Genome-Wide Characterization of Wholly Disordered Proteins in Arabidopsis

Wenfen Long , Liang Zhao , Huimin Yang , Xinyi Yang , Yulong Bai , Xiuhua Xue , Doudou Wang * , Shengcheng Han *

Posted Date: 18 December 2024

doi: 10.20944/preprints202412.1512.v1

Keywords: WDPs; physicochemical properties; expression; PPI; liquid-liquid phase separation; Arabidopsis

 doi:10.20944/preprints202412.1512.v1

Article

Genome-Wide Characterization of Wholly Disordered Proteins in *Arabidopsis*

Wenfen Long [1,†], Liang Zhao [1,†], Huimin Yang [1], Xinyi Yang [1], Yulong Bai [1], Xiuhua Xue [1], Doudou Wang [1,*] and Shengcheng Han [1,2,*]

[1] Beijing Key Laboratory of Gene Resources and Molecular Development, College of Life Sciences, Beijing Normal University, Beijing 100875, China; 202321200015@mail.bnu.edu.cn (W.L.); zhaoliang20181012@163.com (L.Z.); YHM_2024@163.com (H.Y.); yangxinyi2024@sibcb.ac.cn (X.Y.); Byl1605164599@163.com (Y.B.); xiuhuaxue@bnu.edu.cn (X.X.)

[2] Academy of Plateau Science and Sustainability of the People's Government of Qinghai Province & Beijing Normal University, Qinghai Normal University, Xining 810008, Qinghai, China

* Correspondence: wangdoudou9408@163.com (D.W.); schan@bnu.edu.cn (S.H.)

† These authors contributed equally to this work.

Abstract: Intrinsically disordered proteins (IDPs) include two types of proteins: one contains partial disordered regions (IDRs), the other is wholly disordered proteins (WDPs). Extensive studies focused on the proteins with IDRs, but less is known about WDPs because of their difficult to form the folded tertiary structure. In this study, we developed a bioinformatics method for screening WDPs more than 50 amino acids in the genome level and found a total of 27 categories including 56 WDPs in *Arabidopsis*. After comparing with randomly selected 56 structural proteins, we found that WDPs possessed more wide range of theoretical isoelectric point (PI), more negative of Grand Average of Hydropathicity (GRAVY), higher value of Instability Index (II) and lower values of Aliphatic Index (AI). In addition, by calculating the FCR (fraction of charged residue) and NCPR (net charge per residue) values of each WDP, we found twenty WDPs in R1 (FCR < 0.25 & NCPR < 0.25) group, fifteen in R2 (0.25 ≤ FCR ≤ 0.35 & NCPR ≤ 0.35), nineteen in R3 (FCR > 0.35 & NCPR ≤ 0.35), and two in R4 (FCR > 0.35 & NCPR > 0.35). Moreover, the gene expression and protein-protein interaction (PPI) network analysis showed that WDPs perform different biological functions. We also showed that two WDPs, SIS (Salt Induced Serine rich) and RAB18 (a dehydrin family protein) undergo the *in vitro* liquid-liquid phase separation (LLPS). Therefore, our results provide insight into understanding the biochemical characters and biological functions of WDPs in plants.

Keywords: WDPs; physicochemical properties; expression; PPI; liquid-liquid phase separation; Arabidopsis.

1. Introduction

The lock-and-key hypothesis proposed by Emil Fischer highlights the requirement for complementary and matching structures of given substrates to fit to an enzyme [1]. This model solidifies the conception of structure-function continuum. With the development of X-ray crystal diffraction, NMR and other experimental techniques, researchers have realised that the structures of proteins are closely related to their biological functions. However, not all proteins have a highly specific three-dimensional structure when function in biological process [2-6]. Increasing evidences have suggested that disordered conformation and flexible structure hold a key position in the function of proteins [7]. The above findings challenge traditional structure-function paradigm [8]. Under physiological conditions, intrinsically disordered proteins (IDPs) are unable to form specific secondary or 3D structures, but nevertheless participate in various processes. IDPs are divided into two types: proteins with partially disordered regions (IDRs), which confer the conformational

 doi:10.20944/preprints202412.1512.v1

flexibility to IDPs, and wholly disordered proteins (WDPs) [9]. When combined with different molecules, IDPs shift their binding modes and experience a transition from disorder to order, so that they ensemble specific conformations to function [10,11]. The dynamic transformation of conformation enables IDPs to interact with proteins or nucleic acids, trigger liquid-liquid phase separation (LLPS) and regulate membraneless organelles, etc [12,13].

IDPs are widely distributed throughout the biological kingdom and are particularly prevalent in complex eukaryotes, where they account for about 40% of all proteins [14,15]. Previously studies showed that IDPs often form biomolecular polymers via LLPS, dependent on their IDRs [16]. In mammals, aberrant IDP expression has been implicated in many diseases, such as cancers, cardio-cerebrovascular and neurodegenerative diseases [17-19]. For example, as one of identified IDPs, Tau's aggregation contributes to Alzheimer's disease (AD) [20], and phosphorylation modification promotes the phase separation of Tau and accelerates the formation of amyloid protein [21]. In plants, IDPs affect transcriptional regulation and post-translational modification, participate in the process of signal transduction, disease resistance and stress response [22,23]. The Late Embryogenesis Abundant (LEA) protein gene family is a group of disordered proteins which have been widely studied in plants including *Solanum lycopersicum*, *Prunus mume* and *Arabidopsis thaliana* [24-26]. When plants are exposed to extreme conditions, LEA is highly expressed and even ensures plant survival through complete loss of water [27-29]. When transformed into other plants, LEA genes improve the stress resistance of the transgenic plants. For example, overexpression of Barley *HVA1* gene in wheat and *Oryza sativa* enhance their tolerance to water shortage and significantly increase their water utilization [30,31]. In sumarry, IDPs are essential for a wide range of physiological processes in complex biological systems.

In recent years, a variety of computational prediction methods have been developed to characterize IDPs, based on different training purposes and different data sets [32,33]. DisProt has become the gold standard in IDP/IDR annotation, and the number of experimentally validated IDPs in the DisProt has increased substantially [34,35]. DisProt 9.5 contains 2896 entries of IDPs, including relevant data for *Arabidopsis*, mice, yeast and other species. In *Arabidopsis*, one hundred IDPs are verified, among them, twelve members are completely disordered which correspond to the LEA family (including dehydrin), the Calvin cycle protein CP12-2 and the protein COLD-REGULATED 15A [36,37]. However, the classification of IDPs is chaotic. In previous studies, proteins with partial disordered regions have been studied well for their ordered regions pocessing obvious physiological roles, instead, the study of WDPs were challenged by its structural characterisation [38,39]. The properties of WDPs and the distinctions in their expression patterns under different circumstances remain poorly understood. Moreover, the effect of whole-chain disorder on intermolecular interactionand the specific affinity of their binding sites as well as bonds with other molecules need to be further elucidated.

Here, we presents a clear definition of wholly disordered proteins (WDPs) in Arabidopsis, which represent proteins with nearly 100% of highly disorder. A set of criteria was developed based on the Arabidopsis Proteome and predicted structures from AlphaFold and MobiDB, which were used to screen out fifty-six WDPs with disorder degrees of 90% to 100%. A comprehensive evaluation was conducted to assess their physicochemical properties, evolutionary characteristics and expression patterns under four different abiotic stress treatments. To further comprehend the potential biological roles of selected WDPs, we proceeded to predict their interactions with others, and these WDPs underwent GO analysis in the process. In vitro phase separation assay shows that two WDPs, SIS and RAB18, are capable of form LLPS in vitro, which provides insights for future research on these proteins in plants.

2. Materials and Methods

2.1. Screening of Arabidopsis WDPs

We used the Advanced Search function of UniprotKB in Uniprot (https://www.uniprot.org/, Release 2022 _05, Released on: Wed Dec 14 2022) to database and set filtering parameters:

 doi:10.20944/preprints202412.1512.v1

(Proteome_id:UP000006548) AND (length:(xxx TO xxx)) AND (Region:Disordered:(xxx TO xxx)) .Peptide sequences are composed of fewer than 50 amino acids each [40], so we commenced our investigation at amino acid number 0 and proceeded in increments of 50, setting a search interval of 50 amino acids. The length of unstructured area was established by setting the shortest to the longest length of intervals to guarantee the areas of disorder accounted for 90~100%. The retrieved results were then manually screened to ensure they were all consistent with the WDPs' definition. Furthermore, any repeated sequences were removed using the Seqkit software. The disordered protein sequences were subjected to comparative analysis with the Arabidopsis proteome, and the matched sequences underwent multiple sequence alignment using Uniprot-Align. The alignment results indicated the removal of specific short fragments. It should be noted that protein structures predicted by AlphaFold and MobiDB were also combined in order to ensure the totality of WDPs.

2.2. Phylogenetic Analysis of Arabidopsis WDPs

The representative species of algae, mosses, monocotyledonous and dicotyledonous plants were selected for phylogenetic analysis, including Brassica oleracea, Nicotiana tabacum, Oryza sativa, Physcomitrium patens, Chlamydomonas reinhardtii. The proteome data of WDPs from Uniprot-Protome was employed by TBtools Blast in the search for local similarities and evolutionary relationships between Arabidopsis and the aforementioned species.

2.3. Physicochemical Properties Analysis of WDPs in Arabidopsis

A total of 56 structural proteins in Arabidopsis genome were randomly selected from Uniprot, and TBtools The Protein Paramter Calc (ProtParam-based) function was used to calculate the theoretical isoelectric point (PI), Instability index(II), Grand Average of Hydrophathicity (GRAVY) and Aliphatic Index (AI) of proteins. Unpaired t test -Mann-Whitne test (unpaired, two-tail) statistical calculations and plots were performed uniformly in Prism (GraphPad Prism 9.5.0 (730)).

