Article

Integrative Identification of Chloroplast Metabolism-Related RETICULATA-RELATED Genes in Soybean

Qianli Dong [1,*,†] , Lu Niu [2,†], Xiyu Gong [1], Qianlong Xing [1], Jie Liang [1], Jun Lang [1], Tianya Wang [1,*] and Xiangdong Yang [2,*]

1   Key Laboratory of Molecular Epigenetics of Ministry of Education (MOE), Northeast Normal University, Changchun 130024, China; gongxiyu110@nenu.edu.cn (X.G.); xingql983@nenu.edu.cn (Q.X.); liangjie109@nenu.edu.cn (J.L.); langj@nenu.edu.cn (J.L.)
2   Jilin Provincial Key Laboratory of Agricultural Biotechnology, Jilin Academy of Agricultural Sciences, Changchun 130033, China; niulu@jaas.com.cn
*   Correspondence: dongql043@nenu.edu.cn (Q.D.); wangty309@nenu.edu.cn (T.W.); xdyang020918@jaas.com.cn (X.Y.)
†   These authors contributed equally to this work.

Abstract: As a globally important leguminous crop, soybean (*Glycine max* L.) serves as a vital source of edible oils and proteins for humans and livestock. Oils in leaves can help crops combat fungal infections, adapt to temperature changes via fatty acid modulation, and support resource recycling during leaf senescence. However, accumulating oils in leaves is a fundamental challenge due to the need to balance the inherently competing photosynthesis and fatty acid biosynthesis processes within chloroplasts. RETICULATA-RELATED (RER), known to regulate chloroplast function and plastid metabolism in Arabidopsis, plays an essential role in leaf development. Here, 14 non-redundant *GmRER* genes were identified in soybean and phylogenetically classified into four subclades. Most Arabidopsis *RER* genes were evolutionarily preserved as gene duplicates in soybean, except for *GmRER5* and *GmRER6*. RNA secondary structures spanning the coding sequences (CDSs), the 5′- and 3′- untranslated regions (UTRs) of *GmRERs*, displayed exceptional structural plasticity in CDSs, while exhibiting limited conservation in UTRs. In contrast, protein structures retained conserved folds, underscoring evolutionary constraints on functional domains despite transcriptional plasticity. Notably, GmRER4a and GmRER4b represented an exceptional case of high similarity in both protein and RNA structures. Expression profiling across fourteen tissues and three abiotic stress conditions revealed a dynamic shift in expression levels between leaf-predominant and root-enriched *GmRER* paralogs after stress treatments. A comparative transcriptome analysis of six soybean landraces further revealed transcriptional polymorphism in the *GmRER* family, which was associated with the expression patterns of lipid biosynthesis regulators. Our comprehensive characterization of GmRERs may offer potential targets for soybean breeding optimization in overall plant oil production.

Keywords: soybean; RETICULATA-RELATED genes; chloroplast development; RNA/ protein structural analysis; transcriptomics analysis

1. Introduction

Soybean (*Glycine max* L.), a globally vital legume crop, serves as a primary source of plant-derived protein and oil [1–4]. By now, advances in scientific technologies have enabled the identification of a growing number of genes associated with soybean seed

oil content, facilitating improvements in both oil yield and quality [5–11]. Although the global demand for vegetable oils from plants is continuously rising, the cultivation areas and productivity of conventional oil crops remain limited. Therefore, innovative strategies to produce oils from non-seed biomass, with a particular focus on leaf tissues, have been developed [12–15]. Previous studies have demonstrated that oils in soybean leaves fulfill multiple physiological roles. First, they function as subcellular factories producing antifungal compounds such as oxylipins, which protect against pathogens and support seedling survival [16–18]. Second, during leaf senescence, oil accumulation provides transient lipid reserves and precursor molecules critical for nutrient remobilization [16,18,19]. Additionally, leaf oils store bioactive secondary metabolites that aid in stress adaptation and defense mechanisms [17,19]. The fatty acid composition of leaf oil dynamically adjusts to environmental conditions; for instance, elevated levels of polyunsaturated fatty acids (e.g., linolenic acid) at high temperatures enhance thermotolerance [16].

As chloroplasts are central to fatty acid biosynthesis, photosynthesis, and stress signaling [20,21], understanding and exploring the regulatory mechanisms of chloroplast development and identifying key regulatory genes for leaf oil production in soybean are of great importance. The RETICULATA-RELATED (RER) gene family has emerged as a key player in chloroplast maintenance and plastid metabolism [22–24]. Over the past few decades, research endeavors have extensively documented the Arabidopsis RETICULATA-RELATED gene family, which consists of six genes encoding proteins with high sequence similarity to *RE* [22,25,26]. These genes, named *RER1* to *RER6*, are plant-specific and found in most sequenced plant genomes. The RER proteins can be categorized into three pairs (RE-RER1, RER2-RER3, and RER5-RER6) and one single protein, RER4, based on gene structure and amino acid sequence similarity. Most of the RER family proteins contain a conserved plant-specific domain, DUF3411, while some, like RER2, RER3, and RER4, also contain additional transmembrane helices, like TM3, consistent with their membrane localization [22]. RER5 and RER6 retain the conserved DUF3411 domain but are distinguished by an additional C-terminal DUF399 domain, a feature correlating with their localization in the thylakoid lumen [27,28]. The RER family can be divided into four functional modules with distinct roles [22]. For instance, RE and RER3 have redundant roles in downstream biosynthetic pathways of pyruvate, which are crucial for the generation of acetyl-CoA and de novo fatty acid synthesis in leaves; RER3 plays a specific role in embryogenesis in response to auxin; and RER5 and RER6 show functional redundancy in constituting the essential components of thylakoids [22,29].

Gene family expansion through duplication is a hallmark of plant genome evolution, enabling functional diversification and adaptation to environmental challenges [30–32]. Moreover, the interplay between genetic redundancy, regulatory plasticity, and protein structural conservation in polyploids remains poorly resolved [30,33]. Soybean, as a diploidized tetraploid, offers a unique system to study the lineage-specific expansions and subfunctionalization of family genes [33,34]. As described above, AtRER members are known to regulate chloroplast retrograde signaling and stress responses [22,23], the mechanisms by which duplicated RER paralogs partition roles across tissues and stresses are unclear. Thus, the evolutionary history, structural diversity, and functional roles of RER homologs in soybean, a paleopolyploid species with a complex genome shaped by whole-genome duplication events [34,35], need to be illustrated.

Here, we conducted genome-wide identification, structural characterization, and a comparative transcriptome analysis of the GmRER family in soybean. We investigated the evolutionary dynamics of GmRERs diversification, dissected their regulatory architectures, unraveled their spatiotemporal roles in stress adaptation, and displayed their natural transcriptional plasticity in six soybean landraces. Our study revealed how soybean

paleopolyploid history sculpted the GmRER family into functionally compartmentalized subclades, balancing conserved chloroplast metabolic roles with lineage-specific innovations in transcriptional regulation. The findings from our study provide comprehensive and systematic information for further studies of the effects of the GmRER gene family on soybean leaf development regulation, and shed light on dissecting the molecular mechanism of GmRER proteins; hence, they may contribute to the leaf oil breeding of soybean in the future.

2. Results

### 2.1. Identification of the GmRER Gene Family

To identify members of the RETICULATA-RELATED (RER) gene family in soybean (*Glycine max*), we first retrieved the protein sequences of Arabidopsis RE and six RER genes (AtRER1 to AtRER6) from TAIR (https://www.arabidopsis.org/ (accessed on 1 April 2025)), including their conserved DUF3411 domain (PFAM family PF11891). These sequences served as queries for bidirectional BLASTP searches (e-value $\leq 1 \times 10^{-10}$) against the soybean proteome (assembly Wm82.a4.v1 from Phytozome v13) [36], followed by HMMER validation using the PF11891 profile (cutoff: 1×10^{-50}) to ensure domain integrity. Redundant sequences were removed by manual curation and InterProScan confirmation. In total, 14 non-redundant *GmRER* genes were identified (Tables S1 and 1). Notably, duplication type analysis revealed that most GmRER paralogs originated from whole-genome duplication (WGD), whereas GmRER4c was derived from dispersed duplication, suggesting distinct evolutionary origins within the GmRER4 subfamily.

The phylogenetic architecture of soybean GmRERs was reconstructed through a comparative analysis of RER homologs from six species (*Arabidopsis thaliana*, *Glycine max*, *Oryza sativa*, *Zea mays*, *Gossypium hirsutum*, and *Triticum aestivum*). The resulting phylogeny delineated clear monocot–dicot divergence, with homologs from rice, maize, and wheat forming clades, distinct from those of Arabidopsis, soybean, and cotton (Figure 1a). Four major subclades were identified for the 14 soybean GmRER proteins (RE/RER1, RER2/RER3, RER4, and RER5/RER6). The Arabidopsis single-copy orthologs AtRE and AtRER1 corresponded to duplicated soybean GmREa/b and GmRER1a/b paralogs, suggesting post-speciation duplication events in soybean. A notable expansion occurred in soybean RER4 lineage, generating five soybean paralogs, GmRER4a–e. Intriguingly, three GmRER paralogs co-clustered with AtRER2 and AtRER3 in one subclade. Through reciprocal BLASTP analysis, the homology with AtRER2 was excluded, warranting their designation as GmRER3a–c. Although GmRER5 and GmRER6 were phylogenetically aligned together with AtRER5 and AtRER6, bidirectional best-hit BLASTP analysis specified their relationships, with GmRER5 best matching AtRER5, and GmRER6 best matching AtRER6.

Chromosomal localization indicated that *GmRER* genes were localized across seven chromosomes (Chr07, 09, 11, 12, 13, 16, 18), revealing duplication patterns tied to polyploidization (Figure 1b). Tandem duplications on chromosomes 09, 16, and 18 generated the *GmRER3* cluster (*GmRER3a–c*), while *GmRER4* paralogs (*GmRER4a–e*) were scattered on chromosomes 09, 11, and 12 via segmental duplications. Intriguingly, *GmRER5/6*, positioned as sister clades on chromosomes 11 and 18, mirrored the Arabidopsis *AtRER5/AtRER6* chromosome location (Chr.2/Chr.3). In general, the *GmRER* genes were unevenly distributed on chromosomes 07, 09, 13, and 16, but were relatively evenly distributed near the telomeres of chromosomes 11, 12, and 18 (Figure 1b).

Table 1. Orthology-based identification and integrating properties of GmRERs.

