther@mpimp-golm.mpg.de; Tel.: +49-331-5678251
- † We dedicate this paper to the memory of Dirk K. Hincha.



Citation: Ginsawaeng, O.; Gorka, M.; Erban, A.; Heise, C.; Brueckner, F.; Hoefgen, R.; Kopka, J.; Skirydz, A.; Hincha, D.K.; Zuther, E. Characterization of the Heat-Stable Proteome during Seed Germination in Arabidopsis with Special Focus on LEA Proteins. *Int. J. Mol. Sci.* **2021**, *22*, 8172. <https://doi.org/10.3390/ijms22158172>

Academic Editors: Sixue Chen

Received: 8 June 2021

Accepted: 28 July 2021

Published: 29 July 2021

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

Abstract: During seed germination, desiccation tolerance is lost in the radicle with progressing radicle protrusion and seedling establishment. This process is accompanied by comprehensive changes in the metabolome and proteome. Germination of Arabidopsis seeds was investigated over 72 h with special focus on the heat-stable proteome including late embryogenesis abundant (LEA) proteins together with changes in primary metabolites. Six metabolites in dry seeds known to be important for seed longevity decreased during germination and seedling establishment, while all other metabolites increased simultaneously with activation of growth and development. Thermo-stable proteins were associated with a multitude of biological processes. In the heat-stable proteome, a relatively similar proportion of fully ordered and fully intrinsically disordered proteins (IDP) was discovered. Highly disordered proteins were found to be associated with functional categories development, protein, RNA and stress. As expected, the majority of LEA proteins decreased during germination and seedling establishment. However, four germination-specific dehydrins were identified, not present in dry seeds. A network analysis of proteins, metabolites and amino acids generated during the course of germination revealed a highly connected LEA protein network.

Keywords: late embryogenesis abundant; Arabidopsis; seed germination; metabolomics; heat-stable proteome; dehydrins; LEA transcripts; intrinsically disordered proteins

1. Introduction

Seeds are the desiccation-tolerant life stage of most plant species. However, seedlings lose their desiccation tolerance during the early stages of germination [1–3]. In the vegetative stage, with the exception of a small number of resurrection plants, the vast majority of species are desiccation-sensitive (see [4] for a recent review).

During germination, the quiescent embryo resumes growth and development, resulting in seedling establishment under favorable environmental conditions. The process of germination is initiated upon imbibition and has been divided into three phases [5,6]. Phase I consists of a rapid uptake of water which allows the initiation of metabolic activity, such as protein synthesis from stored mRNAs. During Phase II, further water uptake is limited. Phenotypically, testa and endosperm rupture are observed, followed by radicle protrusion, which marks the end of this phase. Gene transcription from newly synthesized mRNAs

and protein synthesis results in further activation of metabolism. In Phase III, also referred to as the post-germination phase, water uptake increases again and seedling development is observed. This includes cell division, elongation of the radicle and mobilization of stored reserves to fuel seedling development before plants become photoautotrophic. After emergence of the radicle, seedling establishment progresses by emergence of hypocotyl and cotyledons and a new growth stage begins with the full opening of the cotyledons [7].

This complex developmental process has been investigated on metabolic [8–11], transcriptomic [10,12–17] and proteomic [18–26] levels. In all cases, distinct regulatory patterns have been observed for different metabolites, transcripts and proteins during germination and seedling development in different species. Among the metabolites that are typically found in high concentrations in dry seeds are sucrose and raffinose-family oligosaccharides (RFO) such as raffinose, stachyose and verbascose (see [27] for review). These sugars are mainly responsible for the formation of a cytosolic glass phase during maturation drying, which protects the cells from damage as water content declines [28–30]. Obviously, this process needs to be reversed during germination to allow full metabolic activity.

In addition to seed storage proteins, late embryogenesis abundant (LEA) proteins are highly abundant in dry seeds [31]. They were first discovered in cotton seeds during the late stages of seed development [32–34]. Besides seeds, they are also found in vegetative plant organs, mainly in response to dehydration stress. For example, it has been shown that LEA proteins can protect enzymes and membranes during freezing and drying [35–38]. LEA proteins are also found in other organisms including algae, anhydrobiotic invertebrates and bacteria [35,37,39]. In plants, LEA proteins have been linked to the desiccation tolerance of seeds as their accumulation coincides with the development of desiccation tolerance in orthodox seeds and their accumulation is decreasing in germinating seeds when they become desiccation sensitive [18,40,41]. In *Arabidopsis*, the transcripts encoding several LEA proteins have also been detected in dry, mature seeds [42]. However, the role of specific LEA proteins during seed germination is not known yet. A detailed temporal proteome analysis during seed germination and seedling establishment will help shed light on their importance during these processes.

Most LEA proteins contain a high fraction of hydrophilic amino acid residues and have been computationally and experimentally classified as intrinsically disordered proteins (IDPs) [38,42–45]. This lack of folding in solution and the low fraction of hydrophobic amino acids make most LEA proteins heat-stable [35–38]. This property has been used, on the one hand, for the purification of recombinant LEA proteins (see, e.g., [45,46] for reviews), and on the other hand, to enrich LEA proteins from seed extracts of *Medicago truncatula* in a heat-stable proteome [41,47]. In *Arabidopsis thaliana*, a total of 51 LEA proteins across nine different groups based on Pfam family domains have been identified, of which 34 were classified as seed-expressed based on the presence of the corresponding mRNAs in dry seeds [42].

Here, we explored the heat-stable proteome in dry and germinating seeds and young seedlings of *Arabidopsis* with a special focus on low abundant proteins. The assignment of heat-stable proteins to functional metabolic categories was investigated. LEA proteins and the abundance of LEA transcripts was analyzed using proteomics and qRT-PCR approaches, respectively. In addition, we investigated the dynamics of the seed metabolome for comparison with previous publications to improve our understanding of the complex metabolic changes taking place during germination and seedling establishment.

2. Results

To investigate the early stages of germination in *Arabidopsis*, seeds were imbibed and then subsequently collected at different time points. We collected dry seeds (DS) and germinating seeds and seedlings (Supplemental Figure S1). All treatments, except DS, also experienced a stratification period, which might have an additional impact on the results. After imbibition, we observed testa rupture (24 HAI; hours after imbibition), elongation of

radicles (36 HAI), greening of cotyledons (48 HAI), hypocotyl and cotyledon emergence (60 HAI) and fully opened cotyledons (72 HAI).

2.1. Dynamics of Seed Metabolites and Amino Acids after Imbibition

As metabolic levels during seed germination of *Arabidopsis* were previously reported [8,10,11], we firstly characterized the molecular status during imbibition on the metabolic level for potential comparisons and additional evaluation of the complex changes.

Small hydrophilic metabolites were analyzed using gas chromatography-mass spectrometry analysis. After excluding known contaminants and potential contaminants identified by hierarchical cluster analysis, the data set contained 100 metabolites (see Supplemental Table S1 for the complete data set). More than half of the metabolites detected in all time points increased in abundance during germination and seedling establishment. Six metabolites (A148003, A225007, erythronic acid, mannitol, uracil and uric acid) were only found in DS.

Principal component analysis (PCA) showed a clear separation between the metabolite composition of DS and all other time points along principal component 1 (PC1), which explained 61.8% of the total variance in the data set (Figure 1A). The different time points after imbibition were mainly separated by PC2, explaining 17.7% of the total variance, indicating that imbibition had the most drastic effect on metabolite composition, compared to further germination. Data obtained from some data points clustered more closely together, suggesting that the metabolite composition was similar in testa ruptured and radicle protruding seeds (24 and 36 HAI) and also before and after cotyledon opening in seedlings (60 and 72 HAI).

To resolve the different kinetic patterns of the metabolites (Supplemental Figure S2), we subjected the data of those that were detected at all time points (70% of the data set) to k-means clustering (Figure 2A). We observed six clearly resolved kinetic patterns, four of which showed an overall increase in metabolite content over 72 HAI. Cluster 1 and cluster 4 represented metabolites with a gradual increase over time from DS to 72 HAI with a higher increase in cluster 4. Similarly, metabolites in cluster 5 were highly increased from 24 HAI onwards. Cluster 6 was characterized by a higher abundance of the respective metabolites already in DS and more gradual changes over 72 HAI. The largest cluster 2 represented metabolites with only small fluctuations, especially at early time points after imbibition. The smallest cluster 3, contained metabolites with a high abundance in dry seeds and a steep decrease after 24 HAI or 36 HAI.

Metabolites of the six clusters were assigned to nine metabolite classes (Figure 2B). The most diverse clusters were cluster 1 (acids, unknowns (MSTs; mass spectral tags), N-compounds, phosphates, polyols and sugars) and cluster 2 (acids, MSTs, N-compounds, phenylpropanoids, sugar conjugates and sugars). Sugars was the only class found in all clusters, with the highest proportion in cluster 5 (40%). Acids were detected in almost every cluster except for cluster 4. MSTs, found in all clusters except for cluster 3, was the class with the highest number of members in the clustered data set (42.9%). Phenylpropanoids were exclusively observed in cluster 2.

The abundance of 20 amino acids in DS, germinating seeds and seedlings was measured using reverse High-Performance Liquid Chromatography (HPLC) (Supplemental Table S2). A PCA plot of amino acid concentrations (Figure 1B) showed a similar pattern to the PCA of metabolites. A separation of DS samples from other samples collected 24–72 HAI could be observed by PC1, which explained 72.6% of the total variance. Mainly PC2, explaining 15.9% of the total variance, separated samples collected after seed germination. The changes in amino acids during seed germination clustered into four groups according to their kinetic patterns (Figure 3). It revealed that most amino acids (75%) increased in abundance over time from DS to seedlings (Figure 3A). Amino acids with increased abundance clustered into cluster 2 and also cluster 3 with the largest number of members (55%) and a more pronounced increase (Figure 3B). The abundance of amino acids in cluster 1 slightly fluctuated during DS-to-seedling transition. Lastly, the abundance of

γ -aminobutyric acid (gABA) as the only member of cluster 4 decreased drastically 24 HAI and recovered 48 HAI.

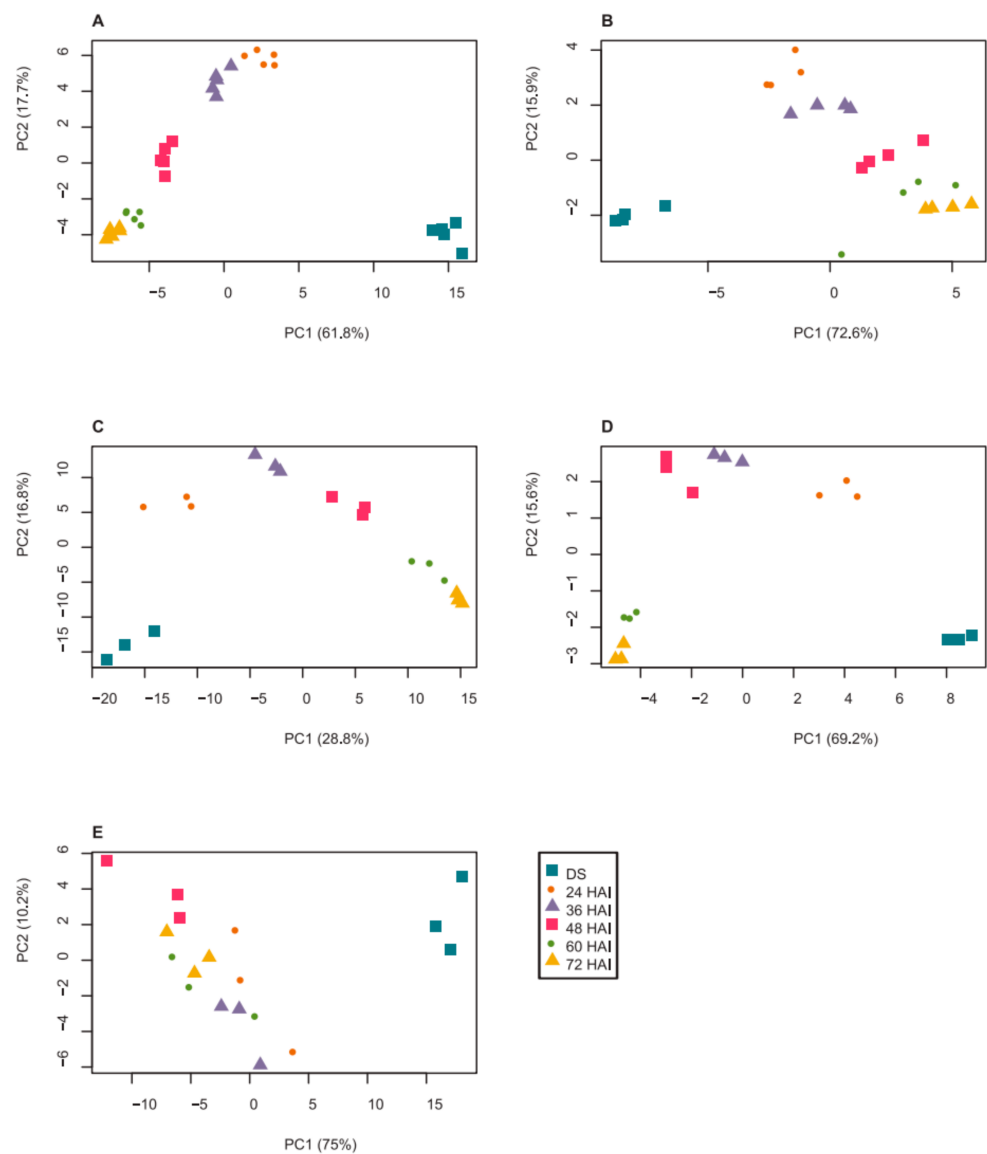


Figure 1. PCA of (A) metabolites, (B) amino acid composition, (C) heat-stable proteome, (D) LEA proteins and (E) expression of genes encoding LEA proteins of dry seeds (DS), germinating seeds and seedlings at different hours after imbibition (HAI). Data are \log_2 -transformed, median-normalized mass spectral intensities ($n = 5$) (A), concentration of amino acid (nmol/gDW) after \log_2 transformation and median normalization ($n = 4$) (B), mass spectral label-free quantification (LFQ) intensities after \log_2 transformation and median normalization ($n = 3$) (C,D) or DW-normalized absolute transcript copy number in \log_2 with median normalization ($n = 3$) (E). Missing values were replaced with 0 prior to the analysis. Samples at different HAI are as indicated by the legend.

2.2. Changes in the Heat-Stable Proteome during Germination and Seedling Development

The strong domination of seed storage proteins in dry seeds [48] makes it difficult to detect lower abundant proteins, e.g., LEA proteins. Thus, the heat-stable proteome was characterized for the detection of lower abundant proteins, especially chaperones, which may be important for seed germination and seedling establishment. We used a heat treatment to remove seed storage proteins and other heat-sensitive proteins before analyzing the composition of the remaining heat-stable proteome in seeds before and after imbibition by Liquid Chromatography-Mass Spectrometry (LC-MS/MS) analysis

of proteolytic peptides. This proteomic approach revealed the presence of 1545 different proteins. After removing known contaminants, proteins represented by only one unique peptide and proteins with missing values in at least two of the three replicates in all time points, 898 proteins remained in the data set (see Supplemental Table S3 for the complete data set). Not all proteins of the heat-stable proteome were detectable at every time point—72.6%, 45.3%, 31.3%, 22.3%, 28.4% and 24.3% of the total number of proteins were missing from the samples collected at DS, 24, 36, 48, 60 and 72 HAI, respectively.