2.4. Calculation of FCR and NCPR Values of WDPs

CIDER (https://157.245.85.131:8000/CIDER/) was directly used to calculate the value of FCR and NCPR according to the protein sequences [41]. Positively charged amino acids include arginine (Arg,R) and lysine (Lys, K), negatively charged amino acids include aspartic acid (Asp, D) and glutamic acid(Glu, E). According to the calculation method provided by Rahul K Daset al. [42], N: total number of amino acid residues in the sequence; N+, N-: positively charged, negatively charged residual base; f+ = N+ / N , f- = N-/ N .

FCR = (f+ + f-); NCPR = | (f+ − f-) |

2.5. Analysis of Expression Patterns and Function Prediction of WDPs

The transcriptome expression data of WDPs was retrieved from TAIR Arabidopsis eFP Browser (https://www.arabidopsis.org/), selecting Abiotic Stress. The data was obtained from AtGenExpress, provided by Kilian using Affymetrix ATH1 microarray [43]. We selected sample data from roots and shoots treated for 0 (control), 3, 6, 12 and 24 h, under cold treatment (4 °C), drought treatment (15-min dry air), salt treatment (150 mM NaCl) and oxidative stress treatment (300 mM D-Mannitol) respectively. Transcripts per million values for these WDPs were log2 transformed and the heatmap was constructed using TBtools.

The interaction data pertaining to WDPs was extracted from STRING (https://cn.string-db.org/) and imported into Cytoscape_v3.9.1, to construct related interaction network and perform GO enrichment analysis.

2.6. In Vitro Phase Separation Assay

Download the full-length CDS sequences of SIS and RAB18 from the TAIR website. Then, replace the rare codons in the sequences with the commonly used codons in E. coli. Next, use the bridge PCR method to fuse the target fragment with the GFP fragment. Subsequently, ligate the

 doi:10.20944/preprints202412.1512.v1

fragment into a prokaryotic expression vector, and then transform the vector into the Transetta(DE3) strain of E. coli. The bacterial culture is grown to an OD_{600} of 0.5-0.6, and IPTG is added to a final concentration of 1 mM. After that, induce at 20 °C, 180 rpm, for 16 h. After collecting the bacteria, lyse the cells with lysis buffer (25 mM Tris, 150 mM NaCl, adjusted to pH 7.4 with HCl), and enrich the protein by affinity chromatography, followed by elution with lysis buffer containing 200 mM imidazole. SIS-GFP and GFP protein was diluted to 10 μM, RAB18-GFP protein was diluted to 20 μM, in a buffer containing 25 mM Tris (pH 7.4), 150 mM NaCl and 10% PEG as a crowding agent. Protein solution (5 μL) was loaded onto a glass-bottom cell culture dish and imaged using a laser scanning confocal microscopy (ZEISS, LSM880). The images presented are of droplets settled on the glass coverslip. For fluorescence recovery after photobleaching (FRAP) assay, GFP was excited at 488 nm and detected at 498-530 nm, and fluorescence intensity detection was performed for every second.

3. Results

3.1. Finding WDPs from the Uniprot Proteme

In accordance with the aforementioned criteria, which stipulate those disordered regions occupy 90%~100% of its total length and structural information provided by AlphaFold and MobiDB, a comprehensive search of the Uniprot was conducted to find all entries related to WDPs. The database of such "wholly disordered" complexes contained 27 categories of 56 WDP members in Arabidopsis (Table 1), accounting for approximately 0.14% of all proteins.

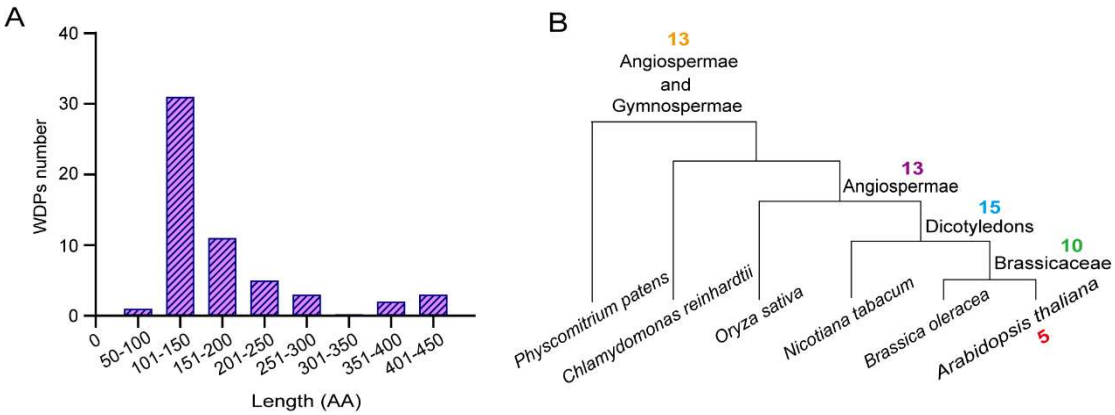
Table 1. Classification of WDPs in *Arabidopsis Thaliana.*

Protein ID	Gene ID	Gene Name	Description
F4I179	AT1G15840	/	Hypothetical protein
Q9M3G8	AT4G11430	/	Hydroxyproline-rich glycoprotein family protein
O82760	AT4G23110	*EFC*	Early flowering and curly leaves
Q8VXY1	AT5G28630	/	Glycine-rich protein
Q9FRL0	AT1G75190	/	Hypothetical protein
Q9LUC7	AT3G14670	/	Hypothetical protein
Q9C7W1	AT1G64370	*PARCL*	Phloem associated rna chaperone-like
Q8L9A7	AT4G27580	/	Phosphatidylinositol transfer SFH5-like protein
Q681J0	AT5G54095	/	Proteoglycan-like protein
Q9SCK5	AT3G49540	*T9C5.130*	Hypothetical protein
Q9LU05	AT5G44610	*PCAP2*	Plasma membrane-associated cation-binding protein 2
A0A1P8ASG6	AT1G04105	/	Hypothetical protein
Q9SXE9	AT1G62480	*CCaP1*	Vacuolar calcium-binding protein-like protein
Q8L7Z6	AT3G54680	/	Proteophosphoglycan-like protein
Q5XV49	AT5G05965	/	Cell wall RBR3-like protein
Q84WZ5	AT2G39855	/	Plant/protein
Q9SWI1	AT2G02950	*PKS1*	Protein phytochrome kinase substrate
Q9M9T4	AT1G14280	*PKS2*	
Q8GXS8	AT1G18810	*PKS3*	
Q9FYE2	AT5G04190	*PKS4*	
Q9XI29	AT1G15400	*MASS2*	MAPK substrates in the stomatal
Q9SSC1	AT1G80180	*MASS1*	
Q3E9A8	AT5G20100	*MASS3*	
Q9LZM9	AT5G02020	*SIS*	Salt induced serine rich
Q9STG0	AT3G46880	*T6H20.90*	Hypothetical protein
Q9FGU5	AT5G59080	/	Hypothetical protein
Q1G3N4	AT3G55646	/	Tprxl
Q9LSN1	AT3G17160	/	Hypothetical protein
Q9C7Y9	AT1G47970	/	Nucleolin

 doi:10.20944/preprints202412.1512.v1

O04254	AT4G02140	/	Hypothetical protein
Q8GYJ0	AT4G22320	*BCL7A*	BCL-domain homolog
Q9FKA5	AT5G39570	*PLDrp1*	PLD regulated protein
Q9LJV8	AT3G29075	/	Glycine-rich protein
Q9FM74	AT5G55640	*MDF20.8*	Na-translocating NADH-quinone reductase subunit A
Q9LF22	AT5G15600	*SP1L4*	Protein SPIRAL1-like
B3H4F1	AT1G26355	*SP1L1*	
Q9LE54	AT1G69230	*SP1L2*	
Q9SJW3	AT2G03680	*SPR1*	
Q8LGD1	AT4G23496	*SP1L5*	
Q9S7P8	AT3G02180	*SP1L3*	
Q9FL02	AT5G66780	*MUD21.2*	Late embryogenesis abundant protein
P30185	AT5G66400	*RAB18*	Dehydrin protein family
P42758	AT3G50970	*Xero 2*	Dehydrin protein family
Q9M2Q5	AT3G57930	*T10K17.140*	Rho gtpase-activating gaco-like protein
O48526	AT2G42190	*T24P15.10*	
B6IDH8	AT1G58460	*SOFL6*	Hypothetical protein
Q67YG7	AT1G26210	*SOFL1*	Protein SOB FIVE-LIKE
Q9CA45	AT1G68870	*SOFL2*	Protein SOB FIVE-LIKE
F4J6N7	AT3G30580	*SOFL3*	Hypothetical protein
Q9FKQ9	AT5G38790	*SOFL4*	Hypothetical protein
Q8L9K4	AT4G33800	*SOFL5*	Hypothetical protein
Q9LEZ1	AT1G58460	*SOB5*	Hypothetical protein
Q9FH00	AT5G42290	/	Transcription activator-like protein
Q9LPW6	AT1G12830	/	Nucleolin
F4I7D8	AT1G11125	/	Hypothetical protein
O22729	AT1G61170	/	Hypothetical protein

We firstly found that a small part of WDPs can be assigned to known protein families, including PKS (Protein PHYTOCHROME KINASE SUBSTRATE), LEA (Late embryogenesis abundant protein), SPIRAL1 (Protein SPIRAL1-like) and SOFL (Protein SOB FIVE-LIKE). Whereas, the rest WDPs could not to be classified into any known proteins; instead, these proteins were grouped together based on their great similarity and shared conserved areas (Fig S1). Possibly, these homologous WDPs, with similar amino acid substitutions, may adopt analogous folding patterns and perform the same or similar biochemical functions. The lengths of WDPs varied from 99 (SP1L5) to 442 amino acids (PKS2), with a prevalent distribution in the range of 100~200 amino acids (Fig 1A). Chain lengths and sequence features represent important factors determining the compaction degree of IDPs, which might correlate with the propensity to phase separate [44,45]. Such short lengths of these WDPs may lead to unsuccessful amplification of weak intramolecular interactions, accompanied by increasing probability of functional misfolding.



 doi:10.20944/preprints202412.1512.v1

Figure 1. Wholly disordered protein (WDPs) in *Arabidopsis thaliana*. A. Length distribution of WDPs in *Arabidopsis thaliana*. B. Distribution of homologous WDPs in the plants. The counts denote the number of homologous WDPs.