Gene	Arabidopsis Gene ID	Name	Type of Duplicattion	Location (Mb)	Protein Length (aa)	Protein pI	ProteinMW (Kda)	Subcellular Localization
RE	AT2G37860	GmREa	WGD	16.276–16.281	444	4.813	47.64	Chloroplast
		GmREb	WGD	45.830–45.836	446	5.688	47.94	Chloroplast
RER1	AT5G22790	GmRER1a	WGD	35.553–35.558	385	7.5	41.37	Chloroplast
		GmRER1b	WGD	40.846–40.851	387	8.021	41.53	Chloroplast
RER3	AT3G08640	GmRER3a	WGD	38.020–38.022	321	10.196	34.04	Chloroplast
		GmRER3b	WGD	49.235–49.238	350	8.476	36.94	Chloroplast
		GmRER3c	WGD	50.736–50.739	348	9.149	36.92	Chloroplast
RER4	AT5G12470	GmRER4a	WGD	1.403–1.408	365	8.461	38.8	Chloroplast
		GmRER4b	WGD	7.082–7.087	369	9.474	39.38	Chloroplast
		GmRER4c	Dispersed	12.077–12.090	432	9.722	46.96	Chloroplast
		GmRER4d	WGD	0.697–0.702	373	8.944	39.82	Chloroplast
		GmRER4e	WGD	45.552–45.556	377	9.156	40.16	Chloroplast
RER5	AT2G40400	GmRER5	WGD	1.682–1.690	748	8.777	82.64	Plasma membrane
RER6	AT3G56140	GmRER6	WGD	37.799–37.807	749	8.44	82.63	Plasma membrane
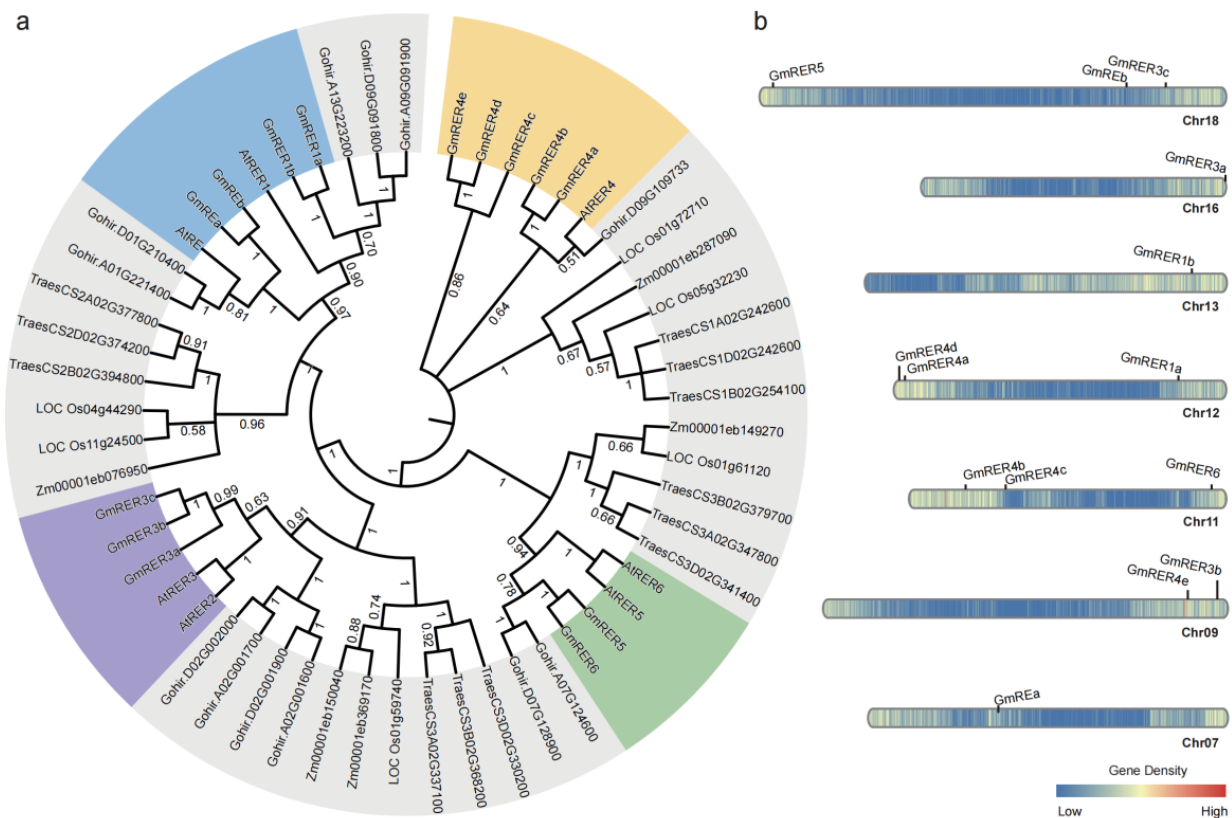
Figure 1. Phylogenetic analysis and genomic distribution of RER family genes in soybean and related species. (**a**) Rooted maximum-likelihood (ML) phylogenetic tree of RER proteins from six species: *Oryza sativa* (rice), *Gossypium hirsutum* (cotton), *Zea mays* (maize), *Triticum aestivum* (wheat), *Arabidopsis thaliana* (Arabidopsis), and *Glycine max* (soybean). Numbers at nodes indicate percentage of replicate trees (500 replicates) in which associated taxa are clustered together, representing branch support confidence. Soybean names are labeled as *GmRERs* and Arabidopsis names as *AtRERs*, while identifiers (IDs) from rice, cotton, maize, and wheat are retained as annotated in their respective genomes. Four major subclades containing GmRERs and AtRERs were represented in colors. (**b**) Chromosomal distribution and regional duplication of 14 *GmRER* genes in *Glycine max*. Locations of 14 *GmRER* genes on soybean chromosomes (07, 09, 11, 12, 13, 16, 18) are shown according to Wm82.a4.v1 of soybean genome annotation. Color bar represents gene density distribution on chromosomes.

2.2. Architectural Diversity of GmRERs in Gene Structures and Conserved Motifs

The structural plasticity and functional diversification of GmRERs were unveiled by an in-depth analysis of exon–intron architectures and conserved motifs. As shown in Figure 2a, the *GmRERs* exhibited subclade-specific structural patterns. Notably, *GmRER3* paralogs exhibited divergent exon-intron architectures. *GmRER3a* and *GmRER3c* were intronless, in accordance with their Arabidopsis ortholog *AtRER3* [22], while *GmRER3b* retained a single intron. Members of the *GmRER1* and *GmRER5/6* subclades retained conserved seven-exon structures, resembling their Arabidopsis counterparts [22], whereas the expanded *GmRER4* subclade diverged (Figure 2a). *GmRER4a/b/d/e* retained eight exons, but *GmRER4c* acquired a ninth exon via intron retention in the 3′ UTR, which may potentially influence transcript stability or alternative splicing. This gene structural divergence of *GmRERs* suggests an evolutionary interplay between genetic drift and selection during soybean domestication.
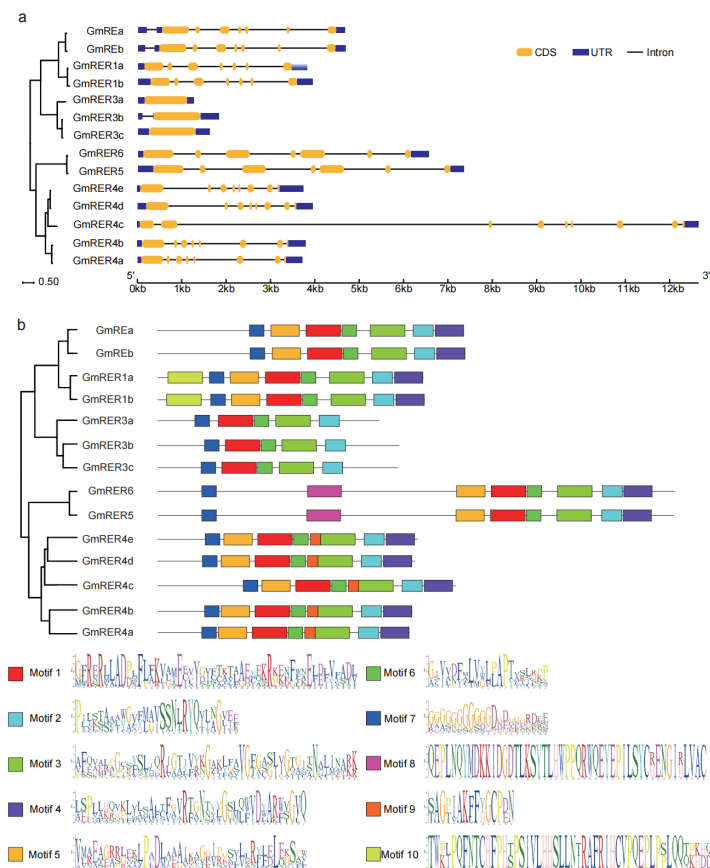
Figure 2. Gene structure and motif distribution of GmRERs. (**a**) The gene structures of *GmRER*s were described using the GSDS tool. The yellow rectangles represent the coding sequences (CDS), the blue rectangles indicate the untranslated regions (UTRs), and the straight lines denote the introns. (**b**) The motif distributions within GmRER proteins were identified using the MEME tool. The GmRER proteins are drawn to scale. The position of each colored block indicates the location of a motif with a matching sequence.

Based on MEME-based motif profiling, 10 conserved motifs (five universal, five subclade-specific) were identified (Figure 2b, Table S2). GmRER paralogs within a subclade shared similar motif content, which was the case for their phylogenetic relationships (Figure 1a). The distinct motif organization patterns across subclades of GmRERs suggested their probable functional divergence. Motifs 1–3 and 6–7 were universal, suggesting that their encoding domains are essential for functions of GmRERs (Figure 2b). Intriguingly, while GmRER3 paralogs formed a tightly clustered clade in the phylogenetic tree, they uniquely lacked motif 5, which was otherwise conserved in other subclades. Coupled with the intronless configuration of *GmRER3a* and *GmRER3c* in gene structures (Figure 2a), this implied that GmRER3 paralogs might have distinct regulatory strategies in transcription. Furthermore, GmRER4 paralogs uniquely retained motif 9, which was predicted to be an MYB-binding site, indicating they could be potential targets for MYB transcription factors.

2.3. RNA Secondary Structure and Protein Structure Prediction of GmRERs

With the purpose of understanding the potential mechanisms of GmRERs in post-transcriptional regulation and functional performance, we analyzed their RNA secondary structures and 3D protein structures. To increase the reliability of the prediction outcomes, the RNA secondary structures of 5′- and 3′- UTRs and CDS regions were analyzed separately (Figure S1). Then, we quantified the similarity of the RNA structures using RNAforester-based comparisons. The heatmaps in Figure 3a–c indicate that the RNA secondary structures of *GmRER* paralogs were remarkably different from each other. Among

these structures, 5′ and 3′ UTRs were slightly conserved across the *GmRER* family genes (Figure 3a,c), while CDS regions displayed distinct paralog-specific structural divergence (Figure 3b), suggesting selective pressure on post-transcriptional regulatory mechanisms. Specifically, *GmRER4* paralogs formed a clustered group with homologous RNA architectures (CDS region); *GmRER3* paralogs, despite lacking introns, exhibited exceptional CDS structural plasticity, implying evolutionary dynamics independent of coding sequence conservation (Figure 3b).
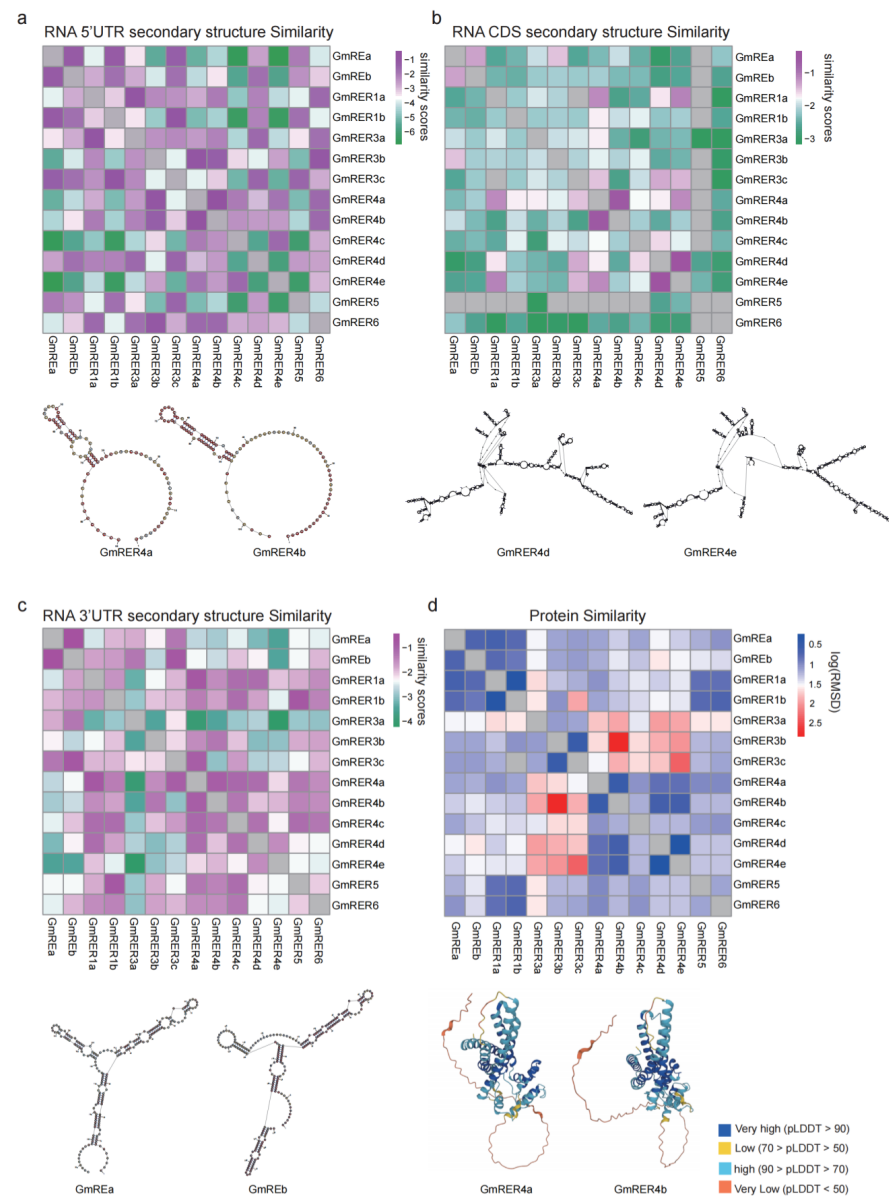
Figure 3. The structural similarity of GmRER mRNA and proteins. (**a–c**) The RNA secondary structure similarity of GmRER mRNA in the 5′ UTR (**a**), CDS (**b**), and 3′ UTR (**c**) regions. The values represent similarity scores calculated by RNAForester, with higher scores indicating greater similarity (purple) and lower scores indicating lower similarity (green). (**d**) The protein structure similarity of GmRERs, represented by log(RMSD) values calculated using PyMOL. Lower log(RMSD) values indicate higher similarity (blue), while higher values indicate lower similarity (red). Structure similarities which cannot be calculated are labeled in gray, and below the heatmaps are examples of some of the most conserved structure pairs.