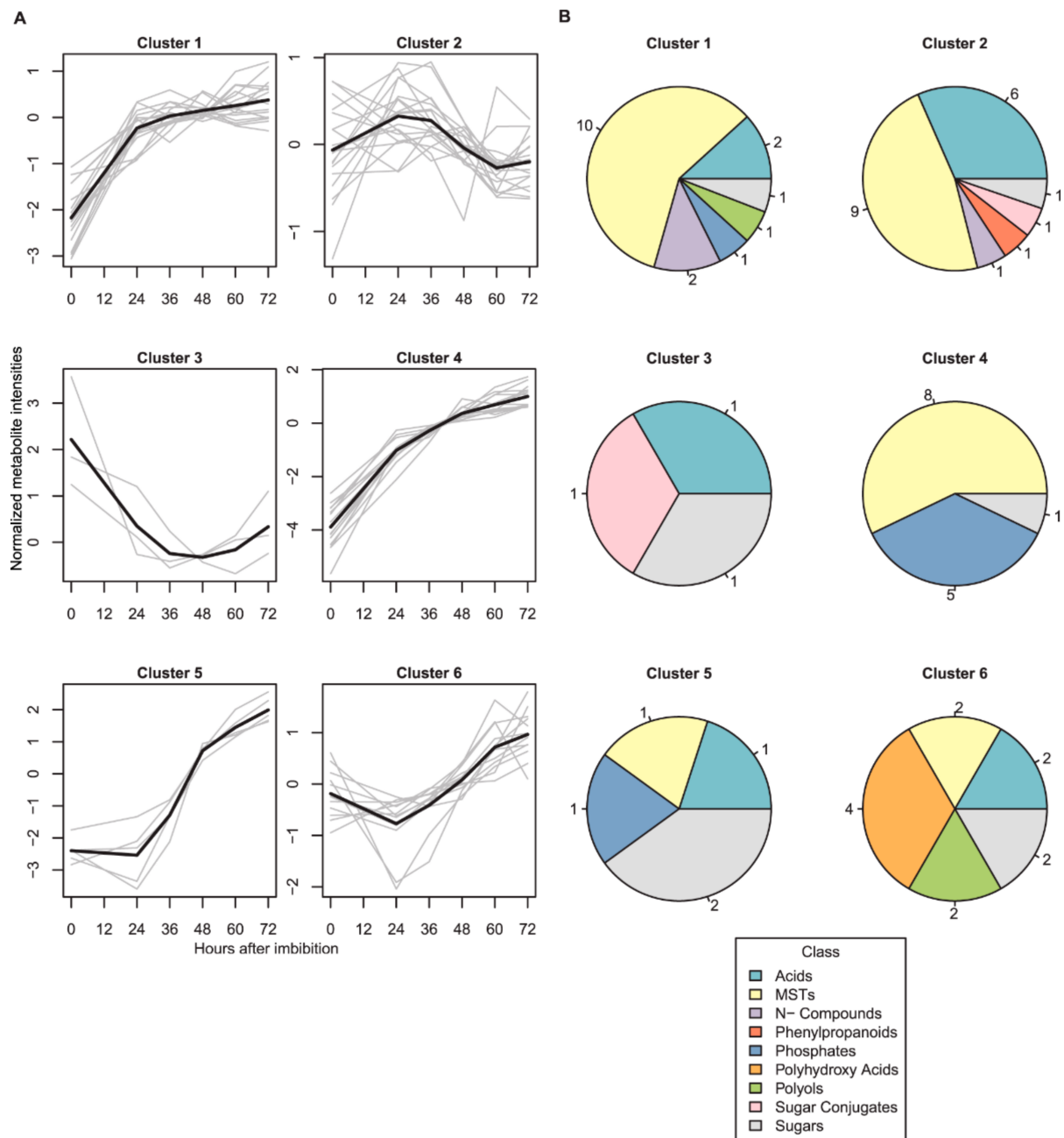


Figure 2. Patterns of metabolite abundance as determined by k-means clustering (clusters 1–6) (see Supplemental Figure S2) (A). Gray lines represent individual metabolites while black lines show the mean of each cluster. (B) Distribution of metabolite classes in each cluster. Amino acids were considered separately. Each chemical class is represented by a unique color as shown (MSTs = mass-spectral tags). Numbers around the pie charts represent the numbers of class members.

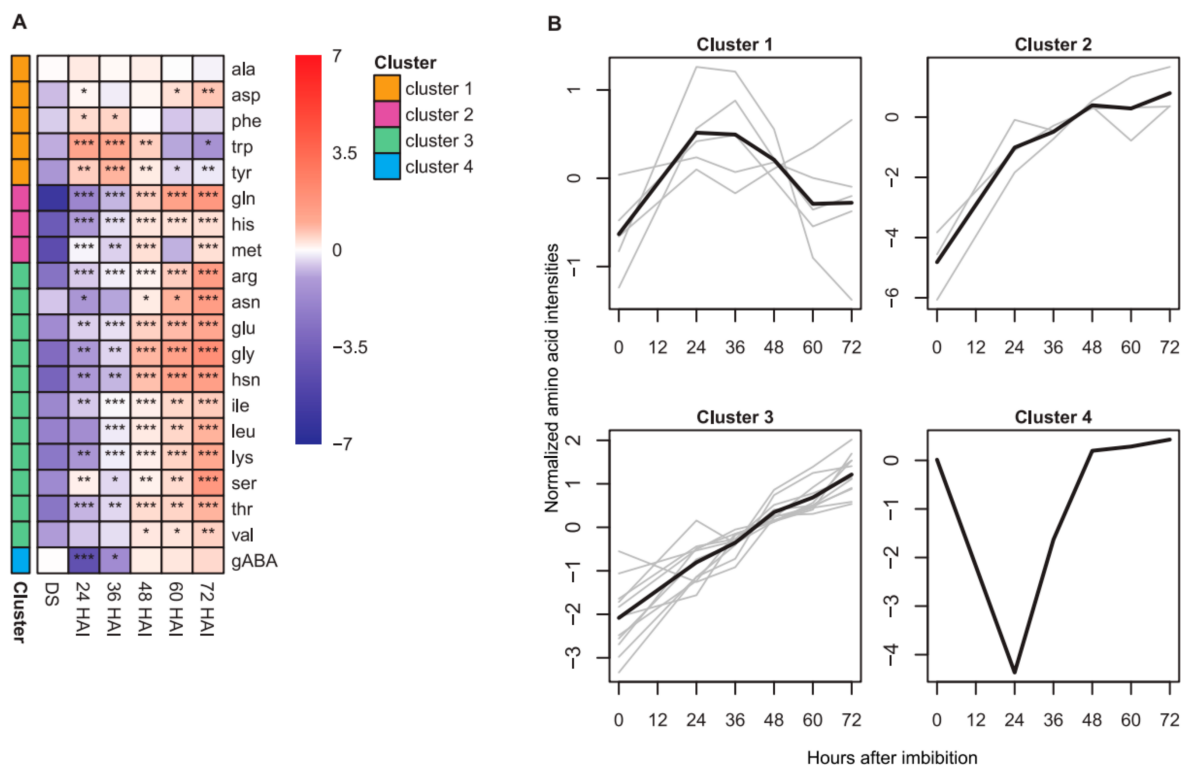


Figure 3. Amino acid levels in dry seeds (DS), germinating seeds and seedlings and their kinetic patterns. **(A)** Abundance of amino acids in DS, germinating seeds and seedlings collected at the indicated time points after imbibition. The color gradient represents mean concentration of amino acid (nmol/gDW) after \log_2 transformation and median normalization ($n = 4$). Levels of significance comparing the amino acids in germination seeds and seedlings to DS (unpaired two-sided t -test) are indicated by asterisks (* $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$). Amino acids are listed alphabetically within their clusters (cluster 1–4) as determined by k-means clustering. Each cluster is indicated by a unique color. **(B)** Patterns of amino acid abundance as determined by k-means clustering. Gray lines represent individual metabolites while black lines show the mean of each cluster. gABA— γ -aminobutyric acid.

PCA revealed a similar separation of the heat-stable proteins from different time points after imbibition (Figure 1C) as observed for amino acids (Figure 1B). PC1 separated DS from germinating seeds and seedlings, explaining 28.8% of the total variance in the proteomic data set, while PC2 (16.8% of the total variance) separated samples at different time points after imbibition. Data obtained from seedlings 60 and 72 HAI clustered closely together, suggesting that protein composition remained similar in seedlings once cotyledons had emerged.

The heatmap of 898 heat-stable proteins (Supplemental Figure S3) revealed that most of the proteins had one directional change in abundance, i.e., either increased or decreased during germination and seedling establishment. Some clusters showed also a transient increase for 24 to 48 HAI. The number of detected proteins in each time point is shown in a Venn diagram using proteins from DS, 24 HAI, 36 together with 48 HAI and 60 together with 72 HAI data sets (Figure 4A, Supplemental Table S4). The total number of detected proteins is increasing over time. The two time points with the least number of common proteins were DS and 60 and 72 HAI (193 proteins). On the contrary, the highest common protein number was found when overlapping 36 and 48 HAI proteins and 60 and 72 HAI proteins (613 proteins). After excluding eight proteins, which were only present in either one of the combined time points (marked in yellow in Supplemental Table S4a), we were left with 182 common heat-stable proteins that were found throughout germination and early seedling development. Among them, CpHsc70-1 (AT4G24280) and LEA20 (AT2G40170) had the highest increase and decrease in abundance in 72 HAI seedlings compared to DS, respectively (Supplemental Table S3).

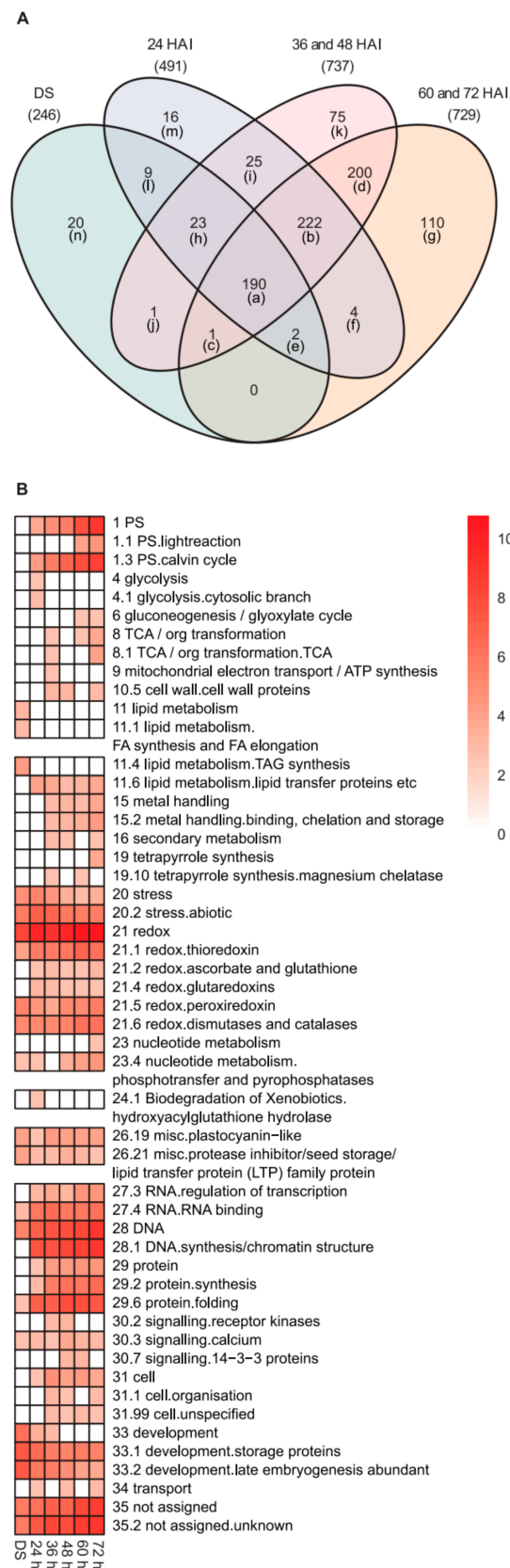


Figure 4. Heat-stable proteins in dry seeds (DS), germinating seeds and seedlings. **(A)** Venn diagram showing common and unique heat-stable proteins at the indicated time points after imbibition. **(B)** Functional enrichment of MapMan terms in the heat-stable proteome of DS, germinating seeds and seedlings. The white/red color gradient represents z-scores calculated from *p*-values ($p < 0.01$).

To explore the assignment of the 898 heat-stable proteins to main metabolic pathways and other processes we performed a functional enrichment analysis (Figure 4B, Supplementary Table S5) as statistical measure of whether a particular functional group (bin) of proteins contains more up- or down-regulated proteins than expected if all detected proteins would be equally distributed among all bins. We observed significant enrichment of 23 MapMan bins in DS, germinating seeds and seedlings. Nine bins were enriched in all observed time points; stress (20) with sub-bin abiotic stress (20.2), redox (21) with sub-bins thioredoxin (21.1), peroxiredoxin (21.5), dismutases and catalases (21.6), misc-related proteins including plastocyanin-like (26.19) and inhibitor/seed storage/lipid transfer protein (LTP) family protein (26.21), RNA binding (27.4), DNA (28), protein folding (29.6), signaling calcium (30.3), development-related proteins including storage protein (33.1) and late embryogenesis abundant proteins (33.2) and not assigned (35 and 35.2). Among these, redox, DNA and not assigned bins were highly enriched throughout the observation window. RNA binding (27.4) and protein folding (29.6), on the other hand, were highly enriched after 24 HAI.

Proteins from some bins and sub-bins were absent in DS but found in germinating seeds and young seedlings. These bins were PS (1) with sub-bin Calvin cycle (1.3), lipid transfer proteins (11.6), ascorbate and glutathione (21.2) and glutaredoxins (21.4), regulation of transcription (27.3), DNA synthesis/chromatin structure (28.1), protein (29) and its sub-bin synthesis (29.2) and cell (31). Enrichment of PS and Calvin cycle proteins clearly increased with HAI where it reached its peaks in young seedlings, coinciding with the transition from heterotrophic seeds to autotrophic seedlings. Most bins were enriched at more than one observed time point. However, proteins belonging to lipid metabolisms, (11) including FA synthesis and elongation (11.1) and TAG synthesis (11.4), were enriched only in DS while the enrichment of tetrapyrrole synthesis (19) and nucleotide (23) proteins was specific to seedlings 72 HAI.

Additionally, kinetic patterns within the heat-stable proteome were investigated for the 182 proteins, which were common among DS and all samples after imbibition using k-means clustering. Clustering revealed seven patterns of changes in protein abundance (Supplemental Figure S4A) which contained proteins assigned to 21 bins (Supplementary Figure S4B). The five bins with the highest proportion were: not assigned (17.6%), proteins (synthesis/post-translational modification/degradation/folding) (12.6%), redox (11.0%), development (9.90%) and stress (8.24%).

Among the clusters with a larger number of members (clusters 2, 3, 4 and 5), no single functional bin dominated. Excluding cluster 6 with only four members, the most common bins found across all clusters were stress (six clusters) and proteins (five clusters). Some bins were shared among clusters with the same protein abundance pattern. Heat-stable proteins that belonged to the PS, proteins and stress bins were found in all clusters with an overall increase in protein abundance after imbibition (clusters 1, 2 and 4). On the other hand, development was the common bin among clusters with a decrease in abundance (clusters 3, 6 and 7).

2.3. The Heat-Stable Proteome Exhibited an Equal Distribution of Intrinsically Ordered and Disordered Proteins

Since it is well-established that most LEA proteins are intrinsically disordered proteins [38] and that these proteins are heat-stable [49], we analyzed these sequences within the heat-stable proteome to predict disorder content in each protein using the MFDp2 online tool [50] (Table 1, Supplemental Table S3). Of the 898 heat-stable proteins, 701 amino acid sequences were retrieved from the UniProtKB/Swiss-Prot database (Supplemental Table S3). Various degrees of disorder content were found in the heat-stable proteins, from 0% disorder (12%) to 100% (10.8%). Hence, an almost equal fractioning of fully ordered and fully intrinsically disordered proteins (IDP) was observed. Proteins with a predicted disorder content between 0 and 10% comprised the biggest group in our data set (20.5%) followed by the group with a predicted disorder content of 10–20% (15.8%). The protein number tended to be smaller with increasing degree of disorder. In fact, proteins with at

least 50% disorder content comprised only 21.8% of the whole data set, clearly indicating that disorder is not a prerequisite for heat stability. Likewise, small and large proteins were found in all disorder categories. The mean size of proteins was similar among most categories. The smallest and largest means were found in 70–80% and 90–100% disorder predicted protein, respectively (199.1 vs. 461.4 residues).