To understand the evolutionary relatedness of WDPs in plant lineages, we conducted a synteny analysis of the WDPs among *Arabidopsis thaliana*, *Brassica oleracea*, *Nicotiana tabacum*, *Oryza sativa*, *Physcomitrium patens* and *Chlamydomonas reinhardtii*. The results were classified and displayed based on the presence of homologous protein clusters in various plants (Fig 1B, Table S1). Fig 1B presented that there were five specific WDPs (F4I179, Q9M3G8, EFC, A0A1P8ASG6, and CCaP1) in *Arabidopsis Thaliana*, 10 WDPs in Brassicaceae, 15 WDPs in Dicotyledons and 13 WDPs in Angiospermae. It's worth noting that Angiospermae and Gymnospermae shared 13 common WDPs. WDPs have several sequence repeats and generally evolve at a comparatively faster rate, which underlines that the repetitive segments increase with organism complexity and may be shaped by intense evolutionary activities [46,47]. Species-specific WDPs may emerge early and subsequently be retained over the process of evolution and species differentiation.

3.2. Comparison of Physicochemical Properties Between WDPs and Structural Proteins

To gain deeper insight into specific features of WDPs and structural proteins, these WDPs were systematically compared with fifty-six ordered proteins with 100~600 amino acids (Table S2), which were randomly selected from UniProt Proteome. We first assessed four chosen features, the Theoretical pI (PI), Grand Average of Hydropathicity (GRAVY), Instability Index (II), and Aliphatic Index (AI).

The majority of WDPs were found to be weakly acidic or neutral, and their distribution of PI was more polarized (Fig 2A). The average PI values of WDPs were 6.770, with a mean PI of 6.380. Structural proteins exhibited a smaller PI variation range, from 5 to 7, with mean and average PI values of 7.700 and 7.400, respectively. The median GRAVY value of structural proteins was -0.244 and the mean of WDPs was -0.232 (Fig 2B). The GRAVY value indicates the protein-water interactions, and a negative GRAVY value implies a hydrophilic protein [48]. GRAVY values of WDPs were found to be negative, and the average value was -1.175, aligning with WDPs' hydrophilicity. II is one of the primary methods for predicting in-vivo protein stability based on protein structures [49]. The higher the index, the more unstable the protein is. In this analysis, the median II value of structural proteins was 39.87 and the mean value of WDPs was 40.636, demonstrating that the structural proteins were more stable (Fig 2C). The calculations categorized almost all WDPs as unstable proteins, with an II of above 40, except for XERO2, whose II was -0.99. This consistent with previous studies that XERO2, a dehydrin, might rely on conserved K-segments motifs to interact with specific targets, different from other WDPs that may require folding to be functional [50]. Relatively, XERO2 could even retain its disordered state at severe loss of water [51]. In terms of AI values, WDPs have lower values of AI compared with other structural proteins. Higher AI values are synonymous with higher thermostability [52], demonstrating that WDPs were less thermostable and poor in resistance of high temperatures (Fig 2D). Under a narrow range of temperature conditions, the WDPs were stable. Based on the above observations from theoretical studies, WDPs' overall features were different from structural proteins.

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 18 December 2024 doi:10.20944/preprints202412.1512.v1

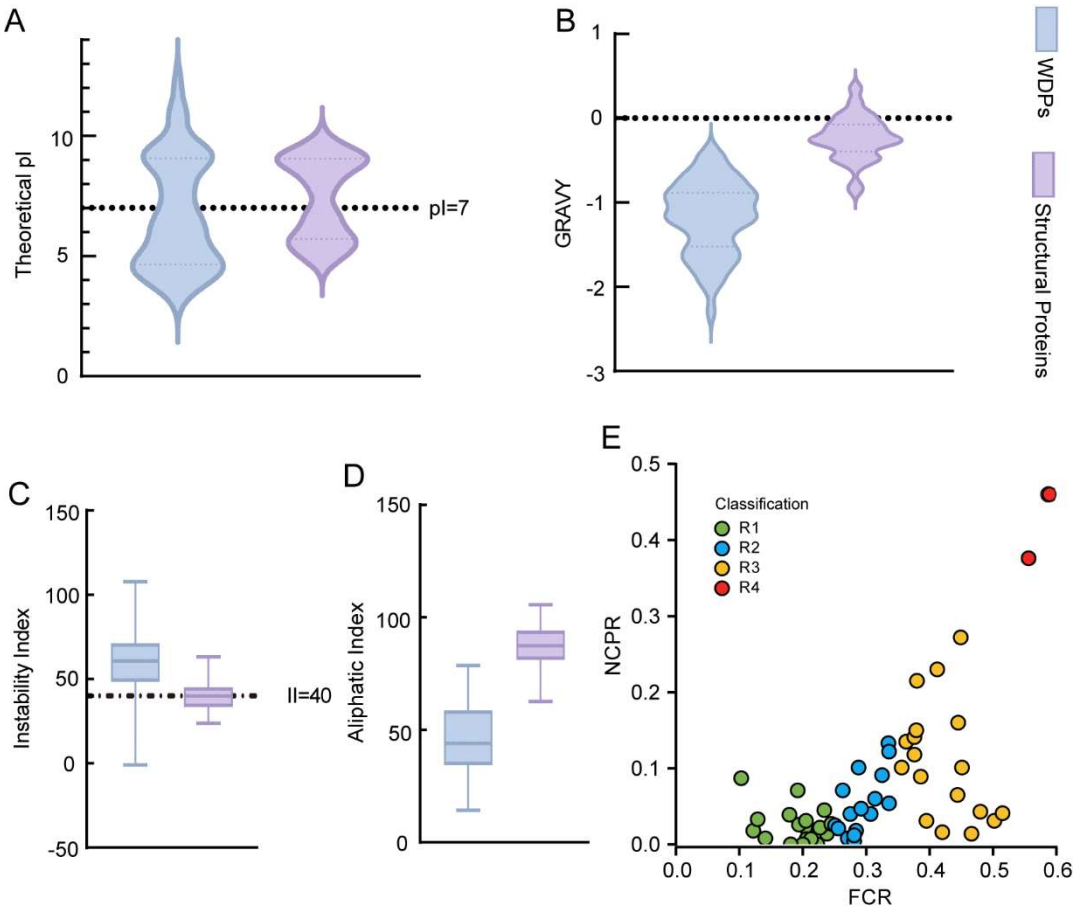


Figure 2. Comparison of physicochemical properties between the WDPs and 56 structural proteins. A. Comparison of average PI. B. Comparison of Total average hydrophilicity (GRAVY). C. Comparison of Instability index (II). D. Comparison of Aliphatic Index. E. Classification of Arabidopsis WDPs based on FCR and NCPR values. R1: FCR < 0.25 & NCPR < 0.25, R2: 0.25 ≤ FCR ≤ 0.35 & 0.25 ≤ NCPR ≤ 0.35, R3: FCR > 0.35 & NCPR ≤ 0.30, R4 & R5: FCR > 0.35 & NCPR > 0.30.

IDPs are highly dynamic and contain hydrophilic residues such as charged and polar amino acids [53,54]. the FCR (fraction of charged residue) and NCPR (net charge per residue) values of IDPs contribute as determinants of corresponding conformation, leading to the classification of IDPs into five different classes [55]. We calculated FCR and NCPR values of these WDPs and adopted the functional classification scheme (Fig 2E, Table S3). Interestingly, the WDPs were distributed at roughly similar levels in these categories except for R4, which had only two proteins, and R5, which had no distribution. Fig 4 showed that twenty WDPs belonged to R1, which corresponds to proteins with globule or tadpole-like conformations [56]. Nineteen WDPs belonged to R3, where proteins tend to have coil-like, hairpin-like or mixed conformations [57]. There were fifteen WDPs in R2, with conformations likely to be mixed between R1 and R3 [58]. What determines IDPs conformation is chain-solvent, chain-chain and solvent-solvent interactions, and the swollen coil prefers a good solvent [59-61]. Consequently, compact and roughly spherical WDPs promote favorable contacts with poor solvent, as the ions drawn into the core interact with the compact interior. These findings recapitulate the fact that flexible state-switching of WDPs will enhance their adaptability to outer environment, highlighting their pivotal positions in biological progress of all living cells.

3.3. Expression Profiles of the WDPs in Response to Abiotic Stresses

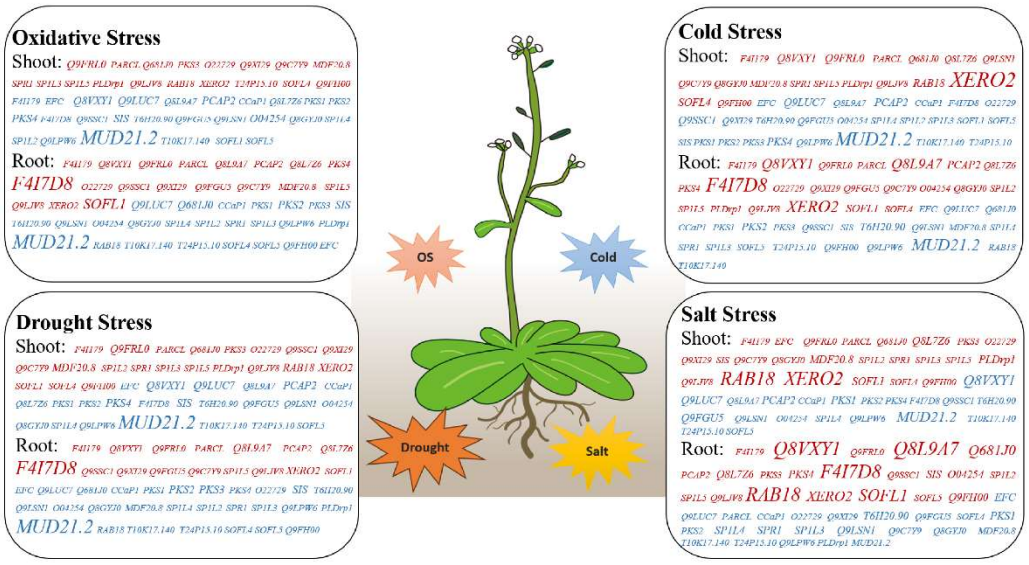
Adverse factors pose substantial challenges to geographic distributions and genetic diversity of plants as well as their morphological traits, and previous research manifests that IDPs are widely involved in the adaptation of plants to environmental stimuli [62]. The functional identification of WDPs in Arabidopsis require analysis of their expression in different tissues and response to various

 doi:10.20944/preprints202412.1512.v1

stresses. Therefore, we studied the expression levels of WDPs in shoots and roots under four different abiotic treatments. Since transcriptome data for twelve WDPs are not available, we herein focused on the remaining forty-four WDPs.