Protein structures of GmRERs from the AlphaFold protein structure database were obtained (Figure S2), and the local accuracy of most of them had high confidence values (pLDDT value > 70, colored in blue). Also, we calculated the log(RMSD) values to compare

the similarity of protein structures using PyMOL. Unlike the RNA secondary structures, the protein structures were highly conserved across GmRER paralogs (Figures 3d and S2). For example, GmRER1 paralogs retained conserved α-helical topologies, despite lineage-specific variations in RNA structural organization. Likewise, GmRER4 paralogs with highly divergent RNA structures (CDS regions) maintained nearly identical β-sheet-dominated folds in their DUF3411 functional domains. Notably, GmRER4a and GmRER4b shared remarkably similar protein structures and RNA structures (Figure 3a,d), suggesting strong functional constraints across GmRER4 paralogs.

2.4. Modular Functionality of GmRER Paralogs via Subcellular Specialization and Lineage-Specific PPI Networks

For further information on the principal function for GmRER proteins, the physico-chemical properties and predicted subcellular localization of GmRERs are summarized in Table 1. Consistent with the phylogenetic relationships, most of the GmRER paralogs shared similar protein pI (isoelectric point), protein MW (molecular weight) and chloroplast localization, except for GmRER5/6. Then, we conducted a STRING-based protein–protein interaction (PPI) analysis to understand the role of GmRERs in the regulatory network of soybean growth and development (Table S3). As shown in Figure 4a, the PPIs of GmRERs can be divided into three clusters. Cluster I consisted of GmRE paralogs, GmRER1 paralogs, GmRER4 paralogs, and their interacting proteins. Cluster II was composed of GmRER3 paralogs and their interacting proteins. Clusters I and II shared some weak interactions with each other. However, the third cluster had no interaction with other GmRERs; it was formed of GmRER5 and GmRER6 along with their interacting proteins.
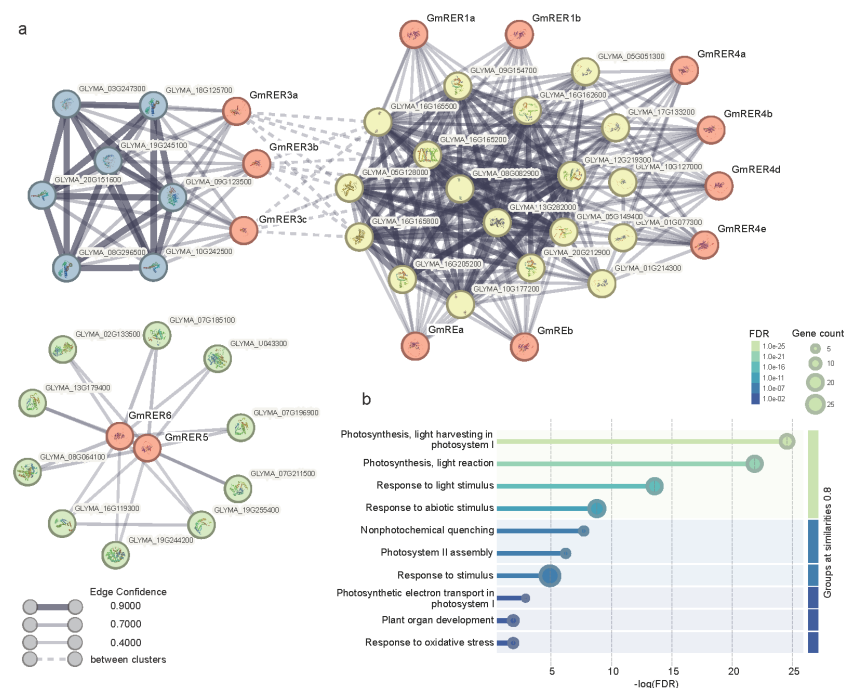
Figure 4. Protein–protein interactions of GmRERs and GO analysis of interacting proteins. (**a**) Protein–protein interaction networks of GmRERs were constructed based on STRING predictions. The nodes of GmRER proteins are labeled in red, and the nodes of other interacting proteins are labeled in yellow, blue, and green, according to the clusters they belong to. Proteins with known or predicted structures are specified in the nodes. Edge width and color represent the scores for edge confidence. (**b**) The lollipop plot displays the GO analysis results of GmRER interacting proteins. The x-axis represents the −log(FDR) values, showing the significance level, and the y-axis shows the significantly enriched GO terms.

Gene Ontology (GO) analysis of all of the interacting proteins illustrated that the top three significantly enriched GO terms were "photosynthesis, light harvesting in photosystem I", "Photosynthesis, light reaction", and "Response to light stimulus" (Figure 4b), which were in line with the previously reported function of RERs in Arabidopsis [22]. Besides photosynthesis-related GO terms, we found that GmRERs can also interact with proteins functionally enriched in "Response to stimulus", "Nonphytochemical quenching", "Plant organ development", and "Response to oxidative stress" (Figure 4b), which corroborated the well-documented regulatory roles of RERs in *Arabidopsis thaliana* [20,22,24].

Subcellular localizations of GmRER paralogs can explain their functional modules properly (Tables 1 and S3, Figure 4). Specifically, chloroplast-localized GmRE/RER1/RER4 paralogs interacted directly with photosynthesis-related proteins (such as Glyma.08G082900, Glyma.05G128000, and Glyma.16G165500). Notably, GmRER3 targeted chloroplasts and also interacted with transcriptional machinery proteins (such as Glyma.18G125700, Glyma.19G245100, and Glyma.10G242500), which may bridge the chloroplast–nuclear regulation critical for coordinating transcription in photosynthetic processes. However, plasma membrane-localized GmRER5/6 interacted with redox networks proteins (Glyma.07G185100, Glyma.19G244200, and Glyma.19G255400) separately, predicting their roles in oxidative stress mitigation.

2.5. Spatiotemporal Expression Dynamics Define Functional Classes of GmRERs

Chloroplasts, integral to numerous critical metabolic processes in plants [37,38], enable plants to adapt to environmental changes via retrograde signaling pathways between plastids and the nucleus [39–41]. Thus, we compared the expression profiles of the *GmRERs* across diverse tissues in various developmental stages and stress conditions, using three published transcriptome datasets from previous studies [42–44]. The expression patterns of the *GmRER* paralogs across diverse tissues and stress conditions delineated the functional dichotomy, which can be categorized into two distinct classes with opposing regulatory strategies (Figure 5). Class I genes, comprising *GmRER3a*, *GmRER4c-e*, *GmRER5*, and *GmRER6*, were predominantly expressed in photosynthetic tissues, such as young leaves and cotyledons (Figure 5a). In Figure 5b, after salt treatment, the expression levels of Class I members were rapidly down-regulated in soybean leaves, reaching minimal expression levels within 4 h. While in roots, Class I members maintained basal-level expression under salt stress. Interestingly, *GmREa/b* were clustered into the Class I expression pattern under salt stress, and they ectopically activated leaf-silencing expression (Figure 5a,b). In contrast with salt stress, drought elicited a delayed response from Class I members in leaves, with transcripts sharply declining after 5 days but surging during recovery (Figure 5c). Moreover, submergence stress uniquely induced sustained up-regulation of *GmRER5/6* in leaves and ectopic activation of *GmRER5/6* in roots (Figure 5c).

Class II genes, including *GmREa/b*, *GmRER1a/b*, *GmRER3b/c*, and *GmRER4a/b*, displayed inverse tissue expression patterns compared with those of Class I genes, and were constitutively expressed in seeds, pods, and roots (Figure 5a). According to Figure 5b,c, the expression of Class II genes was stress-induced. Salt stress moderately up-regulated the expression of *GmRER4b* in roots at 48 h, while drought stress triggered the root-specific induction of *GmRER3b/c* expression. Submergence disrupted the tissue-specific expression of *GmRER1a/b* and induced its ectopic activation in leaves (Figure 5c). In summary, the stress-induced reciprocal expression shifts between leaf-predominant and root-enriched *GmRER* paralogs underscore their opposing regulatory strategies for photosynthesis and energy metabolism.
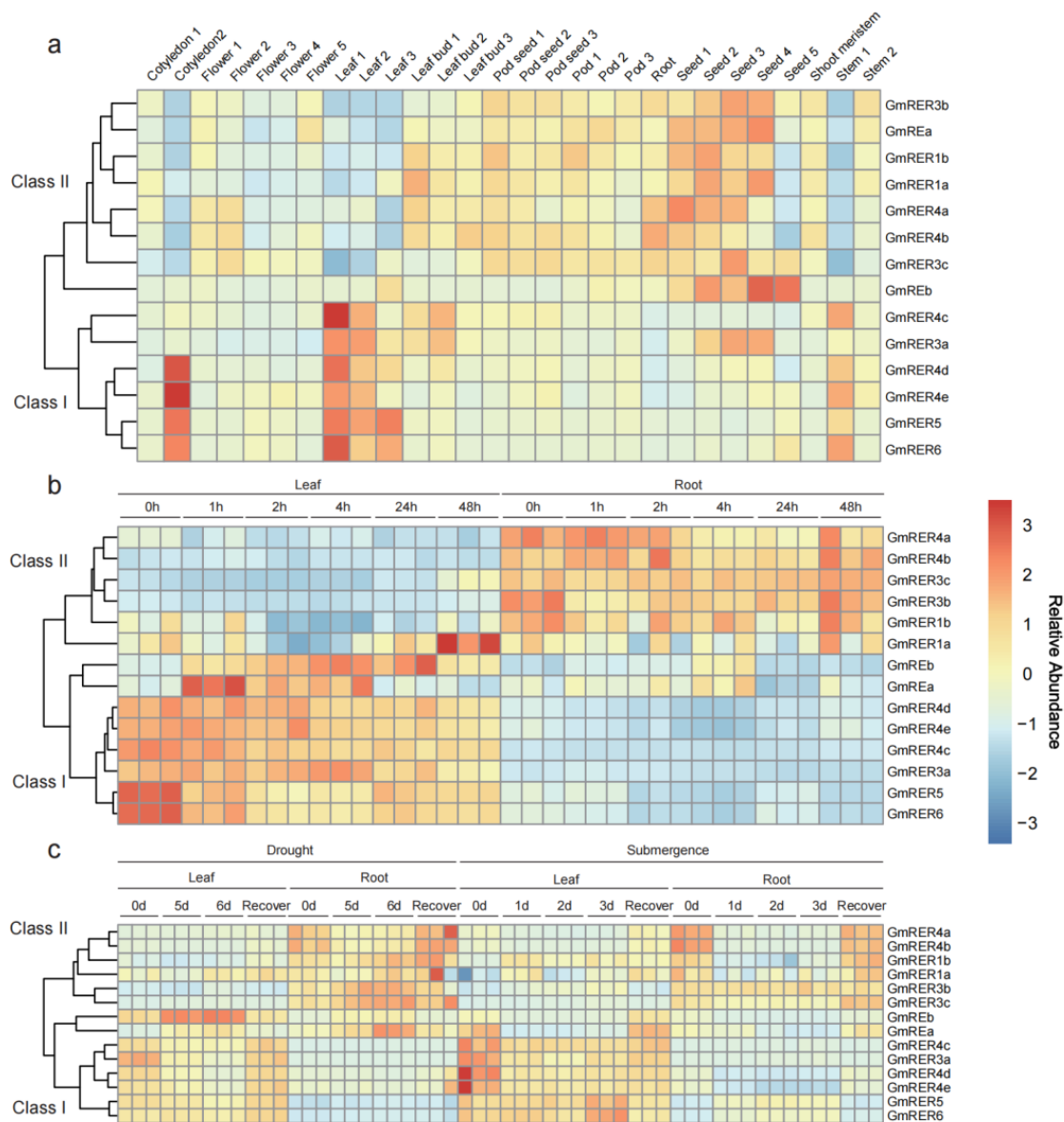
Figure 5. Expression patterns of *GmRERs* in various tissues, developmental stages, and stress conditions. (**a**) Expression of *GmRERs* in cotyledons, roots, stems, leaves, flowers, pods, and seeds at different developmental stages (PRJNA238493). (**b**) Expression of *GmRERs* in leaves and roots under salt stress at six time points (0, 1, 2, 4, 24, and 48 h) with three biological replicates (PRJNA246058). (**c**) Expression of *GmRERs* in leaves and roots under drought and submergence stress at multiple time points (drought stress: 0d, 5d, 6d, recovery; submergence stress: 0d, 1d, 2d, 3d, recovery) with three biological replicates (PRJNA574626). In heatmap, red indicates high expression, and blue indicates low expression.