Table 1. Heat-stable proteins with different predicted disorder content. A total of 701 protein sequences were retrieved from UniProtKB/Swiss-Prot. The prediction was performed by MFDp2. The table shows protein counts and other information (Min = minimum, Max = maximum, Mean of length of residues in each group of predicted disorder content).

Predicted Disorder Content (%)	Counts	Min	Max	Mean
		(Residues)		
0	84	80	843	255.6
0–10	144	56	990	342.3
10–20	111	82	974	310.5
20–30	97	69	976	348.9
30–40	62	87	987	341
40–50	50	72	953	309.8
50–60	27	98	724	294.2
60–70	18	93	772	303.4
70–80	9	129	367	199.1
80–90	7	111	577	277
90–100	16	131	891	461.4
100	76	62	956	284.9

To investigate if the grade of disorder is related to functional categories, the protein abundance for proteins with 0% and 100% disorder content was exemplarily assigned to MapMan bins using TAIR10 annotations [51] during the investigated germination time-course (Figure 5, Supplemental Table S6). While there was no distinct pattern found among fully folded proteins in terms of protein abundance among the bins identified, we could see some pattern among the fully disordered proteins. For example, there was an increasing abundance within 72 HAI for fully disordered proteins in the following bins: cell, protein, signaling and transport. The majority of these proteins was undetected in DS and early stages of germination, and therefore, a significance test to protein abundance in DS was often impossible due to missing values at this condition. Significant decrease in abundance of 100% disordered proteins within 72 HAI was found in the bins development and hormone metabolism.

2.4. LEA Protein Dynamics during Seed Germination

From the 898 heat-stable proteins in the data set, 29 LEA proteins in eight of the nine LEA Pfam groups [42] were identified (Supplemental Table S7). Only members of the Pfam LEA_3 group were not detected. The Pfam group with the largest number of members was LEA_4 (37.9%) followed by dehydrin (20.7%) and seed maturation protein (SMP) (10.3%). Out of 29 LEA proteins in the heat-stable proteome, 17 sequences were retrieved from the UniProtKB/Swiss-Prot database. The length of these LEA proteins ranged between 97 and 479 amino acids. Most of them (76.5%) were predicted by MFDp2 to have 100% disorder content (Supplemental Table S7). Only four of them were predicted with a smaller fraction of disorder content: LEA21 with 16.4%, LEA22 with 12.1%, LEA31 with 24% and LEA32 with 48.8%.

To observe whether the high proportion of fully disordered LEA proteins is only found in the heat-stable proteome, we investigated sequences of LEA proteins which were reported to be seed-expressed [42] but were not detected in our data set (Supplemental Table S8). We were able to retrieve 10 of these from the UniProtKB/Swiss-Prot database and found that 30% of them were predicted to be 100% disordered. The missing LEA

protein with the lowest disorder content was LEA50 with 0%, followed by LEA1 (3.3%) and LEA27 (13.9%).

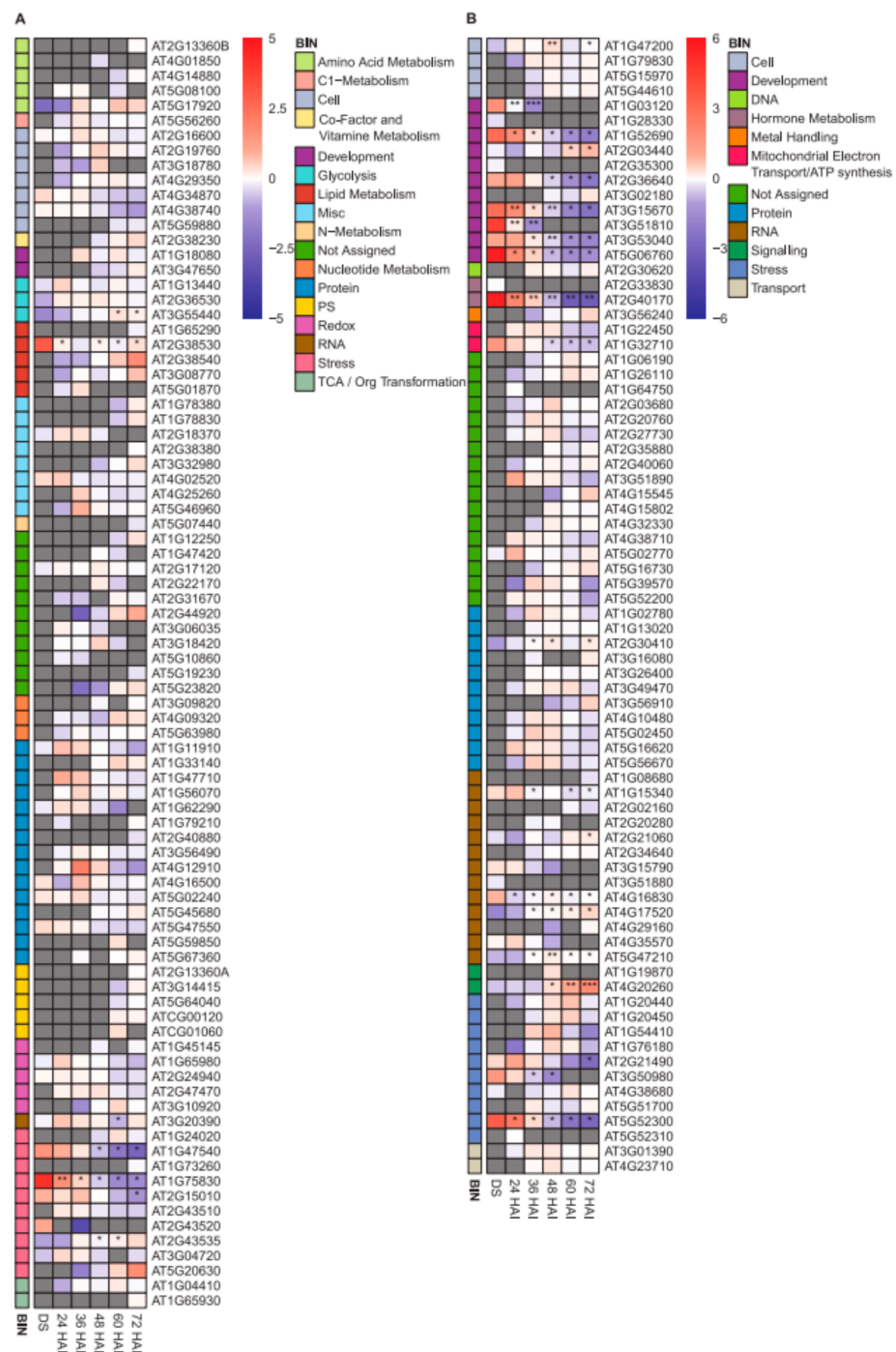


Figure 5. Abundance of proteins with (A) 0% and (B) 100% disorder content in the heat-stable proteome in dry seeds (DS), germinating seeds and seedlings. The color gradient represents the log₂-median transformed means of mass spectral LFQ intensities ($n = 3$). Gray boxes represent missing values at the indicated time point. Levels of significance comparing protein abundance in germinating seeds and seedlings to DS (unpaired two-sided t -test) are indicated by asterisks ($* p < 0.05$; $** p < 0.01$; $*** p < 0.001$). Proteins are assigned to MapMan bins, each bin is indicated by a unique color as shown. The transcript AT2G13360 mapped to two bins and was therefore marked with A and B. For additional information, see Supplemental Table S6.

A PCA of the LEA protein data set (Figure 1D) showed a clear separation between samples from different time points. Similar to the heat-stable proteome data set, PC1 explained 69.2% of the total variance and separated DS from all time points after imbibition, while PC2 (15.6% of the total variance) separated germinating seeds and young seedling samples.

Most of the detected LEA proteins (79.3%) were found in DS (Figure 6A) and most of these showed a significant reduction in abundance after imbibition that became more pronounced with time. Some of these LEA proteins (34.5% and 37.9%) were no longer detected at 60 and 72 HAI. Noticeably, three LEA proteins (LEA18 from LEA_1 and LEA21 and LEA22 from the AtM Pfam groups) were specific to DS. On the other hand, six LEA proteins (LEA4, LEA5, LEA8 and LEA10 from the dehydrin, LEA12 from the LEA_4 and LEA26 from the LEA_2 Pfam groups) became only detectable in germinating seeds from 24 or 36 HAI onwards. Thus, they were referred to as germination-specific. LEA12, however, was already undetectable again 60 HAI, indicating a very specific function during early germination.

To observe patterns of changes in LEA protein abundance, k-means clustering was applied to 13 LEA proteins found across all time points (Figure 6B). Although members of all clusters were reduced after imbibition, three clusters of LEA proteins from different Pfam groups with different kinetics were identified (Figure 6C). LEA_4 proteins were the most commonly found in the cluster data set with 69.2%. Proteins in cluster 1 showed a constant gradual decrease. Proteins in the largest cluster 2 (53.8% of cluster data set) showed a steady abundance up to 24 HAI, followed by a gradual decrease except for LEA16. Proteins in cluster 3 largely decreased in abundance up to 60 HAI with little further change over the last 12 h. LEA20 was the LEA protein with the largest reduction in abundance between DS and 72 HAI among all clustered proteins.

2.5. Transcript Abundance of Genes Encoding LEA Proteins

While we detected 29 LEA proteins in the heat-stable proteome, transcripts encoding 34 LEA proteins were previously detected by qRT-PCR in dry Arabidopsis seeds [42]. Here, we investigated the transcript levels of these LEA genes in DS and after imbibition along with genes encoding LEA proteins in the heat-stable proteome that were not among those 34. Therefore, we investigated in total 41 LEA genes. The abundance of LEA transcripts was expressed in absolute copy number of transcripts normalized to dry weight (DW) of samples used for RNA extraction (copy/gDW). Copy numbers of LEA transcripts were highest in DS, with a maximum average of 3.38×10^{13} copy/gDW for LEA20, and lowest in seedlings 72 HAI, with a minimum average of 9.47×10^6 copy/gDW for LEA50 (see Supplemental Table S9 for the complete data set). We detected transcripts for LEA1, LEA6, LEA9, LEA13, LEA27, LEA33, LEA38, LEA41, LEA45, LEA49, LEA50 and LEA51 (for LEA27, LEA38 and LEA41 with higher abundance at 36 and 60 DAI) even though no corresponding protein was detected (Figure 7A).

A PCA of the LEA transcript data (Figure 1E) showed a separation between DS and all time points after imbibition by PC1, which explained 75% of the total variance. However, unlike for the LEA protein data (Figure 1D), there was no clear separation of the time points after imbibition by PC2, suggesting that the kinetics of LEA transcript appearance were less defined than those of the proteins.

K-means clustering showed three different transcript abundance patterns of the observed LEA transcripts (Figure 7B). Due to presence of missing values, LEA5, LEA12 and LEA45 were excluded from k-means clustering. All LEA Pfam groups were represented across the three clusters (Figure 7C). Cluster 1 comprised LEA transcripts that increased in abundance after imbibition. The increase was gradual up to 36 HAI, and then transcript levels generally remained constant. The transcript of LEA38 was the only one in cluster 1, which did not encode germination-specific LEA proteins as described above. Cluster 2 contained LEA transcripts with similar abundance over time and only small fluctuations. Cluster 3 contained transcripts with reduced abundance after imbibition with the lowest abundance at 48 HAI. This cluster represented the majority of the observed LEA transcripts

(68.4% of cluster data set) and was dominated by LEA_4 transcripts (38.5%). Cluster 3 was the most diverse cluster consisting of LEA transcripts from 7 Pfam groups, only missing members from LEA_2 and LEA_3 groups, which were both detected in cluster 1 and cluster 2 (Figure 7C). Cluster 3 was the only cluster with LEA transcripts from AtM, LEA_1, LEA_5 and SMP Pfam groups.

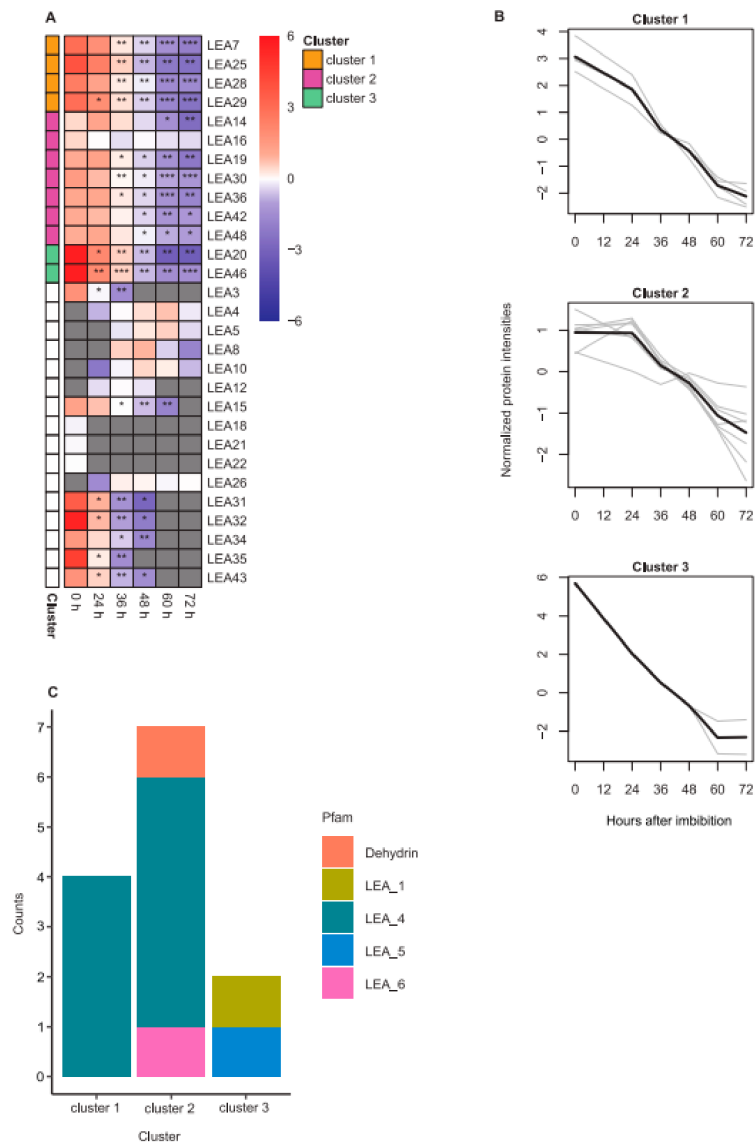


Figure 6. LEA protein abundance in dry seeds (DS), germinating seeds and seedlings and their kinetic patterns and Pfam distributions. **(A)** LEA protein abundance in DS and at different time points after imbibition. The color gradient represents the \log_2 -median transformed means of mass spectral LFQ intensities ($n = 3$). Gray boxes represent missing values at the indicated time point. Levels of significance comparing LEA protein abundance in germinating seeds and seedlings to DS (unpaired two-sided t -test) are indicated by asterisks (* $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$). LEA proteins are listed according to their LEA ID and corresponding clusters (cluster 1–3) as determined by k-means clustering. Each cluster is represented by a unique color. LEA proteins excluded from k-means clustering are depicted in white. **(B)** Patterns of LEA protein abundance as determined by k-means clustering. Gray lines represent individual proteins while black lines show the mean of each cluster. **(C)** Counts of LEA proteins in different Pfam groups in each cluster. Each Pfam group is represented by a unique color.