It was found that after exposure to oxidative stress, cold, drought, and salt treatment, about half of the WDPs showed increased expression, while the rest were observed to be inhibited. These were visualized in Fig 3A and Fig S2. Plants tend to adopt similar strategies to survive different stresses, such as stomatal closure in response to drought or salinity, and thus exhibit similar molecular mechanisms [63,64]. Certain WDPs displayed consistent expression patterns across different tissues, like the downregulation of CCaP1, PKS1, PKS2, T10K17.140, Q9LPW6, T6H20.90, SP1L4, Q9LUC7 and MUD21.2 as well as the upregulation of XERO2, Q9FRL0, Q9LJV8 and SP1L5 under all stress conditions (Fig 3A-B). As a dehydrin belonging to the highly hydrophilic LEA family ubiquitous in plants, XERO2 was induced in response to stress, indicating its role as a positive regulator in the process [65]. The consistent expression trends of these WDPs suggested their crucial roles in plant growth and development. While some genes showed negligible expression, others like RAB18 and XERO2 may have partially overlapping influences on various adaptive processes to stimuli, as shown in Fig S2.

A



B

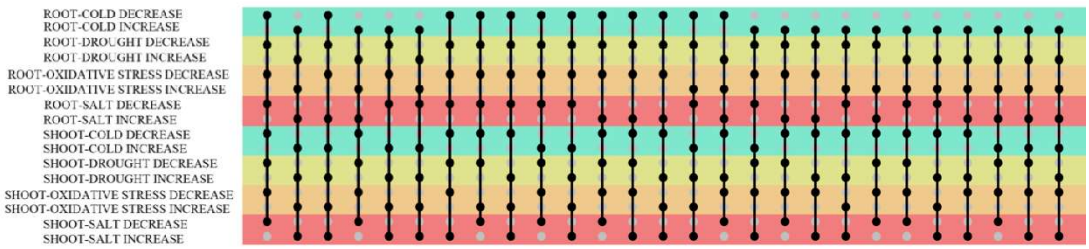


Figure 3. Expression patterns of the screened WDPs under abiotic treatment. A. Collection of up-regulated (red) and down-regulated (blue) WDPs in shoots and roots under abiotic stresses. Log2 (foldchange)≥0 was used as the threshold to determine the upregulation or downregulation. The larger the font, the larger the value of Log2 (foldchange). (b) Selection of related stress-responsive WDPs in shoots and roots.

 doi:10.20944/preprints202412.1512.v1

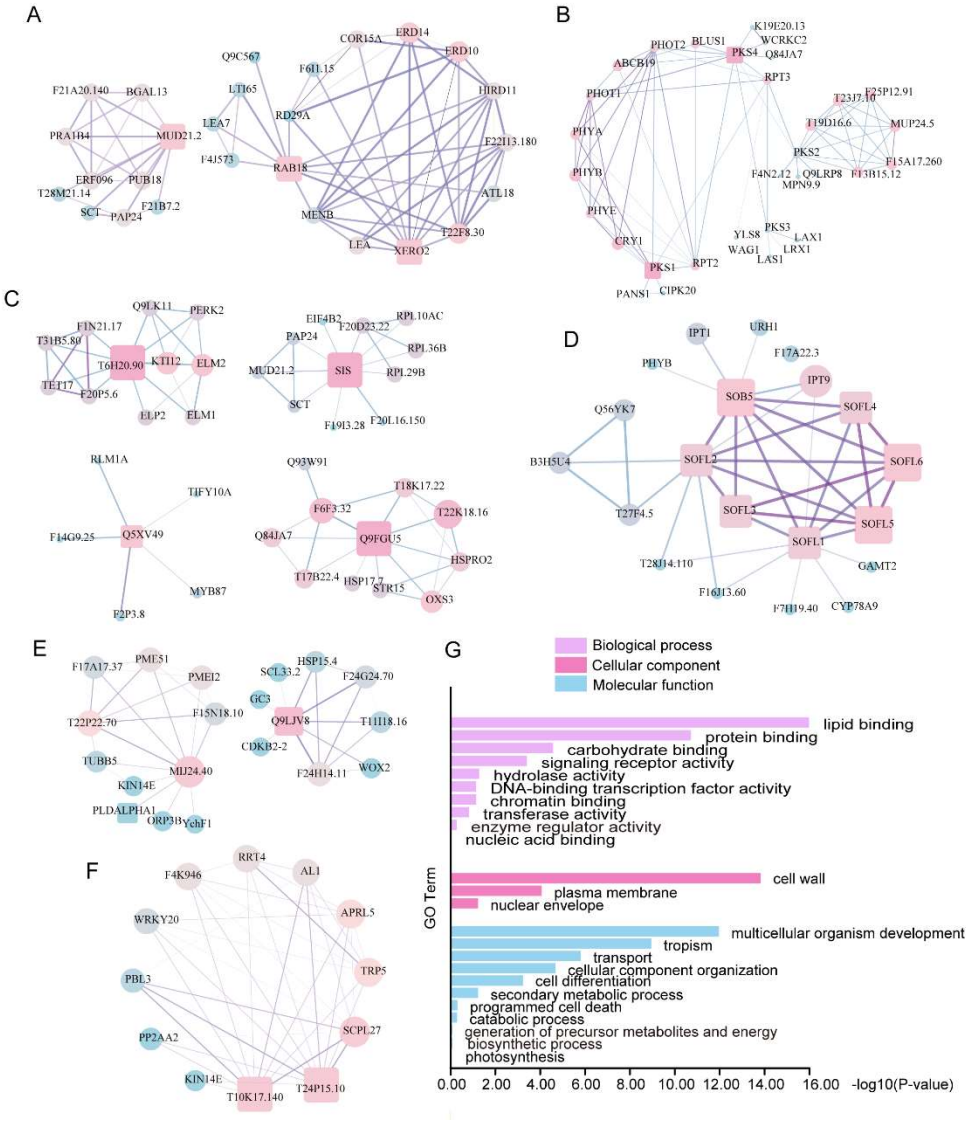


Figure 4. Protein-protein interaction networks of WDPs. A. LEA family proteins. B. PKS family proteins. C. SIS-like family proteins. D. SOB sndSOB5-LIKE family proteins. E. PLDALPHA1. F. Rho-GTPase activating gacO-like family proteins. G. GO enrichment analysis of WDPs and their interacted proteins.

The varying expression levels of some WDPs indicated their complex and diverse patterns regulation. The expression of SP1L2, SP1L5, PLDrp1, SOFL1 and SOFL4 in roots were induced under 4˚C, implying that these proteins were regulated to participate in the chilling tolerance process, thus improving resistance (Fig S2 A). However, the mRNA expression levels of some genes, like protein Q9SSC1 and MUD21.2, gradually decreased with time extension. In the shoots, MDF20.8, SPR1, SP1L5, PLDrp1 and RAB18 showed increased expression under cold stress (Fig S2 A). Under drought stress, the expression of PARCL, SP1L5, Q9C7Y9, Q9SSC1 and PCAP2, in roots exhibited increase (Fig S2 B). MUD21.2 decreased, conversely. It's possible that these proteins may be involved in essential processes like carbohydrate and energy metabolism to cope with specific drought conditions. In shoots after being treated for 24h, PARCL, PLDrp1, SP1L3, Q9SSC1, and RAB18 showed upregulation, while Q8VXY1 and MUD21.2 exhibited decrease (Fig S2 B). Plants usually synthesize ABA to induce the closure of stoma, and therefore prevent water loss by transpiration [66]. These facts imply some WDPs' possible roles as transcription factors relative to the ABA signaling pathway, such as PLDrp1. In addition, some WDPs exhibited inconsistent fluctuation throughout the stress-triggered process, indicating that they may have unique functions. These WDPs may be

 doi:10.20944/preprints202412.1512.v1

expressed in response to specific stimuli, as a result of specific conformational switchover to bind with target molecules.

3.4. Functional Prediction of WDPs

We further generated protein-protein interaction networks using the STRING database together with Cytoscape to decipher the physiological roles of these WDPs (Fig 4, Fig S3). Lack of the relevant data, 18 of screened WDPs had no corresponding statistics.

Gene Ontology (GO) functional enrichment analysis revealed that the rest 38 WDPs and their interacting proteins were mainly associated with multicellular organism development, tropism, transport cellular component organization and cell differentiation et.al (Fig 4G, Table S4). Some WDPs had been functionally characterized and classified into specific classes. The representative PKS protein family, whose members were WDPs including PKS1, PKS2, PKS3 and PKS4, is involved in negatively regulating the signal transduction of photopigments and affect the developmental morphology of roots and leaves [67,68]. SIS interacted with members of the RPL family, which plays an important role in plant immunity and functions in the auxin pathway to optimize Arabidopsis growth and defense [69]. The interaction relationship between SIS and MUD21.2 implied that SIS may act as a transcription factor to regulate the transcriptional process of genes related to stress tolerance [70,71]. Interestingly, SIS and MUD21.2 shared similar expression patterns, except for the opposite expression profiles in response to salt treatment (Fig 3A). This observation may indicate their opposing functions in variable salinity.