2.6. Cis-Regulatory Modules Conveyed Roles of GmRERs in Stress and Development

As is well known, the expression patterns of genes are controlled by their promoters [45]. Thus, we performed an analysis of 2 kb promoter regions across *GmRER* family genes, and 52 categorized *cis*-regulatory elements were identified (Figure 6). Based on their previously reported functional annotations, these *cis*-elements were partitioned into three major classes, including plant growth/development (23 categories, 44.2%), abiotic/biotic stress (18 categories, 34.6%), and phytohormone-responsiveness (11 categories, 21.1%).

|	AAGAA-motif	ARE	as-1	CCAAT-box	DRE core	LTR	MBS	MYB	MYB recognition site	Myb-binding site	MYB-like sequence	STRE	TCA	TC-rich repeats	TGA-box	W box	WRE3	WUN-motif	ABRE	ABRE3a	ABRE4	CGTCA-motif	ERE	GARE-motif	P-box	TATC-box	TCA-element	TGACG-motif	TGA-element	3-AF1 binding site	A-box	ACE	AE-box	AT1-motif	Box 4	CAT-box	CCGTCC_motif	CCGTCC-box	chs-CMA1a	chs-CMA2a	circadian	GA-motif	GATA-motif	G-box	GCN4 motif	GT1-motif	HD-Zip 1	I-box	MRE	MYC	O2-site	Sp1
GmREa	5	2	1	0	0	0	0	2	0	1	0	2	2	0	0	0	0	4	0	0	1	2	0	0	0	0	0	1	0	0	0	3	0	3	0	0	0	0	0	0	0	0	2	4	0	0	0	0	0	10	0	0
GmREb	1	2	0	0	0	0	1	2	0	0	0	1	0	0	0	0	0	2	3	0	0	4	0	0	0	0	0	0	0	0	0	0	0	0	13	0	0	0	1	0	1	0	0	3	1	0	0	0	0	4	0	0
GmRER1a	3	3	0	0	0	4	0	5	0	1	2	0	0	1	0	0	0	1	0	0	0	0	1	0	0	0	0	0	0	0	0	0	2	1	0	0	0	0	0	0	1	0	0	2	0	0	0	0	0	7	0	0
GmRER1b	4	6	0	0	1	0	0	4	0	0	0	2	0	1	0	1	2	3	0	0	0	0	2	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	1	0	1	5	0	0	0	1	0	3	1	0	0	0
GmRER3a	2	2	0	0	0	0	2	6	0	0	2	1	1	1	0	0	0	2	0	0	0	0	2	0	0	0	0	1	0	0	3	0	1	1	1	2	0	0	4	1	0	1	8	0	0	0	0	0	0	0	0	0
GmRER3b	0	0	0	0	0	0	1	2	0	0	0	1	0	1	0	4	2	1	1	0	3	0	0	0	0	0	0	0	0	1	5	0	0	0	1	0	1	1	2	2	0	0	0	0	0	0	0	0	0	0	0	0
GmRER3c	1	3	3	0	0	1	0	2	0	0	1	0	2	1	0	0	0	3	0	0	3	3	0	0	0	0	3	1	0	0	9	2	0	0	1	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	1
GmRER4a	0	2	1	0	0	0	0	4	0	0	2	2	5	0	0	0	0	1	0	0	0	0	1	4	0	0	0	0	1	0	0	1	0	0	0	1	0	2	0	3	0	0	1	0	0	2	3	1	0	0	0	0
GmRER4b	2	2	0	1	0	0	0	3	1	0	1	1	0	0	0	0	0	2	0	0	0	0	1	0	0	0	0	0	1	0	0	0	2	0	0	0	0	0	2	0	0	0	0	0	0	1	0	2	0	0	0	0
GmRER4c	0	2	3	0	0	2	1	1	0	0	0	1	3	1	1	0	2	1	2	0	0	3	2	0	1	0	1	3	0	1	0	0	9	0	0	0	1	0	0	1	2	3	0	4	0	1	0	3	4	0	0	0
GmRER4d	1	5	2	0	0	0	0	2	0	1	0	1	2	0	1	0	0	2	2	1	0	0	2	0	0	0	1	0	0	0	8	1	0	0	1	0	1	1	0	2	0	6	0	0	0	0	0	0	0	0	0	0
GmRER4e	2	7	2	0	0	0	0	4	0	1	2	0	0	2	0	1	1	0	2	3	0	0	0	0	2	2	0	0	2	0	1	0	0	1	0	0	2	0	1	1	3	0	0	0	0	0	0	0	0	0	0	0
GmRER5	1	0	0	0	0	0	1	4	0	0	2	1	0	0	1	1	1	1	1	1	0	2	0	0	0	0	1	3	0	0	1	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
GmRER6	1	3	0	0	0	0	3	0	0	1	1	0	3	0	1	1	0	2	1	1	0	2	0	0	0	0	2	0	5	0	0	0	3	0	1	2	1	2	0	0	5	0	0	0	0	0	0	0	0	0	0	0

Figure 6. Statistical analysis of *Cis*-acting elements in the promoter regions of *GmRERs*. *Cis*-elements are categorized into three main classes: abiotic and biotic stress, phytohormone-responsiveness, and plant growth and development. The numbers represent the occurrence of each *cis*-element in the promoter regions of the genes. The deeper the red color, the higher the frequency of occurrence.

Combined with our expression data (Figure 5), stress-responsive *cis*-regulatory motifs were found to be enriched in *GmRERs* with tissue-specific expression (Figure 6). For instance, root-expressed *GmRER1b* harbored three ABREs (abscisic acid-responsive elements), which was reported to be related to drought induction [46], while predominantly leaf-expressed *GmRER5/6* uniquely combined two ethylene-responsive elements (EREs) with four TC-rich repeats, which was proven to be linked to salt stress adaptation [47]. For phytohormone-responsive modules, *GmRER3a* and *GmRER3b* carried both ABA-responsive ABREs (one–two motifs) and SA-responsive TCA elements (one motif each), while *GmRER3c* retained only SA-responsive TCA motifs (two motifs), enabling partitioned drought–pathogen defense strategies [48]. In contrast, GmRER5/6 promoters combined ethylene-responsive elements (EREs; two motifs per promoter) with photoperiod-associated circadian motifs, suggesting that they might both involved in light-entrained ethylene signaling during hypoxia recovery [49]. Moreover, development-related motifs diverged sharply. Specifically, *GmRER4c/d* contained eight–nine Box-4 elements (light-responsive) and one–three G-box motifs (light-responsive), and *GmRER3a* retained one HD-Zip-I motif, which may potentially be linked to chloroplast differentiation [50]. Additionally, although *GmREa* retained 10 MYC motifs, 4 G-box motifs (light responsiveness), and 2 MYB sites, barely any obvious expression changes in *GmREa* were found under the tested conditions in our expression data (Figure 5), suggesting that transcriptional regulatory suppression may occur despite its functional *cis*-element repertoire.

2.7. Expression Patterns of GmRERs in Six Northern Spring Soybean Landraces

Previous studies have demonstrated that genetic diversity bottlenecks occurred during soybean domestication and subsequent improvement [51,52]. Comprehensive identification and characterization of natural genetic variations are therefore essential for advancing soybean breeding programs [53–56]. To investigate expression variability within the *GmRER* gene family across natural collections, we conducted comparative transcriptome analyses of six northern spring soybean landraces, including Xiaoli Moshidou (L1), Baimaoshuang (L2), Tiejia Silihuang (L3), Heidadou (L4), Zhouye (L5), and Aqi Manjinhuang (L6). These accessions exhibited pronounced phenotypic variation in leaf growth and development at the vegetative stage 1 (V1 stage; Figure 7a).
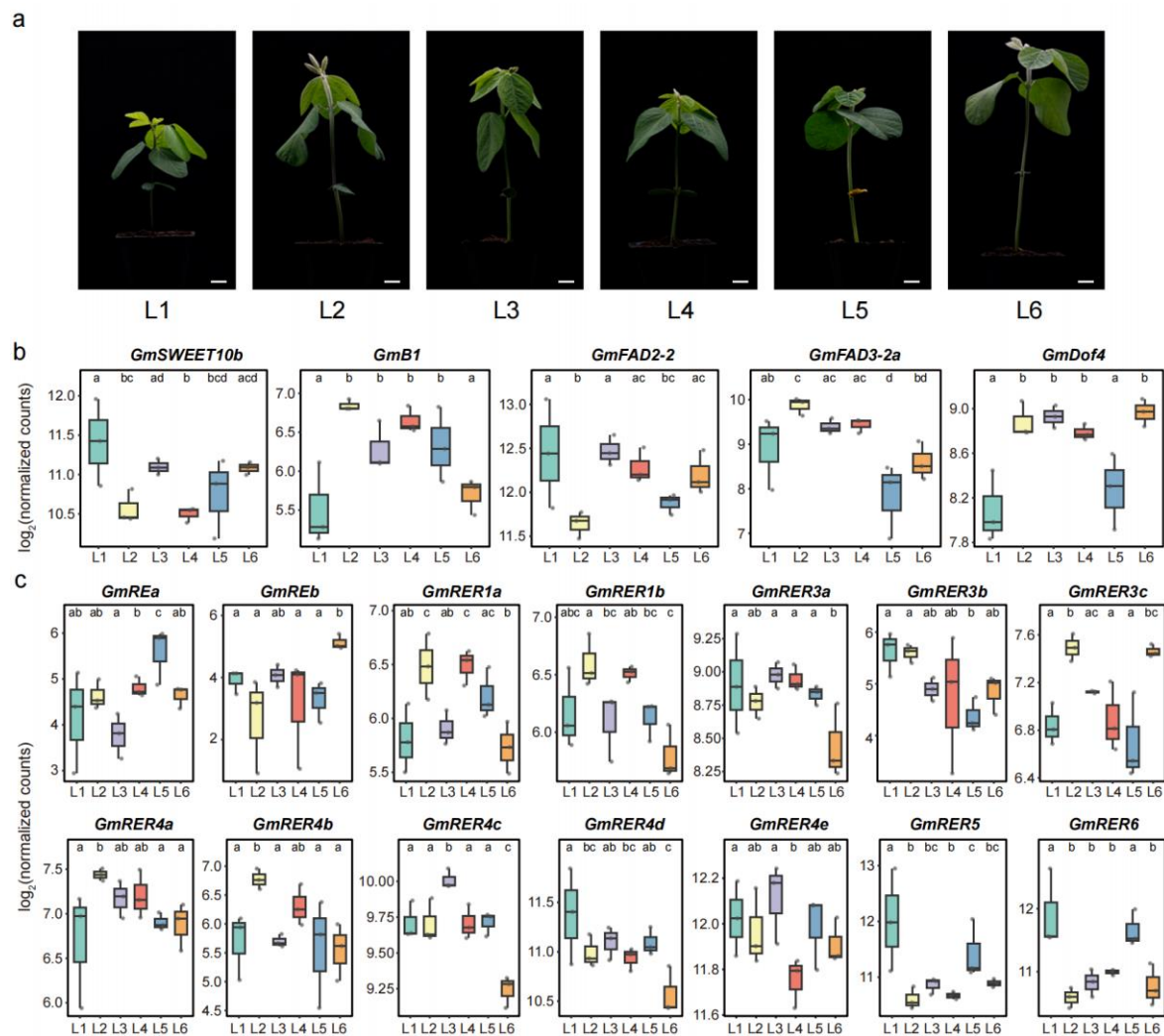
Figure 7. Expression levels of *GmRER* family genes in six northern spring soybean landraces and their seedling phenotypes. (**a**) The seedling phenotypes of the six landraces at the soybean V1 stage. The representative images of accessions labeled L1 to L6 correspond to the following landraces (left to right): Xiaoli Moshidou (L1), Baimaoshuang (L2), Tiejia Silihuang (L3), Heidadou (L4), Zhouye (L5), and Aqi Manjinhuang (L6). Scale bars = 1.2 cm. (**b**) The expression patterns of five known oil biosynthesis-associated genes (*GmSWEET10b*, *GmB1*, *GmFAD2-2*, *GmFAD3-2a*, and *GmDof4*) across the six landraces. The data represent Log2-transformed normalized counts derived from DESeq2. (**c**) Differential expression analysis of GmRER family genes in the six landraces. For panels (**b**,**c**), statistical significance was determined by pairwise comparisons using DESeq2's Wald test. Lowercase letters above the boxplots indicate significant differences: landraces sharing the same letter are not significantly different, while distinct letters denote statistically divergent expression ($p < 0.05$).