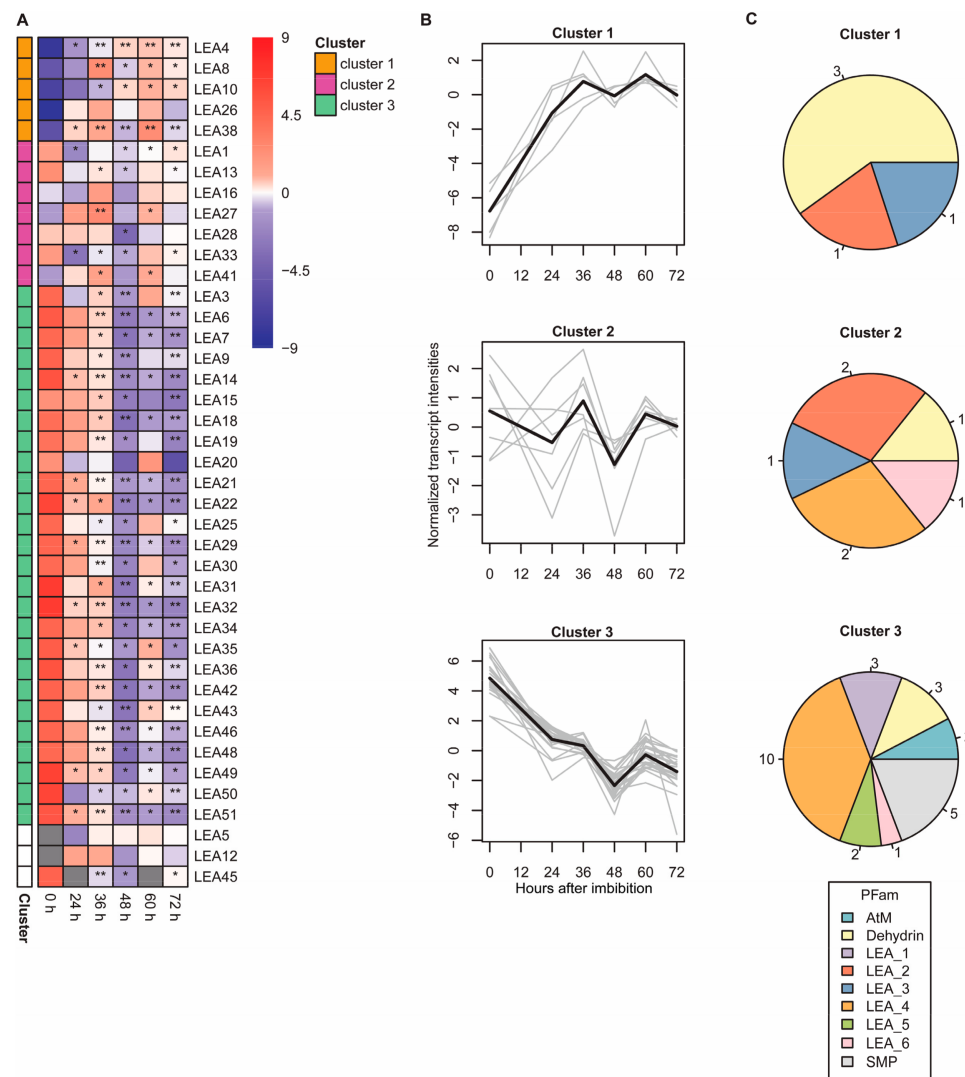


Figure 7. Transcript levels of genes encoding LEA proteins in dry seeds (DS), germinating seeds and seedlings after imbibition. **(A)** LEA transcripts at different time points after imbibition. The color gradient represents means of DW-normalized absolute transcript copy number in \log_2 with median normalization ($n = 3$). Gray boxes represent missing values at the indicated time point. Level of significance comparing LEA transcript abundance in germinating seeds and seedlings to DS (unpaired two-sided t -test) are indicated by asterisks (* $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$). LEA transcripts are listed according to their LEA ID and corresponding clusters as determined by k-means clustering. Each cluster is represented by a unique color. LEA transcripts excluded from k-means clustering are depicted in white. **(B)** Patterns of LEA transcript abundance as determined by k-means clustering. Gray lines represent individual transcripts while black lines show the mean of each cluster. **(C)** Distribution of Pfam of LEA transcripts in each cluster. Each Pfam is represented by a unique color. Numbers in pie charts represent the numbers of Pfam members.

Since we determined both LEA protein and transcript abundances, we were able to investigate the correlation between mRNA and protein levels. We only investigated the correlations between 23 LEA proteins and their transcripts when detected in at least four of the six observed time points. Significant correlations were found in nine of these twenty-three cases, which came from four different Pfam groups of LEA_1, LEA_4, dehydrin and SMP (Table 2). The highest correlation between the abundance of a protein and the corresponding transcript was found for LEA32 from SMP Pfam group. Of the nine significantly correlated pairs of proteins and transcripts, only the germination-specific dehydrins LEA4 and LEA10 were upregulated, while the other seven were downregulated

during germination. A delay in protein accumulation following the transcript accumulation was observed for some LEA proteins. For example, for LEA10, a delay of at least 24 h has been found between the increase in transcript and protein detection.

Table 2. Pearson correlation between the abundance of LEA proteins and their corresponding transcripts. LEA proteins with missing means mass spectral LFQ intensities in more than two observed time points were excluded from analysis. Levels of significance of the correlations are indicated by asterisks (** $p < 0.01$; *** $p < 0.001$).

LEA ID	Pfam (Hundertmark and Hincha, 2008)	Pearson Correlation	p value
LEA4	dehydrin	0.703 **	0.010
LEA5	dehydrin	−0.091	0.839
LEA7	LEA_4	0.669 **	0.009
LEA8	dehydrin	0.052	0.873
LEA10	dehydrin	0.789 **	0.005
LEA14	dehydrin	0.212	0.577
LEA15	LEA_6	0.581	0.061
LEA16	LEA_6	−0.169	0.664
LEA19	LEA_4	0.461	0.131
LEA20	LEA_5	0.250	0.507
LEA25	LEA_4	0.409	0.191
LEA26	LEA_2	0.050	0.873
LEA28	LEA_4	0.070	0.839
LEA29	LEA_4	0.695 **	0.006
LEA30	LEA_4	0.297	0.393
LEA31	SMP	0.839 **	0.005
LEA32	SMP	0.934 ***	0.0003
LEA34	dehydrin	0.813 **	0.006
LEA36	LEA_4	0.127	0.715
LEA42	LEA_4	0.346	0.310
LEA43	LEA_4	0.777 **	0.009
LEA46	LEA_1	0.824 ***	0.0004
LEA48	LEA_4	0.243	0.507

2.6. Protein–Metabolite Network Analyses Reveal a LEA Cluster with High Connectivity

As the majority of LEA proteins common for all time points showed a fast decrease during seed germination and seedling establishment, we were interested in other proteins and metabolites following the same pattern to identify highly correlated protein or protein–metabolite networks associated with seed germination. With this approach, we wanted to integrate omics data for an interpretable network for the discovery of molecular interactions previously not discovered for heat-stable proteins under these conditions. A protein–metabolite network analysis was performed, including all heat stable proteins (182) together with all metabolites (70) and amino acids (20) common for all six time points. For reduction in edges, a cut off ($r \geq 0.99$ and a p -value < 0.05) was chosen (Supplemental Table S10). The identified network included 91 proteins (including 11 LEA proteins), 42 metabolites and 13 amino acids with a varying range of correlations and a high connectivity level (Supplemental Figure S5). LEA proteins (colored in purple) had multiple relationships to each other and revealed positive connections building a small sub-cluster with LEA7, LEA19, LEA25, LEA28, LEA29, LEA30, LEA36, LEA42 and LEA48. Among them, LEA7, LEA25, LEA28, LEA29 and LEA30 were identified as network hubs with a high level of inter-connectivity represented by eight to eleven edges. These LEA proteins were closely correlated with several heat-stable proteins as the seed storage proteins SESA1, a bi-functional inhibitor/lipid-transfer protein/seed storage 2S albumin superfamily protein (AT4G30880), the ABA-inducible dehydration-responsive protein RD29B/LTI65, the eukaryotic elongation factor 5a (ELF5a) and a protein of unknown function (AT1G16850). The large subunit of Rubisco (RBCL) displayed several negative correlations with LEA

proteins (LEA25, LEA29, LEA42) and other members in this sub-network. Only three metabolites were involved in this cluster, namely phosphoric acid, 1,5-anhydro-sorbitol and an unknown analyte (A167004).

This large LEA cluster was connected to another large cluster including mainly metabolites of the central metabolism, most of the amino acids, sugar phosphates, *myo*-inositol-phosphate and phosphoglyceric acid via a positive correlation to LEA20. LEA20 together with LEA46 was more closely correlated with several other heat-stable proteins and metabolites. For LEA20, correlations were found with the seed storage protein SESA4, the Low-Molecular-Weight Cysteine-Rich 67 (LCR67) involved in pathogen response and an unknown protein (AT3G12960). LEA46 showed also positive connections to LCR67, a cytochrome c oxidase (AT1G32710) and a serine protease inhibitor (AT2G38870), possibly involved in biotic stress response. LEA46 showed in contrast to the other LEA proteins a high number of negative correlations with metabolites and amino acids, including mannose-6-phosphate, *myo*-inositol-phosphate, isoleucine, glutamine and threonine.

3. Discussion

Seed germination is a complex process, which is accompanied by comprehensive changes and a concerted action of molecular players represented by gene expression, protein abundance and metabolite levels. Dry seeds with a quiescent metabolic state undergo a transition to proliferative metabolic state for plant propagation [22]. Molecular omics changes during seed germination in *Arabidopsis* were studied before but not much attention was directed to variations of the heat-stable proteome including intrinsically disordered proteins and especially changes in LEA proteins and their correlation network with metabolic changes.

3.1. Degradation of Storage Compounds to Central Metabolites Contributes to the Transition from Heterotrophic to Autotrophic Growth

Fifty-six known and forty-three unknown metabolites were identified over a time frame of 72 HAI, which is comparable to other publications on *Arabidopsis* [8,11], wheat [16] and tomato seeds [17]. Following a previous classification [11], metabolite profiles showed major shifts between three developmental stages: (1) the dry state, (2) initial stages of germination (24 HAI and 36 HAI) and (3) greening and fully opened cotyledons (48 HAI to 72 HAI). Similar to our findings, a PCA clearly distinguished metabolite profiles of wheat embryo and endosperm after 24 h of germination from the profiles at 0 h [16]. Six metabolites, namely fumaric acid, galactinol, sucrose, raffinose, 2,5-dihydroxybenzaldehyde and 3,4-dihydroxybenzoic acid, revealed high levels in DS and were decreased during seed germination and seedling establishment, underlining their role as storage reserves and protective substances for dry seeds [30]. Sucrose or raffinose family oligosaccharides (RFOs) are the most abundant non-reducing sugars in dry mature seeds [36] and serve as storage reserves during germination, resulting in their depletion for energy production and heterotrophic growth before the transition to autotrophic growth [10]. A continuous decrease in sucrose and galactinol as precursor of RFO during seed to seedling transition in *Arabidopsis* was reported previously [10,11]. Sucrose was additionally suggested as a signaling metabolite during seed germination as sucrose levels showed correlations with 129 transcripts including high numbers of sucrose-responsive genes [10].

Galactinol was described as marker for seed longevity in Brassicaceae and tomato [52]. An overexpression of galactinol synthase of *Cicer arietinum* in *Arabidopsis* led to higher seed longevity by reducing reactive oxygen substance (ROS) accumulation [53]. On the other hand, galactinol synthase 1 was reported as negative regulator of seed germination [54]. For fumaric acid, an increase during seed maturation was shown, especially during desiccation, resulting in high levels in dry seeds [8].

The two metabolites 2,5-dihydroxybenzaldehyde and 3,4-dihydroxybenzoic acid, which were present in DS, 24 HAI and 36 HAI, have been, to our knowledge, not described in seed germination studies before. 3,4-dihydroxybenzoic acid (protocatechuic acid) was widely detected in cereals and legumes and other species [55,56]. It affects

growth of tobacco callus and organs, with small concentrations inducing root growth while high concentrations inhibited shoot and root growth [57]. Furthermore, positive correlations between antioxidant enzyme activity and 3,4-dihydroxybenzoic acid were found in *Miscanthus* [58].

Most of the other identified metabolites, sugars and sugar phosphates, amino acids and most organic acids were increased with proceeding germination following the activation of biosynthetic processes for growth and development. They were found previously within 24 h germination or till opening of the cotyledons [8,11].

Consistent with other studies, intermediates of the TCA cycle were rather reduced during germination whereas metabolites involved in glycolysis, such as sucrose-derived sugar-phosphates, were increased, supporting the finding that the main energy source for seed development is glycolysis [8,11,16].

A release of amino acids anticipating radicle protrusion and seedling growth was described to start at 24 HAI [22]. Canonical Correlation Analysis (CCA) of metabolite changes during 6 h of seed germination in tomato revealed close correlations of characteristic combinatorial changes in a larger set of metabolites but not individual metabolites with germination parameters. This larger set included metabolites of central metabolic pathways as sucrose, glucose, fructose, metabolites of the TCA cycle as fumarate and succinate and in addition glycerol-3-phosphate, the sugar alcohols *myo*-inositol and galactinol and several amino acids [17]. Altogether, the metabolic status of the here-described experiment was similar to previously published studies, which allows us to compare our data on the heat-stable proteome with previous publications.

3.2. Thermal Stability of Proteins Is Related to Biological Processes

Whereas metabolic shifts during seed germination were previously investigated, less is known about the heat-stable proteome including low abundant proteins at different stages of seed germination. A total of 898 heat-stable proteins were identified in seeds over 72 h of germination in our study, which was the equivalent of around 30% of 2967 proteins with high melting temperature (T_m) or no melting point (nonmelters) in *Arabidopsis* [59]. Shifts in protein abundance of the heat-stable proteome during seed germination over time differed from that of metabolites. Proteome samples from all time points were clearly separated from each other (except 60 HAI and 72 HAI), whereas metabolite profiles of 24 HAI and 36 HAI were closely clustered.

Heat-stable proteins were enriched in the following bins at all time points: stress, redox, RNA binding, DNA, protein folding and development-related proteins including storage and late embryogenesis abundant proteins. The functional bin photosynthesis with the sub-bin Calvin cycle was highly enriched from 24 HAI. Previously, heat-stable proteins were described to be enriched for ribosomal, RNA-binding and protein biosynthesis processes [60], and functional enrichment analysis of heat-stable *Arabidopsis* proteins by thermal protein profiling (TPP) pointed to bins related to protein folding, carbon fixation and the proteasome [61], thus revealing an overlap with our findings. Interestingly, an enrichment of proteins involved in carbon fixation was observed including RuBisCo subunits [61], as we also discovered in our data set.

For human cell lines, it was suggested that differences in thermal stability might reflect differences in the activity of biological processes [59]. It was also hypothesized that heat-induced cell damage is mainly caused by the denaturation of a relatively small set of functionally essential “hub” proteins [60]. Following our functional enrichment analysis, this hypothesis can be confirmed as heat-stable proteins were enriched in biological processes important for ongoing seed germination and greening of the cotyledons as photosynthesis, Calvin cycle, abiotic stress and redox reactions.

A comparison of the here-discovered heat-stable proteome of 898 proteins of the *Arabidopsis* accession Col-0 with 256 proteins previously identified with two-dimensional gel-electrophoresis in the accession *Landsberg erecta* (Ler) showed an overlap of 149 proteins, revealing a heat-stable percentage of 58% in the Ler data set [22]. The overlap

included a large amount of heat-stable proteins involved in biological processes, e.g., several seed storage proteins as cruciferins (CRA1, AT5G44120; CRU2, AT1G03880; CRU3, AT4G28520) belonging to the most abundant cruciferins in DS and serving as nitrogen and amino acid source for seedling development. Additionally, abundant 12S cruciferins were hypothesized to be predominant ROS scavengers in seeds being highly exposed to oxidative stress [22].