The correlation between CCaP1 and NOP10 was noted. NOP10 has emerged as a critical regulator in the modification of spliceosome small nuclear RNAs and stabilization of telomerase [72]. CCaP1, known as cytosolic Ca^{2+} binding protein in Arabidopsis, presented a common decrease faced with external stimulation (Fig 3A) [73]. Localized to the plasma membrane, CCaP1 interacts with the plasma membrane H-ATPases AHA1/AHA2 and functions as a regulator [74]. Taken together, the discovered CCaP1 and NOP10 interaction probably underlies the switchover mechanism related to stress response. PARCL showed upregulation under drought and oxidative stresses. Considering its colocalization with RNA in phase-separated condensates, PARCL probably forms liquid-liquid phase separations within cells in response to stimulus[75]. EFC might bind growth factors and regulate transcription, as its interacting proteins F21P8.10, SAC2, HDG3 belong to proteins widely involved in chromatin function and epigenetic modifications[76]. Q3E9A8 and Q9XI29, interacting with YDA, may act as a molecular switch to regulate the development of fertilized eggs in the MAPK pathway and regulate the development of stomata [77,78].

Overall, WDPs seem to be linked with a variety of cellular pathways enriched with transient interactions with diverse molecules, mostly participating in cell division and stress resistance. These may be common features of WDPs, as a result of the preference of specific amino acids and their plastic structures.

3.5. WDPs Tends to Undergo Liquid-Liquid Phase Separation

Previous studies have established that intrinsically disordered regions (IDRs) are a typically unique and common feature of eukaryotic proteins undergoing liquid-liquid phase separation (LLPS) [79]. In our study, we identified 56 wholly disordered proteins, which means there is a high possibility of undergoing phase separation. Thus, we selected two proteins, one is SIS from SIS-like family proteins, another is RAB18 from LEA family proteins, to test if they can undergo phase separation. We obtained proteins in *Escherichia coli* by optimizing the codons of SIS and RAB18 for better expression and then connecting the proteins to the GFP tag (Fig S4). Proteins that undergo phase separation form droplets in an appropriate buffer solution [79]. Using laser confocal microscopy, we observed that two proteins formed smooth and round droplets in a buffer (25 mM Tris pH 7.4,150 mM NaCl, 10% PEG), while GFP, used as a negative control, failed to form droplets under the same conditions (Fig 5A). Time-lapse microscopy also recorded that several small droplets merged into one large droplet in just over 10 seconds (Fig 5B, Video 1 and Video 2). In addition, we used FRAP to monitor the molecular dynamics of SIS-GFP or RAB18-GFP within the droplets (Fig

 doi:10.20944/preprints202412.1512.v1

5C and Fig 5E). As demonstrated by the fluorescence redistribution after photobleaching (FRAP) assay, there was a robust molecular exchange between the bleached region and the surrounding area for both SIS-GFP and RAB18-GFP (Video 3 and Video 4). Specifically, the recovery rate for SIS-GFP was 71.63% (Fig 5D), while for RAB18-GFP it was 89.34% (Fig 5F). In summary, our results indicated that SIS and RAB18 have the ability to form LLPS in vitro.

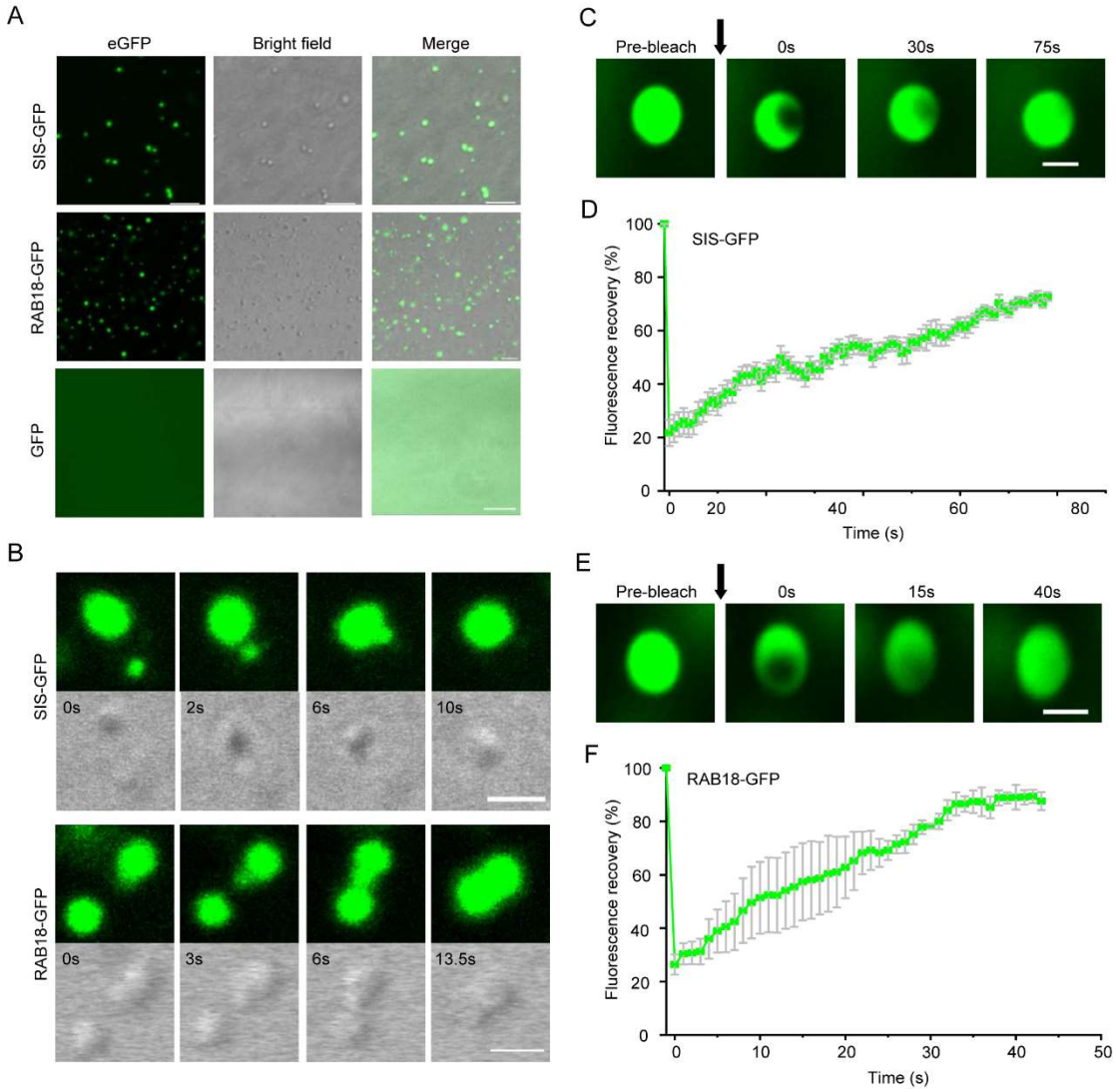


Figure 5. SIS of SIS-like family proteins and RAB18 of LEA family proteins exhibit liquid-liquid phase separation characteristics in vitro. A. SIS-GFP and RAB18-GFP form droplets in vitro. Scale bar, 5 μm. B. Time-lapse microscopy showing fusion dynamics of SIS-GFP and RAB-GFP droplets. Time 0 indicates the time of start recording. Scale bar, 2 μm. C and E. FRAP assay showing dynamic property of SIS-GFP or RAB18-GFP droplets. The black arrow indicate time of photobleaching. Scale bar, 5 μm. D and F. Quantification data of FRAP assays for SIS-GFP or RAB18-GFP droplets. Data are presented as mean ± SD (n = 10).

4. Discussion

In this study, we restricted the definition of WDPs with proteins of highly disordered degrees and screened all WDPs in Arabidopsis. The applied criteria simultaneously considered impacts of the amino acid composition and structural information. The set of WDPs included known plant-specific PKS family, homologous proteins related to MAPK pathway, LEA and SOFL family responding to various abiotic stresses, and other homologous proteins. Our selection results are in accordance with previous and novel IDP researches, such as the inclusion of RAB18, XERO2 and PARCL. Belonging to the disordered dehydrin, RAB18 and XERO2 share same repeat units and are

 doi:10.20944/preprints202412.1512.v1

strongly induced when exposed to cold [80]. In addition, SAXS experimental results also reveal a high degree of disorder in PARCL [81].

The physicochemical properties of WDPs were significantly different from those of structural proteins. They exhibited severe PI polarization, strong hydrophilic ability, and low fatty amino acid content. Based on FCR and NCPR values, most of WDPs belonged to the R1 and R3 categories, with a small portion falling into the R2 region. Their conformations were mostly linear molten globule-like conformations or hairpin structures. These findings suggest that proteins with less compact conformation tend to carry a relatively large number of charges, resulting in the presence of hydrophobic residues and charge polarization [82]. Meanwhile, disorders affect proteins differently when interacting with others. These characteristics enable WDPs to bind other molecules through strong electrostatic forces, recruiting other biomolecules to achieve phase separation, and thus achieve functional diversification and instantaneous action. With a high content of polar amino acids, WDPs may bind more water under physiological conditions, aiding plants in surviving in their external environment. In addition to their role in resistance pathways, WDPs may be involved in various biological processes, such as acting as a microtubule binding protein, regulating the cytoskeleton and cell division.

The evolution of species has been characterized by extensive gene loss [83]. As species complexity increased, so did WDPs. These WDPs have been preserved from early evolution and newly evolved with the increasing complexity of species, allowing them to adapt to environmental changes and achieve functional diversification. Such facts collectively explain why some WDPs have been preserved over the long course of evolution. Interestingly, some WDPs existed only in specific species, possibly resulting from plant differentiation from lower to higher complexity, and from simple to complex, as cells undergo different functional diversification.