We generated six RNA-seq datasets from the first trifoliate leaves of the six landraces, which were aligned to the cultivated soybean reference genome (*Glycine max* Williams 82, assembly version a4v1, Table S4). Prior to cross-landrace comparisons, the RNA-seq data quality was validated through Spearman's correlation analysis, demonstrating high inter-replicate concordance ($R > 0.95$) across three biological replicates per accession (Figure S3). First, we analyzed the expression levels of five well-known genes involved in lipid biosynthesis, including *GmSWEET10b* [57], *GmB1* [58], *GmFAD2-2* [5], *GmFAD3-2a* [6], and *GmDof4* [7]. These five genes exhibited distinct expression patterns among the six landraces. Among them, three genes (*GmSWEET10b*, *GmFAD2-2*, and *GmFAD3-2a*) showed relatively more frequent significant differences in the pairwise lan-

drace comparisons (Figure 7b). These findings indicate expression polymorphism in key regulatory genes associated with soybean oil biosynthesis among the studied landraces.

Then, we evaluated the expression variations in *GmRER* genes among the six landrace accessions. As shown in Figure 7c, most *GmRER* genes predominantly displayed landrace-specific expression divergence. Within the *GmRER3* group, *GmRER3a* and *GmRER3b* displayed limited expression variation, whereas *GmRER3c* exhibited statistically significant differences in expression across landraces. Similar divergence patterns were observed for *GmRE* paralogs and *GmRER4* paralogs, with distinct expression trajectories among landraces (Figure 7c). Strikingly, *GmRER1* paralogs (GmRER1a/b) both showed prominent expression divergence and highly similar expression patterns among the six landrace accessions. In addition, pairwise comparisons revealed that paralog groups diverged more substantially in their expression profiles than *GmRER* members within the groups. Thus, the expression variation dynamics within and between paralog groups differed substantially among GmRER family members across landrace accessions.

3. Discussion

3.1. Evolutionary Conservation and Functional Dynamics in the GmRER Family

The RER gene family has been evolutionarily conserved across land plants to regulate chloroplast development and lipid metabolism [22,25,26], yet lineage-specific expansions in crops like soybean have sculpted unique functional configurations (Figure 1a) [30,35]. Our genome-wide characterization revealed fundamental divergences between soybean and Arabidopsis in the RER family composition. Most strikingly, while Arabidopsis harbors both RER2 and RER3 clades [22], soybean exclusively retains three GmRER3 paralogs clustered with the Arabidopsis RER2-RER3 branch (Figure 1a). Reciprocal BLASTP analysis excluded the homology to AtRER2, indicating either functional divergence or loss of GmRER2 paralogs in soybean. This absence may reflect selective pressure favoring GmRER3 paralogs optimized for legume-specific environmental adaptations. Notably, the intronless configuration of *GmRER3a/c* (versus single-exon *AtRER3*; Figure 2a), facilitates rapid transcriptional responses to drought (Figure 5c), aligning with soybean susceptibility to water deficits during vegetative growth [16,59].

In contrast to Arabidopsis' single *RER4* copy [22], soybean has retained five *GmRER4* paralogs via paleopolyploidy-driven segmental duplications (Figure 1). This expansion likely evolved to support soybean heightened demand for lipid-mediated stress adaptation and resource allocation in leaves [16,18]. Structural conservation in core DUF3411 domains (Figure 3d), alongside promoter diversification (MYB motif enrichment; Figure 6), enables *GmRER4* paralogs to balance constitutive roles in vascular nutrient transport (Figure 5a) with stress-responsive transcription plasticity (Figure 4b). Such subfunctionalization aligns with the dosage balance hypothesis [60], where polyploid/paleopolyploid species preferentially retain duplicated metabolic regulators to maintain stoichiometric relationships in specialized pathways, like nitrogen fixation or lipid reserves during senescence [61,62].

These evolutionary trajectories have practical implications for soybean breeding. The drought-inducible *GmRER3a/c* and hypoxia-responsive *GmRER5/6* identified here (Figure 5) offer candidate loci for breeding climate-resilient soybeans. However, functional redundancy among paralogs poses challenges for gene editing, in that knocking out single *GmRER4* members (e.g., GmRER4d) may not yield phenotypic changes unless multiple paralogs are targeted [63]. Future studies should employ multiplex CRISPR systems to dissect additive effects and prioritize hub paralogs (e.g., high-expression *GmRER4e*) for metabolic engineering.

3.2. Structural Diversification Coupled with Regulatory Plasticity in GmRERs

Take *GmRERs* as an example. Soybean has fostered structural innovations that decouple RNA-level plasticity from protein functional conservation (Figure 3). Meanwhile, RNA secondary structures of *GmRER3* and *GmRER4* paralogs diverged markedly in their CDS regions (Figure 3b), while their protein core folds remained conserved (Figure 3d). The preservation of core protein structures despite RNA heterogeneity likely buffered soybean against genomic perturbations while permitting regulatory diversification. For example, *GmRER4* paralogs acquired MYB-responsive promoters (Figure 6) without disrupting DUF3411 domain integrity, enabling tissue-specific lipid mobilization during leaf senescence, which may serve as a critical adaptation for nutrient recovery in soybean's determinate growth habit [16,18,62].

Divergent subcellular localization further partitions paralog functions. Unlike chloroplast-anchored GmRER1/RER4 paralogs, *GmRER5/RER6* targeted the plasma membrane (Table 1) and were involved in redox regulation through interactions with NADP-oxidoreductases (Figure 4). This subcellular rewiring parallels their ethylene-responsive promoter motifs (Figure 6), suggesting neofunctionalization to mitigate hypoxia-triggered oxidative stress in soybean roots, a rare but devastating challenge in submerged fields [24,64]. Such compartment-specific adaptations highlight how structural diversification enables paralogs to resolve conflicting evolutionary pressures (e.g., stress tolerance vs. developmental growth).

3.3. Functional Implications of GmRERs in Leaf Lipid Metabolism

Our results unveil soybean's hierarchical resource allocation strategy under stress, mediated by the spatiotemporal polarization of *GmRER* paralogs (Figure 5), Class I genes (mainly expressed in photosynthetic tissues; Figure 5a) are transiently suppressed during acute stress to conserve energy (e.g., salt-responsive *GmRER5/6* down-regulation in leaves; Figure 5b), reflecting a growth–defense tradeoff common in annual crops [65]. Conversely, Class II paralogs (root-enriched *GmRER4a/b*; Figure 5c) prioritize stress mitigation through sustained activation, mirroring the "root-first" drought response strategy observed in other legumes [66–69]. Additionally, this dynamic modulation of leaf fatty acids, such as the heat-induced accumulation of linolenic acid [16], may complement GmRER-mediated stress responses. Intriguingly, Class I and II genes exhibited ectopic activation in non-native tissues under abiotic stress, e.g., GmREa/b and *GmRER1a* in leaves, and *GmRER5a/b* in roots (Figure 5), implying latent regulatory flexibility critical for soybean resilience.

The transcriptional polymorphism observed across northern spring soybean landraces (Figure 7c) suggests that artificial selection has leveraged *GmRER* diversity for local adaptation. For example, the parallel expression patterns of *GmRER4a* and *GmFAD3-2a* (an enzyme responsible for the synthesis of α-linolenic acid) in six landraces (Figure 7) imply co-regulated lipid-remobilization pathways. It is noteworthy that the expression profiling of the six landraces identified in this study was performed with three biological replicates per genotype. Hence, the possibility cannot be ruled out that the observed variations in gene expression may have originated from intra-cultivar variations [70–73]. For soybean breeding trials, enhancing the polyunsaturated fatty acid content (e.g., linolenic acid under high temperature) via *GmRER4a* editing could simultaneously improve thermotolerance and oil-based defense metabolites. However, the incomplete annotation of lipid-associated interactors (Figure 4a) currently limits network-level insights. Integrative multi-omics (e.g., lipidomics or ChIP-seq) studies are needed to map regulatory hierarchies of GmRERs and identify master regulators of soybean leaf oil metabolism.

4. Materials and Methods

4.1. Growth Conditions

The seeds of six landrace genotypes (L1–L6) from soybean plants that had undergone more than five consecutive generations of self-pollination were selected and surface-sterilized with 75% ethanol for 30 s followed by two rinses with RO water before cultivation. The seeds were sown individually in pots (one plant per pot), which were filled with sterilized soil and placed in a growth chamber under controlled conditions: 18 h light/6 h dark photoperiod, with day/night temperatures maintained at 25 °C/22 °C, respectively. After the plants reached full development of the first trifoliate leaf, part of each first trifoliate leaf was harvested and immediately frozen in liquid nitrogen. Three independent biological replicates per genotype (corresponding to three separate pot-grown plants) were collected and stored separately at −80 °C for subsequent RNA extraction and transcriptome analysis. It is critical to acknowledge that the use of triplicate biological replicates per genotype was limited because of the intra-cultivar variations.

4.2. Genome-Wide Identification of GmRER Genes

The soybean (*Glycine max* cv. Williams 82) genome (assembly Wm82.a4.v1) was retrieved from Phytozome v13 [36]. To identify RER homologs, the protein sequences of seven Arabidopsis RER genes (AtRE and AtRER1-AtRER6) were extracted from TAIR (https://www.arabidopsis.org/ (accessed on 1 April 2025)) and subjected to InterProScan (version 5.70-102) to confirm the presence of the DUF3411 (PF11891) domain [74]. These sequences were then used as queries for BLASTP searches (e-value $\leq 1 \times 10^{-10}$) against the soybean proteome. To investigate genomic synteny relationships, the BLASTP matches were subsequently analyzed through MCScanX (with default parameters) to detect collinear chromosomal regions [75]. Gene duplication events were further characterized using the built-in gene classification algorithm in MCScanX. Candidate sequences containing the full-length DUF3411 domain were further validated via HMMER v3.3.2 using the PF11891 HMM profile (cutoff: 1×10^{-50}) [76]. Non-redundant candidates were assigned systematic names based on phylogenetic clustering (e.g., GmRER1a, GmRER3b) following established nomenclature guidelines.

4.3. Molecular Evolution Analysis and Chromosomal Locations

Multiple sequence alignment of RER proteins from six species (*Arabidopsis thaliana*, *Glycine max*, *Oryza sativa*, *Gossypium hirsutum*, *Zea mays*, and *Triticum aestivum*) was performed using ClustalW in MEGA12 with the default parameters [77]. To ensure alignment integrity, gaps and ambiguous sites were treated using the complete deletion method, removing all positions containing gaps or missing data prior to downstream phylogenetic analysis. A maximum-likelihood (ML) phylogenetic tree was constructed in MEGA12. The bootstrap consensus tree inferred from 500 replicates was taken to represent the evolutionary history of the taxa analyzed, where branches corresponding to partitions reproduced in less than 50% of replicate trees were collapsed. The percentage of replicate trees in which the associated taxa were clustered together (500 replicates) was shown next to the branches. The initial tree for the heuristic search was selected by choosing the tree with the superior log-likelihood between a Neighbor-Joining (NJ) tree and a Maximum Parsimony (MP) tree. The NJ tree was generated using a matrix of pairwise distances computed using the p-distance. The MP tree had the shortest length among 10 MP tree searches, each performed with a randomly generated starting tree. The final dataset encompassed 56 amino acid sequences with 796 positions in the final dataset. Evolutionary analyses were conducted in MEGA12 utilizing up to 4 parallel computing threads. Orthology relationships were

validated via bidirectional best-hit BLASTP (BBH; e-value $\leq 1 \times 10^{-10}$) and visualized using TBtools (version 2.056) [78].