In the early germination phases, the de novo synthesis of key enzymes of the glyoxylate cycle, such as isocitrate lyase (AT3G21720), aconitase 3 (AT2G05710) and malate synthase (AT5G03860), was reported [22], pointing to lipid remobilization. All of them were also included in the heat-stable proteome of Col-0. Malate dehydrogenase (AT1G04410), converting malate into oxaloacetate for amino acid synthesis pathways or conversion to phosphoenolpyruvate by phosphoenolpyruvate carboxykinase 1 (AT4G37870), was also found in both studies.

Furthermore, heat-stable proteins common in both studies are involved in antioxidant defense and detoxification: iron/manganese superoxide dismutase (AT3G56350), 1-cysteine peroxiredoxin (AT1G48130), catalase 2 (AT4G35090), thioredoxin-dependent peroxidase 1 (AT1G65980), ferritin 2 (AT3G11050) and monodehydroascorbate reductase 6 (AT1G63940). Generation of ROS takes place during water uptake, which might have negative effects on proteins and other cell components [62]. In dry pea seeds, superoxide dismutase, catalase, ascorbate peroxidase, dehydroascorbate reductase and glutathione reductase were detected as being involved in radical scavenging during imbibition [62].

The restart of the cellular metabolism during germination involves elements of the cytoskeleton, including actin and tubulin, and consequently, tubulin 3 (AT5G62700), actin 2 (AT3G18780) and actin 7 (AT5G09810), which were identified in both data sets and are heat stable. Finally, proteins involved in stress response as heat shock proteins (HSP70, AT3G12580; HSC70.5, AT5G09590; HSP60.2, AT2G33210) [22] overlapped between the studies, underlying the thermal stability of proteins involved in important biological processes during germination. The classification of heat-stable proteins into main metabolic processes and defense responses and the enrichment in functional bins representing biological processes point to mobilization of storage proteins and lipids, an activated heterotrophic and later autotrophic metabolism and several stress response mechanisms during seed germination.

3.3. An Equal Distribution of Fully Ordered and Fully Intrinsically Disordered Proteins Characterized a Part of the Heat-Stable Proteome of Germinating Seeds

The loose structure of intrinsically disordered proteins (IDP), e.g., LEA proteins, contributes to their heat-stability when globular proteins aggregate [63]. Often, IDPs are enriched after heating, e.g., when comparing heat-stable proteins with the whole proteome of soybean [64]. An increased content of IDPs was also reported in the heat-stable protein fraction of imbibed radicles of *Medicago truncatula* [41]. In a study compiling an atlas of the thermal stability of 48,000 proteins across 13 species ranging from archaea to humans, nonmelters were most strongly enriched in disordered regions [59].

Nevertheless, a surprisingly high percentage of low disorder content was found for the heat-stable proteome in our study, independent of protein length. In contrast, a study estimating the relative content of intrinsic protein disorder in 96 plant proteomes reported an inverse relationship between the degree of intrinsic protein disorder and protein length [65]. A high intrinsic disorder does not seem to be the main requirement for the heat stability of proteins. A more complex relationship between disorder and thermal protein stability may exist [59]. Recently, features as molecular weight, hydrophobicity, charged versus polar (CvP) bias and protein helix and sheet composition were shown to be highly correlated with thermal stability of Arabidopsis proteins. In addition, cytoplasmic protein concentration, interactions of small molecules and cellular localization were suggested to have an effect [59,61], whereas protein abundance was controversially discussed as either not a good predictor of thermal stability or important for stability [59]. Thermal stability of proteins can be additionally changed by phosphorylation [66] or the redox state [67].

While no functional category was favored in proteins with 0% disorder content, 100% disordered proteins of the heat-stable proteome during seed germination and seedling establishment were mainly assigned to the functional MapMan bins development, DNA, protein, RNA and stress. The disorder-related functions of plant proteins were previously related to stress tolerance, transcription regulation, cell cycle regulation, molecular chaperones and developmental regulation [64]. In addition, IDPs were mainly involved in the GO terms regulation of nucleus activities, regulation of metabolic processes, response to biotic and abiotic stimuli and signaling, depicting an overlap with our results even that functional bins were differently classified [65]. When investigating proteins from 13 species, highly unstructured proteins were enriched in nuclear and phosphorylated proteins [59]. On the other hand, nonmelters with low disorder were enriched in transmembrane-domain-containing proteins, proteins with extracellular domains, glycosylated proteins, secreted proteins with signal peptides and proteins containing disulfide bonds [59].

3.4. Dehydrins Were Identified as Seed Germination-Specific LEA Proteins

Within the heat-stable proteome 29 LEA proteins were detected with the vast majority of them having 100% disorder. An analysis of the heat-stable proteome of imbibed radicles or seeds of *Medicago truncatula* aiming to investigate proteins linked to desiccation tolerance identified 11 to 16 LEA proteins [41,47].

The ABA-induced accumulation of LEA proteins in parallel to the accumulation of non-reducing sugars during seed maturation provide resistance to desiccation [36,68]. Seed rehydration imposes severe stress, including leakage of solutes and temporary membrane and organelle damage from free radicals. Consequently, synthesis of proteins and compounds which limit and repair cell damage is prevalently needed [6]. With proceeding water uptake during seed germination and the loss of desiccation tolerance, the abundance of most LEA proteins declines sharply [24,68].

The 10 LEA proteins previously described as seed-expressed in *Arabidopsis* [42] were not detected in the heat-stable proteome of germinating seeds (LEA1, 6, 27, 33, 37, 38, 41, 49, 50, 51), even if 100% disordered (LEA6, 33, 51). Most LEA proteins were detected in DS, including three LEA, specific to DS and decreased during germination. LEA proteins were decreasing and partially remaining during germination in *Arabidopsis* [18] and pea [69] with a protective function postulated for the remaining proteins.

Only a small number of LEA proteins appeared at later time points. LEA12 was only present from 24 HAI to 48 HAI and was previously described as bud specific [42]. Five other LEA proteins were only present from 36 HAI and were defined as germination specific (LEA4-synonym COR47, LEA5-ERD10/LTI45/LTI29, LEA8-HIRD11, LEA10-ERD14, LEA26), most of them belonging to the class of dehydrins and predicted to be 100% disordered. Three of the five germination-specific LEA proteins were previously described as stress induced [42], which suggests the possibility of contributing to stress tolerance during the development of seedlings.

The protein abundance of LEA4 and LEA10 was highly correlated with the related transcript levels, whereas for the other germination-specific LEAs (LEA5, 8, 26), transcripts were present earlier compared to proteins, suggesting the possibility of a fast de-novo synthesis during seed germination.

Dehydrins contain highly conserved stretches of 7-17 residues that are repetitively scattered in their sequences, the K-, S- and Y-rich segments, and accumulate during seed maturation and in response to abiotic stress as cold, dehydration, osmotic stress or ABA [70]. Some of them are phosphorylated in response to stress and are then capable of calcium binding (e.g., ERD14). Dehydrins are able to bind to phospholipids and thereby to modulate membrane properties [39]. In most studies, their abundance is reduced during seed germination (for a review, see [68]). Nevertheless, these proteins were also shown to be involved in seed germination and development [71]. LEA5/ERD10 is important for completion of seed development as *erd10* mutants revealed abnormal shape and reduced

germination [71]. Results on the expression of *ERD10* in DS are controversial, from no expression (our results and [72]) to detection in DS [71].

LEA4/COR47, LEA5/ERD10 and LEA10/ERD14 were previously identified as members of protein complexes. Homodimeric and heterodimeric interactions were verified for LEA4/COR47 and LEA5/ERD10 in the cytosol of tobacco cells by bimolecular fluorescence complementation, but also heterodimeric associations between LEA51/RAB18 or PIP2B, and these dehydrins took place in the cytosol [73,74]. For LEA4/COR47, the transcription factor WRKY63, involved in seedling establishment and growth, was identified as an upstream regulator, pointing to a possible role in seedling development [75].

LEA8/HIRD11 has RNA binding activity, containing an RNA G-quadruplex structure (RG4) in its 3'UTR, which folds with guanine-rich sequences [76]. These complexes play a role in post-transcriptional regulation of gene expression, plant development and growth modulation [76]. An *HIRD11* mutant showed reduced root growth and retarded shoot growth.

LEA26, a germination-specific LEA beside LEA12 not belonging to the class of dehydrins, was shown to function as a protector (chaperon) of cellular components from stress, as this protein prevented the inactivation of the enzyme lactate dehydrogenase (LDH) during freezing and stabilized membrane structures [77].

All of these germination-specific, stress-related dehydrins in addition to their specific functions might protect germinating seeds from environmental stresses after the reduction in osmoprotective substances such as raffinose and other protecting LEA proteins. As an example, they were able to provide cold stress resistance to highly watered seed [68].

3.5. LEA Proteins Build a Highly Correlated Cluster in the Seed Germination Specific Common Protein–Metabolite Network

The construction of a protein–metabolite network helped to develop a deeper understanding of their pattern during seed germination, even though only heat-stable proteins and metabolites were included common for all investigated time points. The discovery of a tight LEA protein cluster among all common proteins during germination is mainly based on a common decrease in abundance of all LEA proteins at the same time during imbibition, but is still mentionable, as only very little proteins follow the same pattern. For that reason, the following discussion of this network focuses mainly on the roles of the specific LEA proteins in DS and very early germination rather than for seedling establishment. Several LEA proteins were identified as network hubs, including LEA7, LEA25, LEA28, LEA29 and LEA30, all belonging to the Pfam group LEA_4.

As a high number of metabolites in the network were unknown, not many conclusions could be drawn regarding common protein–metabolite accumulation patterns that might reflect interactions. Nevertheless, the LEA proteins of the LEA cluster were separated from almost all other metabolites. All of these LEA proteins play important roles for seed protection in the dry state also represented by connections to seed storage proteins and a seed storage 2S albumin superfamily protein. LEA7 in leaves had a protective function on enzymes during freezing and drying [78] and its structural transitions upon drying were modulated by the presence of membranes [79]. A connection of several LEA proteins to the ABA-inducible dehydration-responsive protein RD29B/LTI65 might be related to the induction of LEA proteins by ABA [80,81]. In *nced2569* (9-cis epoxy-carotenoid dioxygenase), a quadruple mutant with ABA deficiency, expression of genes encoding LEA28, LEA29, LEA48 and LEA42 was strongly downregulated together with 1-Cys peroxiredoxin1 (Per1), all involved in our LEA cluster [81]. The negative connectivity to RBCL was also interesting, which might point to a protective role of LEA proteins as chaperones for this protein before the induction of photosynthesis during development.

Only LEA20 and LEA46 showed connectivity to a cluster, including metabolites and amino acids of the central metabolism and were only weakly connected to the LEA cluster. LEA20 from Pfam group LEA_5, also named EM6, is induced by ABA and salt [42] and mutant analysis revealed that LEA20 is required for normal seed development in *Arabidopsis* [82]. LEA20 has a role in water binding and loss during embryo maturation,

as mature seed from these mutants lose more water than wild-type seeds during air drying and show an accelerated acquisition of desiccation tolerance [83].

The hub LEA46 (Pfam group LEA_1) with high connectivity might have a protective function, e.g., for enzymes as cytochrome-c-oxidase involved in respiration. The overexpression of LEA46 conferred drought tolerance to severe stress [84]. This might be due to the chaperone-like activity after conformational transition from disorder to alpha-helix folding [85]. Negative correlation of abundance patterns of LEA46 with that of isoleucine, glutamine and threonine might point to a mechanism switching from protection of enzymes during the dry state to securing biosynthesis of crucial metabolites during the developmental stages. Threonine serves as precursor of isoleucine, important for root development, as isoleucine deficiency impairs root development in *Arabidopsis* [86]. Furthermore, all three amino acids might serve as precursor for glutathione, important as antioxidants and for detoxification [23].

A demonstration of the functional relevance of the identified putative interactions will be necessary in the future to fully understand the importance of the putative protein-metabolite network during seed germination.

4. Materials and Methods

4.1. Plant Materials and Experimental Conditions

Arabidopsis thaliana (accession Col-0) plants were grown on soil in a climate chamber with 20 °C/60% relative humidity (RH) during the day and 6 °C/70% RH during the night in a 14 h/10 h light/dark cycle with of 180 $\mu\text{mol m}^{-2} \text{s}^{-1}$ light intensity. After a week, plants were moved to a phytotron with an 8 h light period at a light intensity of 200 $\mu\text{mol m}^{-2} \text{s}^{-1}$, 20 °C, 60% RH and 16 °C during the 16 h night at 75% RH for another week before pricking. After pricking, plants were kept at these conditions for two additional weeks to gain more leaf biomass before being transferred to long-day conditions in a greenhouse (16 h light at 21 °C, 8 h night at 19 °C and 50% RH during day and night) until seed harvest. Plants were bagged after siliques matured to minimize seed loss. After bagging, plants were left to dry for two weeks in the greenhouse. Seeds were collected in the bags and stored in the dark at 15 °C/15% RH for three weeks before they were used for experiments.

Samples for metabolomics, proteomics and transcript analysis were taken from dry seeds (DS), germinating seeds and young seedlings (24, 36, 48, 60 and 72 HAI). Seeds (except for DS) (75 mg of DS per replicate, 3–5 replicates per time point) were sterilized with 70% ethanol for 2 min, followed by 9% bleach solution containing 0.02% Triton X-100 for 17 min and five washings with sterile water. Seeds were then sown on blue germinating paper (grade 190, 300 g/m², Sartorius, Göttingen, Germany) on agar plates containing half-strength MS medium without sucrose. Seeds on plates were stratified at 4 °C in the dark for 24 h before transfer to a growth chamber (22 °C, 16 h light period). Germinating seeds and seedlings were collected 24, 36, 48, 60 and 72 HAI by flash freezing with liquid nitrogen in a mortar. After homogenizing dry seed samples using mixer mill MM200 (Retsch, Haan, Germany) and germinating seeds and seedlings with mortar and pestle, fresh weight (FW) of the samples was determined before extractions for further analysis. Dry weight (DW) of samples was determined from 60 mg FW of plant material after drying overnight at 4 °C in a freeze dryer.

4.2. Metabolite Profiling

Aliquots of small polar metabolites were prepared from five biological replicates of each time point using 60 mg of ground fresh material. The extraction used water/methanol:chloroform with lipid partitioning as previously described [87]. ¹³C₆-sorbitol (Sigma-Aldrich, Taufkirchen, Germany) was added during the extraction and used as internal standard. The analysis, using 160 μL from the upper polar phase, dried overnight in a vacuum concentrator, was carried out using gas chromatography after derivatization with 40 μL methoxylation reagent in pyridine and 80 μL silylation-mix for tri-methyl-silylation

in split 1/30 and splitless mode with 1 μL injection volume coupled to electron impact ionization-time of flight-mass spectrometry (GC/EI-TOF-MS) [87]. ChromaTOF software (LECO Instrumente GmbH, Mönchengladbach, Germany) was used to process the acquired chromatograms. Identification of metabolites was conducted using TagFinder [88], the NIST08 software (<http://chemdata.nist.gov/dokuwiki/doku.php?id=start> (accessed on 28 January 2020)) and the mass spectral and retention time index reference collection of the Golm Metabolome Database [89–92]. Statistical analysis was based on mass spectral intensities were normalized to dry weight and $^{13}\text{C}_6$ -sorbitol.

The mass spectrometry metabolomics data have been deposited into the MetaboLights database [93] with the dataset identifier MTBLS2980.