The exact functions of WDPs have yet to be determined by experiments, but transcriptional and GO analysis indicated that these WDPs were widely involved in stress resistance and cell cycles. Mechanisms discovered so far may be relevant for the WDPs to carry out their duties. A small number of proteins, like F4I179 and Q9LUC7, lack of relevant available data and demand further characterization. Due to their flexible conformation, studying how these WDPs function in biological processes presents significant challenges. Phase separation currently appears to be a primary structural form through which intrinsically disordered proteins exert their functions. Hnisz et al. proposed a phase separation model in mammals that explains features of transcriptional control, including the formation of super-enhancers, the sensitivity of super-enhancers to perturbation, and their transcriptional bursting patterns [84]. Subsequently, it emerges that heterochromatin condensation in both animals and plants is orchestrated by phase separation, with the distinction lying in the involvement of species-specific proteins [85]. Moreover, proteins with intrinsically disordered regions, RNA-binding domains, and translation-related proteins dynamically forms stress granules (SGs) through LLPS in response to stress[86]. Recent research breakthroughs have uncovered that Arabidopsis De-Capping 5 (DCP5) detects hyperosmotic stress through molecular crowding and phase separation, forming DOSG stress granules that regulate the translatome and transcriptome to aid in plant osmotic stress adaptation. [87]. Among the DOSG, there were RNA-binding proteins and mRNA, as well as translation initiation factors. In our study, WDPs were found to be responsive to a spectrum of stresses, including cold, salt, oxidative, and drought stress (Figure 3). By analyzing the protein-protein interaction network of SIS, we observed its associations with numerous ribosomal proteins, such as F20D23.22, RPL36B, RPL29B, RPL10AC, as well as the translation initiation factor EIF4B2 (Figure 4C). In the case of RAB18, our investigation revealed significant interactions with various stress-related proteins and transcription factors, including LTI65, HIRD11, ATL18, RD29A, and LEA7 (Figure 4A). Notablely, we documented the in vitro phase separation of SIS and RAB18 (Figure 5). It is reasonable to speculate that SIS or RAB18 may also form stress granules through liquid-liquid phase separation, recruiting a suite of transcription and translation machinery, after sensing external stress. Also, other WDPs in Arabidopsis are potential stress sensors and warrant further investigation.

 doi:10.20944/preprints202412.1512.v1

5. Conclusions

In this paper, we firstly characterized the meaning of WDPs and develop a set of standards for screening Arabidopsis WDPs at genome level. Thorough bioinformatics study we clarify the functional characterization of these WDPs. In vitro experiments revealed that these proteins have great potential to function in a phase separation manner, providing clues for subsequent research on these proteins. Our results provide a valuable foundation for further investigation and functional analysis of WDPs, as well as for studies on the molecular mechanisms implementing various stages of plant development.

Supplementary Material: The following supporting information can be downloaded at the website of this paper posted on Preprints.org.

Author Contributions: Wenfen Long: conceptualization, data curation, software, investigation, methodology, and writing-original manuscript; **Liang Zhao**: data curation and investigation and methodology; **Huimin Yang**: data curation and investigation and methodology; **Xinyi Yang**: data curation and methodology; **Yulong Bai**: data curation and methodology; **Xiuhua Xue**: data curation and methodology; **Doudou Wang**: data curation, software, investigation, methodology, writing-review & editing; **Shengcheng Han**: funding acquisition, conceptualization, project administration, supervision, and writing-review & editing. All authors read and approved the manuscript.

Acknowledgments: This work was supported by the National Natural Science Foundation of China (Grant No. 31970723). The funders had no role in the study design, data collection and analysis, decision to publish, or preparation of the manuscript.

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

References

1. Lichtenthaler, F.W. 100 Years “Schlüssel-Schloss-Prinzip”: What Made Emil Fischer Use this Analogy? *Angewandte Chemie International Edition in English* **1995**, *33*, 2364-2374.
2. Uversky, V.N.; Dunker, A.K. Understanding protein non-folding. *Biochim Biophys Acta* **2010**, *1804*, 1231-1264.
3. Uversky, V.N. Natively unfolded proteins: a point where biology waits for physics. *Protein. Sci.* **2002**, *11*, 739-756.
4. Dyson, H.J.; Wright, P.E. Intrinsically unstructured proteins and their functions. *Nat. Rev. Mol. Cell Biol.* **2005**, *6*, 197-208.
5. Eliezer, D. Biophysical characterization of intrinsically disordered proteins. *Curr. Opin. Struct. Biol.* **2009**, *19*, 23-30.
6. Jensen, M.R.; Zweckstetter, M.; Huang, J.R.; Blackledge, M. Exploring free-energy landscapes of intrinsically disordered proteins at atomic resolution using NMR spectroscopy. *Chem. Rev.* **2014**, *114*, 6632-6660.
7. Wright, P.E.; Dyson, H.J. Intrinsically unstructured proteins: re-assessing the protein structure-function paradigm. *J. Mol. Biol.* **1999**, *293*, 321-331.
8. Nussinov, R.; Liu, Y.; Zhang, W.; Jang, H. Protein conformational ensembles in function: roles and mechanisms. *RSC Chem. Biol.* **2023**, *4*, 850-864.
9. Wright, P.E.; Dyson, H.J. Intrinsically disordered proteins in cellular signalling and regulation. *Nat. Rev. Mol. Cell Biol.* **2015**, *16*, 18-29.
10. Pajkos, M.; Erdős, G.; Dosztányi, Z. The Origin of Discrepancies between Predictions and Annotations in Intrinsically Disordered Proteins. *Biomolecules* **2023**, *13*, 1442.
11. Bhattarai, A.; Emerson, I.A. Dynamic conformational flexibility and molecular interactions of intrinsically disordered proteins. *J. Biosci.* **2020**, *45*.
12. Feng, Z.; Chen, X.; Wu, X.; Zhang, M. Formation of biological condensates via phase separation: Characteristics, analytical methods, and physiological implications. *Journal of Biological Chemistry* **2019**, *294*, 14823-14835.

13. Bondos, S.E.; Dunker, A.K.; Uversky, V.N. Intrinsically disordered proteins play diverse roles in cell signaling. *Cell Commun. Signal.* **2022**, *20*, 20.
14. Galea, C.A.; High, A.A.; Obenauer, J.C.; Mishra, A.; Park, C.; Punta, M.; Schlessinger, A.; Ma, J.; Rost, B.; Slaughter, C.A.; Kriwacki, R.W. Large-Scale Analysis of Thermostable, Mammalian Proteins Provides Insights into the Intrinsically Disordered Proteome. *J. Proteome Res.* **2009**, *8*, 211-226.
15. Uversky, V.N. The Mysterious Unfoldome: Structureless, Underappreciated, Yet Vital Part of any Given Proteome. *Journal of Biomedicine and Biotechnology* **2010**, *2010*, 1-14.
16. Poudyal, M.; Patel, K.; Gadhe, L.; Sawner, A.S.; Kadu, P.; Datta, D.; Mukherjee, S.; Ray, S.; Navalkar, A.; Maiti, S.; Chatterjee, D.; Devi, J.; Bera, R.; Gahlot, N.; Joseph, J.; Padinhateeri, R.; Maji, S.K. Intermolecular interactions underlie protein/peptide phase separation irrespective of sequence and structure at crowded milieu. *Nat. Commun.* **2023**, *14*, 6199.
17. Biesaga, M.; Frigole-Vivas, M.; Salvatella, X. Intrinsically disordered proteins and biomolecular condensates as drug targets. *Curr. Opin. Chem. Biol.* **2021**, *62*, 90-100.
18. Mitrea, D.M.; Mittasch, M.; Gomes, B.F.; Klein, I.A.; Murcko, M.A. Modulating biomolecular condensates: a novel approach to drug discovery. *Nat. Rev. Drug Discov.* **2022**, *21*, 841-862.
19. Zhang, F.; Biswas, M.; Massah, S.; Lee, J.; Lingadahalli, S.; Wong, S.; Wells, C.; Foo, J.; Khan, N.; Morin, H.; Saxena, N.; Kung, S.; Sun, B.; Parra, N.A.; Sanchez, C.; Chan, N.; Ung, L.; Altintas, U.B.; Bui, J.M.; Wang, Y.; Fazli, L.; Oo, H.Z.; Rennie, P.S.; Lack, N.A.; Cherkasov, A.; Gleave, M.E.; Gsponer, J.; Lallous, N. Dynamic phase separation of the androgen receptor and its coactivators key to regulate gene expression. *Nucleic. Acids. Res.* **2023**, *51*, 99-116.
20. Majewski, J.; Jones, E.M.; Vander, Z.C.; Biernat, J.; Mandelkow, E.; Chi, E.Y. Lipid membrane templated misfolding and self-assembly of intrinsically disordered tau protein. *Sci. Rep.* **2020**, *10*, 13324.
21. Murray, A.; Yan, L.; Gibson, J.M.; Liu, J.; Eliezer, D.; Lippens, G.; Zhang, F.; Linhardt, R.J.; Zhao, J.; Wang, C. Proline-Rich Region II (PRR2) Plays an Important Role in Tau–Glycan Interaction: an NMR Study. *Biomolecules* **2022**, *12*, 1573.
22. Hsiao, A. Protein Disorder in Plant Stress Adaptation: From Late Embryogenesis Abundant to Other Intrinsically Disordered Proteins. *International Journal of Molecular Sciences* **2024**, *25*, 1178.
23. Zavaliev, R.; Mohan, R.; Chen, T.; Dong, X. Formation of NPR1 Condensates Promotes Cell Survival during the Plant Immune Response. *Cell* **2020**, *182*, 1093-1108.
24. Hundertmark, M.; Hincha, D.K. LEA (late embryogenesis abundant) proteins and their encoding genes in Arabidopsis thaliana. *BMC Genomics* **2008**, *9*, 118.
25. Du D; Zhang, Q.; Cheng, T.; Pan, H.; Yang, W.; Sun, L. Genome-wide identification and analysis of late embryogenesis abundant (LEA) genes in Prunus mume. *Mol. Biol. Rep.* **2013**, *40*, 1937-1946.
26. Cao, J.; Li, X. Identification and phylogenetic analysis of late embryogenesis abundant proteins family in tomato (Solanum lycopersicum). *Planta* **2015**, *241*, 757-772.
27. Murray, D.T.; Kato, M.; Lin, Y.; Thurber, K.R.; Hung, I.; Mcknight, S.L.; Tycko, R. Structure of FUS Protein Fibrils and its Relevance to Self-Assembly and Phase Separation of Low-Complexity Domains. *Cell* **2017**, *171*, 615-627.
28. Hundertmark, M.; Hincha, D.K. LEA (Late Embryogenesis Abundant) proteins and their encoding genes in Arabidopsis thaliana. *BMC Genomics* **2008**, *9*, 118.
29. Artur, M.A.S.; Zhao, T.; Ligterink, W.; Schranz, E.; Hilhorst, H.W.M. Dissecting the Genomic Diversification of Late Embryogenesis Abundant (LEA) Protein Gene Families in Plants. *Genome Biol. Evol.* **2019**, *11*, 459-471.
30. Sivamani, E.; Bahieldin, A.; Wraith, J.M.; Al-Niemi, T.; Dyer, W.E.; Ho, T.D.; Qu, R. Improved biomass productivity and water use efficiency under water deficit conditions in transgenic wheat constitutively expressing the barley HVA1 gene. *Plant Sci.* **2000**, *155*, 1-9.
31. Xu, D.; Duan, X.; Wang, B.; Hong, B.; Ho, T.; Wu, R. Expression of a Late Embryogenesis Abundant Protein Gene, HVA1, from Barley Confers Tolerance to Water Deficit and Salt Stress in Transgenic Rice. *Plant Physiol.* **1996**, *110*, 249-257.
32. Liu, Y.; Wang, X.; Liu, B. A comprehensive review and comparison of existing computational methods for intrinsically disordered protein and region prediction. *Brief. Bioinform.* **2019**, *20*, 330-346.