4.4. Gene Structure and Conserved Motif Annotation

Exon–intron structures were mapped using the Gene Structure Display Server [79]. Conserved motifs were identified via MEME Suite v5.4.1 with the parameters set to discover 10 motifs (width: 6–50 residues; any number of repetitions) [80]. Motif annotations were inferred by cross-referencing with the InterPro and SMART databases [81]. Transmembrane helices and chloroplast transit peptides (cTP) were predicted using TMHMM v2.0 and TargetP v2.0, respectively [82,83].

4.5. Cis-Element Profiling and Promoter Analysis

A 2 kb upstream sequence from each GmRER transcription start site (TSS) was extracted from the soybean genome. *Cis*-regulatory elements were annotated using Plant-CARE with a significance cutoff of $p < 0.01$ for motif occurrence [84]. Elements were categorized into stress-responsive (e.g., MYB, ABRE), hormonal (e.g., ERE, AuxRR), and developmental (e.g., G-box, HD-Zip) modules. Heatmaps visualizing motif density (elements/kb) were generated using TBtools.

4.6. Protein Interaction Network Construction

Protein–protein interaction (PPI) networks were predicted using STRING v11.5 with a high-confidence interaction score ≥ 0.7 [85]. The networks were imported into Cytoscape v3.9.0 for modular analysis, applying the MCODE plugin to identify densely connected clusters [86]. Functional annotations for interactors were derived from the Gene Ontology (GO) database [87].

4.7. Expression Profiling and Data Visualization

For spatiotemporal analysis of gene expression under different conditions, normalized RNA-seq data (FPKM) were retrieved from the NCBI Sequence Read Archive under accessions PRJNA238493 (tissue atlas) [43], PRJNA246058 (salt stress) [44], and PRJNA574626 (drought/submergence) [42]. HISAT2 v2.2.1 and StringTie v2.2.0 were used for alignment and quantification, respectively [88,89]. Heatmaps were plotted using the R package pheatmap with z-score normalization (row-wise) and hierarchical clustering (Euclidean distance, complete linkage). Expression trends under stress (e.g., suppression at 4 h, recovery at 24 h) were qualitatively assessed based on hierarchical clustering patterns and the manual inspection of fold-changes.

4.8. RNA Extraction and Sequencing

Total RNA were extracted from each plant using the Trizol method, strictly following the manufacturer's protocol. The RNA samples were then sent to a sequencing company for library preparation and sequencing on the NovaSeq 6000 platform. Raw reads were processed to remove adapter contamination, low-quality reads, and reads containing more than 5% ambiguous bases (N). Each sample yielded at least 5 Gb of clean data.

4.9. RNA-seq Data Processing, Mapping, and Differential Expression Analysis

After sequencing, raw RNA-seq reads were aligned to the cultivated soybean reference genome (Glycine max Williams 82, assembly version a4v1) using STAR (v2.7.0d) with default parameters, retaining only uniquely mapped reads for downstream analysis. Mapping efficiency for all six landraces (L1–L6) showed consistently high quality, with the raw read counts per replicate ranging from 16.2 to 19.8 million, unique mapping rates of 86.9–88.7%, and overall mapping rates exceeding 96% (complete data in Table S4). The

gene expression values were filtered to retain only genes with raw counts ≥ 10 in all three biological replicates. Spearman correlation analysis was performed using the cor.test function in R with exact *p*-value calculation disabled for large sample sizes. Scatterplots were generated using ggplot2 (Figure S3).

Normalization and differential expression analysis were performed using DESeq2 with default parameters, including median-of-ratios normalization and dispersion estimation via a negative binomial generalized linear model. Pairwise comparisons between all six landraces were conducted using DESeq2's Wald test, with significance defined as a raw *p*-value < 0.05. The expression values were represented as log2-transformed DESeq2-normalized counts for visualization. Boxplots were generated by using ggplot2, and statistical significance annotations (lowercase letters) were generated using the multcomp package in R. For each gene, a pairwise *p*-value matrix was constructed from all genotype comparisons, and letter groupings were assigned through the general linear hypothesis procedure when *p* < 0.05. Groups sharing the same letter indicate no significant differences.

4.10. RNA Secondary Structure and Protein Similarity Assessment

Full-length mRNA sequences (5′ UTR, CDS, 3′ UTR) were folded using RNAfold v2.4.18 under default thermodynamic parameters [90]. Structural similarity scores between paralogs were computed using RNAforester with the Sankoff algorithm [91]. For protein similarity assessment, three-dimensional structures were predicted using AlphaFold2 in ColabFold with default templates and multiple sequence alignments (MSAs) generated from UniRef30 (2021-09) [92,93]. Structurally aligned models were subsequently analyzed in PyMOL v2.5 to determine the root-mean-square deviation (RMSD) values as quantitative measures of structural similarity [94].

5. Conclusions

This study systematically characterized the GmRER gene family in soybean, which is associated with chloroplast development and plastid metabolism. We further elucidated its evolutionary trajectory, structural diversification, and functional plasticity. A total of 14 *GmRER* genes were identified and classified into six paralog groups through phylogenetic analysis. The chromosomal distribution highlighted the contributions of segmental and tandem duplications to GmRER family expansion. Architectural dissection revealed subclade-specific gene structures and conserved motifs, suggesting functional divergence. Moreover, despite divergent RNA secondary structures, the GmRER protein folds remained evolutionarily constrained, highlighting strong functional conservation overriding transcriptional plasticity. The expression profiling and promoter analysis of *GmRERs* implicated their opposing regulatory strategies for photosynthesis and energy metabolism. Remarkably, substantially different expression levels within and between paralog groups highlight potential genetic diversity within the *GmRER* gene family during soybean domestication, offering potential targets for breeding optimization. It is important to know that intra-cultivar variations may constitute a confounding factor underlying the expression divergence in our RNA-seq analysis. Collectively, our findings may not only provide critical insights into the chloroplast-localized trade-off between photosynthesis and lipid biosynthesis elucidated via GmRERs, but also offer a possibility for developing high-oil soybean cultivars.

Supplementary Materials: The following supporting information can be downloaded at: https://www.mdpi.com/article/10.3390/plants14101516/s1, Figure S1: Predicted RNA secondary structure of the GmRER genes; Figure S2: Protein Structures of GmRERs; Figure S3. Spearman correlation analysis between biological replicates; Table S1: List of RER genes in soybean; Table S2: Sequences and

Descriptions of MEME analyzed motifs in GmRERs; Table S3. Uniprot ID and annotations of GmRER interacting proteins; Table S4. Summary of mapping statistics for the six landrace RNA-seq data.

Author Contributions: Conceptualization, Q.D., T.W. and X.Y.; Data Curation, J.L. (Jun Lang); Formal Analysis, L.N.; Funding Acquisition, Q.D. and T.W.; Investigation, L.N. and X.G.; Methodology, J.L. (Jie Liang); Software, Q.X.; Supervision, X.Y.; Writing—Original Draft, Q.D., T.W. and X.Y.; Writing—Review and Editing, L.N., X.G., Q.X., J.L. (Jie Liang) and J.L. (Jun Lang). All authors have read and agreed to the published version of the manuscript.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The RNA-seq data produced in our study have been submitted to the NCBI SRA database and can be found under the following accession number: PRJNA1246969. The other RNA-seq datasets analyzed in this study are publicly available in the NCBI Sequence Read Archive under the following accession numbers: PRJNA238493 (tissue-specific expression atlas) [43], PRJNA246058 (salt stress experiments) [44], and PRJNA574626 (drought and submergence experiments) [42]. The predicted protein structures generated via AlphaFold can be found at https://alphafold.com/ (accessed on 11 January 2025). The phylogenetic trees, motif annotations, and expression heatmaps have been included as figures in the manuscript.

Conflicts of Interest: The authors declare no conflicts of interest.

Abbreviations

The following abbreviations are used in this manuscript:

ABA	Abscisic acid
ABRE	Abscisic acid-responsive element
CDS	Coding sequence
CRISPR	Clustered Regularly Interspaced Short Palindromic Repeats
ERE	Ethylene-responsive element
FPKM	Fragments Per Kilobase of transcript per Million mapped reads
GO	Gene Ontology
MYB	MYeloBlastosis transcription factor family
PPI	Protein–protein interaction
RER	RETICULATA-RELATED
RMSD	Root-mean-square deviation
UTR	Untranslated region
WGD	Whole-genome duplication

References

1. Graham, P.H.; Vance, C.P. Legumes: Importance and Constraints to Greater Use. *Plant Physiol.* **2003**, *131*, 872–877. [CrossRef]
2. Hartman, G.L.; West, E.D.; Herman, T.K. Crops That Feed the World 2. Soybean—Worldwide Production, Use, and Constraints Caused by Pathogens and Pests. *Food Sec.* **2011**, *3*, 5–17. [CrossRef]
3. Pagano, M.C.; Miransari, M. The Importance of Soybean Production Worldwide. In *Abiotic and Biotic Stresses in Soybean Production*; Miransari, M., Ed.; Academic Press: San Diego, CA, USA, 2016; pp. 1–26. ISBN 978-0-12-801536-0.

4. Tian, Z.; Nepomuceno, A.L.; Song, Q.; Stupar, R.M.; Liu, B.; Kong, F.; Ma, J.; Lee, S.-H.; Jackson, S.A. Soybean2035: A Decadal Vision for Soybean Functional Genomics and Breeding. *Mol. Plant* **2025**, *18*, 245–271. [CrossRef]

5. al Amin, N.; Ahmad, N.; Wu, N.; Pu, X.; Ma, T.; Du, Y.; Bo, X.; Wang, N.; Sharif, R.; Wang, P. CRISPR-Cas9 Mediated Targeted Disruption of FAD2–2 Microsomal Omega-6 Desaturase in Soybean (*Glycine max*.L). *BMC Biotechnol.* **2019**, *19*, 9. [CrossRef] [PubMed]

6. Hoshino, T.; Watanabe, S.; Takagi, Y.; Anai, T. A Novel GmFAD3-2a Mutant Allele Developed through TILLING Reduces α-Linolenic Acid Content in Soybean Seed Oil. *Breed. Sci.* **2014**, *64*, 371–377. [CrossRef] [PubMed]

7. Wang, H.-W.; Zhang, B.; Hao, Y.-J.; Huang, J.; Tian, A.-G.; Liao, Y.; Zhang, J.-S.; Chen, S.-Y. The Soybean Dof-Type Transcription Factor Genes, GmDof4 and GmDof11, Enhance Lipid Content in the Seeds of Transgenic Arabidopsis Plants. *Plant J.* **2007**, *52*, 716–729. [CrossRef]

8. Liu, Y.-F.; Li, Q.-T.; Lu, X.; Song, Q.-X.; Lam, S.-M.; Zhang, W.-K.; Ma, B.; Lin, Q.; Man, W.-Q.; Du, W.-G.; et al. Soybean GmMYB73 Promotes Lipid Accumulation in Transgenic Plants. *BMC Plant Biol.* **2014**, *14*, 73. [CrossRef]

9. Zhang, Y.-Q.; Lu, X.; Zhao, F.-Y.; Li, Q.-T.; Niu, S.-L.; Wei, W.; Zhang, W.-K.; Ma, B.; Chen, S.-Y.; Zhang, J.-S. Soybean GmDREBL Increases Lipid Content in Seeds of Transgenic Arabidopsis. *Sci. Rep.* **2016**, *6*, 34307. [CrossRef]

10. Manan, S.; Ahmad, M.Z.; Zhang, G.; Chen, B.; Haq, B.U.; Yang, J.; Zhao, J. Soybean LEC2 Regulates Subsets of Genes Involved in Controlling the Biosynthesis and Catabolism of Seed Storage Substances and Seed Development. *Front. Plant Sci.* **2017**, *8*, 1604. [CrossRef]

11. Lu, X.; Li, Q.-T.; Xiong, Q.; Li, W.; Bi, Y.-D.; Lai, Y.-C.; Liu, X.-L.; Man, W.-Q.; Zhang, W.-K.; Ma, B.; et al. The Transcriptomic Signature of Developing Soybean Seeds Reveals the Genetic Basis of Seed Trait Adaptation during Domestication. *Plant J.* **2016**, *86*, 530–544. [CrossRef]

12. Marchive, C.; Nikovics, K.; To, A.; Lepiniec, L.; Baud, S. Transcriptional Regulation of Fatty Acid Production in Higher Plants: Molecular Bases and Biotechnological Outcomes. *Eur. J. Lipid Sci. Technol.* **2014**, *116*, 1332–1343. [CrossRef]