4.3. Amino Acid Analysis

Dried 160 μL of the upper polar phase extracted for metabolite profiling were used for the analysis. Four biological replicates per time point were resuspended in 65 μL of 0.1 M HCl and centrifuged at 14,000 rpm at 4 $^{\circ}\text{C}$ for 15 min. The measurement of amino acids was carried out using reverse HPLC as previously described [94] with small adjustments in LC gradient. Amino acid content was calculated based on the peak area of the mass fragments normalized to DW of the sample.

4.4. Proteomic Analysis of Heat-Stable Proteins

Extraction of total soluble protein and extraction of the heat-stable proteome (95 $^{\circ}\text{C}$, 10 min) were performed as described [41] from three biological replicates from each time point. Protein concentration was determined using Bradford reagent (B6916, Sigma-Aldrich, Taufkirchen, Germany)/(Bio-Rad Laboratories, Hercules, CA, USA) and bovine serum albumin as a standard. Then, 100 μg of protein per sample was mixed with 100 μL of 8 M urea in 10 mM Tris-HCl, pH 8.0. Samples were loaded on filter columns (Microcon-30 kDa Centrifugal Filter Unit with Ultracel-30 membrane). Columns were washed with 8 M urea, 10 mM Tris-HCl, pH 8.0. Next, proteins were reduced using 10 mM DTT in 8 M urea, 10 mM Tris-HCl, pH 8.0 and alkylated using 27 mM iodoacetamide in the same buffer. Samples were mixed at 600 rpm for 1 min and then incubated without mixing for an additional 5 min. Subsequently, columns were washed using 8 M urea, 10 mM Tris-HCl, pH 8.0 and centrifuged. Proteins in the columns were digested for 14 h using mass spec grade Trypsin/Lys-C mix (Promega, Madison, WI, USA). The resulting peptides were desalted on C18 SepPack columns (Teknokroma, Barcelona, Spain) and eluted with 800 μL 60% acetonitrile, 0.1% trifluoroacetic acid, dried in the speed vacuum concentrator and stored at -80°C . Measurements were performed on a Q Exactive HF high-resolution mass spectrometer (Thermo Scientific, Waltham, MA, USA) coupled to an ACQUITY UPLC M-Class System (Waters, Milford, MA, USA). Peptide samples were loaded onto an Acclaim PepMap RSLC reversed-phase column (75 μm inner diameter, 25 cm length, 2 μm bead size (Thermo Scientific, Waltham, MA, USA) at a flow rate of 0.4 $\mu\text{L min}^{-1}$ in 3% (*v/v*) acetonitrile, 0.1% (*v/v*) formic acid. Peptides were eluted by an acetonitrile gradient from 3% to 80% (*v/v*) over 120 min at a flow rate of 0.5 $\mu\text{L/min}$. Peptide ions were detected in a full scan from mass-to-charge ratio 300 to 1600 at resolution of 60,000. MS/MS scans were performed for the top ten MS scans with the highest MS signal (ddMS2 resolution of 15,000, AGC target 3e5, isolation width mass-to-charge ratio 1.4 *m/z*, relative collision energy 30). Peptides for which MS/MS spectra had been recorded were excluded from further MS/MS scans for 20 s.

Raw data were processed using MaxQuant software [95] and the *A. thaliana* TAIR10 annotations (Arabidopsis TAIR database Version 10, The Arabidopsis Information Resource, www.Arabidopsis.org (accessed on 27 February 2019)) in combination with the search engine Andromeda [96]. The settings for MaxQuant analysis were set as follows: trypsin and lysine selected as digesting enzyme, two missed cleavages allowed, fixed modification was set to carbamidomethylation (cysteine) and oxidation of methionine was set as variable modification. Spectra were also searched against a decoy database of the *A. thaliana*

proteome and results were filtered to obtain a FDR below 1% on the protein level. The “label-free quantification” and “match between runs” options were selected. A minimum peptide length of six amino acids was used. The quantification was performed for proteins with a minimum of one unique and one razor peptide. Known contaminants, such as keratins, and proteins, which were identified with only one unique peptide, were removed from further analysis. Statistical analysis was based on mass spectral label-free quantification (LFQ) intensities.

The mass spectrometry proteomics data have been deposited into the ProteomeX-change Consortium via the PRIDE [97] partner repository with the dataset identifier PXD027546.

4.5. qRT-PCR Analysis

Total RNA was extracted from three biological replicates of each time point. The extraction was performed as described [98], with minor modifications. During the first step of RNA extraction, six artificial RNAs from the ArrayControl RNA Spikes kit (Ambion, Austin, TX, USA) were added to the samples as internal standard to allow calculation of the absolute number of transcript copies [99,100]. The number of copies of the different RNA spikes per extract were as follows: Spike 1 (6.08×10^9), Spike 2 (1.52×10^9), Spike 3 (4.56×10^8), Spike 4 (1.14×10^8), Spike 5 (2.94×10^7) and Spike 8 (7.60×10^5). RNA concentration and quality were assessed using a Nanodrop One UV/VIS spectrophotometer (Thermo Scientific, Waltham, MA, USA). Genomic DNA in the RNA samples were removed using RapidOut DNA removal kit (Thermo Scientific, Waltham, MA, USA). Absence of DNA was confirmed by qPCR using primers designed for an intron sequence of *rbcs* (Supplemental Table S11) [101]. All qPCR reactions used 2× SYBR Green Master Mix Reagent (Applied Biosystems, San Francisco, CA, USA) in optical 384-well plates using 7900HT Fast Real-Time PCR System (Thermo Scientific, Waltham, MA, USA). cDNA was synthesized using PrimeScript RT reagent kit with the use of Oligo dT primers (Takara Bio, Kusatsu, Japan). To check the quality of cDNA, qPCR was performed using primers designed for 5' and 3' ends of *GAPDH* (Supplemental Table S11) [99]. All samples were found with good quality, as $|Ct5'-Ct3'|$ of *GAPDH* was between 0–1.

To observe the dynamics of LEA transcripts at different HAI, we used primers specific for 41 LEA genes and the six RNA spikes. This included 34 LEA genes that were previously identified as expressed in dry seeds [42] and eight additional genes encoding LEA proteins identified by our proteomics analysis. Sequences of all primers targeting LEA transcripts except LEA35 and RNA spikes have been published previously [42,100] and are listed in Supplemental Table S11. The absolute copy number of each LEA transcript was calculated based on the linear correlation of the CT values and \log_{10} copy number of RNA spikes in all samples. Extrapolation based on the linear correlation was carried out for transcripts with a copy number of more than 6.08×10^9 . Absolute copy numbers were normalized to sample DW prior to statistical analysis.

4.6. Statistical and Other Analyses

All statistical analyses were performed with RStudio version 1.2.5033 [102]. Prior to performing Principal Component Analysis (PCA) and unpaired two-sided *t*-test, data imputation was performed if a metabolite was detected in three out of five and a protein or transcript was detected in two out of three replicates per time point. The imputation was performed using half the minimum value of either DW-normalized mass-spectral intensities (for metabolites), LFQ intensity (for heat-stable proteins) or DW-normalized absolute copy number (for LEA transcripts). Analytes with more than 40% missing value at a time point were not considered for further evaluation. All data were \log_2 -transformed and median-normalized for statistical analyses.

In addition, for the heat-stable protein data set, prediction of disorder content of the proteins was carried out using the MFDp2 online tool [50]. The protein sequences of 701 proteins from our data set were available on the UniProtKB/Swiss-Prot database [103].

Unique and common proteins in each time point were shown in a Venn diagram from “VennDiagram” package version 1.6.20. Moreover, functional enrichment of heat-stable proteins was performed according to bins assigned by MapMan using TAIR10 annotations [51]. Enrichment of MapMan bins in the heat-stable proteome was analyzed based on LFQ intensities, which were added with 1 after imputation, thus allowing \log_{10} transformation of proteins with previous intensities of 0 without affecting the intensities of other proteins. CorTo v.1.0.3 (<http://www.usadellab.org/cms/index.php?page=corto> (accessed on 6 January 2021)) was used to analyze the enrichment by conducting Fisher’s exact test followed by the Benjamini and Hochberg correction with a p value of 0.01 as a threshold. Z-scores were calculated from p values using the inverse normal cumulative distribution function.

PCA was performed using the “pcaMethods” package version 1.74.0 [104]. For PCA, missing values were replaced with 0 and mean values were Pareto-scaled and mean-centered using the probabilistic method. Abundance of analytes from DS to 72 HAI were shown in heat maps which illustrated the \log_2 -median transformed mean of each analyte at each observed time point. Only the missing values in the heat map showing the abundance of all heat-stable proteins were replaced with 0 to allow clustering using Euclidean clustering on the heat map, otherwise, the missing values remained. Clustering for other heat maps was performed based on the k-means clustering approach by excluding metabolites, proteins or transcripts with missing values from k-means clustering. Clusters were further characterized based on classes of metabolites, protein bins as assigned by MapMan [51] and Pfam groups of LEA proteins [42]. Unpaired two-sided t -test was performed to test for significant differences between abundance at DS and subsequent time points. Pearson correlation tests were performed between the abundance of LEA proteins and their corresponding transcripts for those with protein and transcript detection from at least four of the six investigated time points. All p values were adjusted using the Benjamini–Hochberg procedure [105] for false discovery rate correction.

4.7. Network Analysis of Common Metabolites, Amino Acids, Heat-Stable Proteins

Mean of \log_2 -transformed and median-normalized data of metabolites, amino acids and heat-stable proteins which could be detected in DS and samples from all observed time points after imbibition were included in a network analysis using Cytoscape version 3.7.2 [106] with the threshold cut-off of Benjamini–Hochberg adjusted $p < 0.05$ and Pearson correlation $r \geq |0.99|$, as described previously [107]. Network statistics of the undirected protein–metabolite interaction network revealed a power law with $R^2 = 0.782$, indicating that the network is scale-free.

Supplementary Materials: The following are available online at <https://www.mdpi.com/article/10.3390/ijms22158172/s1>. Figure S1: Seeds and seedlings collected after imbibition. Figure S2: Relative metabolite abundance in dry seeds (DS), germinating seeds and seedlings and their kinetic patterns and class distributions. Figure S3: Relative abundance of heat-stable proteins in dry seeds (DS), germinating seeds and seedlings collected at the indicated time points after imbibition. Figure S4: Kinetic patterns of 182 heat-stable proteins common in dry seeds (DS), germinating seeds and seedlings at indicated time points after imbibition. Figure S5: Network analysis of heat-stable proteins with metabolites and amino acids common in dry seeds (DS), germinating seeds and seedlings. Table S1: Mass-spectral intensities of metabolites normalized to the internal standard $^{13}\text{C}_6$ -sorbitol and dry weight in dry seeds (DS), germinating seeds and seedlings at indicated time points after imbibition. Table S2: Amino acid content normalized to dry weight (nmol/gDW) in dry seeds (DS), germinating seeds and seedlings at indicated time points after imbibition. Table S3: LFQ intensities, annotation and disorder content of heat-stable proteins in dry seeds, germinating seeds and seedlings at indicated time points after imbibition. Table S4: Proteins in different section of the Venn diagram (Figure 4A) with their TAIR10 annotation. Table S5: Functional enrichment of MapMan bins of heat-stable proteins found in dry seeds, germinating seeds and seedlings at indicated time points after imbibition. Table S6: LFQ intensities, annotation and disorder content of fully folded (0% disorder) and fully disordered (100% disorder) heat-stable proteins in dry seeds,

germinating seeds and seedlings at indicated time points after imbibition (see Figure 5). Table S7: LFQ intensities and disorder content of LEA proteins found in heat-stable proteome in dry seeds, germinating seeds and seedlings at indicated time points after imbibition. Table S8: Disorder content of seed-expressed LEA proteins not found in the heat-stable proteome of dry seeds, germinating seeds and seedlings. Table S9: Copy number of LEA transcript normalized to dry weight (copy/gDW) in dry seeds, germinating seeds and seedlings at the indicated time points after imbibition. Table S10: Pearson correlations and Benjamini–Hochberg adjusted *p* value of analytes included in the network. Table S11: List of primers used for RNA quality check and qRT-PCR.

Author Contributions: Conceptualization, D.K.H., E.Z.; methodology, O.G.; M.G.; A.E., R.H., J.K., A.S.; investigation, O.G., M.G., A.E., C.H., F.B.; data curation, O.G., M.G., A.E., C.H., F.B., R.H., J.K., A.S., E.Z.; writing—original draft preparation, O.G., E.Z.; writing—review and editing, O.G., E.Z., with contribution of all authors; visualization, O.G.; supervision, D.K.H., E.Z.; funding acquisition, D.K.H., O.G. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the Max-Planck-Society. O.G. was funded by a Ph.D. scholarship of the Royal Thai Government.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: All data are available in the Supplementary Materials. In addition, the proteomics data are available at ProteomeXchange via the accession number PXD027546. The metabolomics data are available at MetaboLights via MTBLS2980.

Acknowledgments: We thank Jessica Alpers and Ines Fehrle (Max Planck Institute of Molecular Plant Physiology) for technical support, Virginie Mengin for advice on RNA spikes and Stephanie Schaarschmidt (Max Planck Institute of Molecular Plant Physiology) for support and advice with R programming.

Conflicts of Interest: The authors declare no conflict of interest. The funders had no role in the design of the study; in the collection, analyses or interpretation of data; in the writing of the manuscript, or in the decision to publish the results.

References

1. Maia, J.; Dekkers, B.J.W.; Provar, N.J.; Ligterink, W.; Hilhorst, H.W.M. The Re-Establishment of Desiccation Tolerance in Germinated *Arabidopsis thaliana* Seeds and Its Associated Transcriptome. *PLoS ONE* **2011**, *6*, e29123. [[CrossRef](#)] [[PubMed](#)]
2. Daws, M.; Bolton, S.; Burslem, D.F.R.P.; Garwood, N.C.; Mullins, C.E. Loss of desiccation tolerance during germination in neo-tropical pioneer seeds: Implications for seed mortality and germination characteristics. *Seed Sci. Res.* **2007**, *17*, 273–281. [[CrossRef](#)]
3. Buitink, J.; Vu, B.L.; Satour, P.; Leprince, O. The re-establishment of desiccation tolerance in germinated radicles of *Medicago truncatula* Gaertn. seeds. *Seed Sci. Res.* **2003**, *13*, 273–286. [[CrossRef](#)]
4. Oliver, M.J.; Farrant, J.M.; Hilhorst, H.W.; Mundree, S.; Williams, B.; Bewley, J.D. Desiccation Tolerance: Avoiding Cellular Damage During Drying and Rehydration. *Annu. Rev. Plant Biol.* **2020**, *71*, 435–460. [[CrossRef](#)] [[PubMed](#)]
5. Bewley, J.D. Seed germination and dormancy. *Plant Cell* **1997**, *9*, 1055–1066. [[CrossRef](#)]
6. Nonogaki, H.; Bassel, G.; Bewley, D. Germination—Still a mystery. *Plant Sci.* **2010**, *179*, 574–581. [[CrossRef](#)]
7. Boyes, D.C.; Zayed, A.M.; Ascenzi, R.; McCaskill, A.J.; Hoffman, N.E.; Davis, K.R.; Grolach, J. Growth stage-based phenotypic analysis of *Arabidopsis*: A model for high throughput functional genomics in plants. *Plant Cell* **2001**, *13*, 1499–1510. [[CrossRef](#)]
8. Fait, A.; Angelovici, R.; Less, H.; Ohad, I.; Urbanczyk-Wochniak, E.; Fernie, A.R.; Galili, G. *Arabidopsis* Seed Development and Germination Is Associated with Temporally Distinct Metabolic Switches. *Plant Physiol.* **2006**, *142*, 839–854. [[CrossRef](#)]
9. Sreenivasulu, N.; Usadel, B.; Winter, A.; Radchuk, V.; Scholz, U.; Stein, N.; Weschke, W.; Strickert, M.; Close, T.J.; Stitt, M.; et al. Barley Grain Maturation and Germination: Metabolic Pathway and Regulatory Network Commonalities and Differences Highlighted by New MapMan/PageMan Profiling Tools. *Plant Physiol.* **2008**, *146*, 1738–1758. [[CrossRef](#)]
10. Allen, E.; Moing, A.; Ebbels, T.M.; Maucourt, M.; Tomos, A.D.; Rolin, D.; Hooks, M.A. Correlation Network Analysis reveals a sequential reorganization of metabolic and transcriptional states during germination and gene-metabolite relationships in developing seedlings of *Arabidopsis*. *BMC Syst. Biol.* **2010**, *4*, 62. [[CrossRef](#)]
11. Silva, A.T.; Ligterink, W.; Hilhorst, H.W.M. Metabolite profiling and associated gene expression reveal two metabolic shifts during the seed-to-seedling transition in *Arabidopsis thaliana*. *Plant Mol. Biol.* **2017**, *95*, 481–496. [[CrossRef](#)]
12. Nakabayashi, K.; Okamoto, M.; Koshihara, T.; Kamiya, Y.; Nambara, E. Genome-wide profiling of stored mRNA in *Arabidopsis thaliana* seed germination: Epigenetic and genetic regulation of transcription in seed. *Plant J.* **2005**, *41*, 697–709. [[CrossRef](#)]