 doi:10.20944/preprints202412.1512.v1

33. Necci, M.; Piovesan, D.; Tosatto, S. Critical assessment of protein intrinsic disorder prediction. *Nat. Methods* **2021**, *18*, 472-481.
34. Theillet, F.X.; Kalmar, L.; Tompa, P.; Han, K.H.; Selenko, P.; Dunker, A.K.; Daughdrill, G.W.; Uversky, V.N. The alphabet of intrinsic disorder: I. Act like a Pro: On the abundance and roles of proline residues in intrinsically disordered proteins. *Intrinsically Disord Proteins* **2013**, *1*, e24360.
35. Roesgaard, M.A.; Lundsgaard, J.E.; Newcombe, E.A.; Jacobsen, N.L.; Pesce, F.; Tranchant, E.E.; Lindemose, S.; Prestel, A.; Hartmann-Petersen, R.; Lindorff-Larsen, K.; Kragelund, B.B. Deciphering the Alphabet of Disorder-Glu and Asp Act Differently on Local but Not Global Properties. *Biomolecules* **2022**, *12*.
36. Quaglia, F.; Mészáros, B.; Salladini, E.; Hatos, A.; Pancsa, R.; Chemes, L.B.; Pajkos, M.; Lazar, T.; Peña-Díaz, S.; Santos, J.; ács, V.; Farahi, N.; Fichó, E.; Aspromonte, M.C.; Bassot, C.; Chasapi, A.; Davey, N.E.; Davidović, R.; Dobson, L.; Elofsson, A.; Erdős, G.; Gaudet, P.; Giglio, M.; Glavina, J.; Iserte, J.; Iglesias, V.; Kálmán, Z.; Lambrughi, M.; Leonardi, E.; Longhi, S.; Macedo-Ribeiro, S.; Maiani, E.; Marchetti, J.; Marino-Buslje, C.; Mészáros, A.; Monzon, A.M.; Minervini, G.; Nadendla, S.; Nilsson, J.F.; Novotný, M.; Ouzounis, C.A.; Palopoli, N.; Papaleo, E.; Pereira, P.J.B.; Pozzati, G.; Promponas, V.J.; Pujols, J.; Rocha, A.C.S.; Salas, M.; Sawicki, L.R.; Schad, E.; Shenoy, A.; Szaniszló, T.; Tsirigos, K.D.; Veljkovic, N.; Parisi, G.; Ventura, S.; Dosztányi, Z.; Tompa, P.; Tosatto, S.C.E.; Piovesan, D. DisProt in 2022: improved quality and accessibility of protein intrinsic disorder annotation. *Nucleic. Acids. Res.* **2022**, *50*, D480-D487.
37. Aspromonte, M.C.; Nugnes, M.V.; Quaglia, F.; Bouharoua, A.; Sagris, V.; Promponas, V.J.; Chasapi, A.; Fichó, E.; Balatti, G.E.; Parisi, G.; Buitrón, M.G.; Erdos, G.; Pajkos, M.; Dosztányi, Z.; Dobson, L.; Conte, A.D.; Clementel, D.; Salladini, E.; Leonardi, E.; Kordevani, F.; Ghafouri, H.; Ku, L.G.T.; Monzon, A.M.; Ferrari, C.; Kálmán, Z.; Nilsson, J.F.; Santos, J.; Pintado-Grima, C.; Ventura, S.; ács, V.; Pancsa, R.; Kulik, M.G.; Andrade-Navarro, M.A.; Pereira, P.J.B.; Longhi, S.; Mercier, P.L.; Bergier, J.; Tompa, P.; Lazar, T.; Tosatto, S.C.E.; Piovesan, D. DisProt in 2024: improving function annotation of intrinsically disordered proteins. *Nucleic. Acids. Res.* **2024**, *52*, D434-D441.
38. Wells, M.; Tidow, H.; Rutherford, T.J.; Markwick, P.; Jensen, M.R.; Mylonas, E.; Svergun, D.I.; Blackledge, M.; Fersht, A.R. Structure of tumor suppressor p53 and its intrinsically disordered N-terminal transactivation domain. *Proc. Natl. Acad. Sci. U. S. A.* **2008**, *105*, 5762-5767.
39. Wright, P.E.; Dyson, H.J. Intrinsically disordered proteins in cellular signalling and regulation. *Nat. Rev. Mol. Cell Biol.* **2015**, *16*, 18-29.
40. Hellinger, R.; Sigurdsson, A.; Wu, W.; Romanova, E.V.; Li, L.; Sweedler, J.V.; Sussmuth, R.D.; Gruber, C.W. Peptidomics. *Nat Rev Methods Primers* **2023**, *3*.
41. Holehouse, A.S.; Das, R.K.; Ahad, J.N.; Richardson, M.O.; Pappu, R.V. CIDER: Resources to Analyze Sequence-Ensemble Relationships of Intrinsically Disordered Proteins. *Biophys. J.* **2017**, *112*, 16-21.
42. Uversky, V.N. The most important thing is the tail: multitudinous functionalities of intrinsically disordered protein termini. *FEBS Lett.* **2013**, *587*, 1891-1901.
43. Kilian, J.; Whitehead, D.; Horak, J.; Wanke, D.; Weinl, S.; Batistic, O.; D'Angelo, C.; Bornberg-Bauer, E.; Kudla, J.; Harter, K. The AtGenExpress global stress expression data set: protocols, evaluation and model data analysis of UV-B light, drought and cold stress responses. *Plant. J.* **2007**, *50*, 347-363.
44. Uversky, V.N.; Santambrogio, C.; Brocca, S.; Grandori, R. Length-dependent compaction of intrinsically disordered proteins. *FEBS Lett.* **2012**, *586*, 70-73.
45. Martin, E.W.; Holehouse, A.S.; Peran, I.; Farag, M.; Incicco, J.J.; Bremer, A.; Grace, C.R.; Soranno, A.; Pappu, R.V.; Mittag, T. Valence and patterning of aromatic residues determine the phase behavior of prion-like domains. *Science* **2020**, *367*, 694-699.
46. Tompa, P. Intrinsically unstructured proteins evolve by repeat expansion. *Bioessays.* **2003**, *25*, 847-855.
47. Verbiest, M.A.; Delucchi, M.; Bilgin, S.T.; Anisimova, M. Beyond Microsatellite Instability: Intrinsic Disorder as a Potential Link Between Protein Short Tandem Repeats and Cancer. *Front Bioinform* **2021**, *1*, 685844.
48. Kyte, J.; Doolittle, R.F. A simple method for displaying the hydropathic character of a protein. *J. Mol. Biol.* **1982**, *157*, 105-132.
49. Gamage, D.G.; Gunaratne, A.; Periyannan, G.R.; Russell, T.G. Applicability of Instability Index for in vitro Protein Stability Prediction. *Protein Pept. Lett.* **2019**, *26*, 339-347.