13. Xu, C.; Shanklin, J. Triacylglycerol Metabolism, Function, and Accumulation in Plant Vegetative Tissues*. *Annu. Rev. Plant Biol.* **2016**, *67*, 179–206. [CrossRef]

14. Pyc, M.; Cai, Y.; Greer, M.S.; Yurchenko, O.; Chapman, K.D.; Dyer, J.M.; Mullen, R.T. Turning Over a New Leaf in Lipid Droplet Biology. *Trends Plant Sci.* **2017**, *22*, 596–609. [CrossRef]

15. Wan, S.; Truong-Trieu, V.M.; Ward, T.; Whalen, J.K.; Altosaar, I. Advances in the Use of Genetically Modified Plant Biomass for Biodiesel Generation. *Biofuels Bioprod. Biorefining* **2017**, *11*, 749–764. [CrossRef]

16. Rennie, B.D.; Tanner, J.W. Fatty Acid Composition of Oil from Soybean Leaves Grown at Extreme Temperatures. *J. Am. Oil Chem. Soc.* **1991**, *68*, 1O6–108. [CrossRef]

17. Park, Y.J.; Lee, J.S.; Park, S.; Kim, Y.J.; Mani, V.; Lee, K.; Kwon, S.J.; Park, S.U.; Kim, J.K. Metabolite Changes in Soybean (Glycine Max) Leaves during the Entire Growth Period. *ACS Omega* **2023**, *8*, 41718–41727. [CrossRef]

18. Shimada, T.L.; Hayashi, M.; Hara-Nishimura, I. Membrane Dynamics and Multiple Functions of Oil Bodies in Seeds and Leaves. *Plant Physiol.* **2018**, *176*, 199–207. [CrossRef] [PubMed]

19. Lagunas, B.; Román, Á.; Andreu, V.; Picorel, R.; Alfonso, M. A Temporal Regulatory Mechanism Controls the Different Contribution of Endoplasmic Reticulum and Plastidial ω-3 Desaturases to Trienoic Fatty Acid Content during Leaf Development in Soybean (*Glycine Max* Cv Volania). *Phytochemistry* **2013**, *95*, 158–167. [CrossRef]

20. Pérez-Bueno, M.L.; Pineda, M.; Barón, M. Phenotyping Plant Responses to Biotic Stress by Chlorophyll Fluorescence Imaging. *Front. Plant Sci.* **2019**, *10*, 1135. [CrossRef]

21. Leister, D. Retrograde Signaling in Plants: From Simple to Complex Scenarios. *Front. Plant Sci.* **2012**, *3*, 135. [CrossRef]

22. Pérez-Pérez, J.M.; Esteve-Bruna, D.; González-Bayón, R.; Kangasjärvi, S.; Caldana, C.; Hannah, M.A.; Willmitzer, L.; Ponce, M.R.; Micol, J.L. Functional Redundancy and Divergence within the Arabidopsis RETICULATA-RELATED Gene Family. *Plant Physiol.* **2013**, *162*, 589–603. [CrossRef] [PubMed]

23. Yang, Y.; Li, L.; Qu, L.-J. Plant Mediator Complex and Its Critical Functions in Transcription Regulation. *J. Integr. Plant Biol.* **2016**, *58*, 106–118. [CrossRef]

24. Kesten, C.; García-Moreno, Á.; Amorim-Silva, V.; Menna, A.; Castillo, A.G.; Percio, F.; Armengot, L.; Ruiz-Lopez, N.; Jaillais, Y.; Sánchez-Rodríguez, C.; et al. Peripheral Membrane Proteins Modulate Stress Tolerance by Safeguarding Cellulose Synthases. *Sci. Adv.* **2022**, *8*, eabq6971. [CrossRef] [PubMed]

25. Lundquist, P.K.; Rosar, C.; Bräutigam, A.; Weber, A.P.M. Plastid Signals and the Bundle Sheath: Mesophyll Development in Reticulate Mutants. *Mol. Plant* **2014**, *7*, 14–29. [CrossRef] [PubMed]

26. Price, C.A.; Weitz, J.S. Costs and Benefits of Reticulate Leaf Venation. *BMC Plant Biol.* **2014**, *14*, 234. [CrossRef]

27. Schubert, M.; Petersson, U.A.; Haas, B.J.; Funk, C.; Schröder, W.P.; Kieselbach, T. Proteome Map of the Chloroplast Lumen of *Arabidopsis Thaliana *. *J. Biol. Chem.* **2002**, *277*, 8354–8365. [CrossRef]

28. Peltier, J.-B.; Emanuelsson, O.; Kalume, D.E.; Ytterberg, J.; Friso, G.; Rudella, A.; Liberles, D.A.; Söderberg, L.; Roepstorff, P.; von Heijne, G.; et al. Central Functions of the Lumenal and Peripheral Thylakoid Proteome of Arabidopsis Determined by Experimentation and Genome-Wide Prediction. *Plant Cell* **2002**, *14*, 211–236. [CrossRef]

29. Pérez-Pérez, J.M.; Candela, H.; Robles, P.; López-Torrejón, G.; del Pozo, J.C.; Micol, J.L. A Role for AUXIN RESISTANT3 in the Coordination of Leaf Growth. *Plant Cell Physiol.* **2010**, *51*, 1661–1673. [CrossRef]

30. Lynch, M.; Force, A. The Probability of Duplicate Gene Preservation by Subfunctionalization. *Genetics* **2000**, *154*, 459–473. [CrossRef]

31. Panchy, N.; Lehti-Shiu, M.; Shiu, S.-H. Evolution of Gene Duplication in Plants. *Plant Physiol.* **2016**, *171*, 2294–2316. [CrossRef]

32. Freeling, M.; Scanlon, M.J.; Fowler, J.E. Fractionation and Subfunctionalization Following Genome Duplications: Mechanisms That Drive Gene Content and Their Consequences. *Curr. Opin. Genet. Dev.* **2015**, *35*, 110–118. [CrossRef] [PubMed]

33. Wang, Y.; Tan, X.; Paterson, A.H. Different Patterns of Gene Structure Divergence Following Gene Duplication in Arabidopsis. *BMC Genom.* **2013**, *14*, 652. [CrossRef] [PubMed]

34. Schmutz, J.; Cannon, S.B.; Schlueter, J.; Ma, J.; Mitros, T.; Nelson, W.; Hyten, D.L.; Song, Q.; Thelen, J.J.; Cheng, J.; et al. Genome Sequence of the Palaeopolyploid Soybean. *Nature* **2010**, *463*, 178–183. [CrossRef]

35. Adams, K.L.; Wendel, J.F. Polyploidy and Genome Evolution in Plants. *Curr. Opin. Plant Biol.* **2005**, *8*, 135–141. [CrossRef]

36. Goodstein, D.M.; Shu, S.; Howson, R.; Neupane, R.; Hayes, R.D.; Fazo, J.; Mitros, T.; Dirks, W.; Hellsten, U.; Putnam, N. Phytozome: A Comparative Platform for Green Plant Genomics. *Nucleic Acids Res.* **2012**, *40*, D1178–D1186. [CrossRef]

37. Heinig, U.; Gutensohn, M.; Dudareva, N.; Aharoni, A. The Challenges of Cellular Compartmentalization in Plant Metabolic Engineering. *Curr. Opin. Biotechnol.* **2013**, *24*, 239–246. [CrossRef]

38. Zoschke, R.; Bock, R. Chloroplast Translation: Structural and Functional Organization, Operational Control, and Regulation. *Plant Cell* **2018**, *30*, 745–770. [CrossRef] [PubMed]

39. Woodson, J.D. Chloroplast Stress Signals: Regulation of Cellular Degradation and Chloroplast Turnover. *Curr. Opin. Plant Biol.* **2019**, *52*, 30–37. [CrossRef]

40. Noctor, G.; Lelarge-Trouverie, C.; Mhamdi, A. The Metabolomics of Oxidative Stress. *Phytochemistry* **2015**, *112*, 33–53. [CrossRef]

41. Chan, K.X.; Phua, S.Y.; Crisp, P.; McQuinn, R.; Pogson, B.J. Learning the Languages of the Chloroplast: Retrograde Signaling and Beyond. *Annu. Rev. Plant Biol.* **2016**, *67*, 25–53. [CrossRef]

42. Tamang, B.G.; Li, S.; Rajasundaram, D.; Lamichhane, S.; Fukao, T. Overlapping and Stress-Specific Transcriptomic and Hormonal Responses to Flooding and Drought in Soybean. *Plant J.* **2021**, *107*, 100–117. [CrossRef] [PubMed]

43. Shen, Y.; Zhou, Z.; Wang, Z.; Li, W.; Fang, C.; Wu, M.; Ma, Y.; Liu, T.; Kong, L.-A.; Peng, D.-L.; et al. Global Dissection of Alternative Splicing in Paleopolyploid Soybean. *Plant Cell* **2014**, *26*, 996–1008. [CrossRef] [PubMed]

44. Belamkar, V.; Weeks, N.T.; Bharti, A.K.; Farmer, A.D.; Graham, M.A.; Cannon, S.B. Comprehensive Characterization and RNA-Seq Profiling of the HD-Zip Transcription Factor Family in Soybean (Glycine Max) during Dehydration and Salt Stress. *BMC Genom.* **2014**, *15*, 950. [CrossRef]

45. Juven-Gershon, T.; Hsu, J.-Y.; Theisen, J.W.M.; Kadonaga, J.T. The RNA Polymerase II Core Promoter—The Gateway to Transcription. *Curr. Opin. Cell Biol.* **2008**, *20*, 253–259. [CrossRef]

46. Narusaka, Y.; Nakashima, K.; Shinwari, Z.K.; Sakuma, Y.; Furihata, T.; Abe, H.; Narusaka, M.; Shinozaki, K.; Yamaguchi-Shinozaki, K. Interaction between Two Cis-Acting Elements, ABRE and DRE, in ABA-Dependent Expression of Arabidopsis rd29A Gene in Response to Dehydration and High-Salinity Stresses. *Plant J.* **2003**, *34*, 137–148. [CrossRef]

47. Zhang, M.; Qin, S.; Yan, J.; Li, L.; Xu, M.; Liu, Y.; Zhang, W. Genome-Wide Identification and Analysis of TCP Family Genes in Medicago Sativa Reveal Their Critical Roles in Na+/K+ Homeostasis. *BMC Plant Biol.* **2023**, *23*, 301. [CrossRef] [PubMed]

48. Parwez, R.; Aftab, T.; Gill, S.S.; Naeem, M. Abscisic Acid Signaling and Crosstalk with Phytohormones in Regulation of Environmental Stress Responses. *Environ. Exp. Bot.* **2022**, *199*, 104885. [CrossRef]

49. Wang, G.-Y.; Shi, J.-L.; Ng, G.; Battle, S.L.; Zhang, C.; Lu, H. Circadian Clock-Regulated Phosphate Transporter PHT4;1 Plays an Important Role in Arabidopsis Defense. *Mol. Plant* **2011**, *4*, 516–526. [CrossRef]

50. Zhao, S.; Gao, H.; Jia, X.; Li, X.; Mao, K.; Ma, F. The γ-Clade HD-Zip I Transcription Factor MdHB-7 Regulates Salt Tolerance in Transgenic Apple (Malus Domestica). *Plant Soil* **2021**, *463*, 509–522. [CrossRef]

51. Hyten, D.L.; Song, Q.; Zhu, Y.; Choi, I.-Y.; Nelson, R.L.; Costa, J.M.; Specht, J.E.; Shoemaker, R.C.; Cregan, P.B. Impacts of Genetic Bottlenecks on Soybean Genome Diversity. *Proc. Natl. Acad. Sci. USA* **2006**, *103*, 16666–16671. [CrossRef]

52. Liu, S.; Zhang, M.; Feng, F.; Tian, Z. Toward a "Green Revolution" for Soybean. *Mol. Plant* **2020**, *13*, 688–697. [CrossRef]

53. Zhou, Z.; Jiang, Y.; Wang, Z.; Gou, Z.; Lyu, J.; Li, W.; Yu, Y.; Shu, L.; Zhao, Y.; Ma, Y.; et al. Resequencing 302 Wild and Cultivated Accessions Identifies Genes Related to Domestication and Improvement in Soybean. *Nat. Biotechnol.* **2015**, *33*, 408–414. [CrossRef] [PubMed]

54. Qi, J.; Liu, X.; Shen, D.; Miao, H.; Xie, B.; Li, X.; Zeng, P.; Wang, S.; Shang, Y.; Gu, X.; et al. A Genomic Variation Map Provides Insights into the Genetic Basis of Cucumber Domestication and Diversity. *Nat. Genet.* **2013**, *45*, 1510–1515. [CrossRef] [PubMed]