13. Silva, A.T.; Ribone, P.A.; Chan, R.L.; Ligterink, W.; Hilhorst, H.W. A Predictive Coexpression Network Identifies Novel Genes Controlling the Seed-to-Seedling Phase Transition in *Arabidopsis thaliana*. *Plant Physiol.* **2016**, *170*, 2218–2231. [[CrossRef](#)] [[PubMed](#)]
14. Bai, B.; Peviani, A.; Van Der Horst, S.; Gamm, M.; Snel, B.; Bentsink, L.; Hanson, J. Extensive translational regulation during seed germination revealed by polysomal profiling. *New Phytol.* **2016**, *214*, 233–244. [[CrossRef](#)]
15. Bai, B.; Van Der Horst, S.; Cordewener, J.H.; America, T.A.H.P.; Hanson, J.; Bentsink, L. Seed-Stored mRNAs that Are Specifically Associated to Monosomes Are Translationally Regulated during Germination. *Plant Physiol.* **2020**, *182*, 378–392. [[CrossRef](#)] [[PubMed](#)]
16. Han, C.; Zhen, S.; Zhu, G.; Bian, Y.; Yan, Y. Comparative metabolome analysis of wheat embryo and endosperm reveals the dynamic changes of metabolites during seed germination. *Plant Physiol. Biochem.* **2017**, *115*, 320–327. [[CrossRef](#)]
17. Kazmi, R.H.; Willems, L.A.J.; Joosen, R.V.L.; Khan, N.; Ligterink, W.; Hilhorst, H.W.M. Metabolomic analysis of tomato seed germination. *Metabolomics* **2017**, *13*, 145. [[CrossRef](#)]
18. Gallardo, K.; Job, C.; Groot, S.P.; Puype, M.; Demol, H.; Vandekerckhove, J.; Job, D. Proteomic Analysis of Arabidopsis Seed Germination and Priming. *Plant Physiol.* **2001**, *126*, 835–848. [[CrossRef](#)]
19. Rajjou, L.; Gallardo, K.; Debeaujon, I.; Vandekerckhove, J.; Job, C.; Job, D. The Effect of α -Amanitin on the Arabidopsis Seed Proteome Highlights the Distinct Roles of Stored and Neosynthesized mRNAs during Germination. *Plant Physiol.* **2004**, *134*, 1598–1613. [[CrossRef](#)] [[PubMed](#)]
20. Fu, Q.; Wang, B.-C.; Jin, X.; Li, H.-B.; Han, P.; Wei, K.-H.; Zhang, X.-M.; Zhu, Y.-X. Proteomic Analysis and Extensive Protein Identification from Dry, Germinating Arabidopsis Seeds and Young Seedlings. *BMB Rep.* **2005**, *38*, 650–660. [[CrossRef](#)]
21. Yang, P.; Li, X.; Wang, X.; Chen, H.; Chen, F.; Shen, S. Proteomic analysis of rice (*Oryza sativa*) seeds during germination. *Proteomics* **2007**, *7*, 3358–3368. [[CrossRef](#)]
22. Galland, M.; Huguet, R.; Arc, E.; Cueff, G.; Job, D.; Rajjou, L. Dynamic Proteomics Emphasizes the Importance of Selective mRNA Translation and Protein Turnover during Arabidopsis Seed Germination. *Mol. Cell. Proteom.* **2014**, *13*, 252–268. [[CrossRef](#)]
23. Gu, J.; Chao, H.; Gan, L.; Guo, L.; Zhang, K.; Li, Y.; Wang, H.; Raboanatahiry, N.; Li, M. Proteomic Dissection of Seed Germination and Seedling Establishment in *Brassica napus*. *Front. Plant Sci.* **2016**, *7*, 1482. [[CrossRef](#)] [[PubMed](#)]
24. Kretzschmar, F.K.; Doner, N.M.; Krawczyk, H.E.; Scholz, P.; Schmitt, K.; Valerius, O.; Braus, G.H.; Mullen, R.T.; Ischebeck, T. Identification of Low-Abundance Lipid Droplet Proteins in Seeds and Seedlings. *Plant Physiol.* **2020**, *182*, 1326–1345. [[CrossRef](#)]
25. Sano, N.; Permana, H.; Kumada, R.; Shinozaki, Y.; Tanabata, T.; Yamada, T.; Hirasawa, T.; Kanekatsu, M. Proteomic Analysis of Embryonic Proteins Synthesized from Long-Lived mRNAs During Germination of Rice Seeds. *Plant Cell Physiol.* **2012**, *53*, 687–698. [[CrossRef](#)] [[PubMed](#)]
26. Sano, N.; Takebayashi, Y.; To, A.; Mhiri, C.; Rajjou, L.; Nakagami, H.; Kanekatsu, M. Shotgun Proteomic Analysis Highlights the Roles of Long-Lived mRNAs and De Novo Transcribed mRNAs in Rice Seeds upon Imbibition. *Plant Cell Physiol.* **2019**, *60*, 2584–2596. [[CrossRef](#)] [[PubMed](#)]
27. Obendorf, R.L. Oligosaccharides and galactosyl cyclitols in seed desiccation tolerance. *Seed Sci. Res.* **1997**, *7*, 63–74. [[CrossRef](#)]
28. Ballesteros, D.; Walters, C. Detailed characterization of mechanical properties and molecular mobility within dry seed glasses: Relevance to the physiology of dry biological systems. *Plant J.* **2011**, *68*, 607–619. [[CrossRef](#)]
29. Ballesteros, D.; Walters, C. Solid-State Biology and Seed Longevity: A Mechanical Analysis of Glasses in Pea and Soybean Embryonic Axes. *Front. Plant Sci.* **2019**, *10*, 920. [[CrossRef](#)]
30. Buitink, J.; Leprince, O. Glass formation in plant anhydrobiotes: Survival in the dry state. *Cryobiology* **2004**, *48*, 215–228. [[CrossRef](#)]
31. Roberts, J.K.; DeSimone, N.A.; Lingle, W.L.; Dure, L., 3rd. Cellular concentrations and uniformity of cell-type accumulation of two LEA proteins in cotton embryos. *Plant Cell* **1993**, *5*, 769–780. [[CrossRef](#)]
32. Dure, L.; Chlan, C. Developmental biochemistry of cottonseed embryogenesis and germination: Xii. Purification and properties of principal storage proteins. *Plant Physiol.* **1981**, *68*, 180–186. [[CrossRef](#)]
33. Dure, L.; Galau, G.A. Developmental biochemistry of cottonseed embryogenesis and germination: Xiii. Regulation of biosynthesis of principal storage proteins. *Plant Physiol.* **1981**, *68*, 187–194. [[CrossRef](#)] [[PubMed](#)]
34. Dure, L., 3rd; Greenway, S.C.; Galau, G.A. Developmental biochemistry of cottonseed embryogenesis and germination: Changing messenger ribonucleic acid populations as shown by in vitro and in vivo protein synthesis. *Biochemistry* **1981**, *20*, 4162–4168. [[CrossRef](#)] [[PubMed](#)]
35. Tunnacliffe, A.; Wise, M.J. The continuing conundrum of the LEA proteins. *Naturwissenschaften* **2007**, *94*, 791–812. [[CrossRef](#)] [[PubMed](#)]
36. Leprince, O.; Pellizzaro, A.; Berriri, S.; Buitink, J. Late seed maturation: Drying without dying. *J. Exp. Bot.* **2016**, *68*, 827–841. [[CrossRef](#)]
37. Hand, S.C.; Menze, M.A.; Toner, M.; Boswell, L.; Moore, D. LEA proteins during water stress: Not just for plants anymore. *Annu. Rev. Physiol.* **2011**, *73*, 115–134. [[CrossRef](#)]
38. Hinch, D.K.; Thalhammer, A. LEA proteins: IDPs with versatile functions in cellular dehydration tolerance. *Biochem. Soc. Trans.* **2012**, *40*, 1000–1003. [[CrossRef](#)]
39. Tunnacliffe, A.; Hinch, D.; Leprince, O.; Macherel, D. LEA proteins: Versatility of forms and function. In *Dormancy and Resistance in Harsh Environments*; Lubzens, E., Cerda, J., Clark, M., Eds.; Springer: Berlin, Germany, 2010; pp. 91–108.

40. Blackman, S.A.; Wettlaufer, S.H.; Obendorf, R.L.; Leopold, A.C. Maturation Proteins Associated with Desiccation Tolerance in Soybean. *Plant Physiol.* **1991**, *96*, 868–874. [[CrossRef](#)] [[PubMed](#)]
41. Boudet, J.; Buitink, J.; Hoekstra, F.A.; Rogniaux, H.; Larre, C.; Satour, P.; Leprince, O. Comparative Analysis of the Heat Stable Proteome of Radicles of *Medicago truncatula* Seeds during Germination Identifies Late Embryogenesis Abundant Proteins Associated with Desiccation Tolerance. *Plant Physiol.* **2006**, *140*, 1418–1436. [[CrossRef](#)]
42. Hundertmark, M.; Hinch, D.K. LEA (Late Embryogenesis Abundant) proteins and their encoding genes in *Arabidopsis thaliana*. *BMC Genom.* **2008**, *9*, 118. [[CrossRef](#)]
43. Knox-Brown, P.; Rindfleisch, T.; Günther, A.; Balow, K.; Bremer, A.; Walther, D.; Miettinen, M.S.; Hinch, D.K.; Thalhammer, A. Similar Yet Different—Structural and Functional Diversity among *Arabidopsis thaliana* LEA_4 Proteins. *Int. J. Mol. Sci.* **2020**, *21*, 2794. [[CrossRef](#)] [[PubMed](#)]
44. Sun, X.; Rikkerink, E.; Jones, W.T.; Uversky, V.N. Multifarious Roles of Intrinsic Disorder in Proteins Illustrate Its Broad Impact on Plant Biology. *Plant Cell* **2013**, *25*, 38–55. [[CrossRef](#)] [[PubMed](#)]
45. Covarrubias, A.A.; Romero-Pérez, P.S.; Cuevas-Velazquez, C.L.; Rendón-Luna, D.F. The functional diversity of structural disorder in plant proteins. *Arch. Biochem. Biophys.* **2020**, *680*, 108229. [[CrossRef](#)] [[PubMed](#)]
46. Graether, S.P.; Boddington, K.F. Disorder and function: A review of the dehydrin protein family. *Front. Plant Sci.* **2014**, *5*, 576. [[CrossRef](#)]
47. Chatelain, E.; Hundertmark, M.; Leprince, O.; Le Gall, S.; Satour, P.; Deligny-Penninck, S.; Rogniaux, H.; Buitink, J. 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.* **2012**, *35*, 1440–1455. [[CrossRef](#)] [[PubMed](#)]
48. Higashi, Y.; Hirai, M.; Fujiwara, T.; Naito, S.; Noji, M.; Saito, K. Proteomic and transcriptomic analysis of *Arabidopsis* seeds: Molecular evidence for successive processing of seed proteins and its implication in the stress response to sulfur nutrition. *Plant J.* **2006**, *48*, 557–571. [[CrossRef](#)]
49. Receveur-Brechot, V.; Bourhis, J.-M.; Uversky, V.N.; Canard, B.; Longhi, S. Assessing protein disorder and induced folding. *Proteins: Struct. Funct. Bioinform.* **2005**, *62*, 24–45. [[CrossRef](#)] [[PubMed](#)]
50. Mizianty, M.J.; Peng, Z.; Kurgan, L. Mfdp2: Accurate predictor of disorder in proteins by fusion of disorder probabilities, content and profiles. *Intrinsically Disord Proteins* **2013**, *1*, e24428. [[CrossRef](#)]
51. Thimm, O.; Blasing, O.; Gibon, Y.; Nagel, A.; Meyer, S.; Krüger, P.; Selbig, J.; Müller, L.A.; Rhee, S.Y.; Stitt, M. mapman: A user-driven tool to display genomics data sets onto diagrams of metabolic pathways and other biological processes. *Plant J.* **2004**, *37*, 914–939. [[CrossRef](#)]
52. Vidigal, D.D.S.; Willems, L.; Van Arkel, J.; Dekkers, B.J.; Hilhorst, H.W.; Bentsink, L. Galactinol as marker for seed longevity. *Plant Sci.* **2016**, *246*, 112–118. [[CrossRef](#)]
53. Salvi, P.; Saxena, S.C.; Petla, B.P.; Kamble, N.U.; Kaur, H.; Verma, P.; Rao, V.; Ghosh, S.; Majee, M. Differentially expressed galactinol synthase(s) in chickpea are implicated in seed vigor and longevity by limiting the age induced ROS accumulation. *Sci. Rep.* **2016**, *6*, 35088. [[CrossRef](#)]
54. Jang, J.-H.; Shang, Y.; Kang, H.K.; Kim, S.Y.; Kim, B.H.; Nam, K.H. *Arabidopsis* galactinol synthases 1 (AtGOLS1) negatively regulates seed germination. *Plant Sci.* **2018**, *267*, 94–101. [[CrossRef](#)] [[PubMed](#)]
55. Gennari, L.; Felletti, M.; Blasa, M.; Angelino, D.; Celeghini, C.; Corallini, A.; Ninfali, P. Total extract of *Beta Vulgaris* var. Cicla seeds versus its purified phenolic components: Antioxidant activities and antiproliferative effects against colon cancer cells. *Phytochem. Anal.* **2011**, *22*, 272–279. [[CrossRef](#)] [[PubMed](#)]
56. Xu, H.-Y.; Zheng, H.-C.; Zhang, H.-W.; Zhang, J.-Y.; Ma, C.-M. Comparison of Antioxidant Constituents of *Agriophyllum squarrosum* Seed with Conventional Crop Seeds. *J. Food Sci.* **2018**, *83*, 1823–1831. [[CrossRef](#)] [[PubMed](#)]
57. Mucciarelli, M.; Gallino, M.; Maffei, M.; Scannerini, S. Effects of 3,4-dihydroxybenzoic acid on tobacco (*Nicotiana tabacum* L.) cultured in vitro. Growth regulation in callus and organ cultures. *Plant Biosyst. Int. J. Deal. All Asp. Plant Biol.* **2000**, *134*, 185–192. [[CrossRef](#)]
58. Ghimire, B.K.; Hwang, M.H.; Sacks, E.J.; Yu, C.Y.; Kim, S.H.; Chung, I.M. Screening of Allelochemicals in *Miscanthus sacchariflorus* Extracts and Assessment of Their Effects on Germination and Seedling Growth of Common Weeds. *Plants* **2020**, *9*, 1313. [[CrossRef](#)]
59. Jarzab, A.; Kurzawa, N.; Hopf, T.; Moerch, M.; Zecha, J.; Leijten, N.; Bian, Y.; Musiol, E.; Maschberger, M.; Stoehr, G.; et al. Meltome atlas—thermal proteome stability across the tree of life. *Nat. Methods* **2020**, *17*, 495–503. [[CrossRef](#)] [[PubMed](#)]
60. Leuenberger, P.; Gansch, S.; Kahraman, A.; Cappelletti, V.; Boersema, P.J.; von Mering, C.; Claassen, M.; Picotti, P. Cell-wide analysis of protein thermal unfolding reveals determinants of thermostability. *Science* **2017**, *355*, eaai7825. [[CrossRef](#)]
61. Volkening, J.D.; Stecker, K.E.; Sussman, M.R. Proteome-wide Analysis of Protein Thermal Stability in the Model Higher Plant *Arabidopsis thaliana*. *Mol. Cell. Proteom.* **2019**, *18*, 308–319. [[CrossRef](#)] [[PubMed](#)]
62. Wojtyła, Ł.; Garnczarska, M.; Zalewski, T.; Bednarski, W.; Ratajczak, L.; Jurga, S. A comparative study of water distribution, free radical production and activation of antioxidative metabolism in germinating pea seeds. *J. Plant Physiol.* **2006**, *163*, 1207–1220. [[CrossRef](#)] [[PubMed](#)]
63. Tsvetkov, P.; Myers, N.; Moscovitz, O.; Sharon, M.; Prilusky, J.; Shaul, Y. Thermo-resistant intrinsically disordered proteins are efficient 20S proteasome substrates. *Mol. Biosyst.* **2011**, *8*, 368–373. [[CrossRef](#)] [[PubMed](#)]
64. Liu, Y.; Wu, J.; Sun, N.; Tu, C.; Shi, X.; Cheng, H.; Liu, S.; Li, S.; Wang, Y.; Zheng, Y.; et al. Intrinsically Disordered Proteins as Important Players during Desiccation Stress of Soybean Radicles. *J. Proteome Res.* **2017**, *16*, 2393–2409. [[CrossRef](#)] [[PubMed](#)]