50. Gupta, A.; Marzinek, J.K.; Jefferies, D.; Bond, P.J.; Harryson, P.; Wohland, T. The disordered plant dehydrin Lti30 protects the membrane during water-related stress by cross-linking lipids. *J. Biol. Chem.* **2019**, *294*, 6468-6482.
51. Mouillon, J.M.; Eriksson, S.K.; Harryson, P. Mimicking the plant cell interior under water stress by macromolecular crowding: disordered dehydrin proteins are highly resistant to structural collapse. *Plant Physiol.* **2008**, *148*, 1925-1937.
52. Kaur, A.; Pati, P.K.; Pati, A.M.; Nagpal, A.K. Physico-chemical characterization and topological analysis of pathogenesis-related proteins from Arabidopsis thaliana and Oryza sativa using in-silico approaches. *PLoS One* **2020**, *15*, e0239836.
53. Chang, R.; Chen, J.L.; Zhang, G.Y.; Li, Y.; Duan, H.Z.; Luo, S.Z.; Chen, Y.X. Intrinsically Disordered Protein Condensate-Modified Surface for Mitigation of Biofouling and Foreign Body Response. *J. Am. Chem. Soc.* **2022**, *144*, 12147-12157.
54. Campen, A.; Williams, R.M.; Brown, C.J.; Meng, J.; Uversky, V.N.; Dunker, A.K. TOP-IDP-scale: a new amino acid scale measuring propensity for intrinsic disorder. *Protein Pept. Lett.* **2008**, *15*, 956-963.
55. van der Lee, R.; Buljan, M.; Lang, B.; Weatheritt, R.J.; Daughdrill, G.W.; Dunker, A.K.; Fuxreiter, M.; Gough, J.; Gsponer, J.; Jones, D.T.; Kim, P.M.; Kriwacki, R.W.; Oldfield, C.J.; Pappu, R.V.; Tompa, P.; Uversky, V.N.; Wright, P.E.; Babu, M.M. Classification of intrinsically disordered regions and proteins. *Chem. Rev.* **2014**, *114*, 6589-6631.
56. Das, R.K.; Pappu, R.V. Conformations of intrinsically disordered proteins are influenced by linear sequence distributions of oppositely charged residues. *Proc. Natl. Acad. Sci. U. S. A.* **2013**, *110*, 13392-13397.
57. Bianchi, G.; Longhi, S.; Grandori, R.; Brocca, S. Relevance of Electrostatic Charges in Compactness, Aggregation, and Phase Separation of Intrinsically Disordered Proteins. *Int. J. Mol. Sci.* **2020**, *21*.
58. Das, R.K.; Pappu, R.V. Conformations of intrinsically disordered proteins are influenced by linear sequence distributions of oppositely charged residues. *Proc. Natl. Acad. Sci. U. S. A.* **2013**, *110*, 13392-13397.
59. Choi, U.B.; Sanabria, H.; Smirnova, T.; Bowen, M.E.; Weninger, K.R. Spontaneous Switching among Conformational Ensembles in Intrinsically Disordered Proteins. *Biomolecules* **2019**, *9*.
60. Mao, A.H.; Crick, S.L.; Vitalis, A.; Chicoine, C.L.; Pappu, R.V. Net charge per residue modulates conformational ensembles of intrinsically disordered proteins. *Proc. Natl. Acad. Sci. U. S. A.* **2010**, *107*, 8183-8188.
61. Olsen, J.G.; Teilum, K.; Kragelund, B.B. Behaviour of intrinsically disordered proteins in protein-protein complexes with an emphasis on fuzziness. *Cell. Mol. Life Sci.* **2017**, *74*, 3175-3183.
62. Fang, Y.; Xiong, L. General mechanisms of drought response and their application in drought resistance improvement in plants. *Cell. Mol. Life Sci.* **2015**, *72*, 673-689.
63. Agurla, S.; Gahir, S.; Munemasa, S.; Murata, Y.; Raghavendra, A.S. Mechanism of Stomatal Closure in Plants Exposed to Drought and Cold Stress. *Adv. Exp. Med. Biol.* **2018**, *1081*, 215-232.
64. Henry, C.; John, G.P.; Pan, R.; Bartlett, M.K.; Fletcher, L.R.; Scoffoni, C.; Sack, L. A stomatal safety-efficiency trade-off constrains responses to leaf dehydration. *Nat. Commun.* **2019**, *10*, 3398.
65. Shi, H.; Chen, Y.; Qian, Y.; Chan, Z. Low Temperature-Induced 30 (LTI30) positively regulates drought stress resistance in Arabidopsis: effect on abscisic acid sensitivity and hydrogen peroxide accumulation. *Front. Plant Sci.* **2015**, *6*, 893.
66. Hsu, P.K.; Dubeaux, G.; Takahashi, Y.; Schroeder, J.I. Signaling mechanisms in abscisic acid-mediated stomatal closure. *Plant. J.* **2021**, *105*, 307-321.
67. de Carbonnel, M.; Davis, P.; Roelfsema, M.R.; Inoue, S.; Schepens, I.; Lariguet, P.; Geisler, M.; Shimazaki, K.; Hangarter, R.; Fankhauser, C. The Arabidopsis PHYTOCHROME KINASE SUBSTRATE2 protein is a phototropin signaling element that regulates leaf flattening and leaf positioning. *Plant Physiol.* **2010**, *152*, 1391-1405.
68. Yang, Z.T.; Fan, S.X.; Wang, J.J.; An, Y.; Guo, Z.Q.; Li, K.; Liu, J.X. The plasma membrane-associated transcription factor NAC091 regulates unfolded protein response in Arabidopsis thaliana. *Plant Sci.* **2023**, *334*, 111777.

69. Xu, M.; Wang, X.; Liu, J.; Jia, A.; Xu, C.; Deng, X.W.; He, G. Natural variation in the transcription factor REPLUMLESS contributes to both disease resistance and plant growth in Arabidopsis. *Plant Commun.* **2022**, *3*, 100351.
70. Zhang, Y.; Fan, N.; Wen, W.; Liu, S.; Mo, X.; An, Y.; Zhou, P. Genome-wide identification and analysis of LEA_2 gene family in alfalfa (Medicago sativa L.) Under aluminum stress. *Front. Plant Sci.* **2022**, *13*, 976160.
71. Huang, Y.; Li, C.Y.; Qi, Y.; Park, S.; Gibson, S.I. SIS8, a putative mitogen-activated protein kinase kinase kinase, regulates sugar-resistant seedling development in Arabidopsis. *Plant. J.* **2014**, *77*, 577-588.
72. Bohnsack, M.T.; Sloan, K.E. Modifications in small nuclear RNAs and their roles in spliceosome assembly and function. *Biol. Chem.* **2018**, *399*, 1265-1276.
73. Ide, Y.; Tomioka, R.; Ouchi, Y.; Kamiya, T.; Maeshima, M. Transcriptional Induction of Two Genes for CCaPs, Novel Cytosolic Proteins, in Arabidopsis thaliana in the Dark. *Plant Cell Physiol.* **2007**, *48*, 54-65.
74. Wang, J.J.; Gao, J.; Li, W.; Liu, J.X. CCaP1/CCaP2/CCaP3 interact with plasma membrane H(+)-ATPases and promote thermo-responsive growth by regulating cell wall modification in Arabidopsis. *Plant Commun.* **2024**, 100880.
75. Ostendorp, A.; Ostendorp, S.; Zhou, Y.; Chaudron, Z.; Wolffram, L.; Rombi, K.; von Pein, L.; Falke, S.; Jeffries, C.M.; Svergun, D.I.; Betzel, C.; Morris, R.J.; Kragler, F.; Kehr, J. Intrinsically disordered plant protein PARCL colocalizes with RNA in phase-separated condensates whose formation can be regulated by mutating the PLD. *J. Biol. Chem.* **2022**, *298*, 102631.
76. Pignatta, D.; Novitzky, K.; Satyaki, P.; Gehring, M. A variably imprinted epiallele impacts seed development. *PLoS Genet.* **2018**, *14*, e1007469.
77. Bergmann, D.C.; Lukowitz, W.; Somerville, C.R. Stomatal development and pattern controlled by a MAPKK kinase. *Science* **2004**, *304*, 1494-1497.
78. Wang, H.; Ngwenyama, N.; Liu, Y.; Walker, J.C.; Zhang, S. Stomatal development and patterning are regulated by environmentally responsive mitogen-activated protein kinases in Arabidopsis. *Plant. Cell.* **2007**, *19*, 63-73.
79. Alex, S. H.; Birthe, B. K. The molecular basis for cellular function of intrinsically disordered protein regions. *Nat Rev Mol Cell Biol*. **2024**, *25*, 187-211.
80. Welin, B.V.; Olson, A.; Nylander, M.; Palva, E.T. Characterization and differential expression of dhn/lea/rab-like genes during cold acclimation and drought stress in Arabidopsis thaliana. *Plant Mol. Biol.* **1994**, *26*, 131-144.
81. Ostendorp, A.; Ostendorp, S.; Zhou, Y.; Chaudron, Z.; Wolffram, L.; Rombi, K.; von Pein, L.; Falke, S.; Jeffries, C.M.; Svergun, D.I.; Betzel, C.; Morris, R.J.; Kragler, F.; Kehr, J. Intrinsically disordered plant protein PARCL colocalizes with RNA in phase-separated condensates whose formation can be regulated by mutating the PLD. *J. Biol. Chem.* **2022**, *298*, 102631.
82. Abzalimov, R.R.; Frimpong, A.K.; Kaltashov, I.A. Detection and characterization of large-scale protein conformational transitions in solution using charge-state distribution analysis in ESI-MS. *Methods Mol Biol* **2012**, *896*, 365-373.
83. Harris, B.J.; Clark, J.W.; Schrempf, D.; Szollosi, G.J.; Donoghue, P.; Hetherington, A.M.; Williams, T.A. Divergent evolutionary trajectories of bryophytes and tracheophytes from a complex common ancestor of land plants. *Nat. Ecol. Evol.* **2022**, *6*, 1634-1643.
84. Hnisz, D.; Shrinivas, K.; Young, R.A.; Chakraborty, A.K.; Sharp, P.A. A phase separation model predicts key features of transcriptional control. *Cell.* **2017**, 3, 13-23.
85. Wang, N.; Liu, C. Implications of liquid–liquid phase separation in plant chromatin organization and transcriptional control. *Curr. Opin. Gemet. Dev*. **2019**, 4, 59-65.

 doi:10.20944/preprints202412.1512.v1

86. Israel M.L.; Nicolás, E. F.; Itzell, E.H.; Monika, C. Plant Stress Granules: Trends and Beyond. *Front Plant Sci*. **2021**. 8. 722643.
87. Wang, H.Y.; Yang, Q.H.; Zhang, D.; Lu, Y.Y.; Wang, Y.C.; Pan, Y.J.; Qiu, Y.P.; Men, Y.F.; Yan, W.; Xiao, Z.N.; Sun, R.X.; Li, W.Y.; Huang, H.D.; Guo, H.W. A cytoplasmic osmosensing mechanism mediatedby molecular crowding–sensitive DCP5. *Science*. **2024**, 11, 510.