55. Jiao, Y.; Zhao, H.; Ren, L.; Song, W.; Zeng, B.; Guo, J.; Wang, B.; Liu, Z.; Chen, J.; Li, W.; et al. Genome-Wide Genetic Changes during Modern Breeding of Maize. *Nat. Genet.* **2012**, *44*, 812–815. [CrossRef] [PubMed]

56. Hufford, M.B.; Xu, X.; van Heerwaarden, J.; Pyhäjärvi, T.; Chia, J.-M.; Cartwright, R.A.; Elshire, R.J.; Glaubitz, J.C.; Guill, K.E.; Kaeppler, S.M.; et al. Comparative Population Genomics of Maize Domestication and Improvement. *Nat. Genet.* **2012**, *44*, 808–811. [CrossRef]

57. Wang, S.; Liu, S.; Wang, J.; Yokosho, K.; Zhou, B.; Yu, Y.-C.; Liu, Z.; Frommer, W.B.; Ma, J.F.; Chen, L.-Q.; et al. Simultaneous Changes in Seed Size, Oil Content and Protein Content Driven by Selection of SWEET Homologues during Soybean Domestication. *Natl. Sci. Rev.* **2020**, *7*, 1776–1786. [CrossRef]

58. Zhang, D.; Sun, L.; Li, S.; Wang, W.; Ding, Y.; Swarm, S.A.; Li, L.; Wang, X.; Tang, X.; Zhang, Z.; et al. Elevation of Soybean Seed Oil Content through Selection for Seed Coat Shininess. *Nat. Plants* **2018**, *4*, 30–35. [CrossRef]

59. Zhao, X.; Fu, X.; Yin, C.; Lu, F. Wheat Speciation and Adaptation: Perspectives from Reticulate Evolution. *aBIOTECH* **2021**, *2*, 386–402. [CrossRef]

60. Birchler, J.A.; Veitia, R.A. Gene Balance Hypothesis: Connecting Issues of Dosage Sensitivity across Biological Disciplines. *Proc. Natl. Acad. Sci. USA* **2012**, *109*, 14746–14753. [CrossRef]

61. Li, C.; Ng, C.K.-Y.; Fan, L.-M. MYB Transcription Factors, Active Players in Abiotic Stress Signaling. *Environ. Exp. Bot.* **2015**, *114*, 80–91. [CrossRef]

62. He, H.; Zhang, Y.; Xu, S.; Zhang, X.; Yang, X.; Cheng, Y. Genetic Transformation of GmFBX322 Gene and Salt Tolerance Physiology in Soybean. *PLoS ONE* **2024**, *19*, e0307706. [CrossRef] [PubMed]

63. Varshney, R.K.; Bohra, A.; Yu, J.; Graner, A.; Zhang, Q.; Sorrells, M.E. Designing Future Crops: Genomics-Assisted Breeding Comes of Age. *Trends Plant Sci.* **2021**, *26*, 631–649. [CrossRef] [PubMed]

64. Vu, L.D.; Xu, X.; Zhu, T.; Pan, L.; van Zanten, M.; de Jong, D.; Wang, Y.; Vanremoortele, T.; Locke, A.M.; van de Cotte, B.; et al. The Membrane-Localized Protein Kinase MAP4K4/TOT3 Regulates Thermomorphogenesis. *Nat. Commun.* **2021**, *12*, 2842. [CrossRef]

65. Watson, S.J.; Sowden, R.G.; Jarvis, P. Abiotic Stress-Induced Chloroplast Proteome Remodelling: A Mechanistic Overview. *J. Exp. Bot.* **2018**, *69*, 2773–2781. [CrossRef]

66. Ye, H.; Roorkiwal, M.; Valliyodan, B.; Zhou, L.; Chen, P.; Varshney, R.K.; Nguyen, H.T. Genetic Diversity of Root System Architecture in Response to Drought Stress in Grain Legumes. *J. Exp. Bot.* **2018**, *69*, 3267–3277. [CrossRef] [PubMed]

67. Nadeem, M.; Li, J.; Yahya, M.; Sher, A.; Ma, C.; Wang, X.; Qiu, L. Research Progress and Perspective on Drought Stress in Legumes: A Review. *Int. J. Mol. Sci.* **2019**, *20*, 2541. [CrossRef]

68. Gonzalez-Rizzo, S.; Laporte, P.; Crespi, M.; Frugier, F. Legume Root Architecture: A Peculiar Root System. In *Annual Plant Reviews Online*; John Wiley & Sons, Ltd.: Hoboken, NJ, USA, 2018; pp. 239–287. ISBN 978-1-119-31299-4.

69. Wang, Z.; Yung, W.-S.; Gao, Y.; Huang, C.; Zhao, X.; Chen, Y.; Li, M.-W.; Lam, H.-M. From Phenotyping to Genetic Mapping: Identifying Water-Stress Adaptations in Legume Root Traits. *BMC Plant Biol.* **2024**, *24*, 749. [CrossRef]

70. Jiang, G.-L.; Mireku, P.; Song, Q. Utilization of Natural Hybridization and Intra-Cultivar Variations for Improving Soybean Yield, Seed Weight, and Agronomic Traits. *Crop Sci.* **2024**, *64*, 3174–3184. [CrossRef]

71. Rasmusson, D.C.; Phillips, R.L. Plant Breeding Progress and Genetic Diversity from De Novo Variation and Elevated Epistasis. *Crop Sci.* **1997**, *37*, 303–310. [CrossRef]

72. Fasoula, V.A.; Boerma, H.R. Intra-Cultivar Variation for Seed Weight and Other Agronomic Traits within Three Elite Soybean Cultivars. *Crop Sci.* **2007**, *47*, 367–373. [CrossRef]

73. Haun, W.J.; Hyten, D.L.; Xu, W.W.; Gerhardt, D.J.; Albert, T.J.; Richmond, T.; Jeddeloh, J.A.; Jia, G.; Springer, N.M.; Vance, C.P.; et al. The Composition and Origins of Genomic Variation among Individuals of the Soybean Reference Cultivar Williams 821[W][OA]. *Plant Physiol.* **2011**, *155*, 645–655. [CrossRef] [PubMed]

74. Jones, P.; Binns, D.; Chang, H.-Y.; Fraser, M.; Li, W.; McAnulla, C.; McWilliam, H.; Maslen, J.; Mitchell, A.; Nuka, G.; et al. InterProScan 5: Genome-Scale Protein Function Classification. *Bioinformatics* **2014**, *30*, 1236–1240. [CrossRef]

75. Wang, Y.; Tang, H.; DeBarry, J.D.; Tan, X.; Li, J.; Wang, X.; Lee, T.; Jin, H.; Marler, B.; Guo, H.; et al. MCScanX: A Toolkit for Detection and Evolutionary Analysis of Gene Synteny and Collinearity. *Nucleic Acids Res.* **2012**, *40*, e49. [CrossRef] [PubMed]

76. Finn, R.D.; Clements, J.; Eddy, S.R. HMMER Web Server: Interactive Sequence Similarity Searching. *Nucleic Acids Res.* **2011**, *39*, W29–W37. [CrossRef] [PubMed]

77. Kumar, S.; Stecher, G.; Li, M.; Knyaz, C.; Tamura, K. MEGA X: Molecular Evolutionary Genetics Analysis across Computing Platforms. *Mol. Biol. Evol.* **2018**, *35*, 1547–1549. [CrossRef]

78. Chen, C.; Chen, H.; Zhang, Y.; Thomas, H.R.; Frank, M.H.; He, Y.; Xia, R. TBtools: An Integrative Toolkit Developed for Interactive Analyses of Big Biological Data. *Mol. Plant* **2020**, *13*, 1194–1202. [CrossRef]

79. Hu, B.; Jin, J.; Guo, A.-Y.; Zhang, H.; Luo, J.; Gao, G. GSDS 2.0: An Upgraded Gene Feature Visualization Server. *Bioinformatics* **2015**, *31*, 1296–1297. [CrossRef]

80. Bailey, T.L.; Boden, M.; Buske, F.A.; Frith, M.; Grant, C.E.; Clementi, L.; Ren, J.; Li, W.W.; Noble, W.S. MEME Suite: Tools for Motif Discovery and Searching. *Nucleic Acids Res.* **2009**, *37*, W202–W208. [CrossRef]

81. Letunic, I.; Bork, P. 20 Years of the SMART Protein Domain Annotation Resource. *Nucleic Acids Res.* **2018**, *46*, D493–D496. [CrossRef]

82. Krogh, A.; Larsson, B.; von Heijne, G.; Sonnhammer, E.L.L. Predicting Transmembrane Protein Topology with a Hidden Markov Model: Application to Complete Genomes1. *J. Mol. Biol.* **2001**, *305*, 567–580. [CrossRef]

83. Almagro Armenteros, J.J.; Salvatore, M.; Emanuelsson, O.; Winther, O.; von Heijne, G.; Elofsson, A.; Nielsen, H. Detecting Sequence Signals in Targeting Peptides Using Deep Learning. *Life Sci. Alliance* **2019**, *2*, e201900429. [CrossRef] [PubMed]

84. Lescot, M.; Déhais, P.; Thijs, G.; Marchal, K.; Moreau, Y.; Van de Peer, Y.; Rouzé, P.; Rombauts, S. PlantCARE, a Database of Plant Cis-Acting Regulatory Elements and a Portal to Tools for in Silico Analysis of Promoter Sequences. *Nucleic Acids Res.* **2002**, *30*, 325–327. [CrossRef]

85. Szklarczyk, D.; Gable, A.L.; Nastou, K.C.; Lyon, D.; Kirsch, R.; Pyysalo, S.; Doncheva, N.T.; Legeay, M.; Fang, T.; Bork, P.; et al. The STRING Database in 2021: Customizable Protein–Protein Networks, and Functional Characterization of User-Uploaded Gene/Measurement Sets. *Nucleic Acids Res.* **2020**, *49*, D605–D612. [CrossRef] [PubMed]

86. Shannon, P. Cytoscape: A Software Environment for Integrated Models of Biomolecular Interaction Networks. *Genome Res.* **2003**, *13*, 2498–2504. [CrossRef]

87. Ashburner, M.; Ball, C.A.; Blake, J.A.; Botstein, D.; Butler, H.; Cherry, J.M.; Davis, A.P.; Dolinski, K.; Dwight, S.S.; Eppig, J.T.; et al. Gene Ontology: Tool for the Unification of Biology. *Nat. Genet.* **2000**, *25*, 25–29. [CrossRef]

88. Kovaka, S.; Zimin, A.V.; Pertea, G.M.; Razaghi, R.; Salzberg, S.L.; Pertea, M. Transcriptome Assembly from Long-Read RNA-Seq Alignments with StringTie2. *Genome Biol.* **2019**, *20*, 278. [CrossRef]

89. Kim, D.; Paggi, J.M.; Park, C.; Bennett, C.; Salzberg, S.L. Graph-Based Genome Alignment and Genotyping with HISAT2 and HISAT-Genotype. *Nat. Biotechnol.* **2019**, *37*, 907–915. [CrossRef] [PubMed]

90. Lorenz, R.; Bernhart, S.H.; Höner zu Siederdissen, C.; Tafer, H.; Flamm, C.; Stadler, P.F.; Hofacker, I.L. ViennaRNA Package 2.0. *Algorithms Mol. Biol.* **2011**, *6*, 26. [CrossRef]

91. Schirmer, S.; Giegerich, R. Forest Alignment with Affine Gaps and Anchors, Applied in RNA Structure Comparison. *Theor. Comput. Sci.* **2013**, *483*, 51–67. [CrossRef]

92. Jumper, J.; Evans, R.; Pritzel, A.; Green, T.; Figurnov, M.; Ronneberger, O.; Tunyasuvunakool, K.; Bates, R.; Žídek, A.; Potapenko, A.; et al. Highly Accurate Protein Structure Prediction with AlphaFold. *Nature* **2021**, *596*, 583–589. [CrossRef]

93. Mirdita, M.; Schütze, K.; Moriwaki, Y.; Heo, L.; Ovchinnikov, S.; Steinegger, M. ColabFold: Making Protein Folding Accessible to All. *Nat. Methods* **2022**, *19*, 679–682. [CrossRef]

94. Schrödinger, L.; DeLano, W. *PyMOL*, (version 3.1.4.1); DeLano Scientific LLC: Palo Alto, CA, USA, 2020.