65. Zamora-Briseño, J.A.; Pereira-Santana, A.; Reyes-Hernández, S.J.; Cerqueda-García, D.; Castaño, E.; Rodríguez-Zapata, L.C. Towards an understanding of the role of intrinsic protein disorder on plant adaptation to environmental challenges. *Cell Stress Chaperon* **2021**, *26*, 141–150. [[CrossRef](#)] [[PubMed](#)]
66. Savitski, M.M.; Reinhard, F.B.M.; Franken, H.; Werner, T.; Savitski, M.F.; Eberhard, D.; Molina, D.M.; Jafari, R.; Dovega, R.B.; Klaeger, S.; et al. Tracking cancer drugs in living cells by thermal profiling of the proteome. *Science* **2014**, *346*, 1255784. [[CrossRef](#)] [[PubMed](#)]
67. Sun, W.; Dai, L.; Yu, H.; Puspita, B.; Zhao, T.; Li, F.; Tan, J.; Lim, Y.T.; Chen, M.W.; Sobota, R.; et al. Monitoring structural modulation of redox-sensitive proteins in cells with MS-CETSA. *Redox Biol.* **2019**, *24*, 101168. [[CrossRef](#)]
68. Azarkovich, M.I. Dehydrins in Orthodox and Recalcitrant Seeds. *Russ. J. Plant Physiol.* **2020**, *67*, 221–230. [[CrossRef](#)]
69. DeRocher, A.E.; Vierling, E. Developmental control of small heat shock protein expression during pea seed maturation. *Plant J.* **1994**, *5*, 93–102. [[CrossRef](#)]
70. Hundertmark, M.; Buitink, J.; Leprince, O.; Hinch, D.K. The reduction of seed-specific dehydrins reduces seed longevity in *Arabidopsis thaliana*. *Seed Sci. Res.* **2011**, *21*, 165–173. [[CrossRef](#)]
71. Kim, S.Y.; Nam, K.H. Physiological roles of ERD10 in abiotic stresses and seed germination of *Arabidopsis*. *Plant Cell Rep.* **2010**, *29*, 203–209. [[CrossRef](#)] [[PubMed](#)]
72. Kiyosue, T.; Yamaguchi-Shinozaki, K.; Shinozaki, K. Characterization of Two cDNAs (ERD10 and ERD14) Corresponding to Genes That Respond Rapidly to Dehydration Stress in *Arabidopsis thaliana*. *Plant Cell Physiol.* **1994**, *35*, 225–231. [[CrossRef](#)] [[PubMed](#)]
73. Sánchez, I.E.H.; Maruri-López, I.; Graether, S.P.; Jiménez-Bremont, J.F. In vivo evidence for homo- and heterodimeric interactions of *Arabidopsis thaliana* dehydrins AtCOR47, AtERD10, and AtRAB18. *Sci. Rep.* **2017**, *7*, 17036. [[CrossRef](#)]
74. Hernández-Sánchez, I.E.; Maruri-López, I.; Molphe-Balch, E.P.; Becerra-Flora, A.; Jaimes-Miranda, F.; Jiménez-Bremont, J.F. Evidence for in vivo interactions between dehydrins and the aquaporin AtPIP2B. *Biochem. Biophys. Res. Commun.* **2019**, *510*, 545–550. [[CrossRef](#)] [[PubMed](#)]
75. Rushton, D.L.; Tripathi, P.; Rabara, R.; Lin, J.; Ringler, P.; Boken, A.K.; Langum, T.J.; Smidt, L.; Boomsma, D.D.; Emme, N.J.; et al. WRKY transcription factors: Key components in abscisic acid signalling. *Plant Biotechnol. J.* **2011**, *10*, 2–11. [[CrossRef](#)]
76. Yang, X.; Cheema, J.; Zhang, Y.; Deng, H.; Duncan, S.; Umar, M.I.; Zhao, J.; Liu, Q.; Cao, X.; Kwok, C.K.; et al. RNA G-quadruplex structures exist and function in vivo in plants. *Genome Biol.* **2020**, *21*, 226. [[CrossRef](#)]
77. Dang, N.X.; Popova, A.; Hundertmark, M.; Hinch, D.K. Functional characterization of selected LEA proteins from *Arabidopsis thaliana* in yeast and in vitro. *Planta* **2014**, *240*, 325–336. [[CrossRef](#)] [[PubMed](#)]
78. Popova, A.; Rausch, S.; Hundertmark, M.; Gibon, Y.; Hinch, D.K. The intrinsically disordered protein LEA7 from *Arabidopsis thaliana* protects the isolated enzyme lactate dehydrogenase and enzymes in a soluble leaf proteome during freezing and drying. *Biochim. Biophys. Acta BBA Proteins Proteom.* **2015**, *1854*, 1517–1525. [[CrossRef](#)] [[PubMed](#)]
79. Popova, A.; Hundertmark, M.; Seckler, R.; Hinch, D.K. Structural transitions in the intrinsically disordered plant dehydration stress protein LEA7 upon drying are modulated by the presence of membranes. *Biochim. Biophys. Acta BBA Biomembr.* **2011**, *1808*, 1879–1887. [[CrossRef](#)]
80. Zamora-Briseño, J.A.; de Jiménez, E.S. A LEA 4 protein up-regulated by ANA is involved in drought response in maize roots. *Mol. Biol. Rep.* **2016**, *43*, 221–228. [[CrossRef](#)]
81. Chauffour, F.; Bailly, M.; Perreau, F.; Cueff, G.; Suzuki, H.; Collet, B.; Frey, A.; Clément, G.; Soubigou-Taconnat, L.; Balliau, T.; et al. Multi-omics Analysis Reveals Sequential Roles for ABA during Seed Maturation. *Plant Physiol.* **2019**, *180*, 1198–1218. [[CrossRef](#)]
82. Manfre, A.J.; Lanni, L.M.; Marcotte, W.R. The *Arabidopsis* Group 1 Late Embryogenesis Abundant Protein ATEM6 Is Required for Normal Seed Development. *Plant Physiol.* **2006**, *140*, 140–149. [[CrossRef](#)]
83. Manfre, A.J.; LaHatte, G.A.; Climer, C.R.; Marcotte, W.R. Seed Dehydration and the Establishment of Desiccation Tolerance During Seed Maturation is Altered in the *Arabidopsis thaliana* Mutant *atem6-1*. *Plant Cell Physiol.* **2008**, *50*, 243–253. [[CrossRef](#)] [[PubMed](#)]
84. Olvera-Carrillo, Y.; Campos, F.; Reyes, J.L.; Garcarrubio, A.; Covarrubias, A.A. Functional Analysis of the Group 4 Late Embryogenesis Abundant Proteins Reveals Their Relevance in the Adaptive Response during Water Deficit in *Arabidopsis*. *Plant Physiol.* **2010**, *154*, 373–390. [[CrossRef](#)]
85. Cuevas-Velazquez, C.; Saab-Rincón, G.; Reyes, J.L.; Covarrubias, A.A. The Unstructured N-terminal Region of *Arabidopsis* Group 4 Late Embryogenesis Abundant (LEA) Proteins Is Required for Folding and for Chaperone-like Activity under Water Deficit. *J. Biol. Chem.* **2016**, *291*, 10893–10903. [[CrossRef](#)] [[PubMed](#)]
86. Yu, H.; Zhang, F.; Wang, G.; Liu, Y.; Liu, D. Partial deficiency of isoleucine impairs root development and alters transcript levels of the genes involved in branched-chain amino acid and glucosinolate metabolism in *Arabidopsis*. *J. Exp. Bot.* **2012**, *64*, 599–612. [[CrossRef](#)]
87. Erban, A.; Martinez-Seidel, F.; Rajarathinam, Y.; Dethloff, F.; Orf, I.; Fehrle, I.; Alpers, J.; Beine-Golovchuk, O.; Kopka, J. Multiplexed profiling and data processing methods to identify temperature-regulated primary metabolites using gas chromatography coupled to mass spectrometry. *Methods Mol. Biol.* **2020**, *2156*, 203–239. [[PubMed](#)]
88. Luedemann, A.; Strassburg, K.; Erban, A.; Kopka, J. TagFinder for the quantitative analysis of gas chromatography—Mass spectrometry (GC-MS)-based metabolite profiling experiments. *Bioinformatics* **2008**, *24*, 732–737. [[CrossRef](#)]

89. Hummel, J.; Strehmel, N.; Selbig, J.; Walther, D.; Kopka, J. Decision tree supported substructure prediction of metabolites from GC-MS profiles. *Metabolomics* **2010**, *6*, 322–333. [[CrossRef](#)]
90. Kopka, J.; Schauer, N.; Krueger, S.; Birkemeyer, C.; Usadel, B.; Bergmuller, E.; Dormann, P.; Weckwerth, W.; Gibon, Y.; Stitt, M.; et al. GMD@CSB.DB: The Golm Metabolome Database. *Bioinformatics* **2004**, *21*, 1635–1638. [[CrossRef](#)] [[PubMed](#)]
91. Allwood, J.W.; Erban, A.; De Koning, S.; Dunn, W.; Luedemann, A.; Lommen, A.; Kay, L.; Löscher, R.; Kopka, J.; Goodacre, R. Inter-laboratory reproducibility of fast gas chromatography–electron impact–time of flight mass spectrometry (GC–EI–TOF/MS) based plant metabolomics. *Metabolomics* **2009**, *5*, 479–496. [[CrossRef](#)]
92. Strehmel, N.; Hummel, J.; Erban, A.; Strassburg, K.; Kopka, J. Retention index thresholds for compound matching in GC–MS metabolite profiling. *J. Chromatogr. B* **2008**, *871*, 182–190. [[CrossRef](#)]
93. Haug, K.; Cochrane, K.; Nainala, V.C.; Williams, M.; Chang, J.; Jayaseelan, K.V.; O’Donovan, C. MetaboLights: A resource evolving in response to the needs of its scientific community. *Nucleic Acids Res.* **2019**, *48*, D440–D444. [[CrossRef](#)] [[PubMed](#)]
94. Watanabe, M.; Tohge, T.; Balazadeh, S.; Erban, A.; Giavalisco, P.; Kopka, J.; Mueller-Roeber, B.; Fernie, A.R.; Hoefgen, R. Comprehensive Metabolomics Studies of Plant Developmental Senescence. *Adv. Struct. Saf. Stud.* **2018**, *1744*, 339–358. [[CrossRef](#)]
95. Cox, J.; Mann, M. MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification. *Nat. Biotechnol.* **2008**, *26*, 1367–1372. [[CrossRef](#)] [[PubMed](#)]
96. Cox, J.; Neuhauser, N.; Michalski, A.; Scheltema, R.; Olsen, J.V.; Mann, M. Andromeda: A Peptide Search Engine Integrated into the MaxQuant Environment. *J. Proteome Res.* **2011**, *10*, 1794–1805. [[CrossRef](#)]
97. Perez-Riverol, Y.; Csordas, A.; Bai, J.; Bernal-Llinares, M.; Hewapathirana, S.; Kundu, D.J.; Inuganti, A.; Griss, J.; Mayer, G.; Eisenacher, M.; et al. The PRIDE database and related tools and resources in 2019: Improving support for quantification data. *Nucleic Acids Res.* **2019**, *47*, D442–D450. [[CrossRef](#)]
98. Vicient, C.M.; Delseny, M. Isolation of Total RNA from Arabidopsis thaliana Seeds. *Anal. Biochem.* **1999**, *268*, 412–413. [[CrossRef](#)]
99. Piques, M.; Schulze, W.X.; Höhne, M.; Usadel, B.; Gibon, Y.; Rohwer, J.; Stitt, M. Ribosome and transcript copy numbers, polysome occupancy and enzyme dynamics in Arabidopsis. *Mol. Syst. Biol.* **2009**, *5*, 314. [[CrossRef](#)]
100. Flis, A.; Fernández, A.P.; Zielinski, T.; Mengin, V.; Sulpice, R.; Stratford, K.; Hume, A.; Pokhilko, A.; Southern, M.M.; Seaton, D.D.; et al. Defining the robust behaviour of the plant clock gene circuit with absolute RNA timeseries and open infrastructure. *Open Biol.* **2015**, *5*, 5. [[CrossRef](#)]
101. Czechowski, T.; Stitt, M.; Altmann, T.; Udvardi, M.K.; Scheible, W.-R. Genome-Wide Identification and Testing of Superior Reference Genes for Transcript Normalization in Arabidopsis. *Plant Physiol.* **2005**, *139*, 5–17. [[CrossRef](#)]
102. RStudio Team. *RStudio: Integrated Development for R*; R Studio, Inc.: Boston, MA, USA, 2019.
103. UniProt, Consortium. UniProt: A worldwide hub of protein knowledge. *Nucleic Acids Res.* **2019**, *47*, D506–D515. [[CrossRef](#)]
104. Stacklies, W.; Redestig, H.; Scholz, M.; Walther, D.; Selbig, J. pcaMethods a bioconductor package providing PCA methods for incomplete data. *Bioinformatics* **2007**, *23*, 1164–1167. [[CrossRef](#)] [[PubMed](#)]
105. Benjamini, Y.; Hochberg, Y. Controlling the false discovery rate: A practical and powerful approach to multiple testing. *J. R. Stat. Soc. Ser. B* **1995**, *57*, 289–300. [[CrossRef](#)]
106. Shannon, P.; Markiel, A.; Ozier, O.; Baliga, N.S.; Wang, J.T.; Ramage, D.; Amin, N.; Schwikowski, B.; Ideker, T. Cytoscape: A Software Environment for Integrated Models of Biomolecular Interaction Networks. *Genome Res.* **2003**, *13*, 2498–2504. [[CrossRef](#)] [[PubMed](#)]
107. Contreras-López, O.; Moyano, T.C.; Soto, D.; Gutiérrez, R.A. Step-by-Step Construction of Gene Co-expression Networks from High-Throughput Arabidopsis RNA Sequencing Data. *Methods Mol. Biol.* **2018**, *1761*, 275–301. [[CrossRef](#)] [[PubMed](#)]